ADJUSTING TRUE ODDS TO ALLOW FOR VIGORISH

Stephen R Clarke
Swinburne University of Technology
sandkclarke@hotmail.com

Abstract

A mathematical model predicting sporting outcomes produces probabilities that sum to one, whereas the probabilities implied by bookmaker's odds sum to more than one. Vigorish or over-round is the excess probability that supplies the bookies margin, and could typically range from 2% to over 20%. If the probabilities from a mathematical model are to be used for supplying real time odds they need to be adjusted for the bookmaker's vigorish. In the reverse situation, when testing models from past data, bookmaker's odds need to be adjusted so that the implied probabilities sum to one. Schembri et al (2011) discusses two methods – normalisation and equal distribution. However neither of these suitably allow for the fact that the margin on outsiders is usually greater than favourites. A true price of $1.05 can be reduced to $1.03 for less than 2% margin, whereas a $100 true price could be set at $50 for a $50% margin. This paper discusses an alternative approach using a power function to transform probabilities. This was successfully used when supplying real time odds to a leading bookmarker (Clarke, 2007).

Keywords: Betting, over-round

1. INTRODUCTION

In some sporting studies we need to estimate the chances of past or future events. Often the only estimates available are historical bookmaker's odds or prices. When these are converted to probabilities they sum to more than one. This excess probability is known as the over-round O. This supplies the bookies margin since it results in payouts that are less than justified by the true probabilities. An alternative measure used is the vigorish V, or proportion of the amount bet that the bookmaker retains from a balanced book. Thus V = O/ (1 + O), which in casino applications is called the house percentage. In order to estimate the true probabilities the over-round or vigorish needs to be removed so the probabilities sum to one. In this paper we also use R, the expected proportion of amount bet that is returned to the punter, where R = 1 – V.

Two methods have generally been used in the literature. Schembri et al (2011) and Viney (et al) discuss two methods – normalisation and equal distribution. Equal distribution merely subtracts an equal amount from each probability, whereas normalisation reduces each probability by the same proportion. Thus for two outcomes and implied probabilities of P and Q (P+Q > 1, so O =P+Q -1) we have:

Equal distribution results in p = P - O/2, q = Q - O/2

Normalisation results in p = P / (P+Q), q = Q / (P+Q).

So for example in a tennis match where the prices are $4.04 and $1.20, giving implied probabilities of 0.25 and 0.83 for an over-round of 8% (V = 7.5%, R = 92.5%), equal distribution gives true probabilities of 0.21 and 0.79 (fair prices of $4.83 and $1.26), whereas normalisation results in 0.23 and 0.77 (fair prices of $4.37 and $1.30).

There are some problems with such approaches. The normalisation method results in probabilities that are reduced by a common percentage of V (7.5% in the above example). ie the same percentage is taken from all bets. However it is common practice for bookmakers to take a greater margin out of longer priced outcomes. Clearly the equal distribution method achieves this, with the outsider's probability reduced by 16% and the favourite by only 5%. However this often goes too far, and can produce negative probabilities, particularly when there are several outcomes. For example a horse race with 5 runners at $40, $30, $20, $5 and a hot favourite at $1.12 results in an over-round of 20%, which gives non-positive probabilities for the 3 outsiders.

This paper looks at an alternative. While equal distribution is an additive model and normalisation a multiplicative model, here we discuss a power model.

2. ADJUSTING TRUE PROBABILITIES TO ALLOW FOR VIGORISH.

This problem first arose when providing real time odds for a betting company. A regression model was used to produce estimated probabilities for the number of runs scored in an over of cricket. These had to be adjusted to allow for the vigorish required by the bookmaker, and the solution to take an equal percentage off all prices was considered inadequate. If we wish to return only a proportion R of the amount bet to punters, we need to reduce the payout for an event with probability p and a fair price of 1/p to R/p. Since this return must be greater than 1, it could only be viable when p <R. So for example, for a return to punter of 80% (house

percentage 20%) the shortest price we could adjust would be 1/.8 = $1.25 or 4 to 1 on. Roulette can use this system to decrease all fair prices by the same percentage only because it has a high R (97% or 94%) and the shortest price bets are even money.

Unlike roulette, it is common to take a higher percentage from low probability/high payout events. Thus a 1000 to one shot can be given a payout of $500 for a return to punter of only 50%, while a hot favourite with a true price of $1.02 can barely be reduced at all.

One way to implement this is to give a punter the same return for a double as betting on the two individual events.

Let the payout for an event with probability x be P(x).

Then if winnings are placed all up on a bet with probability y, the final payout is P(y) P(x)

Alternatively, since the double has a probability of xy, the payout on that will be P(xy).

So we want P(xy)=P(x)P(y), and the function satisfying this is the power function $P(x) = 1/x^k$ and we need k < 1 so the price is reduced, not increased.

Thus while equal distribution alters probabilities by an additive constant, normalisation by a constant multiplier, the power method raises them by a constant power. k depends on the return to gambler R. Taking logs we get k = -log(P) / log(x). Thus for a 50/50 bet with a return to punter R the payout will be 2R so k = log(2R)/log2. (For a bet with n equally likely outcomes, k = log (nR)/log n)

So for example if the return for a 50% bet is R = 90.0%, payout is 2R= 1.8, and k = log(1.8)/log2 = 0.848. While this is only exact for 2 equally likely outcomes, it can be used as an approximation. Table 1gives the adjusted probabilities and prices obtained using this method for a range of true probability events. The expected return to punters is for any bet is $xP = x.(1/x^k) = x^{1-k}$, equal to 90% for a 50% bet, higher for favourites and less for outsiders.

True Probability x	Fair Price 1/x	Adjusted Probability x^k	Adjusted Price $1/x^k$	Expected Return x^{1-k}
0.01	$100.00	0.02	$49.66	50%
0.05	$20.00	0.08	$12.68	63%
0.1	$10.00	0.14	$7.05	70%
0.2	$5.00	0.26	$3.91	78%
0.25	$4.00	0.31	$3.24	81%
0.3	$3.33	0.36	$2.78	83%
0.4	$2.50	0.46	$2.17	87%
0.5	$2.00	0.56	$1.80	90%
0.6	$1.67	0.65	$1.54	93%
0.7	$1.43	0.74	$1.35	95%
0.8	$1.25	0.83	$1.21	97%
0.9	$1.11	0.91	$1.09	98%
0.95	$1.05	0.96	$1.04	99%
0.99	$1.01	0.99	$1.009	99.8%

Table 1: Adjusted probabilities and prices using the power method with k = 0.848

With 11 outcomes k = log(11R)/log(11) =0.932 for R = 85% and 0.907 for R = 80%. Table 2 shows the resultant prices for a range of bets for a nominal R = 85% and 80%, along with the actual expected percentage return to the punter. Note the returns are close to the expected values for values around the average $11.00 payout. In our application it was felt the adjusted prices obtained were realistic, and the returns were close enough to that expected for the formula to be used in real time rather than many tables for varying R.

Fair	Nominal R = 85% (k= 0.932)		Nominal R = 80% (k=.907)	
Payout	Payout	% return	Payout	% return
P	P^k	P^{k-1}	P^k	P^{k-1}
$40.00	$31.15	78%	$28.38	71%
$33.33	$26.28	79%	$24.05	72%
$28.57	$22.76	80%	$20.91	73%
$25.00	$20.10	80%	$18.53	74%
$22.22	$18.01	81%	$16.65	75%
$20.00	$16.32	82%	$15.13	76%
$18.18	$14.94	82%	$13.88	76%
$16.67	$13.77	83%	$12.83	77%
$15.38	$12.78	83%	$11.93	78%
$14.29	$11.93	83%	$11.15	78%
$13.33	$11.19	84%	$10.48	79%
$12.50	$10.53	84%	$9.88	79%
$11.76	$9.95	85%	$9.35	80%
$11.11	$9.44	85%	$8.88	80%
$11.00	$9.35	85%	$8.80	80%
$10.53	$8.97	85%	$8.46	80%
$10.00	$8.56	86%	$8.07	81%
$6.67	$5.86	88%	$5.59	84%
$5.00	$4.48	90%	$4.30	86%
$4.00	$3.64	91%	$3.52	88%
$3.33	$3.07	92%	$2.98	89%
$2.86	$2.66	93%	$2.59	91%
$2.50	$2.35	94%	$2.30	92%
$2.22	$2.11	95%	$2.06	93%
$2.00	$1.91	96%	$1.88	94%

Table 2: Adjusted prices and expected returns for an 11 outcome event

3. ACTUAL HOUSE PERCENTAGES OBTAINED IN PRACTICE.

Because the house percentage of different bets changes, the overall percentage taken depends on the distribution of the probabilities of the particular outcomes, and the amounts bet. Thus for example, we expect an event where there are equally probable outcomes to have a different return than one in which there are one or two hot favourites and the rest are highly unlikely.

Expected return to punter = (Σ bet size * true prob winning * payout)/(total bet)

= (Σ proportion of pool bet * true prob winning * payout)

If punters bet in the same proportion as the probabilities, we have,

Expected return to punter = Σ x*x* payout, where x is the probability of an outcome

(Note for the constant percentage case, payout = R/x, so expected return to punter = R as required)

Using the power formula, payout = $1/x^k$, so expected return = Σx^{2-k}

It is easily shown using Lagrange Multipiers that the maximum value of this is R when all x's are equal.

So assuming n outcomes all with equal probability of 1/n we finally get

Maximum expected return = $\Sigma x^{2-k} = \Sigma(1/n)^{2-k} = n.(1/n)^{2-k} = (1/n)^{1-k} = R$, since

$k = \log(nR)/\log(n) = 1 + \log(R)/\log(n)$, so $\log(R) = (k-1)\log(n) = \log(1/n)^{1-k}$

Thus the value used for R is in general a maximum and actual returns will be less than this.

In the two outcome example this means the target return is only achieved for two equal opponents. Table 3 shows the true and reduced prices and the expected return to the bookmaker for a balanced book on a two outcome event for a sample of markets. The return to the bookmaker only deviates markedly from the expected percentage if the outsider is less than a 30% chance.

Probabilities of 2 outcomes		True Prices		Reduced prices		Return to Bookmaker
0.5	0.5	$2.00	$2.00	$1.80	$1.80	10.0%
0.4	0.6	$2.50	$1.67	$2.17	$1.54	9.8%
0.3	0.7	$3.33	$1.43	$2.78	$1.35	9.0%
0.2	0.8	$5.00	$1.25	$3.91	$1.21	7.7%
0.1	0.9	$10.00	$1.11	$7.05	$1.09	5.3%
0.02	0.98	$50.00	$1.02	$27.59	$1.02	1.9%
0.01	0.99	$100.00	$1.01	$49.66	$1.01	1.2%

Table 3: Return to bookmaker for a balanced book on a 2 outcome event using Power method to convert true probabilities with a nominal vigorish of 10%.

4. ADJUSTING BOOKMAKERS ODDS TO ALLOW FOR VIGORISH.

The above shows we can use the formula $P = 1/x^k$ to adjust prices of events with true probability x to allow for a return to punter of R, where k = log (nR)/ log(n). This gives adjusted probabilities that give a return to the punter of exactly R when all events are equally likely, but less than this in other cases.

When used in reverse to remove over-round from bookmakers prices P we have adjusted price = 1/x = $P^{(1/k)}$ or Adjusted price = $P^{k'}$ where k' = log(n)/log (nR) where n is the number of outcomes and R is the return to punter = (1- O). However since this only gives the required R for equally likely outcomes, we need to use iteration to produce probabilities that sum to 1. This is easily performed in a spreadsheet.

Consider the prices for a tennis match where published prices are $1.22 and $4.33. This gives implied probabilities of 0.820 and 0.231 for an over-round of 0.051 and return to punter R of 0.952 and so k = 0.929. The equal probability method distributes the 0.051 equally for probabilities of 0.794 and 0.206 (prices of $1.26 and $4.88). The normalisation method takes each probability as a proportion of the total for probabilities of 0.780 and 0.220 (prices of $1.28 and $4.55) Note this is equivalent to increasing each price by the (required) same proportion. The power method initially gives probabilities of .807 and 0.206, but these still sum to more than one. Using iteration to adjust k to correct this, we obtain k = 0.904 and probabilities of 0.802 and 0.198 (or prices of $1.25 and $5.05). Clearly this method increases the prices of the outsiders to a greater extent than the other two methods.

In events with a larger number of competitors such as horse racing, the outsiders are at longer odds and the over-rounds are much greater than in two person events such as tennis. Table 4 shows the three methods applied to a race with 6 runners.

Prices and their Implied probabilities		Calculated True Probabilities: Equal Distribution	Calculated True Probabilities: Normalisation	Calculated True Probabilities: Power	Calculated Fair Prices: Equal Distribution	Calculated Fair Prices: Normalisation	Calculated Fair Prices: Power
$1.15	0.870	0.828	0.696	0.825	$1.21	$1.44	$1.21
$5.00	0.200	0.158	0.160	0.110	$6.31	$6.25	$9.12
$10.00	0.100	0.058	0.080	0.042	$17.12	$12.50	$23.63
$20.00	0.050	0.008	0.040	0.016	$118.97	$24.99	$61.23
$50.00	0.020	-0.022	0.016	0.005	-$46.31	$62.48	$215.54
$100.00	0.010	-0.032	0.008	0.002	-$31.65	$124.96	$558.47
Total	1.250	1.000	1.000	1.000			

Table 4: Comparison of 3 methods of adjusting prices to remove 20% vigorish

With an over-round of 25% or return to punter of 80% the equal probability method breaks down giving negative probabilities for the outsiders. The normalisation method merely reduces all probabilities by 20% (increases prices by 25%). Schembri et. al. (2011) concluded that the normalisation method is less effective when there is a strong favourite, as too much over-round is given to the favourite. The iterative power method with an initial k of 0.876 iterates to k = 0.728 adjusts the favourites to a lesser extent than the normalisation, but adjusts the outsiders much more. The power method thus avoids the problems the equal distribution method has with outsiders, and the over allocation the normalisation method has with favourites.

5. CONCLUSION

A power method can be used to adjust true probabilities to ones that sum to more than one to allow for over-round. Used in reverse requires iteration. The method should be considered, as it more truly allows for the practice of taking a greater percentage out of winning bets on outsiders than favourites.

Acknowledgements

This work originally arose from a project working with Mark Solonsch, and resulted from discussions I had with him on adjusting odds.

References

Clarke, S.R. (2007) Successful applications of statistical modeling to betting markets. In IMA Sport 2007: First International conference on Mathematics in Sport. D. Percy, P Scarf & C Robinson (Eds.) The Institute of Mathematics and its Applications, Salford, UK. pp35-43.

Schembri, A., Bedford, A., O'Bree, B. & Bressanutti, N. (2011). Comparing market efficiency with traditional and non-traditional ratings systems in ATP tennis. *Journal of Quantitative Analysis in Sports* (JQAS),1.

Viney, M., Bedford, A. & Kondo, E. (2013). Incorporating over-round into in-play Markov Chain models in tennis. 15th International Conference on Gambling & Risk-Taking, Las Vegas, USA.