Betting on Excel to enliven the teaching of probability

Stephen R. Clarke
School of Mathematical Sciences
Swinburne University of Technology

Abstract

The study of probability has its roots in gambling problems, and the gaming industry remains a source of interesting applications of many of the methods taught in beginning probability courses. The use of examples from gambling can enhance the interest and usefulness of courses in probability, and equip students to participate intelligently in this form of entertainment. This paper explores how the many statistical functions of Microsoft Excel and utilities such as solver can be used to solve gambling problems. We use a simulation of roulette as a vehicle for experimentation; introduce means as a measure of house percentage; variance as a measure of volatility; Hypergeometric and binomial distributions to calculate chances of winning single bets and of being ahead after many bets; and use solver to balance a sports book. The use of Excel allows key ideas in probability to be explored through the investigation of realistic examples.

Introduction

Probability first arose as an area of study through gambling problems. Gamblers were interested in how to divide the stake when a game was interrupted. Gambling is now a huge industry, with total gambling losses in Victoria, Australia in 1997-8 reaching $3.19 billion. The gaming industry remains a source of interesting applications of many of the methods taught in beginning probability and statistics courses. The use of examples from gambling can enhance the interest and usefulness of statistics courses. Croucher (2000) discusses an introductory course in statistics based solely around gambling and sport. Gambling is becoming pervasive in our society, and an understanding of its principles will assist students to participate sensibly in this form of entertainment, either as providers or consumers. Interest can be heightened if the tedium of calculations can be reduced, and real examples used. Microsoft Excel is a powerful package with many built-in functions that can be used in statistics. This paper explores some statistical examples related to gambling easily implemented within Microsoft Excel.

Gambling is an excellent vehicle to introduce the concepts of probability. The three usual definitions of probability rely on a 'mathematical' definition, based on equally likely outcomes, an 'experimental' definition based on relative frequency of occurrence in repeated trials, and a 'subjective' definition based on degree of belief. Most casino games can be approached via the first, long term profits to the casino or gambler via the second, and most sports betting via the third.

Simulating Roulette

One of the simplest gambling games is Roulette. A roulette wheel has the numbers 1 to 36 with half coloured red and half black, in addition to a green zero (and sometimes a double zero). Various bets such as a straight, split, square, line, street, column, dozen, odd or even, red or black, high or low, can be laid on groups of numbers. Payouts are made as if there are 36 numbers, when in fact there are 37 (or 38 in a double zero wheel). So for example if $1 is bet on a single number and that comes up, the casino pays 35 to1. Thus the punter wins 35 dollars plus his own dollar back. If a $1 bet is made on the High numbers (19-36), and any of those numbers comes up, the casino pays 1 to1 (18 to 18) or even money. Thus the punter wins 1 dollar plus his own dollar back. The site http://www.accidental.com has some information on roulette wheels and a roulette wheel you can operate for no cost.

A roulette wheel is an excellent way to introduce the frequency definition of probability - the probability of an event is the number of outcomes in the event as a proportion of the total number of (equally likely) outcomes. The probability of all bets in roulette can be calculated by simple counting. Thus a bet on the High numbers (19-36) has 18 outcomes, so the probability of the High bet winning is 18/37. A square has four outcomes, so the probability is 4/37.

Insights can be gained into many gaming situations by simulation. This is often the choice when a mathematical analysis is too difficult. However it also gives a vehicle whereby a student can experiment, and gain experience of the long run, without actually putting large sums of money at risk. Excel can be used to simulate a sequence of bets. Students can gain some practical experience of the behavior of random events before discussing issues that arise from that experience. We illustrate with roulette.

This example is to create a simple simulation of a roulette wheel using Excel. It is too complicated to simulate allowing a full range of bets, but some simple ones can be easily implemented. In this case we will simulate betting the same size bet continually on a single number, and on a set of numbers. Students may be able to generalize depending on their Excel skills. Below is a set of instructions as issued to students for simulating roulette.

1. In Cell A1, type your favourite number between 0 and 36. This is the number you will bet on.
2. ColumnC: Head *Bet number*, and generate the sequence 1,2,3...in the column.
3. ColumnD: Head *Winning Number*, and use the formula RANDBETWEEN(0,36) to generate the number appearing on the roulette wheel. You may need to install the Analysis ToolPak to run this function.
4. Column E: Head *Profit*, and use an if statement to generate –1 or 35 depending on whether your number came up.
5. Column F: Head *Cumulative Profit*, and total profit up to now.
6. Copy Formulas down to allow for a betting sequence of 1000 bets.
7. Place in cell A4 your percentage profit/loss, in A6 the largest profit, and in A8 the largest loss.
8. Use the graph wizard to do a scatter plot of the cumulative profit against bet number. Place the graph at the top right of the spreadsheet. Comment on your graph.
9. Repeat several simulations by hitting F9.
10. Add two more columns to allow betting on the even money bet High. Drag the profit column on to your graph. Place in column B the relevant statistics for this bet. Is the even money bet more or less volatile (risky) than the bet on a single number?

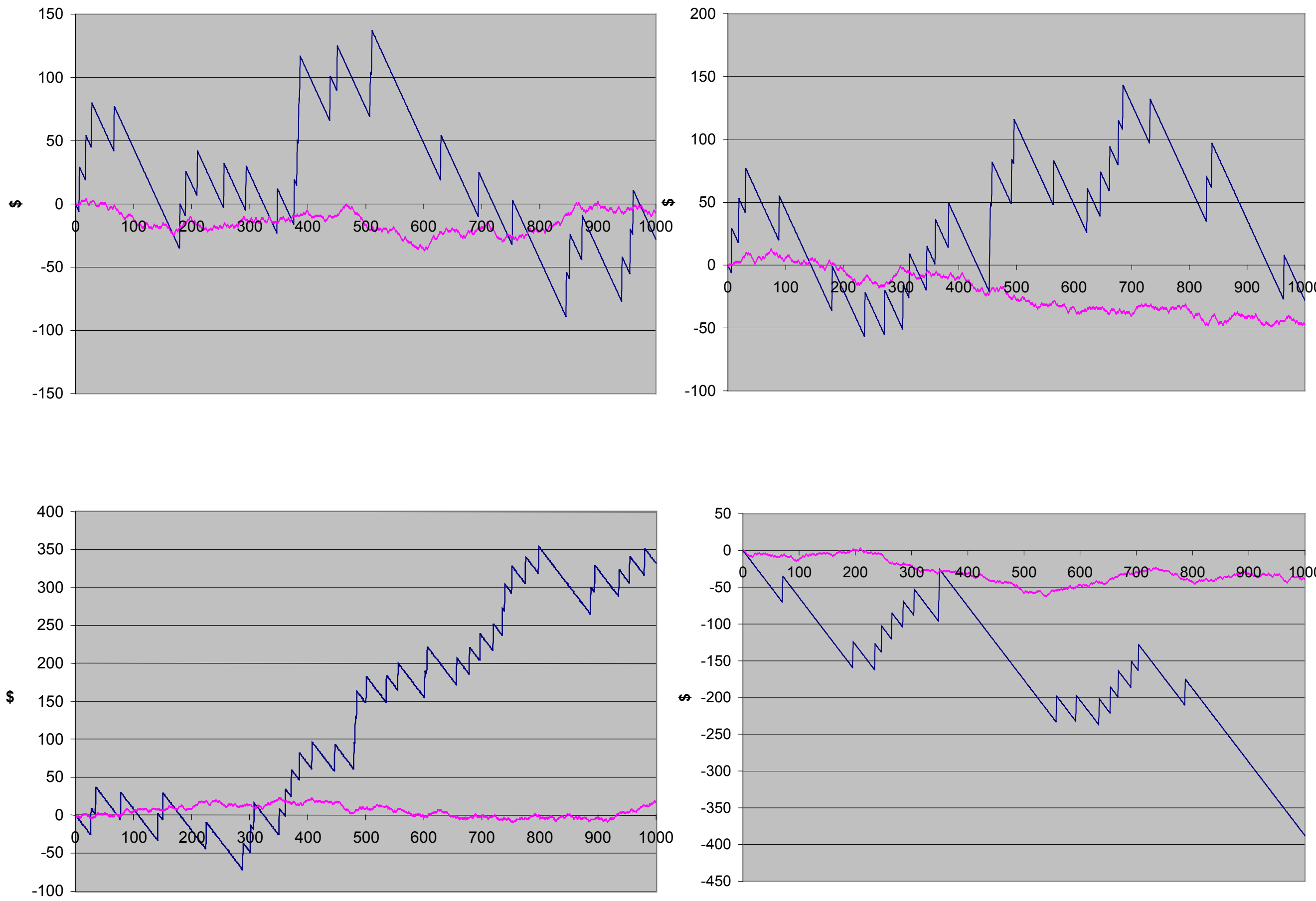

Figure 1. Four simulations of cumulative profit on 1000 bets on a single number and red in roulette.

Figure 1 shows four examples of graphs produced by the above sheet. It is difficult to show 'Typical' graphs as they vary so widely. Students will quickly see the volatility of profit in the short term. Repeatedly hitting F9 gives a surprising variation in outcomes. Runs of good and bad luck characterize the graphs. There are large fluctuations, and even after 1000 bets, a player might be in front. It is obvious that the 'even money' bet is far less volatile than the bet on a single number, and there are many questions that arise naturally out of such simulations. Does a gambler generally win or lose? What is the chance he is ahead after a certain number of spins? What is the chance of winning/losing a given amount? How long does it take to win/lose a given amount? What is overall profit in the long run? Students could investigate these by repeating the above simulation many times, and recording the final profit as given in cell A4, or some other statistic each time. Again Excel has facilities for performing this automatically.

For example, suppose we wish to estimate not only the final amount a player is ahead after 1000 bets, but the maximum amount a player is ahead and behind at any stage in the series of 1000 bets. The data table facility of Excel allows us to simulate the whole sequence of 1000 bets 100 times. The output from one of these sequences showed that betting on 'high' gave an average profit of -$24.70, with an average maximum ahead of $15.40 and average maximum behind of $38.80. Betting on a single number produced -$27.60, $126.60 and $160.80. Such simulations clearly demonstrate that while both bets lose about 2.5%, the number bet is the more volatile (exciting? risky? variable?).

By experimenting with such a simulation, students may come up with their own problems, which might then be investigated via traditional probability calculations.

House percentage and volatility

A random variable is one whose value depends on a random experiment, and probability courses investigate random variables and their expected values. In gambling, as we saw in the above simulation, a random variable of particular interest is the amount returned to a punter or profit made on a bet. The expected profit made by the operator on a $1 bet is the house percentage, and can be considered as the long run price the gambler pays for playing the game. Just as different loans carry different interest rates, the games in a casino can vary markedly in their cost to play (from virtually zero to 40% or 50%). Yet many gamblers are unaware of the house percentages associated with different games or how to calculate them.

Consider a bet of $1 on number 23 at Roulette, which pays 35 to 1 if it wins. With probability 1/37 a player will win and receive $36 (his original bet plus another $35) and with probability 36/37 the player loses and receives nothing. A simple table in Excel can be made with the various outcomes and their respective chances.

Table 1: Return for betting on a single number (23) in roulette.

Outcome	23	Not 23	
Return	36	0	Total
Probability	1/37	36/37	1
Probability x return	36/37	0	36/37

Of interest is the proportion of revenue that is returned to the punter via the various outcomes, which is obtained by calculating the probability of each return times the amount of the return. In this case, there is only one winning outcome, but in many other games, such as lotteries and poker machines, there are several winning outcomes and games are structured differently to appeal to the psychology of players. For example, a lottery may return a high proportion of the prize pool via its main or division 1 prize, whereas a poker machine may return a very low proportion of the amount bet via its biggest jackpot. Similarly, different poker machines may appeal to conservative or reckless gamblers by altering the proportion of stake returned in often occurring small prizes and rarely occurring large prizes. The sum of these proportions gives us the mean or expected value of the return. In the above case this is 36/37. Thus the expected return to a gambler betting on a single number in roulette is 36/37, or about 97.3%. Thus 2.7% is the house percentage - the casino makes on average nearly 3% of everything that is bet on a number. In fact this is the house percentage on virtually every bet at roulette.

There are various alternatives for setting up the table in Excel. It is usually easier to use columns rather than rows to list the various outcomes. If you install the Add-in Analysis VBA toolpack, the sumproduct function allows you to calculate the expected value directly, without calculating the individual values of return times probability. The table could also be calculated using the profit rather than return to the punter. In this case for a $1 bet the punter wins $35 with probability 1/37, and -$1 with probability 36/37, for an expected profit of -$1/37. So the gambler loses on average 2.7%. For a double zero wheel, as is common in the USA, the house percentage doubles to 5.4%.

An alternative to the approach taken in Table 1, which is just as easy to implement in Excel and a bit more flexible, is to list all the individual outcomes from 0 to 36. These all have probability 1/37, and the returns for various bets could be listed. So for a bet on a square, for example, each of the four numbers in the square would return $9 if they came up. Another row could be introduced to allow bets of different sizes, and the total bet and total expected return calculated. This would allow various combinations of bets to be entered, and Excel would immediately give the expected return. For example, there are various systems promulgated, where people bet on several bets simultaneously. These could easily be tested, and shown to have the same expected return of 97.3%. Once such a sheet was constructed, it could easily be converted into a roulette wheel simulation. The randbetween function is used to generate a number between 0 and 36, and if statements are used to convert the probabilities into 1 or 0 depending on whether the outcome matches the generated number.

The spreadsheets can also be modified to calculate the variance of random variables, as shown in Table 2. The SQRT function is used to calculate the standard deviation.

Table 2: Mean and standard deviation of return X for a $1 bet on a single number

x	P(X=x)	xP(X=x)	$(x-\mu)^2$	$(x-\mu)^2P(X=x)$	Std Dev
36	0.0270	0.9730	1226.8926	33.1593	5.84
0	0.9730	0.0000	0.9467	0.9211	Variance
Total	1.0000	0.9730		34.0804	34.08

Note that with careful planning, once a spreadsheet has been created, it can be reused for another problem. Consider the bet on 'High'. This has a return of $2 with probability 18/37, and 0 with probability 19/37. Entering these values in the first and second column of the above sheet, we automatically get the mean and standard deviation of the return for this bet. This is shown in Table 3, where we have used the alternative calculation formula for variance.

Table 3: Mean and standard deviation of return X for a $1 bet on 'high'

x	P(X=x)	xP(X=x)	x^2	x^2 P(X=x)	Std Dev
2	0.4865	0.9730	4.00	1.946	0.9996
0	0.5135	0.0000	0.00	0.000	Variance
Total	1.0000	0.9730		1.946	0.9993

Note that although this bet has the same mean return, the standard deviation is about one sixth of the single number bet. Thus the single number bet is 5.8 times more variable, or more volatile, or 5.8 times more exciting for the gambler and 5.8 times more risky for the operator. (That is why maximum bets on numbers are usually less than the even money bets). A gambler who each week had 100 bets on a number, and kept a record of his returns, would find they were about 6 times more variable than a similar gambler who had 100 bets on 'High' or Red each week.

Spreadsheets similar in design to the above can be used to calculate the mean and variance of return for poker machines.

Keno

Excel has many functions which can be used to calculate probabilities. Two that are often useful in gambling applications are the Hypergeometric and binomial functions. As an example we can look at Keno. There are many versions of Keno, but typically players select from 1 up to 15 numbers from the numbers 1 to 80. Prizes depend on the number of hits in 20 numbers selected randomly from 80. The list of prizes for for tattersalls keno can be found by following the links on the website http://www.tattersalls.com.au/ and they even give relevant formula and chance of winning (but not the house percentage). For example, for a Spot 8 Keno, the prize for matching all 8 numbers is $25000, 7 numbers is $1200 and 6 numbers is $90. The chance of winning these prizes is given as 1 in 230115, 1 in 6,232 and 1 in 422. These Keno tables form a good example to apply the method described above, and the HYPERGEOMETIC function of Excel can be used to calculate the probabilities. Table 4 gives the table produced for the above game. The formula for the probability for matching 8 cell is HYPGEOMDIST(B5,8,20,80), where B5 is the address to the cell on the left. Taking the inverse of the probability figures confirms the chances given in the website. Note the return is a low 51% - in many cases, the returns in Keno are very low.

If the game parameters (number of balls in population, number selected by operator, number of balls selected by player) are entered in separate cells, it is a very simple manner to set up a spreadsheet which will work for all Kenos. The payoffs for the various matches then have to be entered and Excel automatically calculates the chances and returns.

Table 4. Payoff table and probabilities for Keno spot 8

Matches	Probability	Payoff	Probability x Payoff
8	4.3E-06	25000	0.108
7	0.00016	1200	0.192
6	0.00237	90	0.213
5	0.01830	0	0
4	0.08150	0	0
3	0.21479	0	0
2	0.32815	0	0
1	0.26646	0	0
0	0.08827	0	0
Total	1.00000		0.514

Repeated trials

While the house percentage gives the long term average cost of playing a game (or long term profit, from the viewpoint of the operator) we are also interested in the short term. Gamblers usually make a sequence of bets - what are the likely outcomes for the gambler? An operator needs to know the variation about the long term average. It is not much good having a profitable game in the long run, if they go broke in the short term.

If the binomial distribution is covered, the BINOMDIST function of Excel can be used to calculate the chances of being in any position after any number of bets. For example betting on an even money bet, we will be level or behind after 1000 bets if we have won 500 or fewer bets. BINOMDIST(500,1000,18/37,1) gives our chance to be 0.81, so we only have a 19% chance of coming out in front after 1000 bets. Betting on a number, we will be behind if 27 bets or fewer win (since we have returned $36 for each winning bet). BINOMDIST(27,1000,1/37,1) gives .55, so using this strategy we have a 45% chance of being ahead.

As discussed before, these two bets lose the same average amount in the long run - over many simulations they would each lose on average 1000/37 = $27, but the second is more volatile - when you win, you win more, when you lose, you lose more. This could lead to a discussion of variance. Since the profit after 1000 bets is the sum of the 1000 profits on each individual bets, it could also lead to a discussion of the central limit theorem and normality. In this case, parabolic error bands, within which the profit will lie with any degree of confidence, could be put on the graphs.

Odds and bookmaking

In many gaming situations, particularly sports betting, probabilities cannot be calculated but must be estimated using experience. In these cases, the estimated probabilities need to be calculated by working backwards from the prices or odds offered by the operator.

A common way of expressing probabilities, particularly popular in gambling, is by odds. Odds of 2 to 1 against a particular event mean there are 2 chances of the event not occurring and 1 chance of it occurring. Thus the probability of the event is 1/3 or 0.33. Thus odds of x:y against give a probability of y/(x+y). In a betting context, the first figure is the profit you make if you win (the bookmaker's contribution) and the second is the loss if you lose (the punter's contribution). If you place a bet of $1 at 3 to 1 against, and you win, you get back $4 – your original bet of $1 plus the $3 you won. Events with greater than 50% chance are usually expressed as on. Thus odds of 5 to 2 on mean 2 to 5 against or to a probability of 5/7.

Conversion of a bookmaker's odds

Since a bookmaker hopes to make a profit, his odds do not correspond to probabilities – they fail the important criteria of adding to one. In common with Haigh (1999) we will call y/(x+y) the relevant fraction. By adding these up, the amount the relevant fractions exceed one (the overround) can be used to estimate the bookmaker's margin, or percentage.

For example, in a sporting contest, odds of 6 to 4 on and 5 to 4 on might be offered on each competitor. This gives relevant fractions of .6 and .56, for a total of 1.16, and an overround of 16%. Thus the bookmaker's margin is approximately 16%, or more correctly 16/116 = 13.8%.

In many totalisator systems, the price quoted will be the total amount returned to you for a single $1 bet. This is now common in sports betting and many situations where fixed odds are quoted. A price or payout of $2.50 (for a $1) bet means odds of 1.5 to 1 or 3 to 2, and a 'probability' of 1/2.5. In these cases the house percentage is even easier to calculate. Thus a horse race with 10 runners paying $7, $7, $6, $4, $9, $5, $26, $9, $15, and $21 gives a total relevant fraction of 1/7+1/7+1/6+1/4+1/9+1/5+1/26+1/9+1/15+1/21= 1.28 for a bookmaker's margin of 28/128 = 22%

It is a simple exercise to create an Excel spreadsheet that allows the entry of the odds or prices of each possible result in one column, and calculates the relevant fraction (or probabilities in adjacent column) and thus gives the expected return to the bookmaker. Betting markets can be obtained from the Web or the newspaper, and prices entered into the table to find the house percentage

The same method can be used on casino games, provided a set of exclusive and exhaustive outcomes are used. For example, consider the outcomes red, black and zero on a roulette wheel. These outcomes return $2, $2, and $36 for a total 1/2+1/2+1/36 = 37/36 = 102.8 % and a house percentage of 2.8/102.8 = (you guessed it) 2.7%. Once the spreadsheet is constructed, any set of prices can be entered and the house percentage returned immediately.

Such a table can also be used to investigate some betting scams. For example, the bookmaker's percentage if the favourite is nobbled (the probability of the favourite winning becomes zero) might become negative. So a punter who knows the favourite cannot win can make a profit in the long run (and with a sensible choice of bets on all outcomes, be certain of winning in the short term).

Balancing a book

In fact, a bookmaker constructing a book on an event, is not trying to predict the true probabilities of an event occurring, but the proportion of the punters who will bet on a particular outcome. The bookmaker is trying to balance his book, so he will not just make a profit in a long run of events, but will be certain to make a profit on each event no matter what the outcome. This process can also be investigated via a spreadsheet, and some useful insights learnt.

To a spreadsheet storing the outcomes and payouts, add a column that takes the amount bet on each outcome, and calculates the bookmakers return if that outcome occurs. Much like a simple profit and loss spreadsheet, this can be compared with the total amount bet to find the profit on each outcome. The MINIMUM function can be used to create a cell that calculates the bookmaker's worst result. A bookmaker tries to steer money away from his worst result (or any result on which he makes a loss) by adjusting the odds and making that bet less attractive.

Suppose a bookmaker holds $100 on a race. What is the optimal amount for bookmaker to hold on each outcome? Students might try to adjust the amounts bet to ensure the bookmaker's worst result is as good as possible. How does this amount compare to the expected profit? In fact solver is the perfect tool to apply here. Maximise the worst result, by altering the amounts bet, subject to the constraint that the total bet is $100. In this case students find the optimum bets for each outcome are 100/price, the return for each outcome is exactly the same, and the worst result (in fact common result for all outcomes) for the bookmaker is a profit equal to his house percentage. The book is balanced. A Totalisator produces a balanced book automatically, by a similar process, by determining the payouts after all the bets have been placed.

Conclusion

The use of gambling applications can enhance the interest and usefulness of probability and statistics courses. Haigh (1999), Orkin (2000) and Croucher (2001) are examples of texts that could be used with this approach. In addition, the pervasiness of gambling means there is no shortage of real life examples that can be analysed. Real and web casinos, and the betting pages in newspapers would be the first port of call. There are many other interesting classical problems in statistics that can be investigated via gambling applications - betting systems, runs, gambler's ruin problem, optimal bet size. Croucher (2000) and Clarke & Bedford (2002) are examples of interesting applications that arise in the press from time to time. The capabilities of Excel make it a useful tool to use in this context, and allows real problems to be analyzed. With thought in setting up the spreadsheet, new events can be analysed with a minimum of structural alteration and data entry. These spreadsheets can be used to give students an appreciation of the principles, and hopefully the pitfalls, of gambling.

References

Clarke, S. R., & Bedford, A. (2002). Statistics going to the dogs. *Teaching Statistics, 24*(1), 6-9.

Croucher, J. S. (2000). Using probability intervals to evaluate long-term gambling success. *Teaching Statistics, 22*(2), 42-44.

Croucher, J. S. (2001). *The Statistics of Gambling : Understanding Tactics and Risk*. North Ryde: Macquarie University Lighthouse Press.

Haigh, J. (1999). *Taking chances: Winning with Probability*. New York: Oxford University Press.

Orkin, M. (2000). *What are the Odds? Chance in Every day Life*. New York: W.H. Freeman and Company.