The Statistician (1998)
47, *Part* 4, *pp.* 629–641

Using maximum entropy to double one's expected winnings in the UK National Lottery

Simon J. Cox†, Geoffrey J. Daniell and Denis A. Nicole
University of Southampton, UK

[Received May 1997. Revised May 1998]

Summary. We have used the maximum entropy method to estimate the probability of each of the 14 million tickets being chosen by players in the UK National Lottery. As data, we used the numbers of winners in the three-, four- and five-match categories and the total number of tickets sold in each of the first 113 draws. We have computed the marginal distributions for players choosing single numbers and pairs of numbers. A striking conclusion is that players preferentially pick numbers towards the centre of the ticket. By choosing unpopular combinations of numbers, one's expected winnings can be doubled.

Keywords: Expected winnings; Lotteries; Maximum entropy; Models of gambler choice

1. Introduction

For some lotteries the organizers reveal statistical information about players' choices of numbers and combinations of numbers. Such information is of considerable interest in modelling the behaviour of lottery players and can enable players to increase their expected winnings by choosing unpopular combinations of numbers. For the UK National Lottery, however, only the number of prize-winners in each prize category and the total number of players are released for each draw. The numbers of prize-winners in particular categories are random deviates which depend on the probability that players pick sets of numbers. It is easy to demonstrate that the observed numbers of prize-winners are inconsistent with players selecting numbers at random (Haigh, 1995), but it is less easy to deduce details of players' choices from these data. In this paper, we apply the maximum entropy technique to elicit structure in players' choices of numbers starting from the published numbers of prize-winners. Further details of the UK National Lottery are given by Moore (1997) and Haigh (1997).

The maximum entropy method is described in Section 2 and the application to the data in Section 3. The results presented in Sections 4 and 5 show that it is possible to deduce features of players' choice that have hitherto only been seen in detailed information that is occasionally released by lottery organizers. In Section 5 we demonstrate the effect on our results as data from more draws became available, and we give some results indicating how well the model fits later draws in Section 6. The rough calculations in Section 7 indicate that by choosing unpopular

†*Address for correspondence*: Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton, SO17 1BJ, UK.
E-mail: sc@ecs.soton.ac.uk

 0039–0526/98/47629

numbers lottery players can double their expected winnings. Finally in Section 8 we make some general comments about the applicability of maximum entropy to this problem.

2. The maximum entropy method

The best known application of the maximum entropy method is in statistical physics where it can be used, for example, to determine the energy distribution for molecules of a gas; moreover its predictions are in excellent agreement with experiments. Its use in 'pure' statistics problems is less well known but Jaynes (1959, 1983) has persistently argued that its use can be justified as a method of consistent reasoning from incomplete information.

Jaynes's maximum entropy principle says that, if we are forced to assign probabilities p_i using limited information, we should do so by maximizing the entropy of the distribution:

$$S = -\sum_i p_i \log(p_i), \tag{1}$$

subject to the constraints of known expectation values.

In the UK National Lottery, we wish to determine the probability of each of the possible 13983816 tickets being bought, subject to the constraints that the probabilities are consistent with the numbers of winners observed in the draws so far.

Before discussing any technicalities it is instructive to contrast the maximum entropy approach with more traditional methods which start from specific models for the behaviour of lottery players. One simple approach is the 'Poisson' model, in which for every ticket and every draw there is a single constant probability that a player picks a given ticket. This model can be emphatically rejected (Haigh, 1997). Another model attributes a popularity to individual numbers which players then select independently. Again, this model can be demonstrated to be inadequate; Haigh (1997) reviews more sophisticated models. When a satisfactory model has been discovered, we can go on and examine its consequences.

In the maximum entropy method, we deliberately choose to make no assumptions about the system characterized by the probabilities p_i; these are simply numbers determined by the constraints. We need only to assume that these probabilities are constant in time. Of course, by constructing a specific model for the underlying system, one's predictions may be more impressive, but they will be less robust.

This lack of assumptions can be illustrated by considering those lottery tickets that have never won a prize. After 113 draws, there are about 1.5 million such tickets and, using as data only the numbers of prize-winners, we remain completely ignorant of how many players pick these tickets. The maximum entropy method automatically makes the probabilities associated with these all equal. Any predictions based on the maximum entropy probabilities are significantly affected by these tickets.

If we were to add the very reasonable assumption that the draw is fair, so that it was only chance that prevented these tickets from winning, and in addition that the tickets which have not yet won a prize are no different from the others, then we can improve our predictions. It is in the spirit of the maximum entropy method not to make either assumption; it could be that one particular ticket was being bought by a million players, but we do not know this because these players have not yet won anything. If the rumour that more than 10000 people choose the numbers $\{1, 2, 3, 4, 5, 6\}$ is correct, then the second 'very reasonable' assumption just put forward is false; at least one ticket is not 'typical'.

The maximum entropy distribution is summarized by Jaynes (1983) as

> 'the most conservative assignment in the sense that it does not permit one to draw any conclusions not warranted by the data'.

3. Application of maximum entropy to the National Lottery

We wish to determine the probability of each of the tickets being bought, subject to the constraints imposed by the numbers of winners observed in the draws so far. It is convenient to number all possible lottery tickets by a single index t, which is an abbreviation for six numbers, chosen without repetition, from the set of integers $\{1, 2, \ldots, 49\}$. Let $P(t)$ be the probability that a player chooses the ticket labelled t. Players make independent samples t from the distribution $P(t)$. The winning set of numbers drawn in a particular week is denoted by n, using the same abbreviated notation. Let

$$\Delta_r(t, n) = \begin{cases} 1 & \text{if } t \text{ and } n \text{ have exactly } r \text{ numbers in common,} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The expected fraction of players matching exactly r numbers, in a week when the winning numbers are indexed by n, is then given by $f_r(n)$ where

$$f_r(n) = \sum_t \Delta_r(t, n) P(t). \tag{3}$$

In the UK National Lottery the number of winners each week in the categories $r = 3, 4, 5$ is sufficiently large that values of $f_r(n)$ can be derived from published statistics with negligible uncertainty. Suppose that W lottery draws have been made; then values of $f_3(n)$, $f_4(n)$ and $f_5(n)$ are known for W different values of n: $n_1, n_2, \ldots, n_W$. Equation (3) then leads to a set of $3W$ constraints that apply to the distribution $P(t)$.

We have assumed that $P(t)$ is independent of time. This is obviously false in the short term, since, for example, people bet on numbers that have come up, or not come up, recently (Haigh, 1997). However, it is known in lotteries which release information about players' choices that the popularity of individual numbers is remarkably constant over time (Haigh, 1997). Perhaps more interesting is that stationarity is observed despite the information being published! In the long term, the assumption of stationarity is reasonable. With the assumption of stationarity, players make independent samples from $P(t)$; therefore the number of players buying ticket t follows a Poisson distribution with parameter $\mu(t) = P(t)N$, when N tickets are sold.

We have used data from the first 113 weeks, up to January 1997. The data were obtained from the Internet site `http://lottery.merseyworld.com/`, which is updated after each draw. Typical numbers of winners in the three-, four- and five-match categories are 1.1 million, 55000 and 1000 respectively.

Because the number of winners in the jackpot ($r = 6$) and five plus the bonus ball categories is small (5 and 50 respectively), only a rough estimate of the *expected* fraction of winners can be made from this. Although the jackpot results are the only direct evidence of the number of players choosing particular tickets, the number of winners is so small that any conclusions based solely on this are obscured by fluctuations. Contrary to popular opinion, the number of winners in the jackpot category contains little information; it only constrains the number of players picking one particular ticket. We have considered those winning in the five plus the bonus ball category simply as five-match winners. As the total number of five-match winners (including those also matching the bonus ball) is about 1050, the expected fraction is uncertain to 3%. However, a change in the data of this magnitude would not affect our conclusions appreciably. Uncertainties in the expected fraction of three- and four-match winners are correspondingly less.

As explained in Section 2, we find the maximum entropy estimate $\hat{P}(t)$ by maximizing the entropy S from equation (1) subject to the $3W$ constraints (3) and additionally the requirement that $\hat{P}(t)$ be normalized. Introducing Lagrange multipliers λ'_r and ν, we maximize

$$-\sum_t \hat{P}(t)\log\{\hat{P}(t)\} + \sum_{j,r} \lambda_r^j \left\{\sum_t \Delta_r(t, n_j)\,\hat{P}(t) - f_r(n_j)\right\} + \nu\left\{\sum_t \hat{P}(t) - 1\right\} \tag{4}$$

and obtain the analogue of the Boltzmann distribution of statistical mechanics:

$$\hat{P}(t) = \frac{1}{Z}\exp\left\{\sum_{j,r} \lambda_r^j \Delta_r(t, n_j)\right\}, \tag{5}$$

where the constant $Z = \Sigma_t \exp\{\Sigma_{j,r} \lambda_r^j \Delta_r(t, n_j)\}$ normalizes the probability distribution and is the analogue of the partition function.

To find the unknown Lagrange multipliers λ_r^j, we substitute $\hat{P}(t)$ from equation (5) into the constraint equations (3). For W draws, the constraint equations (3) define a set of $3W$ non-linear equations for the Lagrange multipliers λ_r^j:

$$f_r(n_j) = \frac{1}{Z}\sum_t \Delta_r(t, n_j)\exp\left\{\sum_{j,r} \lambda_r^j \Delta_r(t, n_j)\right\}. \tag{6}$$

We solved these using an iterative technique based on Newton's method; see Press *et al.* (1992). In total the computations in this paper took 200 node hours on a parallel supercomputer (IBM SP2), requiring 500 Mbytes of memory for $W = 113$. The procedure converges in 5–8 iterations.

4. Results and discussion

It is well known from lotteries where the organizers release statistical information that players do not pick numbers randomly and $P(t)$ is not uniform; see Moore (1997). That this is also true in the UK lottery can be deduced from the variance in the number of prize-winners from week to week (Haigh, 1997). In the absence of any constraints, the maximum entropy distribution is uniform and corresponds to players choosing their tickets at random. When constraints are imposed, the method provides the most uniform configuration for the probability distribution $\hat{P}(t)$ which is consistent with the data. We might therefore expect that the maximum entropy estimate $\hat{P}(t)$ would not vary much. On the contrary, Table 1 gives a selection of 50 tickets, and our estimate of how many people buy them on average, when 68.5 million tickets are sold. This 'most uniform' distribution identifies some tickets that are bought by over 40 people and some by fewer than 1.

Our estimate for the number of people buying $\{1, 2, 3, 4, 5, 6\}$ is considerably less than the 10000–30000 people said to buy this ticket each week (Haigh, 1997). This is to be expected, since this ticket has (up to draw 113) won only three-match prizes and on only three occasions. The effect of an additional 10000 people buying this ticket slightly changes our estimate for the number of people buying each of the 246820 tickets with three matches on these draws. When this ticket has a four match or higher, we can expect to see a large number of winners if it is really bought by 10000 people each week. With these additional data, the maximum entropy estimate for the number of people buying this ticket will increase.

Although the method produces the 14 million values for $\hat{P}(t)$, from equation (5), it is difficult to assimilate this information. We have therefore looked at the marginal distributions for players choosing individual numbers, and pairs of numbers. It is possible that further inspection of the 14 million values of $\hat{P}(t)$ could reveal more about the strategies used by players in choosing numbers.

In Fig. 1, we show the estimated probability of players picking individual numbers. The most popular number is shown to be 7, which agrees with observations on other lotteries. We also confirm the rumour that there is a smaller probability that numbers greater than about 30 are picked. An equally striking effect is the periodicity in the graph. We believe that this is caused by players tending to pick numbers towards the centre of the lottery ticket.

Table 1. Selection of 50 tickets with our estimate of how many people buy them when 68.5 million tickets are sold, using 113 weeks of lottery data†

Ticket number	*Estimate of how many people buy this ticket*	*Ticket number*	*Estimate of how many people buy this ticket*
26 34 44 46 47 49	0.41	1 7 17 23 32 38	21.74
4 41 42 44 46 49	0.53	3 5 14 18 30 44	23.42
17 41 42 44 46 49	0.53	3 7 9 17 32 42	24.11
26 41 44 46 47 49	0.54	4 17 23 32 38 42	24.12
34 43 44 46 47 49	0.55	7 9 17 32 38 42	25.20
26 34 41 44 47 49	0.56	3 7 17 32 38 42	26.95
4 17 41 42 46 49	0.59	7 15 17 23 32 42	30.69
26 34 35 46 47 49	0.60	2 17 23 32 38 42	31.48
26 28 34 44 47 49	0.61	5 17 23 32 38 42	32.29
25 34 44 46 47 49	0.62	6 7 17 23 32 38	32.80
26 34 40 46 47 49	0.63	4 7 17 23 32 42	33.82
4 17 41 42 43 44	0.63	6 7 17 23 32 42	34.70
26 34 43 46 47 49	0.63	3 7 17 23 32 38	34.88
34 35 44 46 47 49	0.63	2 7 17 23 32 42	35.33
1 4 17 41 42 49	0.66	7 9 17 23 32 38	35.33
1 4 17 41 42 44	0.76	7 8 17 23 32 38	36.59
2 4 17 41 42 44	0.87	7 17 23 32 42 48	39.74
4 41 42 43 44 45	0.91	7 17 23 32 35 38	39.76
7 26 38 44 46 47	0.91	7 13 17 23 32 42	40.60
1 2 3 4 5 6	4.81	7 8 17 23 32 42	40.73
7 17 23 32 33 35	20.14	4 7 17 23 32 38	42.37
3 5 14 18 33 44	20.43	7 9 17 23 32 42	42.66
7 13 17 23 40 42	20.62	7 17 23 32 35 42	44.26
7 8 11 17 23 38	20.68	3 7 17 23 32 42	45.62
7 11 17 32 42 47	20.69	7 17 23 32 40 42	45.62

†The number of people on average buying a ticket, t, is $P(t) \times 68.5 \times 10^6$.

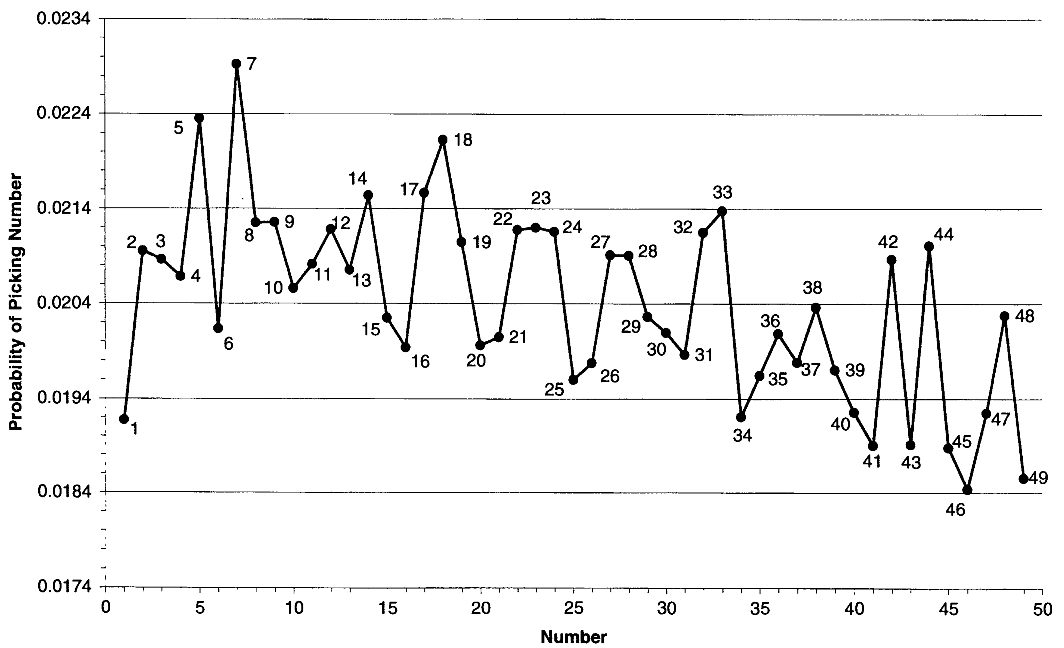

Fig. 1. Estimated probability of players picking individual numbers, based on 113 weeks of data

In Fig. 2, we superimpose the probability distribution for players picking individual numbers on the ticket lay-out, which makes the structure in the picking distribution very clear. The most popular numbers are unshaded, and the least popular are darkly shaded. Higher numbers and those lying around the edges of the ticket can be seen to be less popular. We suggest that 17 and 18 are popular simply because they lie in the middle of the ticket. There is no obvious explanation for the popularity of 5. Our conclusion that players avoid the edges of the ticket is confirmed by published marginal distributions, e.g. Stern and Cover (1989) for the Canadian lottery. Here a similar periodicity is visible, which is consistent with the different lay-out of the ticket.

Our estimate of the distribution with which players choose pairs of numbers is dominated by the way in which individual numbers are picked. This effect can be removed by dividing the marginal distribution for players picking pairs of numbers by the probability that this pair is picked based on sampling without replacement using the marginal probabilities in Fig. 1. The result in Fig. 3 shows that the following structure in players' choice of number pairs is detectable by our method.

(a) The greater density of dark squares at the bottom right-hand side of Fig. 3 indicates that pairs of high numbers, e.g. $\{41, 43\}$, are even less likely to be picked by players than the popularity of the individual numbers suggests.

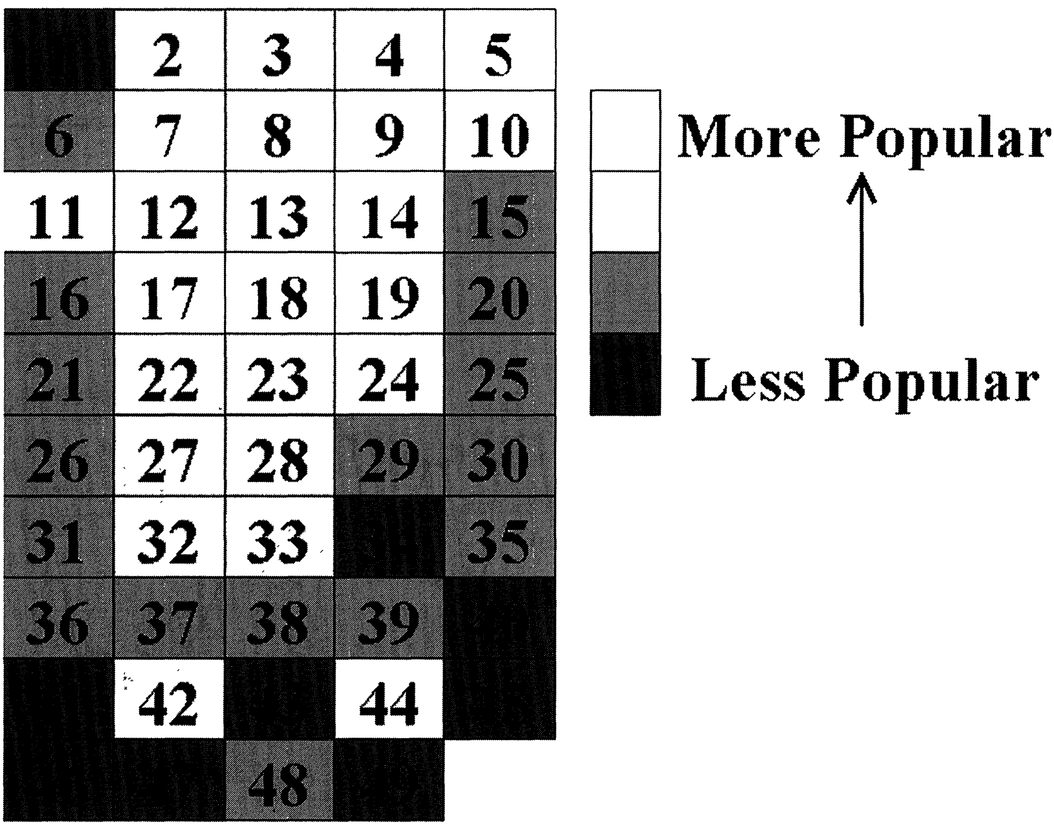

Fig. 2. Estimated probability of players picking each number, laid out as on a lottery ticket: the most popular numbers are unshaded and the least popular are darkly shaded

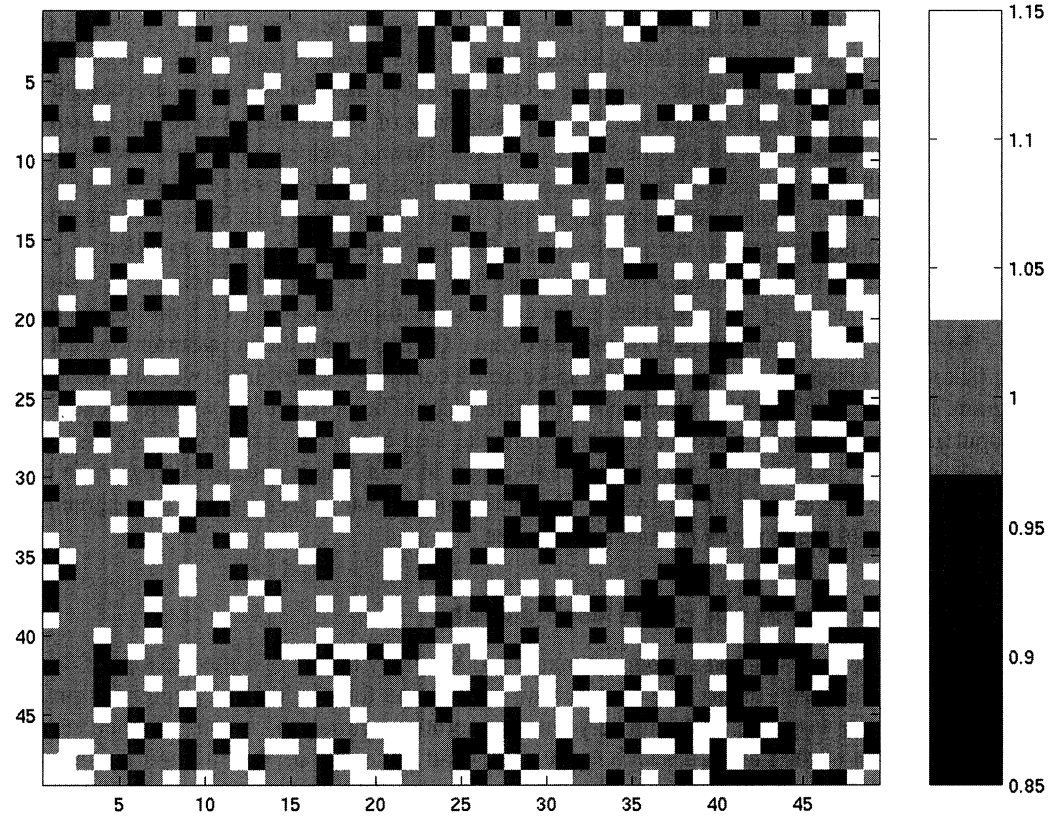

Fig. 3. Probabilities of players picking pairs of numbers, with the effect of the single-number marginal distribution removed: a ratio of less than 1 (darker shading) in the entry for pair *i* with *j* means that the pair of numbers is less popular than would be expected from the individual numbers *i* and *j*

(b) The lighter squares lie roughly parallel to the diagonal and suggest that players tend to space their numbers more uniformly over the range 1–49 than we would expect by chance.

(c) Commentators on the lottery have often suggested that people underestimate the probability that the winning numbers contain two consecutive numbers and may be disinclined to pick tickets which include these. From our values for $\hat{P}(t)$, we can estimate the probability that a player chooses a ticket with at least one consecutive pair. The result is 0.4910, compared with the theoretical value of 0.4952 that such a ticket is drawn; so the predicted effect is not huge. It must be remembered, however, that this single figure is an average over all tickets and any genuine effect is obscured by more dominant effects such as the edge–centre preference and the popularity of individual numbers.

(d) Looking down the leading diagonal of Fig. 3 gives an indication of the popularity of pairs of consecutive numbers with the effect of the single-number marginal distribution removed. 32 pairs of consecutive numbers are less popular than we would expect and only 16 more popular. If these are binomial trials, a split as extreme as this would occur by chance with probability 0.015. Haigh (private communication) has noted that the average number of jackpot winners is low when the draw produces two consecutive numbers. These simple calculations therefore suggest that the effect is worth investigating in more detail.

As is clear from Table 1, we can identify tickets which are bought on average by as few as 0.41 people and as many as 45.6 people. In Fig. 4 we extend the information from Table 1 to show how the popularity of tickets varies. For example, around 500000 identifiable tickets are bought, on average, by between 3.4 and 3.6 players. As a consequence of strict adherence to the maximum entropy principle, tickets can be assigned to two classes: 'losing', which have never won a prize, and 'winning', which have. At the time of writing, there are 1.5 million losing tickets and the data contain no information about how many people buy them. As explained in Section 2, maximum entropy automatically assigns an equal probability that they are bought. From equation (5), this probability is $1/Z$ so they are bought on average by $68.5 \times 10^6/Z = 5.2$ people. The 1.5 million losing tickets appear in Fig. 4 as an extra spike at 5.2 superimposed on the 12.5 million winning tickets that make up the smooth part of the histogram. Given the additional assumption that the losing tickets are not special in any way, the spike at 5.2 could be redistributed over the rest of the histogram. For example, the maximum entropy solution could be obtained for winning tickets and the popularity of the losing tickets estimated from the single-number probabilities by using an independent choice model. It is important to realize that the maximum entropy histogram is only spread out from a single spike at 5.2 by the minimum amount that is necessary to give agreement with the data. The true histogram is necessarily wider.

5. Effect of the number of draws used as data

It is natural to enquire how good are our estimates and how they improve as the number of draws, W, increases. Starting with theory, the best that we could hope for would be asymptotic formulae for the variance for large W. Unfortunately the asymptotic regime will not be reached until a substantial fraction of the tickets has won many prizes and this value of W will require thousands of years of lottery draws! Monte Carlo simulations are also difficult in such a computationally intensive problem.

As a first step towards investigating how our results depend on W we have repeatedly solved

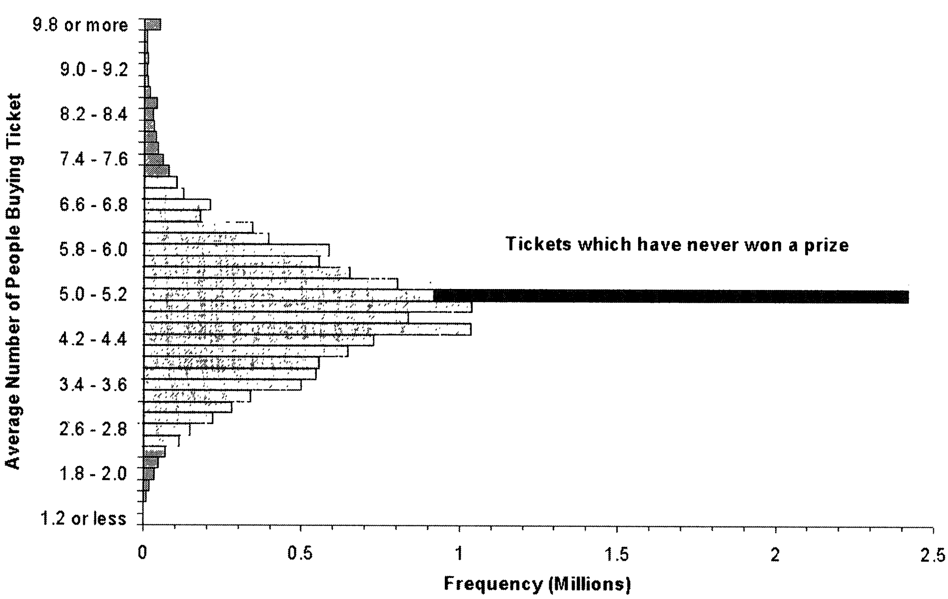

Fig. 4. Histogram of the popularity of tickets, when 68.5 million tickets are sold

equation (6) using data for the first W lottery draws, as W is increased from 1 to 113. These calculations are also computationally expensive.

In Fig. 5 we show a representative sample of how the probabilities of picking various numbers have evolved as more data are added. In interpreting these we must remember that the maximum entropy estimator is highly biased, but the bias is decreasing as W increases. For $W = 0$, all probabilities are equal to $1/49$ and, given the assumption that the ticket buying distribution is stationary, they tend towards a limit as data from more draws are added. It is difficult to separate the effects of increasing W from true non-stationarity, but further work with additional data might reveal something.

The main differences in popularity between numbers have persisted for at least a year, and it is inconceivable that these are not genuine effects. W would have to become much larger before the gross features that are visible in Fig. 5 could be changed.

6. Predicting the number of winners

We claim to identify tickets which are popular or unpopular, and it is important to test the predictions that this implies for the number of winners. The goal of predicting the number of jackpot winners is unattainable because this number has large statistical fluctuations even if the probabilities were known exactly. We have therefore chosen to estimate the *fraction* of three-match winners, which is additionally independent of the number of players.

We have divided the available data into a test set and a data set. The data from draws 1–98 (inclusively) are used to predict the results for the remaining test set of the 15 weeks of data from draws 99–113.

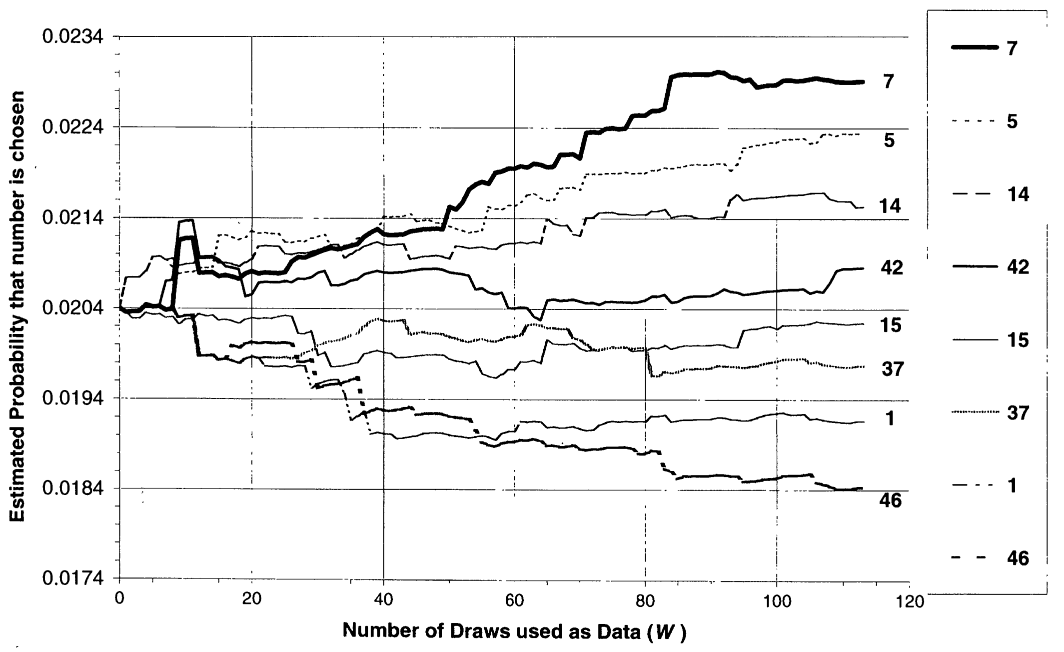

Fig. 5. Evolution of the estimated probability of players picking individual numbers as more draw data are added sequentially

Fig. 6 shows a scatterplot of the actual fraction of three-match winners *versus* our predicted fraction. As explained earlier, the maximum entropy method is necessarily conservative: underestimating the number of players buying popular tickets and overestimating the number buying unpopular tickets. Thus we tend to underestimate an abnormally large number of winners and to overestimate when the true number of winners is abnormally small. The degree of overestimate or underestimate depends on how much is known about the winning ticket for that draw, which in turn depends on how many times the ticket has won prizes before. The points in Fig. 6 are therefore a mixed population with the well-measured tickets lying near a line with a slope of 1, and the less-well-measured tickets off this line.

At present our ability to quantify how well the method can predict the number of three-match winners is limited by several factors. Firstly, the test data contain a mixed distribution of well-measured and poorly measured tickets. Secondly, a very limited amount of data are available, which must be divided up to deduce our conclusions and then to test them. Thirdly, we have not made any additional assumptions about those tickets for which we have little or no data. Our predictions could therefore be enhanced in the short term. However, we believe that this short-term gain is unnecessary in any case, since in the long term the tickets will become better measured as more data accumulate. The appeal of the maximum entropy principle is the lack of endless *ad hoc* assumptions that require additional justification.

7. Long-term returns

Lottery players naturally want to know the expected return from their stake. For a randomly chosen ticket this is the 45p that is returned in prizes. We have shown that there are tickets that are

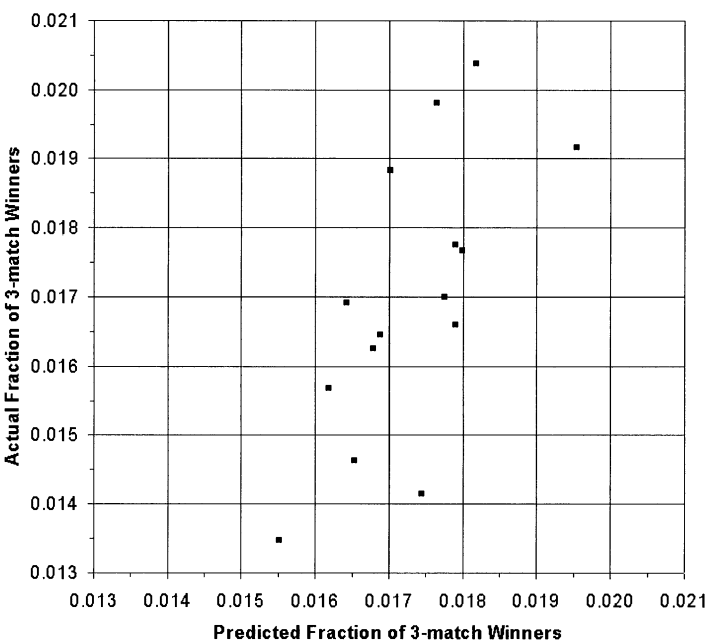

Fig. 6. Prediction of the fraction of three-match winners in weeks 99–113, using data from weeks 1–98 (inclusively)

bought, on average, by one or fewer players, e.g. {26, 34, 44, 46, 47, 49}, and we can consider how much we can expect to win if we buy such a ticket.

We could obtain a Monte Carlo estimate of our long-term winnings by generating samples from $P(t)$, including our extra ticket purchase, and computing the average winnings over all possible draws. Here we present a simpler approach.

It is convenient to estimate separately the long-term return from the jackpot prize, three-match prize and the other prizes. Our long-term expected return from a ticket, $R(t)$, is

$$R(t) = R_3(t) + R_6(t) + R_{\text{other}}(t). \tag{7}$$

The expected return from the guaranteed three-match prize is fixed at $R_3(t) = £0.176$.

The return from the jackpot, $R_6(t)$, can be estimated as follows. In those weeks, on average about every 10^5 years, when a particular ticket t is drawn, the number of players buying this ticket follows a Poisson distribution with parameter $\mu(t) = P(t)N$, when N people enter the draw. If there are i winners in the jackpot category then an additional 'test player' purchasing ticket t will have a probability $\mu(t)^i \exp\{-\mu(t)\}/i!$ of winning a $(1 + i)$th share of the jackpot pool. The factor $1 + i$ results from the fact that the test player is an *additional* player purchasing this ticket. The expected return from the jackpot is therefore

$$R_6(t) = \sum_{i=0}^{\infty} \frac{\rho q}{i+1} \frac{\mu(t)^i}{i!} \exp\{-\mu(t)\} = \frac{\rho q\{1 - \exp(-\mu)\}}{\mu} \tag{8}$$

where q is the probability of matching six numbers (equal to $1/13\,983\,816$). ρ is a typical jackpot pool of £10 million, which is the value observed when 68.5 million tickets are sold.

The third component of the long-term expected return, $R_{\text{other}}(t)$, is more difficult to estimate. However, we can observe in the lottery results so far that, when there are no winners in the jackpot category, the average return from these prizes is as high as 26p, compared with 13p when there are the average five jackpot winners. Thus if our chosen ticket is bought by fewer than one person, then the expected return from these prizes, $R_{\text{other}}(t)$, is almost certain to lie between 13p and 26p.

From Table 1, our estimate of $\mu(t)$ for the ticket {26, 34, 44, 46, 47, 49} is 0.41 giving an expected return between £0.90 and £1.03, which is at least double the expected winnings of £0.45 from a randomly chosen ticket.

A slightly more popular ticket {26, 34, 40, 46, 47, 49}, with $\hat{\mu}(t) = 0.63$, gives an expected return in the range £0.84–0.97. The expected return is likely to be in the upper half of this range, since other similar tickets are also unpopular and any four-, five- or five plus the bonus ball prize from this ticket is likely to be high. In Table 1 we see, for example, {25, 34, 44, 46, 47, 49} for which $\hat{\mu}(t) = 0.62$.

The 'lucky-dip' facility allows players to buy random tickets and this currently accounts for about 13% of total sales. If the allocation of lucky-dip tickets is genuinely random, this gives a floor of 0.57 people per ticket on average. Since this is comparable with our estimated number of buyers for very unpopular tickets, the lucky-dip ought to be allowed for in the calculations.

There are several reasons why our estimate of the expected return from unpopular tickets is likely to be an underestimate.

(a) The maximum entropy estimate of $\mu(t)$ is biased and overestimates the number of players buying unpopular tickets, as explained earlier.
(b) We have ignored the effect of the smaller number of three-match winners which generally occurs when there are few winners in the jackpot category. This implies a larger jackpot prize pool.
(c) In the long run there is an additional return in the jackpot category from occasional

'superdraws', when the jackpot prize pool is guaranteed. This is funded from 5p of the £1 stake. Up to week 113 there have only been three superdraws, though more have occurred following the launch of the midweek draw. Thus in the long term a player's expected return is increased by 5p.

8. Discussion of the maximum entropy method

An axiomatic justification of the maximum entropy method is given in Shore and Johnson (1980); their work can be illustrated in the present context by the following example. Imagine that lottery players choose just *two* numbers in the range $1-M$, and that repeated values are not forbidden. Let p_{ij} $(= p_{ji})$ be the probability that the pair $\{i, j\}$ is selected and suppose that the only data we are given are the marginal probabilities of the two numbers being selected:

$$\sum_j p_{ij} = \theta_i \tag{9}$$

and the problem is to assign values to p_{ij}. Since there is no unique solution we would prefer not to address the problem at all! But, if forced to give an answer, the only sensible one is

$$p_{ij} = \theta_i \theta_j. \tag{10}$$

Any other value for p_{ij} would suggest that there was some correlation between i and j, and we have not been given any data from which this conclusion may be drawn. Having decided that $p_{ij} = \theta_i \theta_j$ is the preferred answer, we can ask whether this realizes an extremum for some function of p_{ij}. It can be shown that the entropy

$$S = -\sum_i p_{ij} \log(p_{ij})$$

has its maximum value subject to constraint (9) when $p_{ij} = \theta_i \theta_j$. Moreover, this function is unique. There is no fundamental reason why the problem of assigning probabilities should be solved by a variational principle, but this argument does suggest that the principle of maximum entropy could be used profitably with more complicated data.

Stern and Cover (1989) have used this argument in reverse; they used the principle of maximum entropy to derive the model $p_{ij} = \theta_i \theta_j$ when the marginal distribution of i and j was given. This seems to be an overcomplicated way of introducing a model which merely says that players pick i and j independently. In the case of the UK National Lottery the primary data that are needed to use such a model are not available. Stern and Cover (1989) indeed suggested that, without being given the marginal probability of picking each number, little can be done to estimate $P(t)$. Our use of the maximum entropy method estimates $P(t)$ directly from the only data available for the UK National Lottery: the observed numbers of prize-winners. By examining these 14 million probabilities, we can begin to see a breakdown of the product model in Fig. 3.

A criticism of the maximum entropy method is that, in the absence of data, it assigns a probability of $1/13\,983\,816$ that a player picks a given ticket (Stern and Cover, 1989). They comment that 'there are several reasons, however, to suspect that the unconditional distribution may not be uniform' and list some possible evidence for bias in players' choices of numbers. We believe that 'unconditional distribution' means the distribution one would assign in the absence of *any* data or evidence *at all*. Without any evidence or data, the symmetry between the tickets demands that we should assign a probability of $1/13\,983\,816$ that a player picks a given ticket.

Our approach has been to take the evidence that players do not buy tickets at random, namely the large fluctuations in the observed numbers of winners, and to derive a ticket buying distribution which is consistent with these fluctuations. Of the many distributions which are consistent with the observed numbers of winners, picking the maximum entropy distribution means that we are choosing in a consistent way. To pick any other distribution would imply that there was some quantitative additional data that we had not used.

9. Conclusions

Using the fraction of winners in the prize categories for matching three, four and five numbers, we have applied the maximum entropy method to determine the probability of each ticket being bought in the UK National Lottery. A striking observation is that people tend to pick numbers towards the centre and top of the ticket lay-out. We have also observed that certain pairs of numbers are less likely to be picked than we would expect from the popularity of the individual numbers. We can identify tickets for which the long-term return is double that of a random ticket.

Acknowledgements

We would like to thank Angeli Thomas and Keith Lloyd for useful discussions and John Haigh for some constructive comments. We also appreciate the efforts of Richard Lloyd in carefully collating the data and placing them on the Internet.

References

Haigh, J. (1995) Inferring gamblers' choice of combinations in the National Lottery. *IMA Bull.*, **31**, 132–136.
———(1997) The statistics of the National Lottery. *J. R. Statist. Soc.* A, **160**, 187–206.
Jaynes, E. T. (1959) *Probability Theory in Engineering and Science*, pp. 110–151. Socony Mobil Oil Company.
———(1983) *Papers on Probability, Statistics and Statistical Physics* (ed. R. D. Rosenkrantz). Dordrecht: Reidel.
Moore, P. G. (1997) The development of the UK National Lottery: 1992–96. *J. R. Statist. Soc.* A, **160**, 169–185.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. R. (1992) *Numerical Recipes in FORTRAN 77*, 2nd edn. Cambridge: Cambridge University Press.
Shore, J. E. and Johnson, R. W. (1980) Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inform. Theory*, **26**, 26–37.
Stern, H. and Cover, T. M. (1989) Maximum Entropy and the lottery. *J. Am. Statist. Ass.*, **84**, 980–985.