A Probabilistic Analysis of the Dutch Lotto[1]

Ruud H. Koning[2] and Thijs Vermaat[3]

SOM Theme F: Interactions between consumers and firms

Abstract

The size of the jackpot of the American powerball lottery that fell in August 2001 has generated a lot of interest in the probability of winning a lottery, both in the United States and elsewhere. Many people dream of becoming rich in a single moment. We calculate the probabilities of winning prizes, and the expected pay-off of the Dutch Lotto.

We also address an important issue: are the numbers and color drawn truly random? We analyze draws of the Lotto from 1974 onwards. It is impractical to test whether each possible draw occurs according to its expected frequency. It is possible, though, to test different implications of the hypothesis that the numbers and colors are drawn randomly. We find that there is no evidence against the hypothesis that the numbered balls are drawn randomly, but the hypothesis that the colored ball is drawn randomly is rejected decisively. Keywords: Lottery, Jackpot, goodness-of-fit.

[1]We thank Ton Steerneman for helpful comments.

[2]Corresponding author, Department of Econometrics, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands, email: ruud@rhkoning.com.

[3]IBIS UvA BV, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands, email: tvermaat@science.uva.nl.

1 Introduction

The size of the jackpot of the American powerball lottery that fell in August 2001 has generated a lot of interest in the probability of winning a lottery, both in the United States and elsewhere. Many people dream of becoming rich in a single moment. In The Netherlands there are three big lotteries that offer this opportunity: the 'Staatsloterij', the 'Postcode loterij', and the 'Lotto'. The three lotteries differ in the way tickets are sold, and how the winning numbers are generated. In this paper we focus on the Lotto only. We calculate the probabilities of winning prizes, and the expected pay-off in Section 3. Before doing so, we give a brief history of the Lotto and its rules in Section 2.

In Section 4 we address an important issue: are the numbers and color drawn truly random? We analyze draws of the Lotto from 1974 onwards. It is impractical to test whether each possible draw occurs according to its expected frequency. It is possible, though, to test different implications of the hypothesis that the numbers and colors are drawn randomly. We end with some concluding remarks in Section 5.

2 The Lotto

The Lotto started on September 1st, 1974 as lottery designed to support the Dutch soccer association. Later, the profits of the Lotto were used to fund sports in general (75%, the Dutch Olympic committee is the main beneficiary), and other charities (25%). The Lotto started with 41 numbered balls. In 1996 the number of balls was increased to 45, and as of 2000 a colored ball was drawn as well. Prizes have varied over time: in 1974 it was possible to win a house or some gold or a videorecorder, nowadays only monetary prizes or new tickets can be won.

The rules of the Lotto are as follows. People participate in the Lotto by buying a ticket or a subscription. By entering a subscription, the participant plays in every draw. The cost of participation and prizes are automatically deducted from and added to the bank account. There are two types of tickets. Jackpot tickets, which offer a chance on the Jackpot, and small tickets, which have smaller prizes (and do not offer a chance on the Jackpot). Jackpot tickets cost €0.68 each and the prizes are listed in Table 1. The Jackpot and the first three prizes are shared if there are multiple winners. The ticket lists six numbers out of 1 to 45, and it also lists a color. There are six possible colors, green, orange, blue, purple, yellow, and red. A participant can choose to fill the six numbers on the ticket himself, or he can buy a ticket on which the numbers are pre-printed. The latter option ensures that the numbers on the lottery ticket are random[4]. A preprinted ticket and an open

[4]It is well known that people are not very good at drawing random numbers. Smaller numbers are more popular than bigger numbers, see on this issue also Haigh (1997) and Haigh (1999). This is important, because the four highest prizes are shared if there is more than one winning ticket.

Figure 1: A pre-printed Lotto ticket (top) with numbers 14, 17, 21, 37, 38, 45 and color blue, and an open ticket (bottom) that can be completed by the participant.

ticket where the participant chooses the numbers himself are shown in Figure 1. Small tickets cost €0.45 each, and the prizes are listed in Table 2. No color can be chosen on a small ticket.

The winning numbers are chosen by a machine, under the supervision of a notary[5]. The machine is of a bowl with balls labelled 1 to 45, and it chooses six balls randomly (without replacement, of course). These balls are the 'regular balls'. Then, it chooses a seventh ball, the so-called 'bonus-ball', so in total seven balls are drawn, six regular balls and one bonus-ball. Finally, another machine has a bowl with 36 balls, six balls of each colors and a colored ball is chosen. A complete draw is an eight-tuple $D = (B_1, B_2, B_3, B_4, B_5, B_6, B_B, B_C)$.

A ticket wins the Jackpot, if it has the numbers of all six regular balls, and if the color is correct. There is no ordering of numbers on the ticket. If fewer than six of the regular balls are on a ticket, it is possible that the bonus ball is on the ticket. In Table 1, K denotes the number of regular balls that are on a ticket, B indicates whether the bonus ball is on the ticket, and C indicates whether the color is is on the ticket. For example, $(K, B, C) = (4, 1, 0)$ means that 4 regular balls are listed on the ticket, that the bonus ball is listed on the ticket, but that the color is not listed on the ticket. The probabilities are obtained from

$$\Pr(K = k, B = b, C = c) = \frac{\binom{6}{k}\binom{39}{6-k}}{\binom{45}{6}} \left(\frac{6-k}{39}\right)^{b} \left(\frac{33+k}{39}\right)^{1-b} \left(\frac{1}{6}\right)^{c} \left(\frac{5}{6}\right)^{1-c} \tag{1}$$

The prizes (in Euros) are listed in Table 1. The last four prizes are not monetary prizes, but new tickets. Small tickets can only win prizes with $C = 0$ (prize 1, 3, etc.). Small tickets are not sold much, but are usually obtained as the 15th or 17th prize in a draw.

The size of the Jackpot is not constant. It starts with €2 million, and it is increased by €250 thousand every time the Jackpot does not fall. If the Jackpot has not fallen during 60 draws[6], it is guaranteed to fall in the 61st draw, even if no ticket has six regular numbers and the color correct. In that case, it may fall on a ticket (or tickets) with five regular numbers correct, and the bonus ball correct.

3 Financial Return of the Lotto

The probabilities that any ticket will win a specific prize are given in Tables 1 and 2. The probability that a Jackpot or a small ticket wins a prize is 0.1753. However, the probability of winning a monetary prize is only 0.0238, so the probability of winning one or more new tickets is 0.1515.

[5]This does not guarantee a valid draw. In 1983 one of the balls did enter the machine properly, however, this draw was invalidated.

[6]There are 60 scheduled draws per year.

Table 1: Prizes and winning probabilities on a Jackpot ticket, (s): prize is shared if there are multiple winners.

	(K,B,C)	Amount	Probability π_j^J
J	$(6,0,1)$	≥ €2 million (s)	$2.046e{-}8$
1	$(6,0,0)$	€500 000 (s)	$1.023e{-}7$
2	$(5,1,1)$	€75 000 (s)	$1.227e{-}7$
3	$(5,1,0)$	€50 000 (s)	$6.139e{-}7$
4	$(5,0,1)$	€1000	$4.665e{-}6$
5	$(5,0,0)$	€500	$2.333e{-}5$
6	$(4,1,1)$	€200	$1.1664e{-}5$
7	$(4,1,0)$	€125	$5.832e{-}5$
8	$(4,0,1)$	€20	$2.158e{-}4$
9	$(4,0,0)$	€12	$1.079e{-}3$
10	$(3,1,1)$	€7	$2.877e{-}4$
11	$(3,1,0)$	€5	$1.439e{-}3$
12	$(3,0,1)$	€3.50	$3.452e{-}3$
13	$(3,0,0)$	€2.50	$1.726e{-}2$
14	$(2,1,1)$	2 Jackpot lots	$2.589e{-}3$
15	$(2,1,0)$	2 small lots	$1.295e-2$
16	$(2,0,1)$	1 Jackpot lot	$2.266e-2$
17	$(2,0,0)$	1 small lot	$1.133e-1$

Table 2: Prizes and winning probabilities on a small ticket, (s): prize is shared if there are multiple winners.

	(K,B)	Amount	Probability π_j^s
1	$(6,0)$	€500 000(s)	$1.228e-07$
3	$(5,1)$	€50 000 (s)	$7.366e-07$
5	$(5,0)$	€500	$2.799e-05$
7	$(4,1)$	€125	$6.998e-05$
9	$(4,0)$	€12	$1.295e-03$
11	$(3,1)$	€5	$1.726e-03$
13	$(3,0)$	€2.50	$2.071e-02$
15	$(2,1)$	2 small lots	$1.554e-02$
17	$(2,0)$	1 small lot	$1.359e-01$

To calculate the expected pay-off on a ticket, we need to know the number of participants because that determines the (expected) size of the Jackpot, and the probability that the first, second, and third prize are shared. In a typical draw, 3.1 million tickets participate[7] and we will use this number in the sequel and denote it by N. Dieker and Tijms (2001) estimate the number of participating tickets by maximum likelihood from data on the size of the Jackpot and the number of winners of the first five prizes. Their estimate is 3.2 million tickets, close to the number we use.

Small tickets are hardly ever sold, and because of lack of data, we assume that people only buy Jackpot tickets. The small tickets that participate in a given draw are then obtained as the 15th and 17th prize on a Jackpot ticket. The equilibrium number of small tickets in a draw is then almost 444 thousand, so approximately 2.656 million tickets are Jackpot tickets[8]. We denote these numbers by N_s and N_J respectively, and the total number of tickets is $N_T = N_s + N_J$.

To estimate the return on a Lotto ticket, we first calculate the expected size of the Jackpot. Assuming that the draws are random, the probability that the Jackpot is won is approximately $\psi_J = (1 - \pi_J^J)^N \approx \exp(-N_J \pi_J^J)$ with π_J^J the probability that the Jackpot falls ($\pi_J^J = 2.04623 \cdot 10^{-8}$ from Table 1). This probability ψ_J is only 0.0529, so with probability 0.947 the Jackpot rolls over to the next draw and is increased with e 250 thousand. The number of tickets that win the Jackpot follows a binomial distribution, with parameters N_J and probability of success π_J^J. Because N_J is large and π_J^J is small, this distribution can be approximated by a Poisson distribution with parameter $N_J \pi_J^J$. Using this, the probability that the Jackpot winning ticket is unique is 0.973, so with probability 0.027 the Jackpot has to be shared, given that it falls. Similar calculations for the first, second, and third prize yield Table 3. Note that the first and third prize may be obtained from either a Jackpot ticket or a small ticket. The relevant entries in Table 3 are calculated from

$$\begin{aligned}\Pr(W_i = w) &= \sum_{w^s \le w} \Pr(W_i = w | W_i^s = w^s) \Pr(W_i^s = w^s) \\ &= \sum_{w^s \le w} \Pr(W_i^J = w - w^s) \Pr(W_i^s = w^s),\end{aligned}$$

where W_i denotes the number of winning tickets of the ith prize, and the superscript indicates the type of ticket.

We see in Table 3 that the probabilities that the main prizes have to be shared are non-negligible. For example, the probability that the third prize is shared by three winners is 0.177! This is also reflected in the last row, where we give the expected pay-off on a winning ticket, given that there is at least one winner. Note that he first prize is e 500 thousand, but given that

[7] Private communication with the Lotto.

[8] The exact numbers are 443795 and 2656205 respectively.

Table 3: Distribution of number of winning tickets.

	Jackpot	1st prize	2nd prize	3rd prize
0	$9.471e-01$	$7.216e-01$	$7.217e-01$	$1.412e-01$
1	$5.148e-02$	$2.354e-01$	$2.354e-01$	$2.764e-01$
2	$1.399e-03$	$3.840e-02$	$3.838e-02$	$2.705e-01$
3	$2.534e-05$	$4.176e-03$	$4.172e-03$	$1.765e-01$
4	$3.444e-07$	$3.406e-04$	$3.401e-04$	$8.639e-02$
more than 4	$3.778e-09$	$2.349e-05$	$2.345e-05$	$4.890e-02$
nominal payoff	$2.000e+06$	$5.000e+05$	$7.500e+04$	$5.000e+04$
exp.payoff	$1.973e+06$	$4.600e+05$	$6.900e+04$	$2.922e+04$

a ticket wins this prize, the owner can expect to collect only €460 thousand because there is a possibility that the prize has to be shared.

The number of draws between successive wins of the Jackpot follows a truncated geometric distribution, with parameter ψ_J and truncation at 60 draws. The maximum size of the Jackpot is €17 million, and the probability that the Jackpot increases to this amount is ψ_J^{60}, approximately 0.0383. The average size of the Jackpot is €6.304 million, and the expected number of draws before it is won is 18.2.

To calculate the expected pay-off per ticket, we need to consider two issues: the top four prizes are shared, and some prizes are tickets. Moreover, some tickets are Jackpot tickets, and other tickets are small tickets. Let us consider the first problem first. Suppose for the moment that the Lotto would offer only the first prize (P_1 = €500 thousand), no other prizes, and that there are Jackpot tickets only. We know the distribution of the number of winning tickets for each prize from Table 3. If there are any winning tickets, then the payout per winning ticket is P_1/W_1^J and the probability of winning a prize is the number of winning tickets divided by the total number of tickets, W_1^J/N_J. The expected return on a specific Jackpot ticket is

$$\begin{aligned}\mathcal{E}T^J = P_1 \Bigg(& \Pr(W_1^J = 1)\frac{1}{N_J} + \Pr(W_1^J = 2)\frac{1}{2}\frac{2}{N_J} \\ & + \Pr(W_1^J = 3)\frac{1}{3}\frac{3}{N_J} + \cdots \Bigg) = \frac{P_1}{N_J}(1 - \Pr(W_1^J = 0))\end{aligned} \tag{2}$$

This expression is the expected total payout, divided by the number of tickets. The fact that in practice there are Jackpot tickets and small tickets, does not change anything except that there are more tickets that can win the prize. The expected pay-off on a ticket is

$$\mathcal{E}T = \frac{P_1(1 - \Pr(W_1 = 0))}{N_J + N_s} \tag{3}$$

and is independent of the type of ticket. Now we take all prizes into account, and of course the expected pay-off on a Jackpot ticket is higher than the expected pay-off on a small ticket. By summing over all prizes, we have

$$\mathcal{E}T^J = \frac{P_J(1-\Pr(W_J=0))}{N_J} + \frac{P_1(1-\Pr(W_1=0))}{N_J+N_s} + \frac{P_2(1-\Pr(W_2=0))}{N_J} + \frac{P_3(1-\Pr(W_3=0))}{N_J+N_s} + \sum_{j\geq 4} \pi_j P_j, \tag{4}$$

where π_j is the probability of winning the jth prize (see Table 1). A similar expression holds for the expected value of a small ticket, where we sum over the possible prizes of a small ticket (Table 2). The payoff of the 15th prize is two Jackpot tickets, so $P_{15} = 2\mathcal{T}^{\mathcal{J}}$, and for the 16th prize we have $P_{16} = 2\mathcal{E}T^s$. Hence, equation (4) and its counterpart for a small ticket yield a system with two equations and two unknowns. Solving gives the expected payoffs of a Jackpot ticket and a small ticket. The expected pay-off of a Jackpot ticket is €0.335 and the one of a small ticket is €0.189. The expected return is 49% and 42% of the cost of a ticket.

4 Randomness of the Lotto

The numbers of the Lotto are drawn by a mechanical device. Hence, it is possible that there is some bias in the drawn numbers. This bias may be caused by fluctuations of the weights of the balls, or by the fact that the balls enter the bowl in a fixed order. In this section we will test whether the numbers and color are drawn randomly. We also discuss some implications of random draws that seem to be not so random.

Clearly, it is not feasible to assess whether the numbers and color of the Lotto are drawn randomly by a direct test. There are 49 million equally likely possible draws, and checking independence between draws and randomness within draws would require an unrealistically high number of draws[9]. Note the difference between independence and randomness here: draws are supposed to be mutually independent, but within a draw, the probability distribution of, say, the second number is not independent of the first number drawn.

Following Haigh (1997), we test some implications of randomness within and independence between draws. Because the machine from which the numbered balls are drawn is physically separate from the machine from which the colored balls are drawn, we assume that these events are independent so we study both draws separately.

We test the following hypotheses:

[9]Our data pertain to actual draws only. Of course, one could observe a large number of draws in a laboratory setting but even then we could generate only 2880 draws per 24 hours, assuming that it is possible to generate a draw every 30 seconds (it takes time to mix the balls).

- Does the marginal frequency distribution of numbers drawn correspond to the one expected under randomness (Subsection 4.1), and is the distribution of numbers drawn uniform over the columns on the form (the bottom part of Figure 1)?
- Do the mean and variance of the sum of the numbers, the distribution of odd and even numbers, and the distribution of the minimum and maximum correspond to those expected under randomness (Subsection 4.2)?
- Does the distribution of waiting times (the number of complete draws before a number is drawn again) correspond to the one expected under randomness, and does the Lotto have a memory (Subsection 4.3)?
- Are the colors drawn randomly from the six possible colors (Subsection 4.4)?

In the remainder of this section, we use data on draws obtained from `www.dutch-lotto.com`. We distinguish between two time periods: 1974–1996, when six balls were drawn from 41 balls, and from 1996 onwards, when six balls are drawn from 45 balls. During both periods, a bonus ball was drawn as well. The draw of the colored ball was introduced in 2000, so the test of subsection 4.4 pertains to that period only. The total number of draws in the 1974–1996 period is 1254, and in the 1996–2002 period we have 372 draws, of which the last 107 draws have a colored ball. We will use M to denote the total number of balls (41 or 45), m the number of balls drawns (including the bonus ball, so m is 7) and N the total number of draws (1254, 372, or 107).

4.1 Randomness of the numbers

Are the numbers drawn in a truly random way in the sense that the probability that each number is drawn is 1 over the number of balls in the bowl? We test this hypothesis using a variant of Pearson's goodness-of-fit statistic. This statistic in its usual form is not applicable here, though. As has been pointed out by Joe (1993), the observations on the frequencies of all numbers are not independent. Even though one may assume independence between draws, there is dependence within draws as the balls are drawn without replacement. Joe (1993) shows that the appropriate test statistic is proportional to Pearson's goodness-of-fit statistic:

$$Q_1 = \frac{M-1}{M-m} \sum_{i=1}^{M} \frac{(F_i - E_i)^2}{E_i}, \tag{5}$$

where E_i is the expected frequency of number i (so $E_i = N\frac{m}{M}$), and F_i is the observed frequency. If the number of draws N is large enough, Q_1 follows a χ^2_{M-1} distribution. The correction factor is smaller than 1, so the usual statistic would underestimate the p-value for the null hypothesis of randomness and reject the null hypothesis too often.

The frequency distributions of the numbers drawn are listed in Table 6 in the Appendix. The frequencies of numbers from the first period vary between 187 and 258, and those of the last period vary between 42 and 71. The expected frequencies are 214.1 and 57.9 respectively. The test statistic (5) is 47.54 (p-value is 0.19) for the 1974-1996 period, and 37.84 for the 1996-2002 period (p-value is 0.73). Hence, these tests do not cast any doubts on the hypothesis of randomness.

A second question is whether the numbers are distributed randomly over the columns on the form, see the bottom part of Figure 1. Let the distribution of numbers in a certain draw over k columns be given by the k-vector C. For example, if C_1 is 2, it means that two of the numbers drawn are from column 1. The probability distribution of the vector C is a multivariate hypergeometric distribution:

$$\Pr(C_1 = c_1, \ldots, C_k = c_k) = \frac{\binom{N_1}{c_1} \cdots \binom{N_k}{c_k}}{\binom{M}{m}}, \tag{6}$$

where N_j are the number of numbers listed in column j. In our case, $N_1 = \ldots = N_5 = 9$. Of course, $\sum c_k = m$, so the elements of the random vector C are dependent. Using this formula, we calculate the expectation μ_C of C (which is, unsurprisingly, $m/k = 1.2$) and the variance matrix of C. This latter matrix is denoted by Σ_C. The elements of Σ_C are given by (Johnson, Kotz, and Balakrishnan (1997)):

$$\begin{aligned} \mathrm{var} C_i &= \frac{m(M-m)}{M^2(M-1)} N_i(M - N_i) \\ \mathrm{cov}(C_i, C_j) &= -\frac{m(M-m)}{M^2(M-1)} N_i N_j \end{aligned}$$

Then we have

$$n(\bar{x}_C - \mu_C)' \Sigma_C^- (\bar{x}_C - \mu_C) \sim \chi^2_{k-1}.$$

In this test statistic, a generalized inverse of Σ_C appears because of the linear dependence of the elements of C ($\sum c_k = m$). Alternatively, the test statistic could be based on any $k - 1$ elements of C (and the corresponding versions of μ_C and Σ_C).

Using the data from 1996 onwards, we find mean frequency vector $\bar{x}_C = (1.169 \quad 1.180 \quad 1.196 \quad 1.218 \quad 1.237)'$. The test statistic is 1.049 and the corresponding p-value is 0.90. There is no evidence against uniformity over the columns. There is no information on the columns in the 1974–1996 data set, so the test cannot be performed for that period.

Table 4: Sample statistics per draw.

	μ_D	$\bar{x}_D$	p-value	σ_D^2	s_D^2	p-value
1974–1996	147	145.8	0.517	833	856.4	0.293
1996–2002	161	159.2	0.523	1020	972.2	0.732

Table 5: Frequency distribution odd numbers within each draw

	obs. 1974-1996	exp. 1974-1996	obs. 1996-2002	exp. 1996-2002
0	5	4.46	1	1.40
1	50	46.85	10	14.07
2	178	187.40	46	54.62
3	356	370.89	122	106.20
4	386	392.71	92	111.79
5	220	222.53	87	63.72
6	54	62.47	14	18.21
7	5	6.69	0	2.01

4.2 Draw Statistics

In every draw, we can calculate a number of statistics, and compare these to the ones we expect under the hypothesis of random draws. We look at four of such quantities derived from all draws: the mean and the variance, and the distributions of the minimum and the maximum.

In an m/M lottery, the sum of the numbers drawn has expectation $\mu_D = m(M+1)/2$ and variance $\sigma_D^2 = m(M+1)(M-m)/12$ under the assumption of randomness. Hence, the sample mean of the observed sum over all draws follows (asymptotically) a normal distribution with mean μ_D and variance σ_D^2, and the scaled sample variance of the sums $(n-1)s_D^2/\sigma_D^2$ follows asymptotically a $\chi^2(n-1)$ distribution. The numerical results are listed in Table 4. In that table we see that the sample means and variances correspond well with their theoretical counterparts, which conclusion is also supported by the p-values.

The distribution of odd numbers is given in Table 5. The number of odd numbers within a draw follows a hypergeometric distribution, with parameters 21, 20, and 7 (1974–1996) and 23, 22, and 7 (1996–2002). The observed frequencies of odd numbers and that distribution are well in agreement for the 1974–1996 period (the p-value of the χ^2 goodness of fit is 0.879. In the second period, however, there seems to be some discrepancy between the empirical and theoretical distributions. Especially five odd numbers are drawn more frequently than expected. The p-value is 0.00569, giving some evidence that the distribution of the number of odd numbers is not hypergeometric.

We also calculate the minimum and maximum of each draw. The probability functions of the minimum and maximum are given by

$$\Pr(\min(B_1,\ldots,B_m) = b) = \frac{m}{M}\frac{\binom{M-b}{m-1}}{\binom{M}{m}}, \qquad b \le M+1-m \tag{7}$$

$$\Pr(\max(B_1,\ldots,B_m) = b) = \frac{1}{M}\frac{\binom{b-1}{m-1}}{\binom{M}{m}}, \qquad b \ge m \tag{8}$$

We compare the fit of these distributions to the empirical frequencies which are listed Tables 8 and 7 in the Appendix. The p-values for the distributions (7) and (8) are 0.639 and 0.532 respectively for the 1974–1996 period. For the 1996–2002 period we find 0.875 and 0.674. We grouped observations such that the expected frequency in each cell was at least 5. The p-values indicate that these tests do not cast any doubts on the hypothesis of randomness either.

4.3 Memory of the Lotto

Do numbers that have come up in a certain draw, return in the next draw? It seems that this is the case indeed, if one looks at the 107 draws in the 2000-2002 data set. In 79 of the 107 draws (74.5%) at least one number of the previous draw was drawn again. In fact, in 17 cases (16.2%) there were numbers recurring from the previous two draws.

The probability of the first event, though, is simply calculated from the distribution of the number of recurring numbers in the second draw, K_2:

$$\Pr(K_2 = k_2) = \frac{\binom{38}{7-k_2}\binom{7}{k_2}}{\binom{45}{7}}, \qquad k_3 = 0, 1, \ldots, k_2 \tag{9}$$

from which we have $\Pr(K_2 \ge 1) \approx 0.722$. This differs hardly from the empirical frequency. After the second draw, there are, conditionally on $K_2 = k_2$, $45 - k_2$ balls that have been drawn at most once, and k_2 balls that have been drawn twice. The conditional probability distribution of K_3, the number of balls that appear in three consecutive draws, is now

$$\Pr(K_3 = k_3 | K_2 = k_2) = \frac{\binom{45-k_2}{7-k_3}\binom{k_2}{k_3}}{\binom{45}{7}}, \tag{10}$$

and the marginal distribution of K_3 is from (9) and (10)

$$\Pr(K_3 = k_3) = \sum_{k_2=0}^{7} \Pr(K_3 = k_3 | K_2 = k_2) \Pr(K_2 = k_2). \tag{11}$$

The probability of at least one number occurring three times in a row is 0.160, again very close to the observed frequency.

Another interesting question is how long it takes before a Jackpot winning draw comes up again. Considering the small probability of a certain draw winning the Jackpot (see Table 1), it is expected that this is a long time. In total, there are $L = 6 \times \binom{45}{6}$ different draws that yield the Jackpot. Suppose we draw K times, what is the probability that all K draws are different? Let the L possible draws be labelled as $D_1, \ldots, D_L$, then the probability distribution of the draws within the K draws is multinomial:

$$\begin{aligned}
\Pr(D_1 = d_1, \ldots, D_L = d_L) &= \frac{K!}{d_1! \cdots d_l!} \pi_1^{d_1} \cdots \pi_L^{d_L} \\
&= \frac{K!}{d_1! \cdots d_L!} \pi^K,
\end{aligned} \tag{12}$$

because all draws are equally likely ($\pi = 1/L$) and $\sum d_i = K$. The probability that all K draws are different is can be calculated as $\Pr(\text{no matches}) = \Pr(D_1 \le 1, \ldots, D_L \le 1)$, that is, every possible draw occurs at most once (and probably not at all, if K is much smaller than L). From (12) we can approximate this probability as

$$\begin{aligned}
&\Pr(D_1 \le 1, \ldots, D_L \le 1) \\
&= \sum_{d_1 \le 1, \ldots, d_L \le 1, d_1 + \cdots + d_L = K} \frac{K!}{d_1! \cdots d_L!} \pi^K = \sum_{d_1 + \cdots + d_L = K} K! \pi^K \\
&= \binom{L}{K} K! \pi^K = \frac{L!}{(L-K)!} \frac{1}{L^K} \\
&= \left(1 - \frac{1}{L}\right)\left(1 - \frac{2}{L}\right) \cdots \left(1 - \frac{K+1}{L}\right).
\end{aligned} \tag{13}$$

We can approximate the logarithm of $\Pr(D_1 \le 1, \ldots, D_L \le 1)$ by

$$\sum_{i=1}^{K+1} \ln\left(1 - \frac{i}{L}\right) \approx -\sum_{i=1}^{K+1} \frac{i}{L} + o(\frac{1}{L}) \approx -\frac{1}{2L}(K+1)(K+2).$$

The approximate the number of draws necessary to ensure that the probability of at least two identical draws is $\frac{1}{2}$, is now obtained by equating this expression with $\ln \frac{1}{2}$. Neglecting the linear term, we obtain

$$K \approx \sqrt{2L \ln 2}. \tag{14}$$

In our case, this boils down to $K \approx \sqrt{2 \times 6 \times \binom{45}{6} \times \ln 2} \approx 8231$, so in 8231 draws, there is an even chance that at least two Jackpot winning draws are identical.

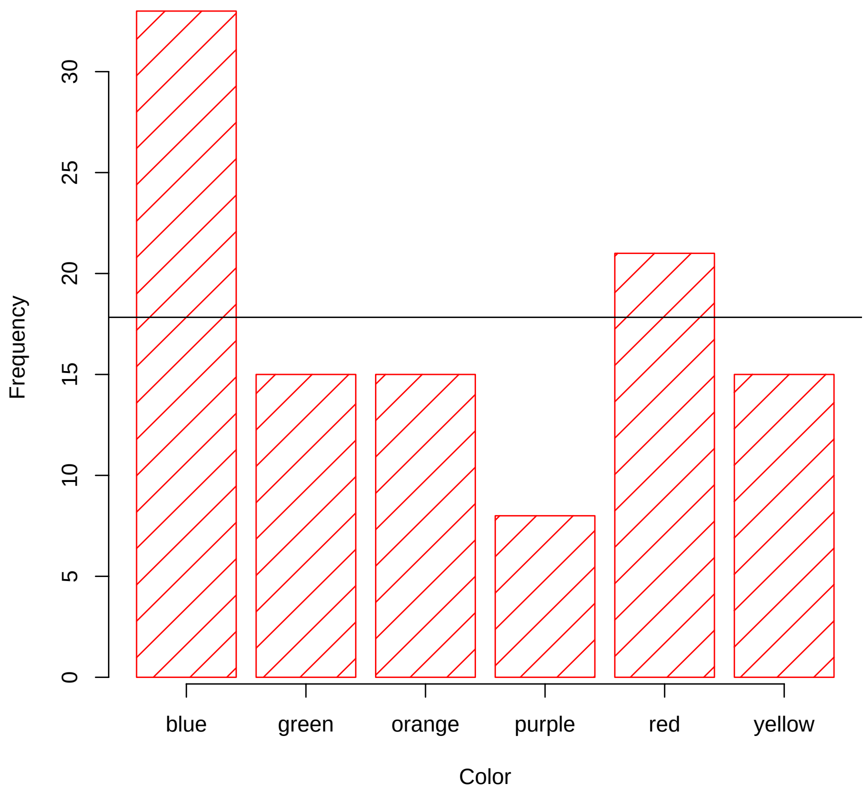

Figure 2: The frequency distribution of the colors.

4.4 Randomness of the colors

In Figure 2 the frequency distribution of the colors is given graphically, and the expected frequency is indicated by the horizontal line. The actual frequencies are again to be found in Appendix A (see Table 9).

Clearly, blue has been drawn most frequently (33 times), and purple has been drawn only eight times. The p-value of the standard Pearson test of the hypothesis that the colors are drawn randomly is 0.00113, far below any reasonable level of significance. Hence, the hypothesis that the colors are drawn randomly, is rejected.

This result of finding evidence *against* randomness is intriguing. Johnson and Klotz (1993) find evidence against randomness in the Lotto America Megabucks Lottery. They explain this by the mechanical mixing device:

> We conjecture that the moderate evidence for nonuniformity is a result of the mechanical mixing process, in which balls enter the urn in sequence (always the same), are mixed, and then are drawn out at the bottom.

However, our evidence against nonuniformity is stronger than theirs. If the quoted reason is the cause of nonuniformity, it may be even more applicable in this case of colored balls since only one ball is drawn, and not multiple balls, some of which have been exposed longer to a mixing process. Clearly, the mixing process could be improved by having the balls enter the bowl in a more random way.

5 Conclusion

In this paper we have calculated the probabilities of winning prizes in the Dutch Lotto, and the expected return on a lottery ticket. The return on a ticket is approximately 49% of the cost of such a ticket. The effect of the possibility of sharing the Jackpot, first, second, or third prize turned out to be non-negligible. A third-prize winner can expect to € 29 thousand, instead of the nominal prize of € 50 thousand. We also calculated the probability distribution of the size of the Jackpot. The probability that the Jackpot increases to its maximum amount (€ 17 million), is 0.0383.

In the second half of the paper we examined whether the numbers are drawn randomly, and—starting from 2000—whether the colored ball is drawn randomly. We found a little bias in the numbers that have been drawn: in the 1996–2002 period the number of odd balls do not conform to the hypergeometric distribution. We also found that the hypothesis that each of the six colors has an equal probability of being drawn, has to be rejected. The color blue was drawn significantly more often than could be expected under uniformity.

All calculations are based on the assumption that the number of participants does not vary with the size of the Jackpot. In future research, we would like to estimate the elasticity of participation with respect to the size of the Jackpot, or the expected return on a ticket.

References

Dieker, E. and H. Tijms (2001). How to estimate the pay-off percentage of the new Dutch Lotto. *STAtOR 2*(1), 9–14. In Dutch.

Haigh, J. (1997). The statistics of the national lottery. *Journal of the Royal Statistical Society A 160*(2), 187–206.

Haigh, J. (1999). *Taking Chances.* Oxford: Oxford University Press.

Joe, H. (1993). Tests of uniformity for sets of lotto numbers. *Statistics & Probability Letters 16*(1), 181–188.

Johnson, N.L., S. Kotz, and N. Balakrishnan (1997). *Discrete Multivariate Distributions.* New York: John Wiley & Sons.

Johnson, R. and J. Klotz (1993). Estimating hot numbers and testing uniformity for the lotto. *Journal of the American Statistical Association 88*(422), 662–668.

A Frequencies

Table 6: Frequency distribution of numbers drawn, with expected fequencies 214.1 (1974–1996) and 57.9 (1996–2002).

Number	1974–1996	1996–2002	Number	1974–1996	1996–2002
1	217	71	24	217	60
2	242	60	25	210	53
3	225	50	26	191	67
4	200	61	27	217	56
5	214	59	28	217	57
6	220	52	29	210	58
7	202	68	30	228	56
8	212	60	31	204	52
9	216	54	32	222	55
10	258	67	33	216	59
11	209	55	34	210	62
12	215	51	35	226	68
13	221	52	36	196	42
14	218	57	37	205	52
15	204	69	38	187	64
16	222	53	39	211	67
17	193	62	40	210	50
18	229	52	41	228	50
19	217	66	42		61
20	210	56	43		57
21	222	54	44		45
22	197	61	45		62
23	210	61			

Table 7: Frequency distribution of the minimum within each draw.

Number	1974-1996	1996-2002	Number	1974-1996	1996-2002
1	217	71	14	22	6
2	205	53	15	10	5
3	158	35	16	7	5
4	121	39	17	9	2
5	107	28	18	5	1
6	97	24	19	3	2
7	64	22	20	3	1
8	49	15	21	3	1
9	52	14	22	1	2
10	49	19	23	0	0
11	29	12	24	0	1
12	28	9	25	1	0
13	14	5	26	0	0

Table 8: Frequency distribution of the maxmimum within each draw.

Number	1974-1996	1996-2002	Number	1974-1996	1996-2002
15	1	0	31	28	2
16	0	0	32	40	7
17	1	0	33	51	9
18	3	0	34	62	8
19	0	0	35	77	12
20	0	0	36	79	13
21	3	0	37	117	17
22	3	0	38	114	20
23	4	1	39	151	29
24	7	0	40	176	22
25	10	2	41	228	27
26	15	4	42		46
27	19	2	43		41
28	17	2	44		39
29	28	6	45		62
30	20	1			

Table 9: Distribution of colors.

	Frequency	Expected freq.
blue	33	17.83
green	15	17.83
orange	15	17.83
purple	8	17.83
red	21	17.83
yellow	15	17.83