A simple card guessing game revisited

Arnold Knopfmacher

The John Knopfmacher Centre for Applicable Analysis and Number Theory
Department of Applied Mathematics
University of the Witwatersrand, P. O. Wits
2050 Johannesburg, South Africa
email: arnoldk@gauss.cam.wits.ac.za
homepage: http://www.wits.ac.za/science/number_theory/arnold.htm

Helmut Prodinger

The John Knopfmacher Centre for Applicable Analysis and Number Theory
Department of Mathematics
University of the Witwatersrand, P. O. Wits
2050 Johannesburg, South Africa
email: helmut@gauss.cam.wits.ac.za
homepage: http://www.wits.ac.za/helmut/index.htm

Submitted: January 4, 2000; Accepted: March 3, 2000.

Abstract

A deck of cards consisting of m red and n black cards is given. A guess is made as to the colour of the top card, after which it is revealed and discarded. To maximise the number of correct guesses one chooses the colour corresponding to the majority of cards remaining in the deck. We determine the expected number of correct guesses with this strategy, as well as the probability of beating an opponent who uses the naive stategy of random guessing.

AMS Subject Classification. 05A15 (primary) 60C05 (secondary)

1 Introduction

A number of authors have considered the following simple combinatorial game; [5, 3, 7, 6]. One starts with a deck of cards consisting of m red and n black cards. A guess is made as to the colour of the top card, after which it is revealed and discarded. To maximise the number of correct guesses one chooses the colour corresponding to the majority of cards remaining in the deck. We determine the expected number of correct guesses with this strategy, as well as the probability of beating an opponent[1] who uses the naive stategy of random guessing. The first of these problems was analysed by Levasseur [3] but only for the case $m = n$. The second question was investigated by Zagier [7], again for the case $m = n$. He obtained an asymptotic answer only, whereas we can derive exact formulas for this probability for general m and n. We also re–derive the probability distribution for the number of correct guesses for a pack of $m + n$ cards, found originally by Sulanke [6].

Whereas the previous authors made use of lattice paths, our approach will be based on recurrence relations and probability generating functions. We also consider a variation of the game in which you and your opponent alternate in guessing the colour of the card. We find that there is a slight advantage to playing second in this variation.

In what follows, m always refers to number of cards of the colour that is predominant, n to the other colour.

2 The number of correctly guessed cards

We begin by re–proving some results of [6].

Let $p(m, n; k)$ denote the probability that, assuming that one has m cards of one colour and n cards of a second colour, with $m \geq n \geq 0$, that one guesses k cards correctly. Sulanke [6], using lattice path arguments, found the following explicit expression:

$$p(m,n;k) = \frac{1}{\binom{m+n}{m}}\left[\binom{m+n}{k} - \binom{m+n}{k+1}\right] \qquad \text{for } m \leq k \leq m+n. \tag{1}$$

One can prove this also using the following recursions. Introducing probability generating functions

$$\varphi_{m,n}(u) = \sum_{m \leq k \leq m+n} p(m,n;k)u^k,$$

one sees the recursions

$$\varphi_{m,n}(u) = u\frac{m}{m+n}\varphi_{m-1,n}(u) + \frac{n}{m+n}\varphi_{m,n-1}(u) \qquad \text{for } m > n \geq 0,$$
$$\varphi_{m,m}(u) = (1+u)\frac{1}{2}\varphi_{m,m-1}(u) \qquad \text{for } m \geq 1,\ \varphi_{0,0}(u) = 1$$

[1]Previous authors [3, 7] used phrases like "How often should you beat your kids?" etc. In accordance with the editor's wishes it was decided to avoid such phrases.

almost immediately. If one defines $\Phi_{m,n}(u) = \binom{m+n}{m}\varphi_{m,n}(u)$ instead, the recursions are nicer, viz.

$$\begin{aligned}\Phi_{m,n}(u) &= u\Phi_{m-1,n}(u) + \Phi_{m,n-1}(u) \qquad \text{for } m > n \geq 0,\\ \Phi_{m,m}(u) &= (1+u)\Phi_{m,m-1}(u) \qquad \text{for } m \geq 1,\ \Phi_{0,0}(u) = 1.\end{aligned}$$

From these, Sulanke's result (1) is readily proved by induction.

Instead of either guessing or using lattice path arguments, we can use a generating function method that deserves perhaps to be known more widely to deal with such recursions with boundary conditions. It is due to Knuth [2, page 537]; a tutorial version of it appears in [4].

If we set

$$\begin{aligned}F(z,x) &= \sum_{m\geq n\geq 0} \Phi_{m,n}(u) z^m x^{m-n},\\ F_0(z) &= \sum_{m\geq 0} \Phi_{m,m}(u) z^m,\\ F_1(z) &= \sum_{m\geq 1} \Phi_{m,m-1}(u) z^m,\end{aligned}$$

then the first and second recursion for $\Phi_{m,n}(u)$ give

$$F(z,x) - F_0(z) = uzx\, F(z,x) + \frac{1}{x}\big[F(z,x) - F_0(z) - F_1(z)x\big]$$

and $F_0(z) = 1 + (1+u)F_1(z)$, respectively. Hence

$$(1 - x + uzx^2)F(z,x) = 1 - x + (1 + u - ux)F_1(z).$$

Now the fact that the power series $F(z,x)$ remains finite at $x = \lambda := 2/(1+\sqrt{1-4uz}\,)$ gives (this is Knuth's trick) $F_1(z) = (\lambda - 1)/(1 + u - \lambda u)$ and finally

$$F(z,x) = \left(\frac{1+\sqrt{1-4uz}}{2} - \frac{1-\sqrt{1-4uz}}{2}u\right)^{-1}\left(1 - \frac{1-\sqrt{1-4uz}}{2}x\right)^{-1}.$$

From this, using the expansion of the geometric series and the identity [1, (5.61)]

$$\sum_{n\geq 0}\binom{2n+k}{n}X^n = \frac{2^k}{\sqrt{1-4X}(1+\sqrt{1-4X}\,)^k}$$

the quantities in (1) can be obtained.

Considering the expected values $T(m,n) = \frac{d}{du}\varphi_{m,n}(u)\big|_{u=1}$, simple telescoping leads to

$$\binom{m+n}{m}T(m,n) = \sum_{k\geq m} k\left[\binom{m+n}{k} - \binom{m+n}{k+1}\right] = m\binom{m+n}{m} + \sum_{k>m}\binom{m+n}{k}.$$

Theorem 1. *The expected number of correct guesses is given by*

$$T(m,n) = m + \frac{1}{\binom{m+n}{m}} \sum_{k=0}^{n-1} \binom{m+n}{k}. \tag{2}$$

Remark. Although this seems to be the nicest expression for the mean, we are able to derive other equivalent ones:

$$T(m,n) = m - 1 + \frac{2^n}{\binom{m+n}{m}} \sum_{k=0}^{n} \binom{m+k-1}{k} \frac{1}{2^k}, \tag{3}$$

$$T(m,n) = m + \frac{2^{n-1}}{\binom{m+n}{m}} \sum_{k=0}^{n-1} \binom{m+k}{k} \frac{1}{2^k}, \tag{4}$$

$$T(m,n) = m + n \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{2^{n-1-k}(-1)^k}{m+1+k}. \tag{5}$$

Since the equivalence of these binomial expressions is not obvious, we prove these equalities. To prove that (2) is equivalent to (3), we must show that

$$\sum_{k=0}^{n} \binom{m+n}{k} = 2^n \sum_{k=0}^{n} \binom{m+k-1}{k} 2^{-k}.$$

In terms of generating functions, we must prove that

$$[x^n] \frac{1}{1-x} (1+x)^{m+n} = [x^n] \frac{1}{1-2x} \frac{1}{(1-x)^m}.$$

This can be seen by Cauchy's integral formula (the substitution $x = t/(1-t)$ is used):

$$\begin{aligned} [x^n] \frac{1}{1-x} (1+x)^{m+n} &= \frac{1}{2\pi i} \oint \frac{dx}{x^{n+1}} \frac{1}{1-x} (1+x)^{m+n} \\ &= \frac{1}{2\pi i} \oint \frac{dt}{t^{n+1}} \frac{1}{1-2t} \frac{1}{(1-t)^m} \\ &= [t^n] \frac{1}{1-2t} \frac{1}{(1-t)^m}. \end{aligned}$$

The equivalence of (3) and (4) is straightforward by writing

$$\binom{m+k}{k} = \binom{m+k-1}{k} + \binom{m+k-1}{k-1}$$

and performing some elementary operations.

For the equivalence of (5) and (4) we use hypergeometric functions:

$$\begin{aligned} n\sum_{k=0}^{n-1}\binom{n-1}{k}\frac{2^{n-1-k}(-1)^k}{m+1+k} &= \frac{n2^{n-1}}{(m+1)}\ {}_2F_1[-n+1,m+1;m+2|\tfrac{1}{2}] \\ &= \frac{n2^{n-1}(n-1)!(m+1)!}{(m+1)(m+n)!}\ {}_2F_1[-n+1,m+1;-n+1|\tfrac{1}{2}] \\ &= \frac{2^{n-1}}{\binom{m+n}{m}}\sum_{k=0}^{n-1}\binom{m+k}{k}\frac{1}{2^k}. \end{aligned}$$

Here, we have used the transformation (5.105) [1].

Remark. As one referee has pointed out, there are shorter proofs of these equivalences. The ones presented above have the advantage of offering some insight as opposed to providing merely mechanical verifications.

It is interesting to note that the formula

$$\sum_{k=0}^{m}\binom{m+k}{k}\frac{1}{2^k}=2^m$$

follows from the discussion of Banach's match box problem, see [1].

It is also not too hard to compute the variance; we start with the second moment:

$$\begin{aligned} &\frac{1}{\binom{m+n}{m}}\sum_{k\ge m}k^2\left[\binom{m+n}{k}-\binom{m+n}{k+1}\right] \\ &= m^2+\frac{1}{\binom{m+n}{m}}\sum_{k>m}(2k-1)\binom{m+n}{k} \\ &= m^2+\frac{2(m+n)}{\binom{m+n}{m}}\sum_{k>m}\binom{m+n-1}{k-1}-\frac{1}{\binom{m+n}{m}}\sum_{k>m}\binom{m+n}{k} \\ &= m^2+\frac{m+n}{\binom{m+n}{m}}\sum_{k>m}\binom{m+n-1}{k-1}+\frac{m+n}{\binom{m+n}{m}}\sum_{k\ge m}\binom{m+n-1}{k}-\frac{1}{\binom{m+n}{m}}\sum_{k>m}\binom{m+n}{k} \\ &= m^2+n+\frac{m+n}{\binom{m+n}{m}}\sum_{k>m}\binom{m+n}{k}-\frac{1}{\binom{m+n}{m}}\sum_{k>m}\binom{m+n}{k} \\ &= m^2+n+\frac{m+n-1}{\binom{m+n}{m}}\sum_{k=0}^{n-1}\binom{m+n}{k}. \end{aligned}$$

Subtracting the square of the expectation, we get:

Theorem 2. *The variance of the number of correct guesses is given by*

$$n-\frac{m+1-n}{\binom{m+n}{m}}\sum_{k=0}^{n-1}\binom{m+n}{k}-\frac{1}{\binom{m+n}{m}^2}\left(\sum_{k=0}^{n-1}\binom{m+n}{k}\right)^2. \quad \square$$

In the symmetric case $m = n$, this can be simplified to

$$m + \frac{1}{4} - \frac{4^{2m-1}}{\binom{2m}{m}^2},$$

in agreement with the earlier result of Sulanke [6].

3 How often will you beat a naive player?

Zagier [7] considered the probability that you beat the naive player (with your superior strategy) and gave the asymptotic answer

$$\frac{1}{2} + \frac{1}{\sqrt{8}} + \mathcal{O}(m^{-1/2})$$

for the symmetric case $m = n$.

Now we can give an exact result: Recall that the probability that you score k by guessing the more likely outcome is given by

$$A(m,k) = \frac{1}{\binom{2m}{m}} \left[\binom{2m}{k} - \binom{2m}{k+1} \right], \qquad m \le k \le 2m.$$

The distribution for the naive player is given by

$$C(m,k) = \frac{1}{4^m} \binom{2m}{k}, \qquad 0 \le k \le 2m.$$

Therefore the probability that you win is

$$\sum_{0 \le k < m} C(m,k) + \sum_{m \le k < j \le 2m} C(m,k) A(m,j).$$

The first, easy sum is

$$\sum_{0 \le k < m} C(m,k) = \frac{1}{2} \left(1 - \frac{\binom{2m}{m}}{4^m} \right).$$

The other sum is

$$\begin{aligned}
\frac{1}{\binom{2m}{m} 4^m} & \sum_{m \le k < j \le 2m} \binom{2m}{k} \left[\binom{2m}{j} - \binom{2m}{j+1} \right] \\
&= \frac{1}{\binom{2m}{m} 4^m} \sum_{m \le k < 2m} \binom{2m}{k} \binom{2m}{k+1} \\
&= \frac{1}{2\binom{2m}{m} 4^m} \sum_{0 \le k < 2m} \binom{2m}{k} \binom{2m}{k+1}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\binom{2m}{m}4^m} \sum_{0\le k<2m} \binom{2m}{k}\binom{2m}{2m-k-1} \\
&= \frac{1}{2\binom{2m}{m}4^m} \binom{4m}{2m-1}.
\end{aligned}
$$

Theorem 3. *The probability that the player guessing the more likely colour beats the one guessing naively is (exactly) given by*

$$
\frac{1}{2} + \frac{1}{2}\frac{\binom{4m}{2m+1}}{\binom{2m}{m}4^m} - \frac{1}{2}\frac{\binom{2m}{m}}{4^m}.
$$

Stirling's formula readily gives the result

$$
\frac{1}{2} + \frac{1}{\sqrt{8}} - \frac{1}{2\sqrt{\pi m}} - \frac{7\sqrt{2}}{64m} + \frac{1}{16\sqrt{\pi}m^{3/2}} + \mathcal{O}(m^{-2}).
$$

Very similarly, we can consider the probability that the game against the naive player is a draw, which turns out to be

$$
\begin{aligned}
\sum_{m\le k\le 2m} C(m,k)A(m,k) &= \frac{1}{\binom{2m}{m}4^m} \sum_{m\le k\le 2m} \binom{2m}{k}\left[\binom{2m}{k} - \binom{2m}{k+1}\right] \\
&= \frac{1}{2\binom{2m}{m}4^m}\left[\sum_{0\le k\le 2m}\binom{2m}{k}^2 + \binom{2m}{m}^2\right] - \frac{1}{2\binom{2m}{m}4^m}\binom{4m}{2m+1} \\
&= \frac{1}{2\binom{2m}{m}4^m}\binom{4m}{2m} + \frac{1}{2\,4^m}\binom{2m}{m} - \frac{1}{2\binom{2m}{m}4^m}\binom{4m}{2m+1} \\
&= \frac{1}{2(2m+1)\binom{2m}{m}4^m}\binom{4m}{2m} + \frac{1}{2\,4^m}\binom{2m}{m} \\
&= \frac{1}{2\sqrt{\pi m}} + \frac{\sqrt{2}}{8m} + \mathcal{O}(m^{-3/2}).
\end{aligned}
$$

4 A variation

Instead of guessing at the same time as the naive player, suppose that the game is now played alternately. The naive player guesses, then the clever player, then the naive player again, and so on. We will see that is makes a (small) difference whether the clever or the naive player starts. We will discuss again the average score of the clever player and start by considering the case where the clever player plays first.

$$
T(m,n) = [\![m+n \text{ even}]\!]\frac{m}{m+n} + \frac{m}{m+n}T(m-1,n) + \frac{n}{m+n}T(m,n-1),\ \ 0\le n<m,
$$

$$T(m,m) = \frac{1}{2} + 2T(m,m-1), \qquad m \geq 1, \quad T(0,0) = 0.$$

($[\![P]\!]$ is 1 if condition P is true, 0 otherwise.) It is quite easy to guess the solution and prove it by induction:

$$\begin{aligned}
T(2m,2n) &= m + \frac{1}{\binom{2m+2n}{2n}} \sum_{k=0}^{n-1} \binom{2m+2n}{2k},\\
T(2m,2n+1) &= m + \frac{1}{\binom{2m+2n+1}{2n+1}} \sum_{k=0}^{n-1} \binom{2m+2n+1}{2k+1} - \frac{m}{2m+2n+1},\\
T(2m+1,2n) &= m + \frac{1}{\binom{2m+2n+1}{2n}} \sum_{k=0}^{n-1} \binom{2m+2n+1}{2k+1} - \frac{n}{2m+2n+1},\\
T(2m+1,2n+1) &= m + \frac{1}{\binom{2m+2n+2}{2n+1}} \sum_{k=0}^{n-1} \binom{2m+2n+2}{2k+1} + \frac{1}{2}.
\end{aligned}$$

Therefore

$$T(m,m) = \frac{m-1}{2} + \frac{4^{m-1}}{\binom{2m}{m}}.$$

Now the case when the naive player starts is briefly sketched.

$$\begin{aligned}
&T(m,n) = [\![m+n \text{ odd}]\!]\frac{m}{m+n} + \frac{m}{m+n}T(m-1,n) + \frac{n}{m+n}T(m,n-1), \;\; 0 \leq n < m,\\
&T(m,m) = 2T(m,m-1), \qquad m \geq 1, \quad T(0,0) = 0.
\end{aligned}$$

And the solution is

$$\begin{aligned}
T(2m,2n) &= m + \frac{1}{\binom{2m+2n}{2n}} \sum_{k=0}^{n-1} \binom{2m+2n}{2k+1},\\
T(2m,2n+1) &= m + \frac{1}{2\binom{2m+2n+1}{2n+1}} \sum_{k=0}^{n} \binom{2m+2n+2}{2k+1},\\
T(2m+1,2n) &= m + \frac{1}{\binom{2m+2n+1}{2n}} \sum_{k=0}^{n} \binom{2m+2n+1}{2k} + \frac{n}{2m+2n+1},\\
T(2m+1,2n+1) &= m + \frac{1}{\binom{2m+2n+2}{2n+1}} \sum_{k=0}^{n-1} \binom{2m+2n+2}{2k+1} + \frac{m+3n+2}{2(m+n+1)}.
\end{aligned}$$

Therefore

$$T(m,m) = \frac{m}{2} + \frac{4^{m-1}}{\binom{2m}{m}}.$$

So one does slightly better (by $\frac{1}{2}$) when one is allowed to play after the naive player. And the sum of the scores in both instances is

$$T(m,m) = m - \frac{1}{2} + \frac{4^m}{2\binom{2m}{m}},$$

as it should be, since the score is only divided into the score obtained at even (resp. odd) steps.

Acknowledgment. The insightful comments of two anonymous referees are gratefully acknowledged.

References

[1] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics (Second Edition).* Addison Wesley, 1994.

[2] D. E. Knuth. *The Art of Computer Programming*, volume 1: Fundamental Algorithms. Addison-Wesley, 1968. Third edition, 1997.

[3] K. Levasseur. How to beat your kids at their own game. *Mathematical Magazine*, 61:301–305, 1988.

[4] A. Panholzer and H. Prodinger. Towards a more precise analysis of an algorithm to generate binary trees : A tutorial. *The Computer Journal*, 41:201–204, 1998.

[5] R. C. Read. Card–guessing with information. A problem in probability. *American Mathematical Monthly*, 69:506–511, 1962.

[6] R. A. Sulanke. Guessing, ballot numbers, and refining Pascal's triangle (9 pages). *http://diamond.idbsu.edu/∼sulanke/recentpapindex.html*, 1995.

[7] D. Zagier. How often should you beat your kids? *Mathematical Magazine*, 63:89–92, 1990.