Math Meets the Bookies:

or

How to win Football Pools
Football Pools

Given \( n \) games (matches) to be played by \( 2n \) teams, a forecast is an \( n \)-tuple consisting of a prediction for each of the \( n \) matches.

In a football pool, each player wagers on a forecast. The first place prize goes to anyone whose forecast matches the actual outcomes of the games (and is split amongst multiple winners). The second place prize goes to those who get only one game wrong (again, split amongst multiple winners). Third and fourth place prizes also exist. Generally, there are no prizes after the fourth.
Football Pools

As the only way to guarantee a 1st place is to bet on all possible forecasts, a more interesting question is:

What is the minimum number of bets needed to guarantee a 2nd place prize?
Football Pools

In the traditional literature, emphasis has been placed on games where there are 3 outcomes; home team wins, away team wins or a tie. (We see the European influence here, since football in Europe means soccer and ties are much more frequent in soccer games than in American football.)

We may increase the generality of the problem and consider $k$ possible outcomes in a game (perhaps thinking of a horse race as a game.) The minimum number of forecasts for $n$ matches with $k$ outcomes per match, so that the actual outcome differs by no more than one match from some forecast of this set will be denoted by: $\sigma(n,k)$.
Example

The Louisville senior boys soccer league has 6 teams. Every Saturday morning 3 matches are played. The teams are evenly balanced, any outcome in a game is possible.

Some of the parents have started a futsball pool. For a dollar, you bet on any forecast. The top prize is 50% of the money generated (about $20) and 2\textsuperscript{nd} prize is 25%. The remainder of the cash is saved for an end of the season party.

There are $3^3 = 27$ possible forecasts on any given Saturday. How would you play in this pool?

As $\sigma(3,3) = 5$, for a $5$ investment you can be guaranteed a 2\textsuperscript{nd} place prize!
Coding Model

We can consider the set of all forecasts as \( V(n,k) \), and then we see that \( \sigma(n,k) \) is the size of the smallest code having covering radius 1 (every forecast is within Hamming distance 1 of some codeword.)

Since a unit ball contains \( n(k-1) + 1 \) elements, the sphere packing bound gives us:

\[
\sigma(n,k) \geq \left\lfloor \frac{k^n}{n(k-1)+1} \right\rfloor.
\]
Other Lower Bounds

In 1970, Rodemich, by carefully analyzing the overlaps of the unit balls, was able to improve the sphere packing bound when \( n \leq k \):

\[
\sigma(n, k) \geq \left\lceil \frac{k^{n-1}}{n-1} \right\rceil \text{ when } n \leq k.
\]

Rodemich Bound

In 1991, van Wee was also able to improve the sphere packing bound in a number of cases (but not all) when \( k \leq n \).
\( \sigma(3,3) \)

\( V(3,3) \) can easily be visualized as a graph whose vertices are the forecasts and two vertices are adjacent iff they have distance one.

(not all edges are drawn ... all vertices on a line are joined to each other)
The example below shows that $\sigma(3,3) \leq 5$. While the Rodemich bound gives $\sigma(3,3) \geq 5$. So we have $\sigma(3,3) = 5$.

(not all edges are drawn ... all vertices on a line are joined to each other)
Exact Solutions

There are relatively few exact values known for $\sigma(n,k)$.

In the last example we obtained an exact value because an upper bound and a lower bound agreed. Lower bounds are obtained by theoretical considerations, while upper bounds are given by examples.

The lack of exact values is a reflection of the fact that very few good lower bounds are known – and this in turn is due to the complexity of the problem, making theoretical insights hard to come by.
Exact Solutions

**Sphere Packing Bound** – Hamming Code solutions:

If a code meets the sphere packing bound, then it must be a perfect code. The only perfect codes with covering radius one, have the same parameters as a Hamming code. This means that $k$ is a power of a prime and $n = (k^r-1)/(k-1)$, and we obtain in these cases (since the Hamming code would be an example):

\[ \sigma(n,k) = k^{n-r}. \]

**Rodemich Bound** – Blokhuis-Lam solutions:

In 1984 these authors gave a construction which meets the Rodemich bound, so if $q$ is a prime power:

\[ \sigma(q+1,qt) = q^{q-1}t^q. \]
Exact Solutions

It is easily seen that $\sigma(2,k) = k$ and combinatorial considerations give the values of $\sigma(3,k) = \left\lceil \frac{1}{2}k^2 \right\rceil$.

The only other exact values that are known were obtained with great effort:

$$\sigma(4,4) = 24 \quad \text{Stanton, Kalbfleisch & Horton (1969)}$$

and

$$\sigma(5,3) = 27 \quad \text{van Lint & Kamps (1967)}$$
## Values of $\sigma(n,k)$

<table>
<thead>
<tr>
<th>n \ k</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>13</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>9</td>
<td>24</td>
<td>46–51</td>
<td>72</td>
<td>115–123</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>27</td>
<td>64</td>
<td>160–184</td>
<td>330–414</td>
<td>606–769</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>71–73</td>
<td>216–256</td>
<td>625</td>
<td>1578–1840</td>
<td>3412–4435</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>156–186</td>
<td>762–992</td>
<td>2722–3125</td>
<td>$6^5+1–9\times6^4$</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>402–486</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>7$^6$</td>
</tr>
</tbody>
</table>

As of 8/18/09, see http://www.sztaki.hu/~keri/codes/ for most current values.
Some theory

We shall establish some of the elementary results concerning the value of $\sigma(n,k)$. These results can be found in: J.G.Kalbfleisch & R.G.Stanton, “A Combinatorial Problem in Matching”, *J.London Math. Soc.* 44(1969), 60-64.

We can view $V(2,k)$ as a lattice or, as we shall now do, as a $k \times k$ chessboard. The codewords can be thought of as “rooks” and the spheres of radius 1 are the cells that the rook can attack (including the cell the rook sits on).
Some Theory

Any set of k rooks that are 1) in a row, 2) in a column, or 3) form a transversal (one in each row and each column) will cover all the cells. So, $\sigma(2,k) \leq k$.

On the other hand, if fewer than k rooks are used on a $k \times k$ board, at least one row, say row i, has no rook in it and at least one column, say column j, has no rook in it. The cell in position (i,j) is not covered by any rook. Thus, $\sigma(2,k) \geq k$.

Thus,

$$\sigma(2,k) = k.$$
$\sigma(3,k)$

The space $V(3,k)$ can be visualized as a 3-dimensional chess board, or a $k \times k \times k$ Rubik's cube. One can also think of it as $k$ copies of a $k \times k$ board (think of the cube as made up of slices).
We will give a geometric covering of $V(3,k)$. First consider the full cube, which we divide naturally into 8 subpieces where the top-left-front piece is an $s \times s \times s$ cube (the A cube) and the bottom-right-back piece is an $(k-s) \times (k-s) \times (k-s)$ cube ($s \geq 1$). Rooks will be placed only in these two subcubes.

Take a Latin Square of order $s$, and think of it as triples $(i,j,t)$ where $t$ is the entry in the $(i,j)$ position of the Latin Square. Now, place a rook in the $(i,j,t)$ positions of the A cube. Repeat this for the other subcube starting with a Latin Square of order $k-s$.

The $s^2 + (k-s)^2 = 2s^2 - 2ks + k^2$ rooks placed this way form a cover of the full cube.
By minimizing this quadratic, we see that the minimum value is obtained when $s = k/2$. Thus, if $k$ is even the minimum is $k^2/2$. While if $k$ is odd the minimum value of $(k^2+1)/2$ is obtained when $s = (k-1)/2$. Therefore:

$$\sigma(3,k) \leq \left\lfloor \frac{k^2}{2} \right\rfloor$$

We shall now show that no fewer rooks can cover all the cells. The argument is the same for $k$ even and odd, but the numerical values are different in the two cases. We present the proof with the values for the $k$ even case given, and (the $k$ odd case in parentheses and red).
If \( k = 2s \) \((k = 2s+1)\) suppose that \(2s^2-1 \ (2s^2 +2s)\) rooks at coordinates \( R_i = (a_i,b_i,c_i) \) cover the space. Let \( a \) be the value that occurs least often in any of the coordinate positions. We may assume that it occurs \( \alpha \) times in the first coordinate and we can order the rooks so that \( a_i = a \) for \( 1 \leq i \leq \alpha \) and \( a_i \neq a \ \forall \ i > \alpha \).

As each of the \( k \) values appearing as a first coordinate appears at least \( \alpha \) times, we have \( k\alpha \leq 2s^2-1 \ (2s^2 +2s) \) and so,

\[
\alpha \leq s-1 \ (s).
\]

Choose \( b \) and \( c \) so that \( b \neq b_i \) and \( c \neq c_i \) for all \( i \leq \alpha \). These choices can be made since there are \( k \) choices and at most \( s \) have been used.
Now the cell at \((a, b, c)\) is not covered by any \(R_i\) for \(i \leq \alpha\).

Suppose that \(r\) of \(b_1, b_2, \ldots, b_\alpha\) are distinct. Each of these \(r\) values appears at least \(\alpha\) times, so \(r\alpha - \alpha\) occurrences will be as \(b_i\)'s for \(i > \alpha\) and the cell at \((a, b, c)\) will not be covered by any of these \(R_i\)'s either.

This cell must then be covered by one of the remaining

\[2s^2 - 1 - \alpha - (r\alpha - \alpha) = 2s^2 - 1 - r\alpha\]

\((2s^2 + 2s - \alpha - (r\alpha - \alpha) = 2s^2 + 2s - r\alpha)\)

\(R_i\)'s.

But, \(b\) can be chosen in \(2s - r\) \((2s + 1 - r)\) ways and \(c\) can be chosen in \(2s - \alpha\) \((2s + 1 - \alpha)\) ways.
Thus we have:
\[(2s-r)(2s-\alpha) \leq 2s^2-1-r\alpha \quad \text{and} \quad (2s+1-r)(2s+1-\alpha) \leq 2s^2 + 2s - r\alpha\]
which leads to
\[2(s-r)(s-\alpha) \leq -1 \quad \text{and} \quad 2(s-r+1)(s-\alpha) \leq r - \alpha - 1\]
which is a contradiction since
\[r \leq \alpha \leq s-1.\]

So the cell at (a,b,c) is not covered. Therefore,

\[
\sigma(3,k) = \left\lfloor \frac{k^2}{2} \right\rfloor
\]
An easily established general upper bound is obtained by:

\[
\sigma(n+1,k) \leq k\sigma(n,k)
\]

Consider \( V(n+1,k) \) as consisting of \( k \) copies of \( V(n,k) \). As each of the \( V(n,k) 's can be covered with \( \sigma(n,k) \) rooks, the result follows.

We should point out that there are cases where this bound is met. \( \sigma(4,2) = 2\sigma(3,2) \) and \( \sigma(5,3) = 3\sigma(4,3) \). Furthermore, no smaller example is known for \( \sigma(6,4) \) and \( \sigma(7,5) \).

The smallest open case: \( 46 \leq \sigma(4,5) \leq 51 \).