Super Bowl Squares: A Different Approach

Ari J. Fitter

Faculty Mentors: Dr. Joseph Nolan and Dr. David Agard

Introduction

The Super Bowl Squares game is played in households across the country during the National Football League (NFL) Championship each year. In this game, players may purchase one or more squares in a 10-by-10 grid. After all squares have been sold, the numbers 0-9 are randomly assigned to the rows and columns as the ones digit of the winner and loser's scores. The person who buys the square that reflects the final score of the real game wins the Super Bowl pool. There are several variations of this game (one popular variation awards a portion of the pot at the end of each quarter of the game). Here is an example of what a 2013 Super Bowl squares pool might look like:

		Philadelphia Eagles									
		0	1	2	3	4	5	6	7	8	9
Denver Broncos	0	AF		DA		AF	DA		AF	JN	JN
	1			DA			DA			JN	JN
	2			DA			DA			JN	JN
	3			DA			DA			JN	JN
	4	AF		DA		AF	DA		AF	JN	JN
	5			DA			DA			JN	JN
	6			DA			DA			JN	JN
	7	AF		DA		AF	DA		AF	JN	JN
	8			DA			DA			JN	JN
	9			DA			DA			JN	JN

This project proposes an interesting variation on the Super Bowl Squares game in which the digits 0 through 9 are pre-assigned as in the example above. In such a case, a few squares will be more highly desired than others. For example, 7-0, 7-7, 10-7, etc. are very likely scores at the end of the first quarter. These squares will win prizes much more often than others. The only way to make the game fair is to make squares of greater likelihood more expensive. Similarly, less likely squares should be less expensive (sometimes drastically so). This project seeks to determine fair prices for all of the combinations of digits in this variation of Super Bowl Squares.

Data

Data[1] for the 2008 through the 2011 NFL seasons were collected from www.pro-football-reference.com. Necessary data included the frequency of each type of scoring play as well as the quarter-by-quarter and final scores of every regular season game of each of those seasons. To allow for an investigation of whether NFL scoring distributions may change substantially over time, the same data were also collected for the 1994-1997 seasons. No large differences in these eras were observed.

Methods

Comprehensive scoring-play data from the 2008-2011 NFL regular seasons allows the empirical estimation of the scoring play distribution shown in the table below:

Distribution of Points Scored on a Scoring Play					
Points	2	3	6	7	8
Probability	0.01	0.39	0.02	0.57	0.01

As one can see, 96% of scores result in either a touchdown with standard extra point (7 points) or a field goal (3 points).

From this distribution, Markov Chain Analysis is applied to advance scoring through multiple scoring plays that would take place within any fixed period of time during the game. The distribution above produces the following transition matrix which reflects the probabilities of moving from an initial score (row 0 – 9) to a new score (column 0 – 9) after a single scoring play.

$$T=\begin{pmatrix} 0 & 0 & .01 & .39 & 0 & 0 & .02 & .57 & .01 & 0 \\ 0 & 0 & 0 & .01 & .39 & 0 & 0 & .02 & .57 & .01 \\ .01 & 0 & 0 & 0 & .01 & .39 & 0 & 0 & .02 & .57 \\ .57 & .01 & 0 & 0 & 0 & .01 & .39 & 0 & 0 & .02 \\ .02 & .57 & .01 & 0 & 0 & 0 & .01 & .39 & 0 & 0 \\ 0 & .02 & .57 & .01 & 0 & 0 & 0 & .01 & .39 & 0 \\ 0 & 0 & .02 & .57 & .01 & 0 & 0 & 0 & .01 & .39 \\ .39 & 0 & 0 & .02 & .57 & .01 & 0 & 0 & 0 & .01 \\ .01 & .39 & 0 & 0 & .02 & .57 & .01 & 0 & 0 & 0 \\ 0 & .01 & .39 & 0 & 0 & .02 & .57 & .01 & 0 & 0 \end{pmatrix}$$

One will notice patterns within this matrix, as for example the chance of the ones digit going from 0 to 3 is the same as the chance of going from 4 to 7 or 8 to 1 (in all cases, 1 field goal must be scored). Furthermore, while the initial matrix reflects a single scoring play, we may take powers of the matrix in order to consider the change in state after multiple scoring plays. For example, the matrix below is the first row of T^3, which estimates the probability distribution for units digit (starting at 0) given three scores:

In this case, the ones digit is most likely to transition from 0 to 1, 3, or 7.

In order to estimate the transition for a given period of time, we must also incorporate the distribution of the number of scoring plays (for both winner and loser). The next table (see next column) reflects the empirical distribution of scoring plays for an entire game.

Methods

Probability of Scoring a Certain Number of Times in a Full Game											
Score	0	1	2	3	4	5	6	7	8	9	10
W %	0.0	0.1	4.4	13.0	23.2	26.3	18.3	9.6	4.2	0.8	0.1
L %	2.0	11.7	24.8	27.8	16.4	10.7	5.2	1.2	0.2	0.0	0.0

As expected the winner will usually score more often than the loser. Similar tables may be produced for earlier points in the game (e.g. if awarding of prizes at the end of each quarter).

Proceeding in analysis, we now define random variables:

d = value of the one's digit for a team's score

k = number of scoring plays by that team in a period

Computationally, using the Law of Total Probability:

$$P(d)=\sum_k P(k)P(d\mid k)$$

$P(k)$ comes from the appropriate distribution of number of scoring plays; $P(d \mid k)$ is the first row of T^k. This formula is applied for both winner and loser. The resulting probability distributions are assumed independent and used to produce cost tables for various prize allocation schemes.

Results

Assuming a $100 pool, the entries in the table below represent the cost for each square if all prize money is awarded at the end of the game.

Cost of Each Square in a Fair Super Bowl Squares Pool for the Whole Game											
W \ L	0	1	2	3	4	5	6	7	8	9	TOT
0	**3.1**	1.2	0.4	2.3	2.2	0.4	1.3	**3.0**	0.6	0.6	15.1
1	2.2	0.9	0.3	1.6	1.6	0.3	0.9	2.2	0.4	0.4	10.8
2	1.0	0.4	0.1	0.7	0.7	0.1	0.4	1.0	0.2	0.2	4.7
3	2.4	0.9	0.3	1.8	1.7	0.3	1.0	2.4	0.5	0.5	11.8
4	**3.0**	1.2	0.3	2.2	2.1	0.3	1.2	**2.9**	0.6	0.6	14.5
5	1.1	0.4	0.1	0.8	0.8	0.1	0.4	1.1	0.2	0.2	5.2
6	1.8	0.7	0.2	1.3	1.3	0.2	0.8	1.8	0.3	0.4	8.8
7	**3.1**	1.2	0.4	2.3	2.2	0.4	1.3	**3.1**	0.6	0.6	15.2
8	1.6	0.6	0.2	1.2	1.2	0.2	0.7	1.6	0.3	0.3	7.8
9	1.2	0.5	0.1	0.9	0.9	0.1	0.5	1.2	0.2	0.2	6.0
TOT	**20.6**	8.0	**2.4**	15.0	14.7	**2.4**	8.6	**20.2**	3.9	4.2	

Highlighted in red are the six most expensive squares; as you can see these scores cost more to purchase than some entire columns. There are also six scores that can be purchased for only a dime each! One might consider purchasing entire rows or columns – and it should not be unexpected that scores ending in 0 or 7 are the most expensive to obtain.

Additional Results

From the basic information discussed in methods, cost tables may be developed for almost any prize allocation scheme. A few more examples are shown below.

Cost of Each Square in a Fair Super Bowl Squares Pool for the First Quarter											
W/L	0	1	2	3	4	5	6	7	8	9	TOT
0	20.1	0.1	0.1	5.7	1.0	0.1	0.8	8.2	0.2	0.1	36.1
1	0.5	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	1.9
2	0.3	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	2.5
3	11.1	0.0	0.1	3.1	0.6	0.0	0.4	4.5	0.1	0.1	23.0
4	4.0	0.0	0.0	1.1	0.2	0.0	0.2	1.6	0.0	0.0	11.1
5	0.3	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	5.5
6	2.5	0.0	0.0	0.7	0.1	0.0	0.1	1.0	0.0	0.0	10.5
7	15.8	0.1	0.1	4.5	0.8	0.1	0.6	6.4	0.1	0.1	35.5
8	0.4	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	8.7
9	0.5	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	9.8
TOT	54.9	1.2	2.3	18.5	6.8	5.2	8.2	29.4	8.5	9.3	

Cost of Each Square for 25% Each Quarter											
W/L	0	1	2	3	4	5	6	7	8	9	TOT
0	8.4	0.7	0.2	3.8	2.0	0.2	1.2	5.3	0.4	0.4	22.2
1	1.8	0.4	0.1	1.2	0.9	0.1	0.5	1.7	0.2	0.2	8.0
2	0.6	0.2	0.0	0.4	0.3	0.0	0.2	0.6	0.1	0.1	4.5
3	5.4	0.6	0.2	2.6	1.5	0.2	0.9	3.6	0.3	0.3	18.2
4	3.7	0.6	0.2	2.2	1.5	0.2	0.9	3.1	0.3	0.3	16.6
5	0.6	0.2	0.0	0.4	0.3	0.0	0.2	0.6	0.1	0.1	7.5
6	2.2	0.4	0.1	1.3	0.9	0.1	0.5	1.8	0.2	0.2	13.5
7	7.4	0.7	0.2	3.5	2.0	0.2	1.2	5.0	0.3	0.4	27.6
8	1.0	0.3	0.1	0.7	0.5	0.1	0.3	0.9	0.1	0.1	12.0
9	1.0	0.2	0.1	0.7	0.5	0.1	0.3	0.9	0.1	0.1	12.8
TOT	31.2	5.0	3.2	19.2	13.9	6.1	11.9	29.6	9.8	11.2	

Future Ideas

Ideas for future work on this project include:

1. Study of the independence assumption. Perhaps the 1994-1997 data may be used for cross-validation to determine whether this assumption is reasonable.
2. Quarters may be different (the last quarter in particular). Use of four distinct transitions (one for each quarter) might allow for investigation of this question.
3. An investigation of the effect of home-field advantage (or perhaps using home/away instead of winner/loser to define the table) would be intriguing.
4. Investigation of a symmetry assumption (i.e. does winner/loser really make much difference?)

Acknowledgement: This research was a product of a summer undergraduate research experience supported by the Center for Integrative Natural Science and Mathematics (CINSAM) at Northern Kentucky University.