The Gambler's Ruin—An Analysis by Spreadsheet

Thomas C. McMillan *James R. Fulmer*
tcmcmillan@ualr.edu jrfulmer@ualr.edu
Department of Mathematics and Statistics
University of Arkansas at Little Rock
72204-1099
USA

Abstract

The "Gambler's Ruin" problem is stated as follows. Two players have a total of N coins, they win or lose a coin based on the outcome of a coin toss, and they continue to flip a coin until one or the other player has won all N coins. We use this problem to illustrate how the spreadsheet can be used to give a detailed analysis of this game, giving numerical answers to the following questions: (1) What is $p(n,t)$, the probability that a game ends in t coin tosses given that one of the players has n coins? (2) What is $q(n,t)$, the probability that a player with n coins loses the game in t coin tosses? (3) What is $p(t)$, the probability that a game will last t coin tosses? We illustrate using the spreadsheet to check our models for internal consistency and for consistency with known results. Our examples illustrate how using the spreadsheet motivates the study of recurrence relations and the advantages of this approach in an introductory discrete mathematics class.

1 Introduction

This paper uses the "gambler's ruin" problem as a vehicle for illustrating how the electronic spreadsheet can be used to solve recurrence relations numerically. It is hoped that this example will illustrate that the spreadsheet is a powerful computational tool whose simplest features can be introduced without too much difficulty. We have used the spreadsheet as a tool for discrete analysis explorations in professional development classes for inservice secondary mathematics teachers. Once students master the fundamentals of the spreadsheet, they can use it to quickly obtain numerical solutions for discrete boundary value problems without having to resort to complicated programming. This approach encourages students to take a natural path to problem solving: State the problem and the desired results clearly, develop a mathematical model (in our examples, discrete boundary value problems), and analyze the model to obtain the desired results. This approach enables a teacher to introduce recurrence relations in the context of solving problems and empowers students to solve those problems even before learning analytic techniques for solving recurrence relations. Even after students have equipped their toolbox with analytical techniques, they can use the spreadsheet to deal with problems that are difficult or impossible to solve analytically. A brief introduction to using the spreadsheet can be found in [3].

The professional development class in which we introduced this problem was a grant-funded class for inservice secondary mathematics teachers, many of whom hold the master's degree. This class explored uses of technology, including spreadsheets, in the classroom. Those teachers who were familiar with the spreadsheet saw it mainly as a financial tool. A few had used the spreadsheet to maintain class records and were familiar with some of the built-in functions (*e.g.* `SUM` and `AVERAGE`). None of the teachers had recognized that the spreadsheet can be used to solve recurrence relations. We used the gambler's ruin problem described here as well as many of the examples described in [3]. Our biggest hurdle was getting the teachers to think recursively. We presented problems to the teachers in worksheets in which a series of questions lead them to a recursive model with boundary conditions. This was followed by a series of questions which guided the entry of the problem into the spreadsheet. As the teachers gained experience, the worksheets became less explicit. The more detailed analyses of the probability distributions described in this paper have not yet been tried in class.

Suppose one player has n coins out of a total of N coins held by two players. They each win or lose a coin based on the outcome of a fair coin toss, and the game proceeds until one player has won all of the coins. This game is referred to as the gambler's ruin. Let $p(n)$ denote the probability that a player with n coins ends up winning all of the coins. It is shown in [1] by solving the boundary value problem $p(n) = \frac{1}{2}p(n-1) + \frac{1}{2}p(n+1)$ for $1 \le n \le N-1$, $p(0) = 0$, $p(N) = 1$ that $p(n) = n/N$, and hence that the game will always end. In [4], it is shown by solving the boundary value problem $E(n) = \frac{1}{2}E(n-1) + \frac{1}{2}E(n+1)$ for $1 \le n \le N-1$, $E(0) = E(N) = 0$ that the expected number of coin tosses, $E(n)$, when one player has n coins is $E(n) = n(N-n)$. Using a spreadsheet to numerically solve these problems is illustrated in [3]. This last paper also begins to explore probability distributions related to the gambler's ruin.

In this paper, we will look more closely at models for probability distributions and describe how the spreadsheet can be used to analyze them. In our discussions, spreadsheet items will be depicted using bold typewriter font, *e.g.* `B10` refers to the cell in column `B`, row `10`.

2 Probabilities

In this section, we analyze probabilities for the length of a gambler's ruin game in which a total of N coins are in play.

2.1 Probability of t Moves from a Given Starting Position

Let $p(n,t)$ denote the probability that a game has t moves (coin tosses), given that one of the players has n coins. Then $p(n,t)$ is modeled by the boundary value problem

$$\begin{aligned} p(n,t) &= \frac{1}{2}p(n-1,t-1) + \frac{1}{2}p(n+1,t-1), \text{ for } 1 \le n \le N-1,\ t \ge 1, \qquad (1)\\ p(0,0) &= 1,\\ p(N,0) &= 1,\\ p(n,0) &= 0, \text{ for } 1 \le n \le N-1,\\ p(0,t) &= 0, \text{ for } t \ge 1, \text{ and}\\ p(N,t) &= 0, \text{ for } t \ge 1, \end{aligned}$$

for $n \in \{0,1,2,\ldots,N\}$ and $t \in \{0,1,2,3,\ldots\}$. The recurrence follows because a player with n coins will see the end of the game in t moves by either losing a coin and proceeding through $t-1$ additional moves, or by winning a coin and proceeding through $t-1$ additional moves, with each of these mutually exclusive events occurring with probability one-half. The reasoning leading to the boundary conditions is straightforward. If a player has no coins, or all of the coins, then the game is over in zero moves with probability one. If a player has some, but not all of the coins, then the probability that the game is over in zero moves is zero. If a player has none of the coins, or all of the coins, then the probability that the game lasts one or more moves is zero.

For a predetermined value of N, our model will be entered into a spreadsheet as an array with column headings `0,1,2,...,N`, and row headings `0,1,2,3,...`. Clearly the spreadsheet cannot accommodate an infinite number of rows, but we can create enough rows to get past the part of the distribution we are interested in. The spreadsheet array will ultimately be populated with the calculated values for $p(n, t)$. For example, the cell under column heading `3` and across from row heading `4` will be $p(3, 4)$, the probability that a game ends in 4 coin tosses for a player with 3 coins. After entering the column headings and the row headings into the spreadsheet, enter the boundary conditions: a `1` in each of the corner cells corresponding to 0 coins or N coins and `0`'s in each of the other cells in the row corresponding to $t = 0$ and the columns corresponding to $n = 0$ and $n = N$.

◇	A	B	C	D	E	F	G
1		0	1	2	3	4	5
2	0	1	0	0	0	0	1
3	1	0	0.5	0	0	0.5	0
4	2	0	0	0.25	0.25	0	0
5	3	0	0.125	0.125	0.125	0.125	0
6	4	0	0.0625	0.125	0.125	0.0625	0
7	5	0	0.0625	0.0938	0.0938	0.0625	0

Figure 1: Spreadsheet for $N = 5$

Figure 1 shows a few rows of this spreadsheet array for calculating $p(n, t)$, $n \in \{0, 1, 2, 3, 4, 5\}$. Note the `1`'s in the corners of the boundary and the `0`'s in the other boundary positions. The array gives visual meaning to the term "boundary values." The 0.5 in cell `C3` is, being the value $p(1, 1)$, easy to calculate and enter by hand. However, the 0.5 you see comes from a calculation using the formula formula `=0.5*B2+0.5*D2` that was entered into cell `C3`. Note, this formula encodes the relation $p(1, 1) = \frac{1}{2}p(0, 0) + \frac{1}{2}p(2, 0)$, one instance of recurrence (1). This formula needs to be explicitly entered in only one of the interior cells. It is entered into the other cells by "drag-copying" it from cell `C3` into the other interior cells. Because the spreadsheet uses relative addressing, each instance of the formula refers to the cells up one, to the left one and up one, to the right one from the cell containing the formula. For example, when the formula in `C3` is copied into `E5`, the resulting formula is `=0.5*D4+0.5*F4`. The numerical solution to the boundary value problem (1) appears instantaneously in the array when the formula in `C3` is copied into the interior cells of the table.

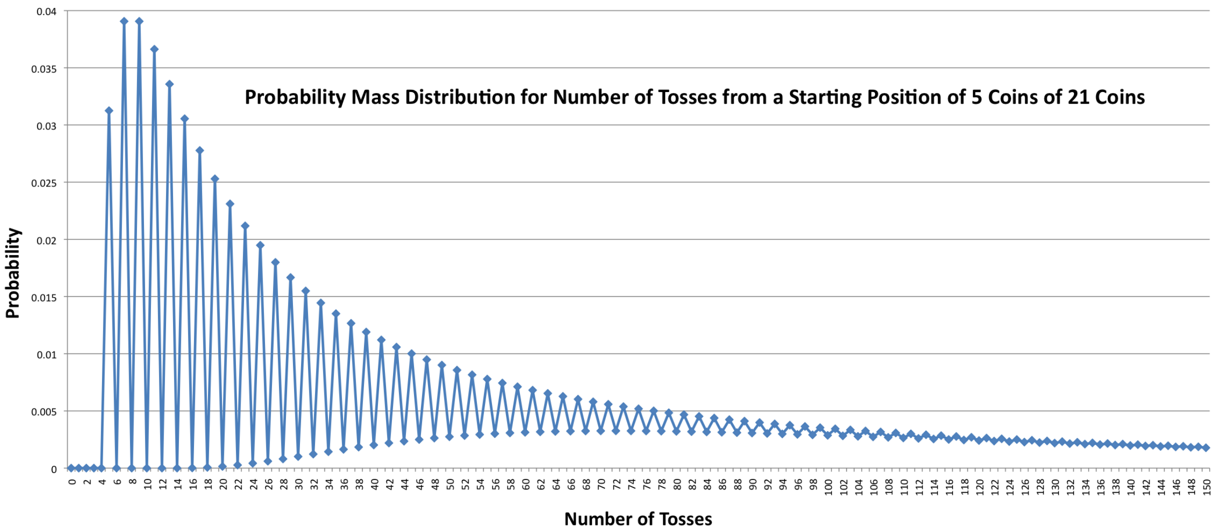

Figure 2: Probability Distribution $p(5, t)$

We continue our discussion assuming we have constructed an array of probabilities $p(n, t)$ when there are $N = 21$ coins in play. Although it is not a guarantee that our reasoning is correct, one check on the validity of our solution is to verify for each value of $n \in \{0, 1, \ldots, N\}$ that

$$\sum_{t=0}^{\infty} p(n,t) = 1.$$

Using a table for $p(n,t)$ with 2000 rows and summing columns, we get partial sums that approximate one, with the maximum error being 2.22×10^{-10}. Figure 2 shows the probability mass function $p(5,t)$, the probability that a game ends in t tosses of the coin for a player with 5 coins. The graph was produced using a spreadsheet's graphing function.

2.2 Probability of a Loss in t Moves from a Given Starting Position

Let $q(n,t)$ denote the the probability that a player with n coins loses the game in t tosses of the coin. The model for $q(n,t)$ is almost identical to that for $p(n,t)$:

$$\begin{aligned} q(n,t) &= \frac{1}{2}q(n-1,t-1) + \frac{1}{2}q(n+1,t-1), \text{ for } 1 \le n \le N-1,\ t \ge 1, \qquad (2)\\ q(0,0) &= 1,\\ q(N,0) &= 0,\\ q(n,0) &= 0, \text{ for } 1 \le n \le N-1,\\ q(0,t) &= 0, \text{ for } t \ge 1, \text{ and}\\ q(N,t) &= 0, \text{ for } t \ge 1, \end{aligned}$$

for $n \in \{0, 1, 2, \ldots, N\}$ and $t \in \{0, 1, 2, 3, \ldots\}$. The only difference is for the boundary condition $q(N,0)$. Whereas $p(N,0) = 1$, $q(N,0) = 0$ because a player with all the coins will lose the game in zero moves with probability zero. The numerical solution for $q(n,t)$ can be found by making a copy of the corresponding array for $p(n,t)$ and changing the boundary condition corresponding to position $(N,0)$ from 1 to 0. Figure 3 shows a spreadsheet graph of $q(5,t)$, the probability of losing in t tosses of the coin from a starting position of 5 coins out of $N = 21$ coins.

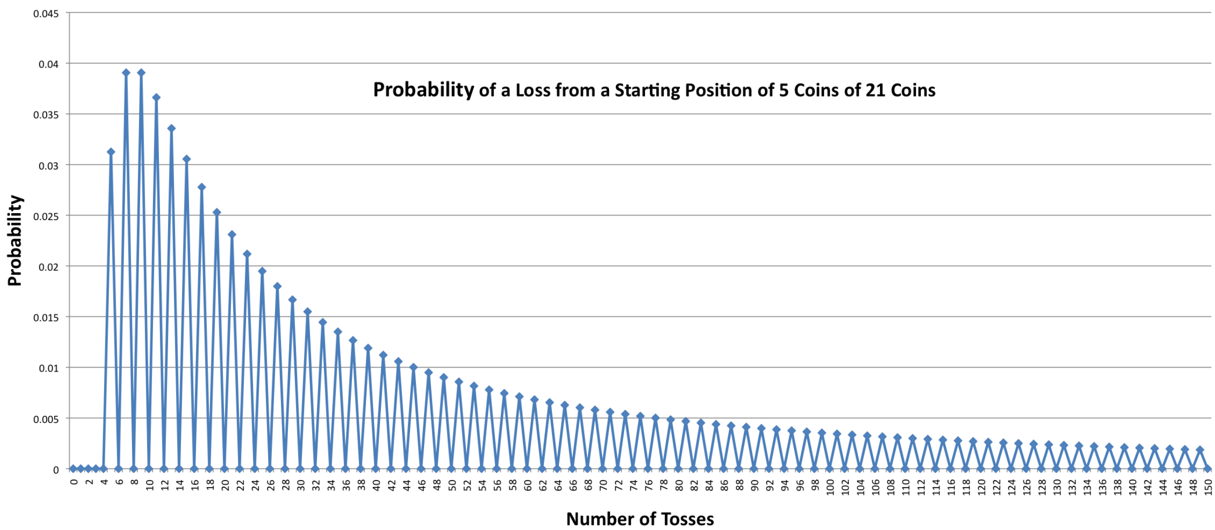

Figure 3: Probability Distribution $q(5,t)$

From the work cited in the introduction, we know that the probability of losing from a starting position of n coins out of a total of N coins is

$$L(n) = \frac{N-n}{N}.$$

This yields a validity check for our model. For each $n \in \{1, 2, \ldots, N\}$ we should have that

$$\sum_{t=0}^{\infty} q(n,t) = L(n) = \frac{N-n}{N}. \tag{3}$$

As we did with p, this identity can be verified in the spreadsheet summing the columns of the array for q to obtain partial sums for these series. In an array with 2000 rows, the partial sums approximate the predicted values, with the maximum error being 1.12×10^{-10}.

2.3 Probability That a Game Ends in t Moves

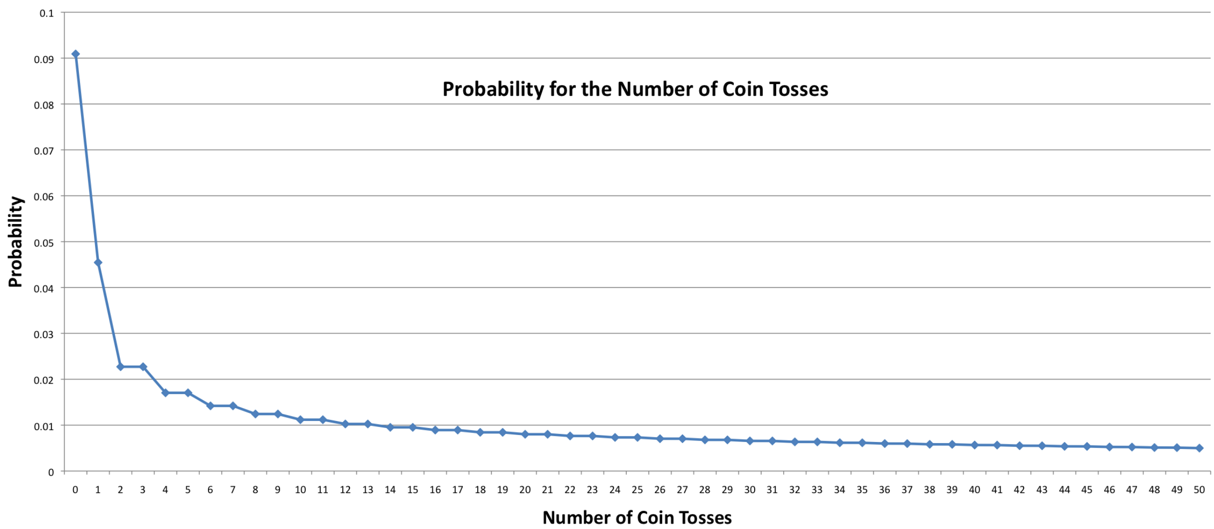

Figure 4: Probability Distribution for the Length of a Game with 21 Coins in Play

Suppose a player's initial starting position is determined at random. That is, suppose the number of coins a player is given is selected at random from $\{0, 1, 2, \ldots, N\}$, with each number having equal probability. We want to find the probability $p(t)$ that the game lasts t moves. For a given value of n, the probability that a player receives n coins is $1/(N+1)$. Therefore the probability that a player receives n coins and the game lasts for t moves is

$$\frac{1}{N+1} p(n,t).$$

It follows that

$$p(t) = \sum_{n=0}^{N} \frac{1}{N+1} p(n,t) = \frac{1}{N+1} \sum_{n=0}^{N} p(n,t).$$

In our spreadsheet containing the array for $p(n,t)$ with 2000 rows, we can create a column containing the values for $p(t)$ by summing the rows of $p(n,t)$ and dividing by $N+1$. Figure 4 is a spreadsheet graph for the probability mass function $p(t)$ when there are 21 coins in play. By summing the column containing $p(t)$, we get a partial sum approximately equal to 1, with an error of 1.35×10^{-10}. This validates that

$$\sum_{t=0}^{\infty} p(t) = 1.$$

2.4 Another Consistency Check

From the work cited in the introduction, we have a formula for the expected length of a game in which one of the players starts with n coins is

$$E(n) = n(N - n).$$

From our array containing values for $p(n,t)$, the probability that a game ends in t moves when a player has n coins, we can construct another array in which the cell in the row corresponding to t and the column corresponding to n contains the value $t \cdot p(n,t)$. By summing the column in this new array that corresponds to n, we get a partial sum approximating the expected length of a game when a player starts with n coins.

$$E(n) = \sum_{t=0}^{\infty} t \cdot p(n,t).$$

In our array with 2000 rows, the partial sums approximating $E(n)$ agree with the predicted values for $E(n)$ with the maximum error being 4.64×10^{-7}.

2.5 Natural Questions

The results and graphs described here are bound to prompt questions. These can typically be answered with some careful thought and, if necessary, by looking for clues in the tables on which the graphs are based. For example, how can we explain the oscillating nature of the graph in Figure 2 of the probability $p(5,t)$ that a game ends in exactly t moves when a player starts with 5 of 21 coins? Relatively straightforward reasoning leads to the conclusion that if a loss occurs in t moves, then $t \geq 5$ and t is odd. Five losses are required to lose the five coins, and cycles of gains and losses that end up with the same number of coins in hand involve an even number of coin tosses. Therefore, if a game ends in a loss in t moves, then t is five plus an even number of moves. Similar reasoning yields that if a game ends in a win in t moves, then $t \geq 16$ and t is even. It will take sixteen wins to gain the remaining 16 coins, and cycles of gains and losses ending in the same number of coins involve an even number of moves. It now follows that, in a game with t moves, a player starting with 5 coins loses the game if $t \geq 5$ is odd, and wins the game if $t \geq 16$ is even. Since a player starting with 5 coins is more likely to lose than win, for odd values of $t \geq 5$, $p(5,t)$ (a loss) will be greater that $p(5,s)$ (wins) for even values of s close to t. This bias towards losses appears to persist as t increases, but its magnitude evidently goes to zero.

Figure 3 is a plot for the probability $q(5,t)$ that a player with 5 coins loses the game in t moves. Its oscillations, for $t \geq 5$ between positive values and zeroes for odd and even values of t, respectively can be explained with reasoning similar to that in the previous paragraph. A striking feature in the graph in Figure 4 of $p(t)$, the probability that a game lasts t moves, is the occurrence of horizontal steps for small values of t. The table used to generate this graph can be examined for clues to an explanation.

2.6 An Appearance of Catalan Numbers

Consider $q(1,t)$, the probability that a player with one coin loses the game in t moves, for $t \geq 2$. If t is even, then $q(1,t) = 0$. If t is odd and not too large, we can reason as follows: In order for a loss to occur in t moves there must be $t-1$ gains and losses to bring us back to holding one coin and then a single losing move to lose that coin. A loss in t moves can be described as a sequence of $t-1$ gains (G's) and losses (L's) followed by a single loss. For example, $GGLGLLL$ describes a loss in 7 moves. In a string S of characters of length t, the *pre-string of length* s, $0 \leq s \leq t$, is the string consisting of the first s characters of S. A string of G's and L's describing a loss in t moves must have the following properties:

1. Every pre-string of length less than t must have at least as many G's as L's. Otherwise, the loss will occur in fewer than t moves.
2. The pre-string of length $t-1$ has the same number of G's as L's, bringing us back to one coin, with one move left.
3. The last character must be an L.

Writing a string with an equal number of G's and L's that satisfies the first two properties for pre-strings is equivalent to writing a string of left and right parentheses in which the parentheses match, a "legal string" of parentheses. For example "$GGLGLL$" is analogous to "(()())". The number of pairs of legal strings of n pairs of parenthesis is the n^{th} Catalan number,

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

The formula for C_n is derived in [5].

In our situation, the number of paths of length $t-1$ that bring us back to one coin, without going to zero coins, is the Catalan number for $n = (t-1)/2$,

$$\frac{1}{\frac{t-1}{2}+1}\binom{t-1}{(t-1)/2} = \frac{2}{t+1}\binom{t-1}{(t-1)/2}.$$

Each of these paths has probability $\left(\frac{1}{2}\right)^{t-1}$, and the probability that the final toss is a loss is $\frac{1}{2}$. It now follows for t odd and not too large that

$$q(1,t) = \left(\frac{1}{2}\right)^t \frac{2}{t+1}\binom{t-1}{(t-1)/2}.$$

This formula can be validated by looking at the first few entries in column `1` of the spreadsheet array for $q(n,t)$.

The reason for specifying that t is not too large is that the amount by which the number of G's in our string of gains and losses can exceed the number of L's must be less than the number of coins in play. If t is large, then some "parenthetically matched" strings of G's and L's will describe games in which more coins are won than there are coins in the game. For example, in enumerating the losing strings for $q(1,41)$ when there are 21 coins in play, we look for pre-strings of length 40 in which the G's and L's are "parenthetically matched" and in every pre-string the number of $G's$ exceeds the number of L's by at most 20. For small values of t, there are not enough moves available to win all of the coins. These restrictions, and similar considerations when analyzing $q(2,t), q(3,t)\ldots, q(20,t)$ lead to counts for strings that are related to the Catalan numbers. We have not yet performed a thorough theoretical analysis of these probabilities.

3 Conclusion

We have examined examples in which fundamental operations in an electronic spreadsheet are used to numerically solve discrete boundary value problems. Using the spreadsheet as we have done here enables the discrete mathematics teacher to introduce recurrence relations in the context of problem solving and mathematical modeling. Students can use recurrence relations to model and analyze problems that on first glance may seem overwhelming. This approach motivates the study of recurrence relations and the need to develop analytical techniques. It also sensitizes the student to the fact that the spreadsheet is a powerful computational tool that can be applied to discrete boundary value problems that are difficult or impossible to solve analytically. We have had success with this approach in a professional development course for inservice secondary mathematics teachers.

References

[1] Honsberger, Ross, *Mathematical Gems I*, The Mathematical Association of America (1973), pp 128–130.

[2] McMillan, Thomas C. "The Birthday Problem: Using an Electronic Spreadsheet in a Gentle Introduction to Discrete Mathematics," *Proceedings of the* 2^{nd} *International Conference on the Teaching of Mathematics*, CD-ROM link, John Wiley & Sons, Inc., July, 2002.

[3] McMillan, Thomas C. "Analysis of Discrete Models Using an Electronic Spreadsheet—A Survey of Examples," *Proceedings of the* 14^{th} *Asian Technology Conference in Mathematics*, ATCM 2009, Beijing, China, December, 2009.

[4] McMillan, Thomas C. "The Gambler's Ruin–How Long Does It Take?", *The AMATYC Review*, Vol. 23, No. 2, (Spring, 2002), pp 33–38.

[5] Roberts, Fred S. *Applied Combinatorics*, Prentice-Hall (1984), pp. 234–236.