Lipping Out and Laying Up

G.H. Hardy and J.E. Littlewood's Curious Encounters with the Mathematics of Golf

Roland Minton

G.H. Hardy (1877-1947) and J.E. Littlewood (1885-1977) are high on any serious list of the most important mathematicians of all time. Individually, they made deep and numerous contributions to analysis and number theory, publishing hundreds of papers each. Together, they formed what many consider to be the greatest mathematical collaboration ever. The Hardy-Littlewood conjectures in number theory contain important insights into the distribution of prime numbers. They revitalized the study of analysis in England, and the rigor that they brought to their work elevated all of British mathematics. A saying, attributed to Harald Bohr as a friend's "joke" but expressing a common belief in the mathematics community, was that in the early 1900's there were only three great English mathematicians: Hardy, Littlewood, and Hardy-Littlewood.

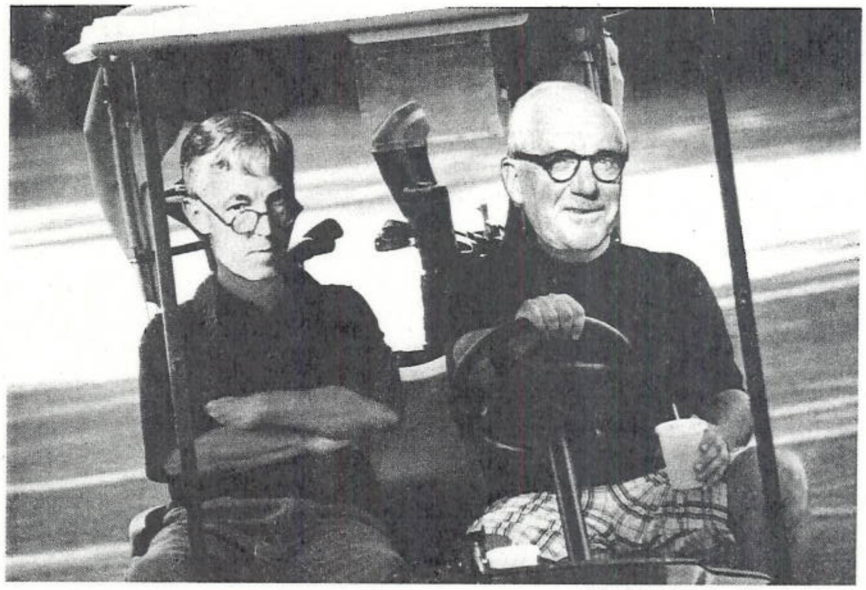

Photo montage by Bruce Torrence

Hardy and Littlewood, pure mathematicians par excellence.

Hardy was outspoken about the beauty and purity of mathematics. In his famous essay, *A Mathematician's Apology*, he wrote, "I have never done anything 'useful'. No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world." This statement is often quoted, and just as often misunderstood. I believe that Hardy's primary intent was to affirm the purity of his motives—that he worked on problems for the beauty of the mathematics and not for personal profit or an illusion of saving the world.

> Because the height function is periodic the ball does not roll to the bottom of the hole. It spirals down some distance, and then spirals back up!

Given this background, Hardy and Littlewood seem unlikely candidates for contributors to the mathematics of sports, which is why I was astonished to discover that each had made a brief mathematical excursion into the world of golf. Hardy's 1945 paper, "A Mathematical Theorem about Golf" tackled the question of whether a bold "go for it" strategy is preferable to taking a more cautious approach, while Littlewood unearthed an oddity of Newtonian mechanics that, among other things, helps explain why Tiger Woods failed to set the course record at the 2007 PGA Championship.

Littlewood's Lip-Out Discovery

One chapter of Littlewood's book, *A Mathematician's Miscellany*, discusses problems from the Mathematical Tripos examination used to rank mathematics graduates of Cambridge University. Littlewood was himself a Senior Wrangler in 1905, this being the term for the candidate with the highest score on the Tripos. Hardy had been fourth wrangler in 1898.

The Tripos problem in question involves a spinning sphere on top of a horizontal cylinder. If displaced slightly, the ball will

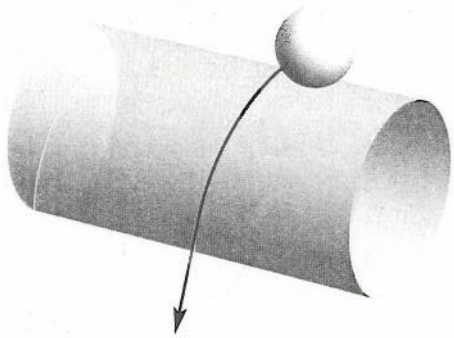

Figure 1. A spinning sphere on a horizontal cylinder follows a helical path before falling off.

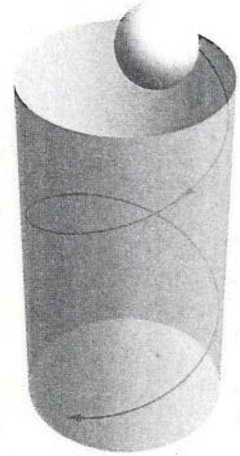

Figure 2. A helical path.

start to roll down the cylinder but will not stick to the outside of the cylinder for long (see Figure 1). The Tripos challenge is to show that the ball follows the path of the helix until it loses contact with the cylinder. Littlewood then asked what would happen if the cylinder were turned upright and the ball rolls along the inside of the cylinder. This could represent a ball rolling around the edge of a golf cup. What is its path? The most reasonable guess is that the ball will spiral down to the bottom of the hole (see Figure 2). You might wonder if the path is a perfect helix or whether gravity stretches out the spiral. However, very few people guess the actual path of the ball.

The mathematical assumptions are that the ball stays in contact with the cylinder and is acted on by gravity. In the two dimensions seen from above the hole, the path is easily described in polar coordinates. The motion is periodic with a constant frequency. The third dimension can be left as a spatial variable representing the height z of the ball above the bottom of the cup (see Figure 3).

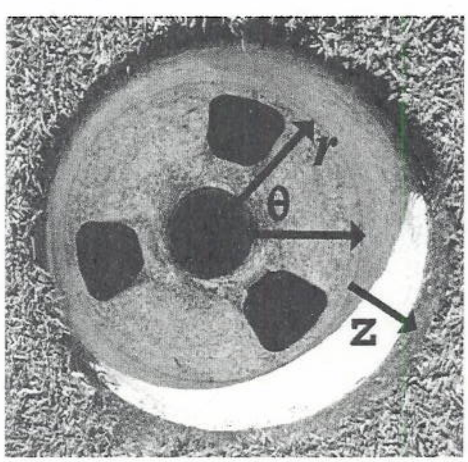

Figure 3. Cylindrical coordinate system.

The result is counterintuitive: the height z is also described by a periodic function! In a helix or stretched helix, the height z would be a strictly decreasing function. Because z is periodic, the ball does not roll to the bottom of the hole. It spirals down some distance, and then spirals back up! (And then back down and then back up and so on.) A full solution is given in "Golfer's dilemma" by Gualtieri, Tokieda, Advis-Gaete, Carry, Reffet and Guthmann in the June 2006 *American Journal of Physics*. They show that the ratio of the period of the vertical motion to the period of the circular motion is $\sqrt{7/2}$.

Mathematically, then, a ball entering the hole could roll down into the hole and then roll up and back out, as shown in Figure 4! The assumption that the ball stays in contact with the inside of the cylinder is not often met on a golf course. However, any experienced golfer can tell you stories about putts that *did* go into the hole and then come back out. The greatest golfer of our time, Tiger Woods, had a putt in the 2007 PGA Championship to shoot a 62, which would have been the lowest score ever shot in a major championship. The putt went most of the way in, did a quick 180° turn and came back out.

The mathematics and physics of a rolling ball show that this can happen. In his *Miscellany*, Littlewood comments that, "Golfers are not so unlucky as they think." Somehow, I doubt this makes Tiger Woods feel any better.

Hardy Handicaps Match Play

The problem proposed by G.H. Hardy is motivated by a hypothetical match between two golfers of "equal" ability. If one golfer is much more consistent than the other, which golfer has the advantage? Another way to state the problem is in terms of "course management." Golfers are often faced with a choice of attempting a risky shot or playing safely. For instance, they may choose to try to hit over a lake directly at the hole, or play safely around the lake but farther from the hole. Is it better to use a cautious strategy or a risky strategy?

Hardy approached this question by imagining a golfer whose shots are either excellent (E), normal (N), or poor (P). The golfer plays a hole with a par of four. A player hitting four normal shots (NNNN) finishes the hole with a score of 4. A poor shot adds one to the required number of shots. Therefore, a player who hits three normal shots and then a poor shot (NNNP) has not finished the hole. Another normal shot (making the shot sequence NNNPN) finishes the hole with a score of 5. By contrast, an excellent shot reduces the required number of shots by one. Thus, a shot sequence of

NNE finishes the hole with a score of 3. Poor shots and excellent shots are essentially inverses of each other. The sequence NPEN finishes the hole with a score of 4. Other examples would be:

Shot Sequence	Score
NPNNN	5
PNNE	4
NEPPPN	6

Notice that a sequence can never end in P; a poor shot always adds one to the sequence. However, a sequence can end in E. A careful understanding of how an ending E affects the score will be the key to our analysis. As shown above, the sequence ENN finishes the hole with a score of 3. However, what score is represented by the sequence ENE? The answer is 3: the score equals the length of the sequence. In a sense, the golfer is "cheated" out of a possible benefit of an excellent shot because ENN and ENE receive the same score. The golfing analogy is that the first two shots (EN) leave the ball close to the hole. A decent (normal) putt from this distance will go in. An excellent putt that goes into the exact center of the hole is enjoyable to watch, but the golfer gets no extra credit for perfection.

The most important assumptions involve the distribution of shots. All shots are independent. One poor shot does not affect the probability that the next shot is excellent. (All golfers wish this were realistic.) The probability of a poor shot is some number p where $0 \le p \le 1/2$. The probability of an excellent shot is the *same* number p. This leaves a probability of $1 - 2p$ for a normal shot. The only difference between one such golfer—let's use the phrase "Hardy golfer" for a golfer playing with these constraints—and another is the value for p. At first glance, all Hardy golfers appear to have equal ability because excellent shots and poor shots have equal probability.

Suppose that golfer C is a Hardy golfer with p-value p_1 and golfer R is a Hardy golfer with p-value $p_2 > p_1$. Golfer C has a higher probability of hitting a normal shot than golfer R, so golfer C is the more consistent golfer (equivalently, the more cautious golfer). Golfer R has a higher probability of hitting either an excellent shot or a poor shot than does golfer C, so golfer R is the more erratic (or risky) golfer. The problem is to determine which golfer, C or R, is more likely to win a match.

Hardy's Analysis

Hardy's "A Mathematical Theorem about Golf" presents the case where $p_1 = 0$, so that golfer C always makes a par 4. Golfer R has a probability p of hitting an E shot (which Hardy calls a *supershot*) and probability p of hitting a P shot (which Hardy cleverly calls a *subshot*). Hardy computes the probability of golfer R winning a hole as $w(p) = 3p - 9p^2 + 10p^3$. This follows from golfer R making a 2 (EE) with probability p^2 and making a 3 (ENN, NEN, NNE, ENE, NEE, EPE or PEE) with probability $3p(1-2p)^2 + 2p^2(1-2p) + 2p^3$.

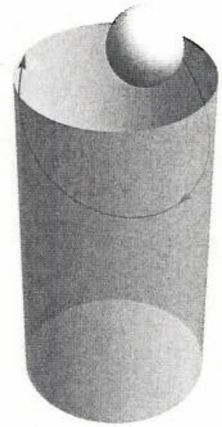

Figure 4. A mathematical lip-out.

He then computes the probability that golfer R loses the hole as $l(p) = 4p - 18p^2 + 40p^3 - 35p^4$. This is most easily obtained by computing $1 - q$, where q is the probability that golfer R makes 2, 3, or 4. A score of 4 can result from four N's with probability $(1-2p)^4$; the sequence NNNE with probability $(1-2p)^3 p$; two P's and two E's (in three possible orders) with probability $3p^4$; a starting sequence of PEN (in six possible orders) followed by either an E or an N with probability $6p^2(1-2p)(1-p)$; or one P and two N's (in three possible orders) followed by an E with probability $3p^2(1-2p)^2$. Summing these gives $q = 1 - 4p + 18p^2 - 40p^3 + 35p^4$ and Hardy's result follows.

Graphs of the functions $w(p)$ and $l(p)$ (see Figure 5) show the probabilities of winning and losing. For all values of p below approximately 0.37, player R is more likely to lose than win. However, $p = 0.37$ is an unrealistic value, since only 26% of the golfer's shots would be normal. The maximum difference between the two curves occurs at approximately $p = 0.09$.

For more realistic values of p, then, the consistent player is more likely to win the hole than the erratic player. Hardy comments that this is at odds with the standard golfing wisdom that an erratic player is better off at match play (counting each hole as a separate contest) than at stroke

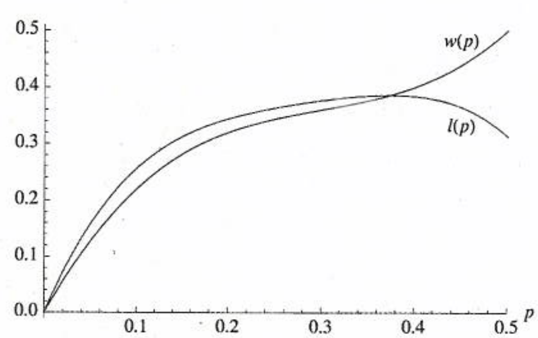

Figure 5. The functions $w(p)$ and $l(p)$.

play (where all strokes are counted for all 18 holes). As we will see, the standard wisdom is actually correct because the consistent player has an even larger advantage in stroke play.

A Mean Calculation

To start to analyze a stroke play match between two Hardy golfers, we can calculate the mean score on a single hole. A reasonable guess is that the mean should be four, since E and P shots are equally likely. This would be correct if it were not the case that some E shots are "wasted." Given that both sequences ENN and ENE produce scores of 3, the second E in the sequence ENE does not improve the golfer's score. For this reason, the mean score for a Hardy golfer with $p > 0$ will be larger than 4.

The key to computing the mean is to follow up on the comment that some E shots are wasted. If w equals the probability that a Hardy golfer's shot sequence has reached a point where an E shot *could* be wasted, then the probability that one *is* wasted is wp. Each wasted E shot raises the golfer's score by 1, so the average score is raised by wp. That is, the mean is $4 + wp$.

We next characterize the set W of (incomplete) sequences for which an appended E would be wasted. For example, the sequences NNN and EPN can be completed with either N or E, so they are in W.

There are two sequences of length 2 in W: EN and NE. The probability of getting such a sequence is $2p(1 - 2p)$. There are two types of sequences of length 3 in W. The maximum number of N's is 3, illustrated by the sequence NNN. This has probability $(1 - 2p)^3$. Two of the N's can be swapped out for a PE combination. All six of the permutations of EPN are in W. The probability of getting one of them is $6p^2(1 - 2p)$.

Generalizing, there are two types of sequences of length k in W for $k \geq 3$. The maximum number of N's is 3, with the remaining $k - 3$ shots being P's. There are

$$\binom{k}{3} = \frac{k(k-1)(k-2)}{6}$$

distinct orders for the N's and P's. The probability of getting a sequence of this type is

$$\frac{k(k-1)(k-2)}{6} p^{k-3}(1-2p)^3.$$

Two of the N's could be swapped out for a PE combination. The sequence has one N and one E, with $k(k - 1)$ distinct orders possible. The probability is $k(k - 1)p^{k-1}(1 - 2p)$.

Summing, the probability of getting a sequence in W is

$$w = 2p(1-2p)\sum_{k=2}^{\infty}\frac{k(k-1)}{2}p^{k-2} + (1-2p)^3\sum_{k=3}^{\infty}\frac{k(k-1)(k-2)}{6}p^{k-3}$$

All that remains is to evaluate the two series, and they can both be evaluated using basic power series rules from first year calculus! Recall the geometric series result

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p} \quad for\ |p| < 1.$$

For us, $0 \leq p \leq 1/2$, so our p-values are within the interval of convergence of the series. We can then take derivatives and get

$$\sum_{k=2}^{\infty} k(k-1)p^{k-2} = \frac{2}{(1-p)^3}$$

and

$$\sum_{k=2}^{\infty} k(k-1)(k-2)p^{k-3} = \frac{6}{(1-p)^4}.$$

Substituting these values for the series into the previous expression for w, we get

$$w = 2p(1-2p)\frac{1}{(1-p)^3} + (1-2p)^3\frac{1}{(1-p)^4} = 1 - \left(\frac{p}{1-p}\right)^4$$

The mean score for a Hardy golfer is thus

$$\text{mean} = 4 + p\left[1 - \left(\frac{p}{1-p}\right)^4\right]$$

which for reasonable values of p is very close to $4 + p$. For example, if $p = 0.1$ the mean is about 4.09998, and if $p = 0.2$ the mean is about 4.1992.

Assuming that all holes are played independently, the mean score for a hole can be extrapolated to the mean score for an entire round. If $p = 0$ and all 18 holes are par 4, the mean score is 72. If $p = 0.1$, the mean increases to about 73.8. If $p = 0.2$, the mean increases to about 75.6. The more erratic the golfer, the higher the mean score.

Different Strokes

Of course, having a lower mean score does not necessarily imply that you will win a majority of your matches. The type of match being played has a strong influence on who wins. In stroke play (also called medal play), each player counts all strokes over 18 holes and the lower total wins. The probabilities of players making different scores on a given hole can be combined to compute the probabilities of a player having a particular score for the entire round. Comparing these probabilities, a Hardy golfer with $p = 0$ will beat a Hardy golfer with $p = 0.1$ in stroke play 63% of the time. The two will tie 10% of the time, and the more erratic ($p = 0.1$) golfer wins 27% of the time.

The percentages are similar but not quite as one-sided in a match between a Hardy golfer with $p = 0.1$ and a Hardy golfer with $p = 0.2$. The more consistent ($p = 0.1$) golfer wins 57% of the time, 6% are tied, and the more erratic golfer wins 37%. This can be illustrated by simulating the scores for two hypothetical Hardy golfers over 20 rounds:

$p = 0.1: 72, 72, 69, 78, 73, 76, 75, 73, 76, 74, 71, 75, 72, 72, 75, 78, 73, 75, 70, 74$

$p = 0.2: 79, 71, 67, 81, 83, 74, 78, 77, 80, 71, 75, 90, 72, 71, 79, 73, 82, 74, 74, 76$

Notice that the more erratic golfer ($p = 0.2$) is actually quite erratic, recording the best score (67) and the worst score (90). The more consistent golfer has the lower score 12 times (60%), the higher score 7 times (35%), and there is one tie.

Match Play and Best Ball

In match play, each hole is a separate contest, and the player who wins the most holes wins the match. This is the competition that Hardy analyzed. Continuing with a comparison between a consistent golfer with $p = 0.1$ and an erratic golfer with $p = 0.2$, the consistent golfer wins 36.3% of the holes, the erratic golfer wins 35.8% of the holes, and 27.9% of the holes are halved. Over an 18-hole match, the consistent golfer wins 46% of the matches, the erratic golfer wins 43% and 11% are tied.

As Hardy commented, the edge goes to the consistent player. However, the edge is relatively small, and the erratic golfer is definitely better off playing match play than playing stroke play. Thus, the common golf folklore that these players are better off competing in match play is correct.

The erratic players have their revenge in competitions involving more players. In "best ball" matches, two teams of two players each compete. The score for a given team on a hole is the minimum of the two scores posted by the players on that team. This competition encourages risk-taking because even if a gamble does not pay off for one player, the other player on the team could still have the best score on a hole.

A team of two erratic Hardy golfers, each with $p = 0.2$, has the advantage over two consistent $p = 0.1$ Hardy golfers. The pair of erratic players wins 38% of the holes while losing only 26%. This team of two erratic players also has the advantage over a team consisting of one erratic $p = 0.2$ Hardy golfer and one consistent $p = 0.1$ Hardy golfer. The "mixed" team loses 35% of the holes and wins 29%. In the best ball format, the more erratic players a team has, the better.

Tournament Jeopardy

An interesting statistical "paradox" arises when a large group of Hardy golfers compete simultaneously in a single tournament: players with the *highest* average scores are most likely to win. If we look back at the simulated list of scores from the previous discussion and imagine that they represent twenty different consistent golfers ($p = 0.1$) and twenty different erratic golfers ($p = 0.2$), four of the top five scores—including the winning score of 67—belong to erratic players. The consistent golfers have a lower mean score, but the much larger variance of the erratic golfers gives them a higher potential for both low and high rounds. You might expect that the consistent players will fare better in a four-round tournament, and this is indeed the case. The longer the tournament is, the less likely it is that the erratic golfer will continue to have a run of good luck.

> An interesting statistical paradox arises when a large group of Hardy golfers compete simultaneously in a single tournament: players with the *highest* average scores are most likely to win.

In *A Mathematician's Apology*, Hardy puts forth a compelling argument that "mathematical reality" is in fact more "real" than the material world of the physicists—which did not deter Hardy from being an avid sports fan (most notably cricket). Whatever his motivations, Hardy was certainly aware that his "Hardy golfers" were, at best, very rough approximations of the real thing. Nevertheless, his model arguably provides a simple basis for the comparison of different strategies in various competitive formats—and yields some surprisingly elegant results.

This is a trade-off that would have satisfied Hardy.

Further Reading

A Mathematician's Apology and *Littlewood's Miscellany* are readily available. Hardy's original paper appeared in the December 1945 issue of *Mathematical Gazette*. The more recent paper "On a Theorem of G.H. Hardy concerning golf" by G.L. Cohen appears in the March 2002 *Mathematical Gazette*. A companion piece related to this article is available online at `http://www.mathaware.org/mam/2010/essays/`.

About the author: Roland Minton is a professor of mathematics at Roanoke College and is co-author of a series of calculus books with Bob Smith. He is currently working on his golf game and on a book about the mathematics of golf, although he confesses to making more progress on the latter.

email: minton@roanoke.edu

DOI: 10.4169/194762110X495416

March Madness to Movies

Tim Chartier, Amy Langville, and Peter Simov

The College of Charleston seniors' class project had a succinct goal: apply math to a real-world problem. Neil Goodson and Colin Stephenson tackled an area of national interest—basketball. They focused on ranking the teams in the NCAA Division I men's basketball tournament. In the end, their model integrated data from more than 5,000 regular season and conference title games and produced a bracket for the 2008 tournament known as March Madness. National interest in this tournament is enormous as reflected by the millions of people who fill in their brackets using their own algorithms, whether it be carefully scrutinizing team line-ups or ranking teams according to a fondness for a school's colors. Word of Goodson and Stephenson's mathematically produced bracket spread and the two were interviewed on National Public Radio. In the course of the discussion, Goodson warned the NPR listeners that their model predicted a first round upset of 6th-seed Southern California by 11th-seed Kansas State. Within hours, reality agreed with the model. The media coverage soon expanded with the seniors discussing their work on a segment of *The Early Show* on CBS.

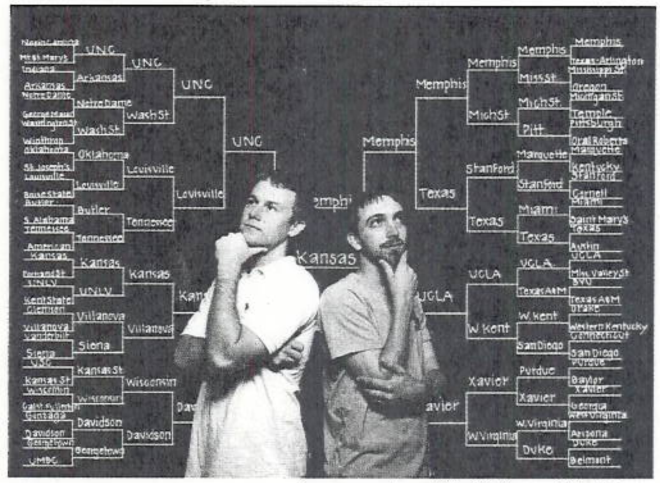

Image courtesy of the College of Charleston.

As undergraduates, Neil Goodson and Colin Stephenson received national media coverage for their mathematical model of the 2008 NCAA men's basketball tournament.

The accuracy of their work in the 2008 tournament continued when all four teams in the Final Four matched the model's predictions. Even further, the model predicted that Kansas would beat Memphis in the final game, although no one could have possibly predicted the buzzer-beating shot that sent the nail-biting game into overtime!

> **Within hours reality agreed with the model's prediction, and soon after the seniors were discussing their work on *The Early Show* at CBS.**

Motivated by this work, let's explore ranking at a deeper level. Note the two ingredients of Goodson and Stephenson's work—a lot of data and an algorithm to produce the rankings. Rather than rank sports teams, since Goodson and Stephenson were so successful in the context of the 2008 NCAA tournament, let's rank movies.

Going to the Movies

The data is a downloadable dataset from Netflix consisting of more than 100 million ratings from over 480,000 randomly-chosen, anonymous users on nearly 18,000 movie titles. The ratings are on a scale from one to five (integral) stars (with five stars being the best rating, and one star the worst) and were collected between October, 1998 and December, 2005. While we could apply the following methods to all movies in the dataset, let's limit ourselves and rank the movies that received the Academy Award for Best Picture between 1994 and 2003 as listed in Table 1. While each film grabbed the coveted Oscar among the many films considered that year, our goal will be to explore how these films stack up against each other.

To answer this question, we will use the ratings of the over 480,000 users in the Netflix dataset who rated at least one of these ten movies. (*Forrest Gump*, *Gladiator*, and *A Beautiful Mind* can be rented with or without extended material. The results that follow correspond to ratings associated with DVDs without extended material.) How do we glean useful information from all these data? This question sits squarely in the field of data mining which is "the science of extracting useful information from large datasets or databases." (See the article by Hand, Mannila, and Smyth in the "Further Reading" section.) This field of computer science is growing in the sciences given the enormous datasets, like the one we are using, produced by modern computational and experimental methods.

Year	Movie Title		Year	Movie Title
1994	Schindler's List		1999	Shakespeare in Love
1995	Forrest Gump		2000	American Beauty
1996	Braveheart		2001	Gladiator
1997	The English Patient		2002	A Beautiful Mind
1998	Titanic		2003	Chicago

Table 1. Movies awarded the Academy Award for Best Picture between 1994 and 2003.

With data in hand, we now need a ranking algorithm. Let's begin with a method that is very easy to implement and rank the movies according to the number of times they were rated; the movie rated the most is ranked the highest. This produces Table 2 and has an obvious flaw. A movie that is rated with one star by 5,000 users would be ranked higher than a movie that was rated with five stars by 4,000 users. This clarifies our goal. We are interested in the quality of the ratings and not simply the tendency of a movie to be rated.

Rank	No. Ratings	Title
1	181,508	Forrest Gump
2	154,832	American Beauty
3	150,592	Gladiator
4	143,668	Titanic
5	135,601	Braveheart
6	113,717	Chicago
7	108,771	A Beautiful Mind
8	101,141	Schindler's List
9	64,957	Shakespeare in Love
10	36,263	The English Patient

Table 2. Top 10 movies selected from films winning the Oscar for Best Picture between 1994 and 2003. Higher ranking implies a higher number of user ratings.

Table 3 shows what happens when we rank the movies according to their average number of stars. *Schindler's List* now tops our list since 101,141 users rated it higher on average than the 181,508 users who rated *Forrest Gump*.

Rank	Avg. Rating	Title
1	4.458	Schindler's List
2	4.300	Forrest Gump
3	4.294	Braveheart
4	4.203	Gladiator
5	3.975	A Beautiful Mind
6	3.963	American Beauty
7	3.867	Shakespeare in Love
8	3.710	Titanic
9	3.594	Chicago
10	3.474	The English Patient

Table 3. Top 10 movies ranked according to their average rating.

This approach also has a potential flaw although possibly not as noticeable in popular films like those that win Academy Awards. Applying this idea to the entire dataset produces Table 4. Suppose another five users watched *Trailer Park Boys: Season 4* and each rated it with one star. This film would plummet out of the top ten whereas the ratings of another ten or even 100 users would hardly impact the top three films in this table.

Rank	Avg. Rating	No. Ratings	Title
1	4.723	73,335	Lord of the Rings: The Return of the King: Extended Ed.
2	4.716	73,422	The Lord of the Rings: The Fellowship of the Ring: Extended Ed.
3	4.702	74,912	Lord of the Rings: The Two Towers: Extended Ed.
4	4.670	7,249	Lost: Season 1
5	4.638	1,747	Battlestar Galactica: Season 1
6	4.605	1,633	Fullmetal Alchemist
7	4.6	25	Trailer Park Boys: Season 4
8	4.6	75	Trailer Park Boys: Season 3
9	4.595	89	Tenchi Muyo! Ryo Ohki
10	4.593	139,660	The Shawshank Redemption: Special Edition

Table 4. Top 10 movies, among all movies in the Netflix dataset, ranked according to their average rating.

The two algorithms we applied in this section demonstrate how simple algorithms can help uncover the types of results we desire, but also reveal their weaknesses. For a better ranking method, let's turn to some of the ideas in Goodson and Stephenson's work.

A Big Bowl of Rankings

Among other methods, Goodson and Stephenson incorporated ranking methods used by the Bowl Championship Series (BCS) system that is used to select NCAA football bowl matchups. One of the computer rating algorithms of the BCS is the Colley method, introduced by Wesley Colley, which modifies one of the simplest and oldest rating systems. Winning percentage rates team i with $p_i = w_i/t_i$, where w_i is the number of wins and t_i is the total number of games played by team i. Although simple and easy to use, this rating method does not factor in the strength of opponents. Defeating the weakest or the strongest opponent results in the same increase in a team's rating, which is arguably unfair. Colley proposed applying Laplace's rule of succession, which transforms the standard winning percentage into

$$r_i = \frac{1+w_i}{2+t_i}. \qquad (1)$$

So, at the beginning of the season every rating is $1/2$, and as the season progresses the ratings deviate above or below this starting point. In fact, the average of all the ratings will remain $1/2$ throughout the season. This insight and the modest modification to the traditional winning percentage formula in (1) will lead us to the Colley method.

Letting l_i be the number of losses for team i, we can decompose the number of games won as

$$w_i = \frac{w_i - l_i}{2} + \frac{w_i + l_i}{2} = \frac{w_i - l_i}{2} + \frac{t_i}{2},$$

and notice that

$$\frac{t_i}{2} = \frac{1}{2}\overbrace{(1+1+\cdots+1)}^{\text{total number of games}} = \left(\frac{1}{2}+\frac{1}{2}+\cdots+\frac{1}{2}\right).$$

Since the ratings hover around 1/2,

$$\frac{1}{2}(\text{total games}) = (\text{sum of opponents' ranks for all games played}),$$

or

$$\frac{1}{2}t_i \approx \sum_{j \in O_i} r_j,$$

where O_i is the set of opponents for team i. This substitution is approximate as the average over all opponents' ratings may not be 1/2 since, for one thing, every team may not play every other team. Substituting this back into our equation for w_i we get

Two movies compete for the highest rating from a user, so a "game" occurs between two movies when a user ranks *both* movies with distinct ratings.

$$w_i \approx \frac{w_i - l_i}{2} + \sum_{j \in O_i} r_j,$$

and because $r_i = (1 + w_i)/(2 + t_i)$,

$$r_i = \frac{1 + (w_i - l_i)/2 + \sum_{j \in O_i} r_j}{2 + t_i}, \qquad (2)$$

giving the Colley method and the interdependence of ratings.

With a little more algebra and matrix notation, (2) can be written compactly as a linear system $C\boldsymbol{r} = \boldsymbol{b}$, where $b_i = 1 + 1/2(w_i - l_i)$ and C is

$$c_{ij} = \begin{cases} 2 + t_i & \text{if } i = j \\ -n_{ij} & \text{if } i \neq j \end{cases}$$

where n_{ij} is the number of times team i plays team j. It can be proven that the matrix C is invertible so, for any season, the Colley method will produce a unique ranking.

When Movies Compete

Returning to movies, let's rank a small sample of four films rated by six users. We first create a matrix U where u_{ij} is user i's rating of movie j with 0 indicating that a user did not rate a movie. So,

$$U = \begin{array}{c} \\ \text{User 1} \\ \text{User 2} \\ \text{User 3} \\ \text{User 4} \\ \text{User 5} \\ \text{User 6} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Movie 1} & \text{Movie 2} & \text{Movie 3} & \text{Movie 4} \end{array} \\ \begin{pmatrix} 5 & 4 & 3 & 0 \\ 5 & 5 & 3 & 1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 2 & 0 \\ 4 & 0 & 0 & 3 \\ 1 & 0 & 0 & 4 \end{pmatrix} \end{array}.$$

In this example, user 1 ranked movies 1 and 2 with a 5 and 4, respectively.

How can we glean win and loss information from these data as required by the Colley method? Two movies compete for the highest rating from a user. So, a "game" occurs between two movies only when a user ranks *both* movies with distinct ratings. We can now form a movie matrix M where m_{ij} is the number of users who ranked movie i higher than movie j. As with the Colley method, we overlook ties although incorporating such information is possible and requires changes in the derivation that led to (2).

It's time for movies 1 and 2 to compete! Only users 1 and 2 rated both movies with user 2 rating them the same. So, $m_{12} = 1$ and $m_{21} = 0$ since user 1 rated movie 1 higher than movie 2. Continuing this process,

$$M = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 2 & 2 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}.$$

From M, we can form the linear system for the Colley method. Let v_j be the column sums of M, and let h_i be the row sums. Then,

$$c_{ij} = \begin{cases} 2 + h_i + v_i & \text{if } i = j \\ -(m_{ij} + m_{ji}) & \text{if } i \neq j \end{cases}$$

and $b_i = 1 + 1/2(h_i - v_i)$.

Using our data, we get

$$C = \begin{pmatrix} 8 & -1 & -2 & -3 \\ -1 & 6 & -2 & -1 \\ -2 & -2 & 7 & -1 \\ -3 & -1 & -1 & 7 \end{pmatrix} \text{ and } \boldsymbol{b} = \begin{pmatrix} 3.0 \\ 2.0 \\ -0.5 \\ -0.5 \end{pmatrix}, \text{ so } \boldsymbol{r} = \begin{pmatrix} .674 \\ .621 \\ .349 \\ .356 \end{pmatrix}.$$

$$M = \begin{pmatrix} 0 & 17954 & 12265 & 32523 & 13801 & 40275 & 24347 & 34482 & 28664 & 18351 \\ 11597 & 0 & 4241 & 14938 & 10704 & 18805 & 10643 & 13494 & 11855 & 9453 \\ 25359 & 19159 & 0 & 30556 & 14843 & 34743 & 23176 & 29735 & 28766 & 21030 \\ 13088 & 11125 & 5153 & 0 & 9352 & 20434 & 11553 & 17342 & 13815 & 10609 \\ 5142 & 4936 & 1660 & 6968 & 0 & 8544 & 4598 & 5827 & 5135 & 4378 \\ 15341 & 12342 & 6958 & 23345 & 11036 & 0 & 12665 & 25672 & 17077 & 12491 \\ 33710 & 22031 & 15510 & 39386 & 16095 & 50733 & 0 & 42196 & 35331 & 26824 \\ 25533 & 18452 & 11037 & 29853 & 14067 & 39911 & 24322 & 0 & 26267 & 19885 \\ 15290 & 13813 & 6779 & 25719 & 12052 & 28219 & 13660 & 22419 & 0 & 13367 \\ 25826 & 18428 & 13258 & 32121 & 13848 & 39495 & 24105 & 34005 & 29868 & 0 \end{pmatrix}$$

Table 5. The rows (and columns) 1 through 10 correspond to the movies *Gladiator, Shakespeare in Love, Schindler's List, Chicago, The English Patient, Titanic, Forrest Gump, American Beauty, A Beautiful Mind,* and *Braveheart.*

The Colley method ranks the films (from highest to lowest) as movie 1, 2, 4, and 3.

If we apply this method to our Oscar winning films data set, the matrix U contains 10 columns and over 480,000 rows. After a flurry of movie to movie competitions we get the matrix M displayed in Table 5.

To check your understanding, you can verify that

$$C_{11} = 2 + h_1 + c_1 = 2 + 222662 + 170886 = 393550$$

and

$$C_{12} = -(m_{ij} + m_{ji}) = -29551 = C_{21}$$

since C is symmetric. Finally, $b_1 = 1 + 1/2(h_1 - c_1) = 25889$. Solving the resulting $Cr = b$ yields the rankings in Table 6.

Rank	Rating	Title
1	.733	Schindler's List
2	.643	Forrest Gump
3	.624	Braveheart
4	.569	Gladiator
5	.497	American Beauty
6	.455	A Beautiful Mind
7	.447	Shakespeare in Love
8	.364	Titanic
9	.357	Chicago
10	.312	The English Patient

Table 6. Top 10 movies, among movies awarded the Oscar for Best Picture between 1994 and 2003, ranked using the Colley method.

While this ranking is similar to Table 3, the underlying method is quite different given the interdependence of ratings in Table 6. A lower ranked film benefits when a user rates it higher than a higher ranked film. This difference in derivation would be more pronounced when we increase the number of movies. ■

Closing Credits

While we have ranked movies with a sports ranking algorithm, a variety of other questions could be tackled with such a large dataset. Suppose I rate *Braveheart* with five stars; what other film might you recommend based on my other movie ratings and all other Netflix user ratings? Netflix recently offered a million dollar prize for accurately producing such a recommendation. The winning team created a recommendation system that predicted ratings just over 10% better than Netflix's recommendation system. Interested? While Netflix recently decided not to run a second competition, according to the official Netflix blog, they "will continue to explore ways to collaborate with the research community…So stay tuned." (See the entry for March 12, 2010 at `http://blog.netflix.com`.) Whatever lies on the research horizon in the science of recommendations and ranking for Netflix and other companies, it will likely involve mathematics, modeling and computer science.

Further Reading

For an introduction to the tools for studying large data sets check out *Principles of Data Mining* by D.J. Hand, H. Mannila and P. Smyth (The MIT Press, 2001). Colley's description of his ranking algorithm appears in "Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained," available online at `http://www.colleyrankings.com/method.html`.

About the authors: *Tim Chartier and Amy Langville are professors of mathematics at Davidson College and the College of Charleston. Their collaborative work often marries mathematics and computer science in a search for meaning in large datasets via clustering or ranking methods. Among their favorite movies are The Princess Bride and Good Will Hunting. Peter Simov began exploring the use of linear algebra in data mining and ranking the day after graduating from Davidson College as a summer research project that involved ranking movies with Netflix ratings data. That fall, he began graduate study at North Carolina State University; among his favorite movies is The Matrix.*

email: tichartier@davidson.edu
langvillea@cofc.edu
prsimov@ncsu.edu

DOI: 10.4169/194762110X495452