A note on the importance of the first goal in a National Hockey League game

Jack Brimberg and W.J. Hurley*

Department of Business Administration,
The Royal Military College of Canada,
Kingston, Ontario, Canada, K7K 5L0
E-mail: Jack.Brimberg@rmc.ca
E-mail: hurley-w@rmc.ca
*Corresponding author

Abstract: In their pregame assessments of which team will win an important hockey game, TV commentators will often point out the importance of scoring the first goal, suggesting that a team improves its chances of winning considerably by scoring it. In this note, we use some simple probability theory to calculate how a team's chance of winning (and losing) changes after the first goal.

Keywords: conditional probability; first goal; NHL; National Hockey League; Poisson distribution; exponential distribution.

Reference to this paper should be made as follows: Brimberg, J. and Hurley, W.J. (2009) 'A note on the importance of the first goal in a National Hockey League game', *Int. J. Operational Research*, Vol. 6, No. 2, pp.282–287.

Biographical notes: Jack Brimberg is a Professor of Operations Research at the Royal Military College of Canada, where he holds cross-appointments in Business Administration and Mathematics. He is an associate member of the Groupe d'Études et de Recherche en Analyse des Décisions, an Adjunct Professor at l'École des Hautes Études Commerciales, Montréal, and a Professional Engineer. He has served as Department Chair at RMC, and Dean of the School of Business, University of Prince Edward Island. His research interests are mainly in discrete and continuous location theory, but also include mathematical programming, heuristic methods, and OR models and applications in general.

W.J. Hurley is a Professor of Operations Research at the Royal Military College of Canada. His research interests are in decision analysis, game theory, military operations research, and the application of operations research to problems in sport.

1 Introduction

The television pregame show for National Hockey League (NHL) games now follows a predictable pattern. Invariably, the announcers will assess how the teams matchup in order to predict who will win. Some commentators will point to the importance of scoring the first goal suggesting that a team improves its chances of winning considerably by scoring it. This is especially true for playoff games which tend to be played more defensively.

We thought it would be interesting to try and measure the importance of the first goal. Consequently we calculate the probability that the team scoring it wins the game. We do this at a number of discrete points characterised by the number of minutes remaining in the game. Our work has the flavour of Gill's (2000) in that he considers the chance of a late-game reversal. It is also related to the literature on the Pythagorean Winning Percentage Formula, the phenomenological law that relates a team's end-of-season winning percentage with the 'points' scored for and against over a season. It was first discovered by the noted sabermatrician Bill James (James, 1983). He posited that a baseball team's end of season Win Ratio (number of games won over the number of games played) is related to the average number of runs per game scored, RS, and given up, RA, by

$$\text{Win Ratio} = \frac{\text{RS}^{\alpha}}{\text{RS}^{\alpha} + \text{RA}^{\alpha}} \tag{1}$$

where α is a parameter. This relationship has been tested extensively in other sports including hockey and it tends to work remarkably well. A number of authors have developed similar laws based on queuing theory. Ryder (2004) has developed one for hockey assuming that the goal-scoring streams of two teams are generated by independent Poisson processes. Miller (2006) assumes independent Weibull distributions for runs scored in a baseball game to derive the Pythagorean Law from first principles.

We also tackle a related issue. In the playoffs, as the third period winds down, commentators sometimes assert that the game is essentially in overtime meaning that the next goal wins. Hence we also measure how big a limb the colour man is stepping out on by calculating the probability he is correct.

2 The importance of the first goal

With team A playing team B, suppose that:

- team A has just scored the first goal of the game
- there are T minutes left in regulation time.

For sports like hockey and soccer where scoring is relatively infrequent, scoring times have been shown to follow a Poisson process. For hockey, this literature would include Mullet (1977), Hurley (1995), Danehy and Lock (1995), Berry (2000) and Gill (2000), and an interesting series of papers on when to pull the goalie (see in particular Morrison and Wheat, 1986; Erkut, 1987; Nydick and Weiss, 1989; Washburn, 1991; Zaman, 2002). Research on times for soccer goals includes Chu (2003), Dixon and Coles (1997) and Lee (1997). Thomas (2006) is an exception; he uses a truncated exponential to take into account the fact that there is a faceoff at centre ice after a goal and hence a short fixed period of time before either team would have a scoring chance. However, for our purpose of modelling the number of goals over a complete game, we feel the Poisson assumption is sufficiently accurate.

The implication is that the number of goals scored over a fixed period of time follows a Poisson distribution. So suppose the number of goals, N_T, scored over the regulation time remaining follows a Poisson distribution with parameter λ. That is, the probability that there are exactly n goals scored in the remaining T minutes is given by

$$P(n) = \Pr(N_T = n) = \mathrm{e}^{-\lambda T} \frac{(\lambda T)^n}{n!}, \quad n = 0, 1, 2, \ldots \tag{2}$$

For the purpose of evaluating the importance of the first goal, we assume that the teams are competitive. Specifically, if a goal is scored, it is just as likely to be scored by team A as by team B. Hence, if $N_T = n$ goals are scored, the probability that team A scores exactly n_A of these, $p_n(n_A)$, follows a binomial distribution:

$$\begin{aligned} p_n(n_A) &= \binom{n}{n_A} \left(\frac{1}{2}\right)^{n_A} \left(\frac{1}{2}\right)^{n-n_A} \\ &= \binom{n}{n_A} \left(\frac{1}{2}\right)^{n}, \quad \text{for } n_A = 0,1,\ldots, n. \end{aligned} \tag{3}$$

(Alternatively, we could use the team's relative standing in the league to estimate the probability it scores the next goal, i.e., the probability of success in each of n independent trials.)

Assuming that team A has a one goal lead and that exactly n goals are scored in the T minutes remaining, it is straightforward to write expressions for the probability that team A wins, $\phi_n(A)$, depending on whether n is odd or even:

$$\phi_n(A) = \begin{cases} \sum_{i=n/2}^{n} p_n(i) & \text{if } n \text{ is even} \\ \frac{1}{2} p_n((n-1)/2) + \sum_{i=(n+1)/2}^{n} p_n(i) & \text{if } n \text{ is odd.} \end{cases} \tag{4}$$

In the case where n is odd, and, of these n goals, team A scores one less goal than B, the game goes into overtime, and, at that point, the teams are equally likely to win. This is the reason for the term, $1/2p_n((n-1)/2)$, in the expression above.

We can now calculate the probability that team A wins after scoring the first goal with T minutes remaining in regulation time, $\phi(T)$, by conditioning on how many goals are scored in the remaining time:

$$\phi(T) = P(0) + \sum_{n=1}^{\infty} \phi_n(A) \cdot P(n). \tag{5}$$

To calculate $\phi(T)$, we need to estimate λ, the parameter of the Poisson distribution generating the goals. To do this, we generated a sample of 300 observations of the time between goals over the 2005–2006 NHL regular season. The average time between goals was 13.56 minutes. Consequently, we estimate λ to be 1/13.56.

Table 1 shows the probability that team A wins for various values of T and two values of λ, $\lambda = 1/13$ and 1/14. Note that these two values bracket the estimate of 1/13.56.

Table 1 Probability that the team scoring the first goal wins for various parameter assumptions

Time left	$\lambda = 1/13$	$\lambda = 1/14$
55	0.688	0.694
45	0.706	0.713
35	0.731	0.738
25	0.766	0.774
15	0.822	0.830
5	0.920	0.925

These values are high. A priori, team A has a 50% chance of winning. However, if it scores first at the 5 minute mark of the first period (there are 55 minutes left to play), it increases its chances of winning to 70%. If this goal is scored later in the game, the probability increases. For instance, if team A scores the first goal with 5 minutes remaining in the second period (25 minutes left to play), its probability of winning goes to almost 80%.

3 When does overtime start?

This same approach can be used to calculate other probabilities that announcers speculate about. For instance, with approximately five minutes to go in the third period of Game 6 of the 2004 Stanley Cup Finals between the Tampa Bay Lightning and the Calgary Flames, with the game tied, the CBC's colour commentator, Harry Neale, remarked that the game was essentially in overtime. It was his view that the next goal scored would win the game. This statement has often been made late in the third period of a tied playoff game.

In such situations announcers are correct if at most one more goal is scored in the remaining regulation time. This happens with probability

$$\Pr(N_T = 0 \text{ or } 1) = e^{-\lambda T}(1 + \lambda T). \tag{6}$$

Table 2 shows values of this probability for various values of T and the two bracketing values for λ; again, these probabilities are high. If an announcer makes the claim with somewhere between 5 minutes and 10 minutes left in the third period, there is about a 90% chance he will be correct.

Table 2 Probabilities that a tied game is effectively in overtime

	Pr ($N_T = 0$ *or* 1)	
T	$\lambda = 1/13$	$\lambda = 1/14$
20	0.545	0.582
15	0.679	0.710
10	0.820	0.839
5	0.943	0.950

4 Summary

Over the last ten years a number of researchers have pointed out the value of using sports problems in a statistics classroom. The interested reader is referred to Albert (2002, 2003) and Cochran (2004) for good examples. We think the problems in this paper are a nice addition to this work. They motivate a number of important topics including the exponential, Poisson, and binomial distributions, probability trees, and the use of conditioning to calculate complex probabilities.

While we have not dealt with soccer explicitly, we note that one could make similar calculations for that sport.

Acknowledgements

We would like to thank the Editor and referees for their helpful comments on an earlier draft of this paper. In addition we thank Tim Corcoran and David Hurley for their careful collection of the data for this analysis.

References

Albert, J. (2002) 'A baseball statistics course', *Journal of Statistics Education*, Vol. 10, Available at www.amstat.org/publications/jse/v10n2/albert.html

Albert, J. (2003) *Teaching Statistics using Baseball*, Mathematical Association of America, Washington DC, USA.

Berry, S. (2000) 'A statistician reads the sports pages: my triple crown', *Chance*, Vol. 13, pp.56–61.

Chu, S. (2003) 'Using soccer goals to motivate the Poisson process', *INFORMS Transactions on Education*, Vol. 3, No. 2, Available at http://ite.pubs.informs.org/Vol3No2/Chu/

Cochran, J.J. (2004) *You Can't Spell SpORts Without O.R., But You can Teach O.R. with Sports!. OR/MS Today*, August, Available at www.lionhrtpub.com/orms/orms-8-04/freducation.html

Danehy, T. and Lock, R. (1995) 'CHODR: using statistics to predict college hockey', *STATS*, Vol. 13, pp.10–14.

Dixon, M.J. and Coles, S.G. (1997) 'Modeling association football scores and inefficiencies in the football betting market', *Applied Statistics*, Vol. 46, pp.165–180.

Erkut, E. (1987) 'More on Morrison and Wheat's – pulling the Goalie revisited', *Interfaces*, Vol. 17, pp.121–123.

Gill, P.S. (2000) 'Late-game reversals in professional basketball, football, and hockey', *The American Statistician*, Vol. 54, pp.94–99.

Hurley, W.J. (1995) 'Overtime or shootout: deciding ties in hockey', *Chance*, Vol. 8, pp.19–22.

James, B. (1983) *Baseball Abstract 1983*, Ballantine, USA.

Lee, A.J. (1997) 'Modeling score in the Premier League: Is Manchester United really the best?', *Chance*, Vol. 10, pp.15–19.

Miller, S.J. (2006) *A Derivation of the Pythagorean Won-Loss Formula in Baseball*, Available at arXiv:math.ST/0509698v4

Morrison, D.G. and Wheat, R.D. (1986) 'Misapplications reviews – pulling the Goalie revisited', *Interfaces*, Vol. 16, pp.28–34.

Mullet, G.M. (1977) 'Simeon Poisson and the National Hockey League', *The American Statistician*, Vol. 31, pp.8–12.

Nydick, R.L. and Weiss, H.J. (1989) 'More on Erkut's More on Morrison and Wheat's pulling the Goalie revisited', *Interfaces*, Vol. 19, pp.45–48.

Ryder, A. (2004) *Win Probabilities: A Tour through Win Probability Models for Hockey*, Available at www.HockeyAnalytics.com

Thomas, A.C. (2006) 'The impact of puck possession and location on ice hockey strategy', *Journal of Quantitative Analysis in Sports*, Vol. 2, No. 1, Accessed at http://www.bepress.com/jqas/Vol2/Iss1/6

Washburn, A. (1991) 'Still more on pulling the Goalie', *Interfaces*, Vol. 21, pp.59–64.

Zaman, Z. (2002) 'Coach Markov pulls Goalie Poisson', *Chance*, Vol. 14, pp.31–35.