Journal of Quantitative Analysis in Sports

Volume 3, *Issue* 3 | 2007 | *Article* 5

Inter-arrival Times of Goals in Ice Hockey

Andrew C. Thomas*

*Harvard University, athomas@stat.harvard.edu

Inter-arrival Times of Goals in Ice Hockey*

Andrew C. Thomas

Abstract

Previous studies have attempted to model goal scoring in sports such as ice hockey as simple Poisson processes. Others (Thomas, 2006) have shown that events within the game of ice hockey are better modelled as a Semi-Markov process determined by puck possession and location. I demonstrate that a similarly defined Semi-Markov process model is well-suited to describe the times between goals scored in NHL hockey, and use this to demonstrate that the scoring of a goal has the effect of shortening the remainder of the game by roughly 20 seconds. This is used to improve previous estimates of the value of a goal scored, calculated as a difference in win probabilities at specific time points during the game.

KEYWORDS: hockey, semi-Markov process, hazard function, survival analysis

*I would like to thank the Harvard Sports Analysis Workshop, in particular Carl Morris, for valuable discussions. I would also like to thank Alan Schwarz for the consultation that produced this line of inquiry, and my anonymous reviewers for very helpful comments.

1 Introduction

It has been suggested that the scoring of goals in single-point running clock games is adequately modelled by a Poisson process, such as in earlier analyses of ice hockey [1][6] and soccer [2]. In the later case, this assumption is supported by the notion that soccer goals are especially rare events, and that the time until the next goal, starting at any particular time point, can be well-approximated by the Exponential distribution.

Goals in ice hockey may also be considered rare events. They are rare when compared to professional basketball, where the total number of scoring events is on the order of 100 in a typical 48 minute game. In comparison, the mean number of goals in a 60-minute hockey game is roughly six. If the time between goals can be adequately modelled as Exponentially distributed, scoring times will be well described as a Poisson process. Another indication that a Poisson model is appropriate is if the distribution of goals within each clock-minute during the game is roughly uniform.

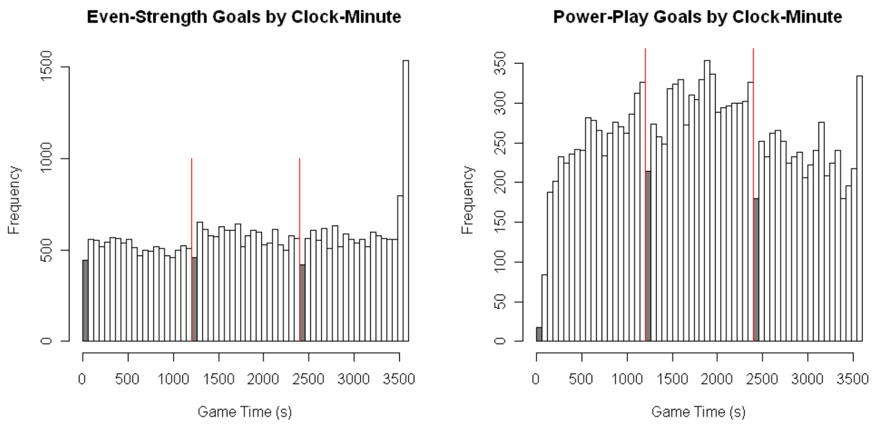
Figure 1 shows the goals scored in each clock-minute of games played in four NHL seasons comprising roughly 4700 games [9]. Three trends are evident:

- The number of goals scored at even strength is of the same magnitude for every minute in the game except the last. This difference is plausibly explained by the strategy of a team losing by one goal to replace their goalie with an extra skater in a desperate attempt to tie the game. This results in more goals scored for both sides; equalizers for one team, empty net goals for the other.

- The beginning of every period sees a lower number of goals than any other, of the even-strength or power play variety. This can be explained by the nature of the game; the first event to occur is a faceoff at center ice, which is near neither team's goal, making a goal less likely to occur.

- The number of power play goals scored at the beginning of a period is lower than for any other minute, most dramatically in the opening two minutes of the game. Since most penalties incurred are two minutes long, the initial shortfall is likely due to the lack of penalties called before the beginning of the first period, a situation not true of the second or third periods. The lower totals there are more readily explained by the

Figure 1: The distribution of goals by minute, in even-strength and power play situations, in NHL hockey games during four seasons of play (4708 games). Vertical lines denote breaks in periods, and dark bars represent the first minute of each period.



initial seconds "lost" by each team after the center-ice faceoff to begin the period.

These anomalies suggest that the simple Poisson model of the game, notably the consequence that goals are scored with an equal rate at any time of the game, is not sufficient to describe the game's dynamics. In particular, since the Poisson model requires that the time between goals is exponentially distributed, a model that uses a generalized form of the exponential to model the time between goals is suggested to improve understanding of the game. If the time until the next goal depends only on the time since the previous goal and the score in the game, this is equivalent to a Semi-Markov process [7], where the states are labelled by the difference in score. It then remains to determine which class of distributions best describe the inter-arrival times between goals and estimate the relevant parameters.

2 Data and Model

Goal-scoring data for the National Hockey League has been accessible through their website for several seasons. I make use of nearly four full seasons of NHL games in the 2002-2003, 2003-2004, 2005-2006 and 2006-2007 seasons,[1] during which 25,786 goals were scored during regulation time. (For the purpose of this study, overtime goals are disregarded, since with only four skaters per team in overtime it's essentially a different game.) This works out to 5.48 goals per sixty minutes - however, the average goals per game in each season varies more widely (5.17, 5.06, 5.96, 5.69 respectively). In particular, a set of rule changes and a stronger enforcement of existing rules were introduced specifically to increase scoring.[2] It therefore makes sense to consider each season separately when fitting the model. Observations are taken to be the time between goals scored, noting which team does the scoring.

In addition, there is a subset of observations that is right-censored. Two types of censoring events effectively break up the play of a game: the end of each 20-minute period, after which play resumes with a faceoff at center ice, and in the endgame when a trailing team substitutes their goaltender for an extra attacking player; the likelihood of a goal being scored is subsequently much higher. Since the fraction of censored observations is non-negligible, survival analysis is used to investigate goal-scoring rates. This requires the consideration of the full likelihood

$$L(X, Y|\theta) = f(X|\theta)S(Y|\theta),$$

where the likelihood is composed of the density function for all observed data X, and of the survival function $S(t) = 1 - F(t)$ for all right-censored data Y.

I consider two possible distributions for the time between goals. The first is the Weibull distribution, commonly used in many survival analyses for its versatility; the second is a distribution where the instantaneous failure rate begins small, then approaches a constant, which by definition is New Better

[1] 1095, 1217, 1222 and 1174 games were used respectively; omitted games were due to the inability to access the games of record.

[2] Preliminary investigation shows that the jump in goals scored starting in 2005 is largely attributable to increased power play scoring. As I consider power play goals to be scored homogeneously throughout the game (as seen in Figure 1), I do not explicitly model these changes.

than Used[3]. I refer to this as the Plateau-Hazard (PH) distribution. The former takes the Exponential distribution as a special case; the latter will be chosen such that the Exponential is a limiting case.

2.1 Pulling The Goaltender

If one team is trailing in the game, the coach may choose to replace the goaltender with an extra attacking player. The risk in this case would be that the opposing team would score a goal on an empty net and extend their lead, so the timing of the exchange is usually limited to the last minute or two of regulation time, when the team's likelihood of scoring a tying goal under ordinary circumstances is already quite small.

During this period, the likelihood of one team scoring is markedly increased, since the trailing team has one extra offensive player and their net is unguarded. The situation is quite different from an ordinary game situation, and so is taken into account separately: when a goaltender is pulled, an indication is made that the next interval is being played with an empty net. In the data being considered, a goaltender was pulled in 64% of games played. In 34% of those games a goal was scored with one net empty, and of those goals 30% were scored by the team that pulled their goaltender.

The data set, however, does not directly specify whether the substitution is being made because the team must catch up, or because the goalie was pulled due to an impending penalty against the opposing team - a circumstance in which the opponents cannot score, since if they make contact with the puck, or are scored upon, play immediately halts. So in this case, I restrict instances of when the goalie is pulled to be only those in the last two-and-a-half minutes of the game, in which the likelihood that a switch is due to a trailing score is much higher than if a penalty call is pending.

Instances of the goalie being pulled are included as the censoring of the next goal that would be scored under regular game circumstances, and represent the shift in the system to empty-net status. If a tying goal or a goal by the leading team is scored, the goalie is replaced and play continues as before.

3 Measurement: Goal Scoring Rates as a Function of Game Score

Goals are grouped according to the absolute difference in score just before the goal was scored; that is, the opening goal of a game is in group 0, and a goal which puts the score at $5-0$ is in group 4. Goals scored when the score difference was 4 or greater were aggregated.

The data are defined in terms of fully observed data,

$$X_{ij}, i \in 0,1,2,3,4, j \in [1,...,m_i],$$

and right-censored data,

$$Y_{ij}, i \in 0,1,2,3,4, j \in [1,...,n_i],$$

where m_i and n_i are the total counts of observed and missing data for each group i.

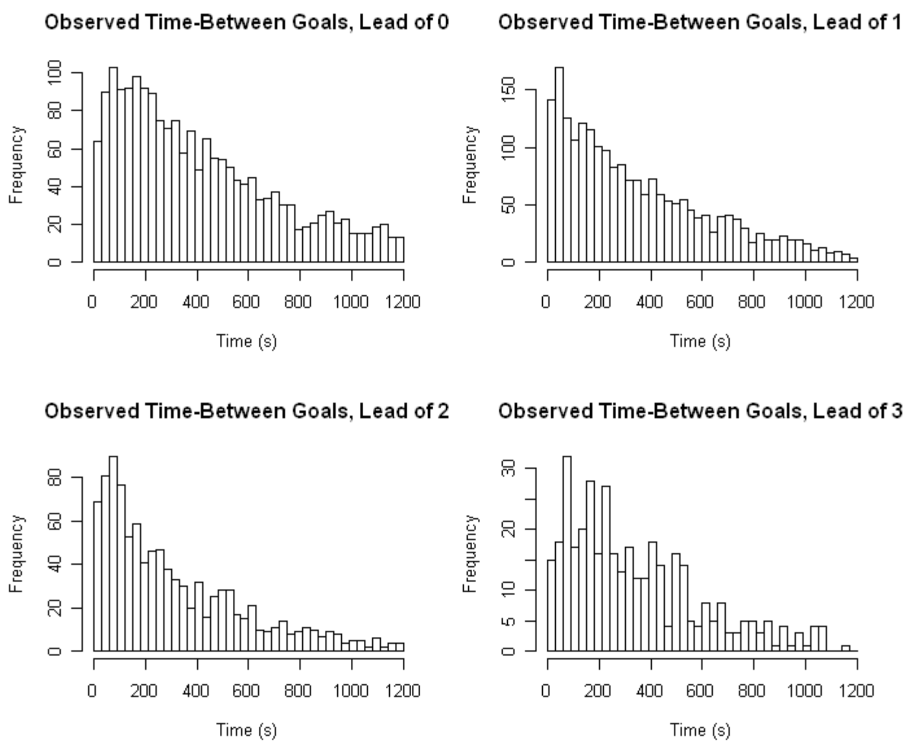
Preliminary inspection shows that goals are not scored quickly in succession - that is, scoring times are not maximized near zero, as expected in the Exponential distribution. The shape of the graphs, however, suggests that there is some perturbation on the Exponential that can describe the scoring time. I therefore choose two distributions that are generalizations of the Exponential distribution in this investigation.

3.1 Candidate Distributions

For a renewal process, all intervals between events must be independent, though not necessarily identically distributed. As play resumes with a face-off at center ice after each goal, I assume that a team's performance before a goal is scored is determined entirely by the state of the system after the last goal occurred - namely, the difference in goals - as well as whether there is an empty net.

A distribution used to model the time between goals must account for the lower-than-expected probability of short scoring times, suggesting that candidates should have low hazard functions at earlier times. Two strong candidates are generalizations of the Exponential distribution:

Figure 2: Scoring times in NHL play, stratified by the difference in score just before scoring. While the curves look exponential in nature, lower counts near zero suggest otherwise.



The Weibull Distribution

The Weibull is defined by the hazard function

$$h(t|k,\lambda) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1}, t \geq 0,$$

which is identical to that for $Exp(\lambda)$ if $k = 1$, and New Better than Used if $k \geq 1$. Note that if k is slightly above 1, the hazard function will start at zero, rise quickly, and then rise very slowly as time increases. The resulting probability functions are

$$S(t) = e^{-\left(\frac{t}{\lambda}\right)^k}, f(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} S(t).$$

This distribution has the consequence that the hazard rate increases without bound until a goal is scored; that is, it does not approach an asymptotic scoring rate. This random variable can be simulated by noting the definition by representation

$$W(k,\lambda) \sim (X/\lambda)^k$$

where $X \sim Exp(1)$.

The Plateau-Hazard Distribution

This distribution has the hazard function

$$h(t|r,w) = r(1 - e^{-t/w}), t \geq 0,$$

and is motivated by the idea that after a faceoff, the system is determined to be in a state where goal-scoring is not immediately expected - in particular, one team has possession of the puck on their half of the ice. The goal-scoring rate asymptotically approaches an equilibrium level r after an interval of time, the magnitude of which is measured by w. The resulting probability functions are

$$S(t) = e^{-rt+rw(1-exp(-t/w))}, f(t) = r(1 - e^{-t/w})S(t).$$

Draws from the PH distribution are obtained using acceptance-rejection sampling, using the proposal distribution $T_p \sim Exp(r)$.

An interesting observation is that, in the case where $w \ll 1/r$, the expected value of the Plateau-Hazard is roughly the expected value of $Exp(r)$ plus the warm-up time w (see Appendix).

3.2 Model Fitting Results

The models were evaluated using Bayesian estimation, for each season, each difference in score, and whether or not one goaltender has been pulled. Prior distributions for each parameter were chosen to be exponentially distributed with very high mean (10^6 in practice) and hence very high variance, which are naturally constrained to be positive. Each probability space was explored using the Metropolis algorithm. (See Appendix for summaries of fitting estimates.) Model comparison was performed through Bayes Factor Analysis, by numerically integrating over the probability spaces of both models [4]. The threshold for strong preference for one model is taken to be a Bayes factor of 1000 or greater (or, conversely, 0.001 or less). A full table of Bayes factors is in the Appendix; in summary, the Plateau-Hazard distribution is strongly preferred over the Weibull unless the difference in score is three or more, or one net is empty.

There is good reason to suspect that this sort of model comparison is inappropriate when the faction of censored data is high, as in the cases with high difference in score or empty-net data. The prime motivation of the Plateau-Hazard model is to show that the scoring rate reaches a maximum after a time. Estimates of the warm-up time are on the order of 20 seconds, with a maximum scoring rate just above one goal every six minutes. Since the net is rarely empty with more than two minutes remaining, the majority data that would be informative about asymptotic behaviour are missing.

4 Model Validation: Win Probabilities

I apply the scoring model to the problem of goal valuation. This has been addressed in other articles by using a Poisson process model [1][6], and determining the probability of winning a game conditioned on the difference in score. The assumption has been made throughout the analysis that all teams score at the same rate given the state of play (their net empty, the opponent's net empty, no empty nets).

The quantities of interest, the probability of winning given score differential and time, can be estimated by simulating a large number of games (in this case, the number of games played in each season) and noting which team won the game based on the score at various selected time points. Games are simulated by the generation of scoring times drawn from the plateau-hazard distribution, with the goal awarded to either team with equal probability. The game ends when a goal scored is generated to have taken more time than how much remains in the game.

By comparison, estimates under the Poisson model are calculated directly using the observed scoring rate as the parameter of interest, with scoring rates unaffected by the goal differential.

I choose not to explicitly model when in the game a goaltender will be pulled. Instead I draw directly from the game samples if the difference in score is one goal or two. I assume that with less than three minutes to go, overcoming a three-goal deficit is so unlikely that its inclusion in the win probability equation would be futile, and would likely result from a freak occurrence, such as severe injuries to the starting and backup goaltenders respectively. If a goal is scored with an empty net, I use the aforementioned estimate that the leading team scores 70% of the time under those circumstances.

All estimates calculated with this method, for each model, are given in Table 1. Both the Plateau-Hazard and Poisson models predict values that are highly consistent with the observed data. In particular, the match between reality and the PH model for a two-goal deficit with one minute remaining is encouraging; it suggests that the effective time lost due scoring the first catch-up goal reduces the odds of the tying goal being scored. (This observation is purely speculative, since the errors on each quantity effectively wipe out any significant difference between win probabilities.)

5 Application: Goal Value

Now that I have a model-based estimate of win probabilities, an estimate for the value of a goal can be produced by subtracting adjacent values.

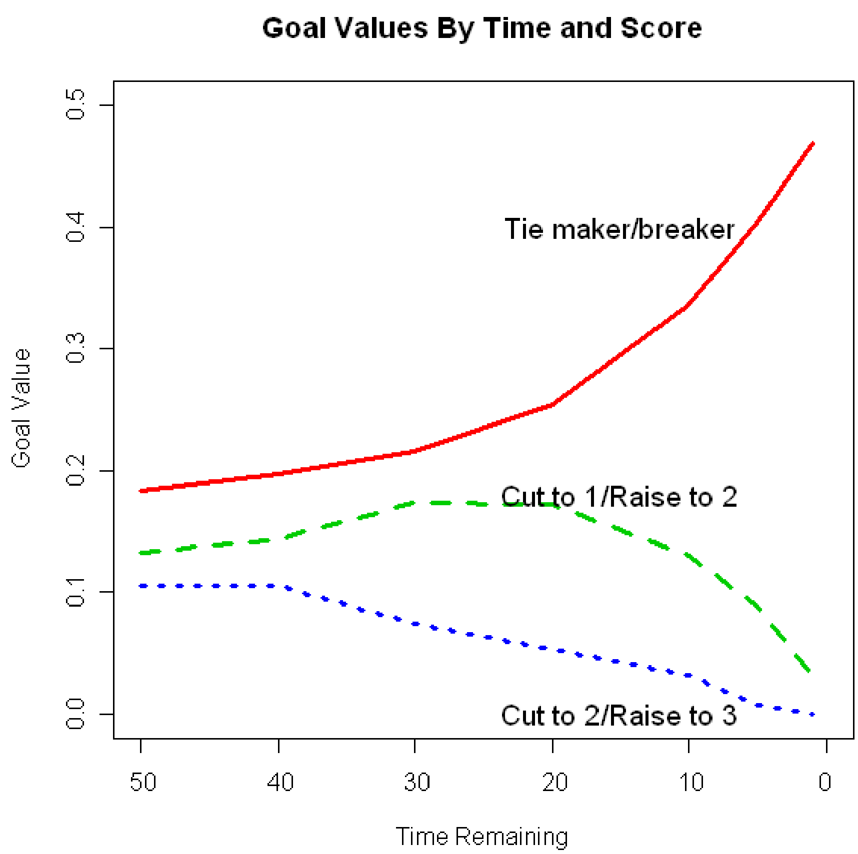
Time left	Lead	P(win),True	P(win),PH	PH SD	P(win),Pois	Pois SD	P(win),Naive Pois
40	1	0.6663	0.6866	0.0198	0.6878	0.0176	0.6935
	2	0.8077	0.8361	0.0236	0.8322	0.0213	0.8419
	3	0.9519	0.9282	0.0286	0.9258	0.0299	0.9311
20	1	0.7144	0.7523	0.0183	0.7535	0.0158	0.7627
	2	0.9104	0.9073	0.0142	0.9058	0.0146	0.9168
	3	0.9806	0.9737	0.0100	0.9713	0.0118	0.9775
10	1	0.8146	0.8202	0.0141	0.8237	0.0155	0.8376
	2	0.9582	0.9580	0.0073	0.9551	0.0083	0.9661
	3	1.0000	0.9920	0.0057	0.9913	0.0054	0.9949
5	1	0.8686	0.8797	0.0105	0.8787	0.0124	0.9013
	2	0.9868	0.9843	0.0045	0.9821	0.0055	0.9888
	3	1.0000	0.9980	0.0021	0.9978	0.0026	0.9991
1	1	0.9461	0.9498	0.0075	0.9470	0.0081	0.9764
	2	0.9978	0.9981	0.0018	0.9969	0.0025	0.9994
	3	1.0000	1.0000	0.0000	0.9999	0.0005	1.0000

Table 1: For the 2005-2006 season, the observed and model-estimated probabilities that a team with the indicated lead will win a game if a certain amount of time remains. The models considered are the Plateau-Hazard and Poisson, which take the empty net into account, and the Naive Poisson, which does not. Both models assume an equal level of ability between teams. Means and standard error estimates come from simulating the season 100 times, considering a single season as the base unit.

Define *goal value* as the change in the probability of victory caused by the goal. This idea has been considered to evaluate pitchers in baseball since the studies of Lindsey [5], by comparing the probabilities of victory before and after a pitcher's appearance in the game. For example, a game-tying goal with 5 minutes remaining shifts the probability of winning for the scoring team to 0.5 from 0.097, an increase of 0.403. The maximum score for a single goal is 0.5, for a game-tying or go-ahead goal made just before the end of regulation time.

As seen in Figure 5, the highest value goals in a game are scored near the end of regulation time that tie the score, preventing a loss in regulation, or break a tie, gaining a regulation win. This view is especially consistent with the strategy of replacing the goaltender with an offensive player in the last minutes of regulation time in an attempt to tie the game, since the value of an opposing goal is far less than that achieved by scoring the tying goal.

Figure 3: Each line represents the value of a goal given the score of the game and the remaining time in regulation. Note that a tying or go-ahead goal is always more valuable than a lead-growing or -shrinking goal.



Discussion

I have wondered for some time whether valuing goals later in games - in so-called 'clutch' situations - is indeed a useful enterprise. The Poisson paradigm suggests that goals are uniformly distributed throughout a game, and in a high-scoring game, a tying goal must have been proceeded by others which figure equally in the final score - scorekeeping does not discriminate between goals scored in the sixth or sixtieth minute of a game, and neither do the leagues that give postseason awards for scoring and goaltending based on raw summary statistics. In the end, the decision to define an official "clutch factor" proportional to a goal's value is not a scientific decision but a standard to be considered by each hockey league.

One useful product of the above methodology is to consider the reverse direction. Rather than measuring scoring rates to determine probabilities of victory, one could affect rule changes designed to set scoring rates which would produce a desired probability of victory at one point in the game, such as increasing the size of the ice surface or the net, which would likely increase the number of scoring chances and the success rate respectively. There has been a perceived economic interest in games where a two-goal lead after two periods does not result in almost-certain victory, in terms of advertising revenues for television broadcasts. Rules that would ensure a higher rate of scoring (though unspecified) would produce that effect, but being able to translate a desired 'competitiveness' factor into an average number of goals per game would be ideal.

References

[1] Griffel, David. "A Poisson Model for Goal Scoring in Hockey and Some of Its Applications". Harvard University Senior Thesis (unpublished). [1995]

[2] Hirotsu, Nobuyoshi & Wright, Michael. "Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions". Journal of the Operational Research Society, 53(1). [2002]

[3] Hollander, Myles & Proschan, Frank. "Testing whether new is better than used". The Annals of Mathematical Statistics, 43(4). [1972]

[4] Kass, Robert E. & Raftery, Adrian E. "Bayes Factors". Journal of the American Statistical Association, Vol 90, Issue 430. [1995]

[5] Lindsey, George. "An Investigation of Strategies in Baseball", Operations Research, 11(4). [1963]

[6] Morrison, D.G. "On the optimal time to pull the goalie: A Poisson model applied to a common strategy used in ice hockey." TIMS Studies in Management Science, Vol. 4. [1976]

[7] E. Pekoz, S.M. Ross. "Estimating the Mean Cover Time for a Semi-Markov Process Via Simulation". Probability in the Engineering and Informational Sciences, 11, p. 267-271. [1997]

[8] Thomas, Andrew C. "The Impact of Puck Possession and Location on Ice Hockey Strategy". Journal for Quantitative Analysis in Sports, 2(1). [2006]

[9] NHL scoring data for each of the four seasons is available through NHL.com. Missing games are due to three principal causes: the relevant files are missing, the files are corrupted and their full game records are incomplete and therefore omitted, or the games had yet to be played at time of this writing.

A Tables of Calculated Quantities

PH Goal Difference	$r \times 10^{-3}$	95% Interval	w	95% Interval	Mean Time	$1/r$
0	1.57	(1.50,1.64)	22.8	(16.1,30.5)	659	637
1	1.60	(1.53,1.67)	19.3	(13.9,25.6)	644	625
2	1.67	(1.56,1.78)	21.5	(13.1,31.3)	619	599
3	1.71	(1.55,1.89)	21.9	(11.17,35.9)	606	585
4+	1.68	(1.49,1.88)	26.2	(11.35,46.3)	620	595
Empty Net						
1	7.97	(6.33,10.00)	7.68	(2.56,15.61)	133	125
2	8.39	(6.90,10.00)	5.18	(1.47,10.39)	124	119

Table 2: Bayesian estimates for Plateau-Hazard parameters, used in describing the time between goals, during the 2006-2007 season. The parameter r is the asymptotic scoring rate; w represents the warm-up time parameter. 95% confidence intervals are the central intervals as drawn from the Metropolis algorithm. Note that the difference between the mean scoring time and $1/r$ is roughly w.

Weibull Goal Difference	λ	95% Interval	k	95% Interval	Mean Time
0	668	(641,696)	1.104	(1.067,1.143)	642
1	648	(624,673)	1.111	(1.076,1.147)	622
2	625	(593,658)	1.118	(1.063,1.167)	600
3	617	(570,664)	1.104	(1.024,1.179)	596
4+	642	(575,717)	1.056	(0.963,1.150)	629
Empty Net					
1	124.1	(105.1,150.9)	1.267	(1.076,1.46)	115
2	125.7	(105.5,153.6)	1.12	(0.9614,1.282)	121

Table 3: Bayesian estimates for Weibull parameters, used in describing the time between goals, during the 2006-2007 season. λ represents the mean of an Exponential distribution which has been raised to a power k.

Regular Play	Tie Game	Lead of 1	Lead of 2	Lead of 3	Lead of 4+
2002-2003	2×10^8	5×10^4	8000	1.97	0.005
2003-2004	5×10^7	1×10^7	32	0.019	0.011
2005-2006	1×10^6	5×10^8	2.7	9.38	0.03
2006-2007	2×10^8	4×10^8	1240	2.88	0.77
Empty Net	Tie Game	Lead of 1	Lead of 2	Lead of 3	Lead of 4+
2002-2003	-	0.003	2.47	-	-
2003-2004	-	0.035	3.27	-	-
2005-2006	-	0.003	0.003	-	-
2006-2007	-	0.014	0.031	-	-

Table 4: Bayes Factors for comparing the goodness of fit of the Plateau-Hazard distribution against the Weibull distribution, with even prior odds of each model. Higher numbers indicate a preference for the Plateau-Hazard. When both goaltenders are on the ice, the difference in score is 0 or 1, the evidence is very strong for the PH; with 2, two seasons show very strong preference for the PH; with 3, both distributions are equally suited, and with 4 or more the Weibull is preferred. When the net is empty, the Weibull distribution is largely preferred.

Time left	Lead	P(win),True	P(win),PH	PH SD	P(win),Pois	Pois SD	P(win),Naive Pois
50	1	0.6458	0.6725	0.0181	0.6702	0.0187	0.6743
	2	0.8368	0.8133	0.0240	0.8091	0.0292	0.8156
	3	0.9211	0.8893	0.0622	0.9039	0.0628	0.9092
40	1	0.6663	0.6866	0.0198	0.6878	0.0176	0.6935
	2	0.8077	0.8361	0.0236	0.8322	0.0213	0.8419
	3	0.9519	0.9282	0.0286	0.9258	0.0299	0.9311
30	1	0.6946	0.7139	0.0182	0.7125	0.0186	0.7206
	2	0.8408	0.8622	0.0171	0.8653	0.0179	0.8750
	3	0.9607	0.9505	0.0198	0.9470	0.0184	0.9545
20	1	0.7144	0.7523	0.0183	0.7535	0.0158	0.7627
	2	0.9104	0.9073	0.0142	0.9058	0.0146	0.9168
	3	0.9806	0.9737	0.0100	0.9713	0.0118	0.9775
10	1	0.8146	0.8202	0.0141	0.8237	0.0155	0.8376
	2	0.9582	0.9580	0.0073	0.9551	0.0083	0.9661
	3	1.0000	0.9920	0.0057	0.9913	0.0054	0.9949
5	1	0.8686	0.8797	0.0105	0.8787	0.0124	0.9013
	2	0.9868	0.9843	0.0045	0.9821	0.0055	0.9888
	3	1.0000	0.9980	0.0021	0.9978	0.0026	0.9991
4	1	0.8880	0.8947	0.0096	0.8923	0.0119	0.9176
	2	0.9887	0.9876	0.0048	0.9867	0.0049	0.9924
	3	1.0000	0.9990	0.0017	0.9986	0.0019	0.9995
3	1	0.8998	0.9120	0.0101	0.9081	0.0107	0.9354
	2	0.9926	0.9913	0.0041	0.9907	0.0047	0.9954
	3	1.0000	0.9995	0.0012	0.9992	0.0015	0.9998
2	1	0.9183	0.9298	0.0093	0.9251	0.0102	0.9549
	2	0.9961	0.9957	0.0026	0.9945	0.0034	0.9978
	3	1.0000	0.9997	0.0010	0.9996	0.0010	0.9999
1	1	0.9461	0.9498	0.0075	0.9470	0.0081	0.9764
	2	0.9978	0.9981	0.0018	0.9969	0.0025	0.9994
	3	1.0000	1.0000	0.0000	0.9999	0.0005	1.0000

Table 5: A win probability table for various selected times remaining.