Win Probabilities

a tour through win probability models for hockey

Hockey Analytics
www.HockeyAnalytics.com

Introduction

There are a large number of winning probability models out there. These have mainly grown up in the baseball world. Some work well for hockey. Some require adaptation. I have also developed some new approaches that are designed specifically for hockey.

The basic reasons for the differences between hockey and baseball are:

- unlike the case with baseball, goal scoring in hockey is a Poisson process,
- variations in scoring are lower in hockey than in baseball, and
- hockey permits ties.

Later I will discuss the importance of these differences. But keep them in mind as I take a brief tour through the methods.

What you will see below is that, in general:

Pr(Win)	*= f(GFg, GAg, U)*	
GFg	*= GF / GP*	*Goals For per Game*
GAg	*= GA / GP*	*Goals Against per Game*
U	*= Unknown Driver(s) of Pr(Win)*	

In any sporting event, one wins by having more credits than debits. In some sports the accounting system is complex. In hockey it is not. Hockey teams win by scoring more goals than they allow. In a single game it is an absolute truth that the team that scores more goals (has a positive goal differential) is the winner. Over the course of a season, however, a team with a positive goal differential only has a **tendency** to win more than it loses. A team that wins half of its games 5-0 and loses the other half 3-2 clearly (a) has a positive goal differential but (b) is only a .500 team.

“U” is tricky to track down. It may relate to intangible abilities in close games. It may relate to the higher “moments” (e.g. variance) of GF and GA. It may be that U is a null set. But GF and GA are by far the biggest drivers of Pr(Win) and I will be sticking with them in this paper.

I think there is actually more interesting work to be done a layer or more down on the “derivatives” (contributing parts) of GF and GA:

$GF = f(X_1, X_2, X_3, \ldots)$
$GA = f(X_1, X_2, X_3, \ldots)$

Understanding the derivatives of GF and GA might give us a better understanding of winning than chasing down U. But I won’t be going there either.

Win probability formulae can be developed from two perspectives. They are either empirical or they are rooted in theory. I will show you a theoretical approach, but first I will cover the empirical approaches. They can be put into 4 basic groups: linear vs. non-linear and static vs. dynamic.

Static Linear Methods

Static linear methods all are based in some way on goals for (GF) minus goals against (GA) or, alternatively, GF per game (GFg) minus GA per game (GAg). The most generic linear form is:

Pr(Win) = K + F x GFg – A x GAg
K, F, A are constants.

This is the regression form of win percentage predictors. Using regression techniques you can theoretically generate a formula that does not weight GF and GA equally. However, every regression I have ever run has given GF and GA about equal weight. Here are my regression results from 1945-46 through 2003-04:

Pr(Win) = 0.501 + (.1456 x GFg - .1458 x GAg)

The use of regression may deny conformity with an underlying theory. In this case the result says than an average team may have a Pr(Win) other than .500. That can't happen. You can force the intercept (K) to be .500 and should when doing such work. If you do that on the same data you get

Pr(Win) = 0.500 + (.1457 x GFg - .1457 x GAg)

This regression explains about 93.4% (R-Square = .934 [1]) of winning since WWII. This is awfully good for a two factor formula. It is an even better result if you consider this to be a one factor formula. The one factor is the Goal Differential (GDg = GFg - GAg) which, of course, requires F = A. The next general form does this:

Pr(Win) = .5 + S x GDg
S = Slope

Clearly a good S is .1457. One common form of this method is:

Pr(Win) = .5+ GDg / GPW

[1] *R-Square is the square of the correlation coefficient of two ordered data sets. It measures the degree to which variation in one data set is explained by variation in the other. In all of my R-Square references one of the data sets will be win percentages in the NHL from 1945-46 to 2003-04 and the other will be the win probability model.*

where GPW is Goals Per Win (GPW is the reciprocal of the slope). In this static linear method, GPW is "constant". It can be constant over all teams and all of time or simply constant across all teams in a given season. If you want GPW to be constant over history, it should be in the 6 to 7 range (1/.1457 = 6.86). The "right" GPW depends on your definition of "all time". The game changes and has its eras.

Although linear methods work pretty well for winning percentages between .300 and .700, they clearly don't work in the tails. A linear method is going to predict that a team with a goal differential of more than 50% of league average goals will win more than 100% of its games. Pr(Win) is clearly not linear in the tails. Teams with very high winning percentages tend to significantly under-perform a linear Pr(Win) prediction. Teams with very low winning percentages tend to significantly over-perform a linear Pr(Win) prediction.

Static Non-Linear Methods

There are also non-linear methods that use constants. Earnshaw Cook was the first to actually publish a Pr(Win) formula for baseball. For hockey it looks like:

Pr(Win) = 484 x GF / GA

This is non-linear because of the division. An obvious problem with this formula is that the average team is predicted to have a .484 winning percentage. This is correctible, at the expense of fit with the data, by using a constant of .500. If you do that, you still have a problem when GF / GA > 2.

Another static non-linear method belongs to Bill Kross. For hockey this would be:

If GF < GA
Pr(Win) =.5 x GF / GA

If GF > GA
Pr(Win) = 1 – .5 x GA / GF

These two methods do not do a good job of reflecting reality. They break down at the extremes. Cook's formula produces a Pr(Win) over 1 with a goals ratio over 2, although it doesn't allow a sub-zero Pr(Win). The Kross formula simply does not provide a very accurate estimation, at least in comparison to other methods, although it does bound Pr(Win) between 0 and 1.

While all of the formulae considered so far will work decently with normal teams in normal scoring contexts, we need methods that work outside of the normal range. There are real .750 teams, and there are teams that play in a significantly larger or smaller than average scoring context. And if we want to apply these methods to individuals at all, we definitely need a more versatile method.

The Pythagorean Relationship

Which brings us to Pythagoras. Bill James' formula, as interpreted for hockey, is:

$Pr(Win) = GF^2 / (GF^2 + GA^2)$

This model takes its name from the Pythagorean Theorem which states that, for a right angle triangle with legs of length A and F and a hypotenuse of length C:

$C^2 = F^2 + A^2$

For example, for a triangle with legs of length 3 and 4, the hypotenuse is of length 5:

$5^2 = 4^2 + 3^2$

GF and GA can be expressed either per game or in aggregate. When GF = 4 and GA = 3 this results in:

$Pr(Win) = 4^2 / (4^2 + 3^2) = .640$

I think of this as the "mass" of the GF divided by the "mass" of the total goals.

The method has an obvious graphical interpretation. If you rotate the triangle (see right) and use some trigonometry you can show that Pr(Win) is W / C where W is the point on C where a perpendicular line intersects the opposite corner of the triangle:

$cos(a) = F / C = 4 / 5$

but also

$cos(a) = W / F$

therefore

$W / F = F / C$

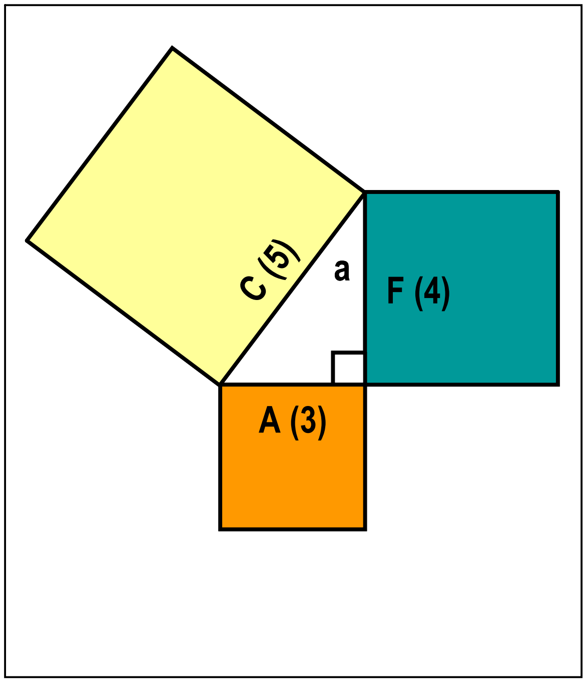
$W = F^2 / C$

$W / C = F^2 / C^2$

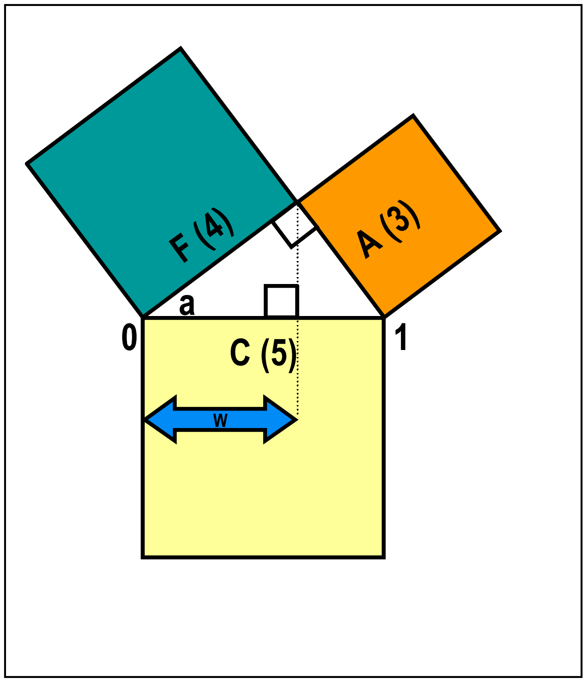
$= F^2 / (F^2 + A^2)$

$= Pr(Win)$

The Pythagorean Relationship

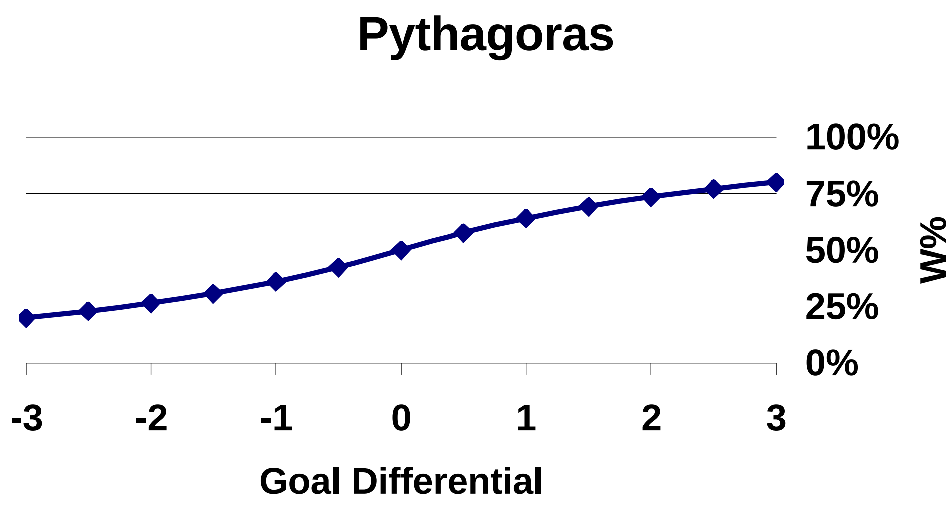


Pythagorean Trigonometry



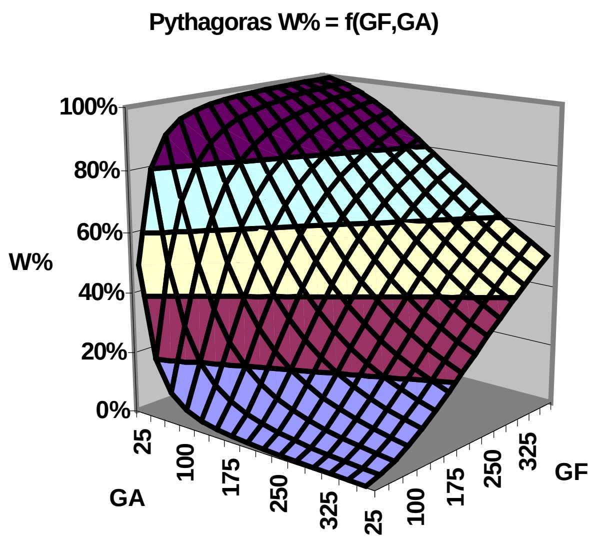
Which means that Pr(Win) can be also determined graphically. It is the percentage of C that is represented by W.

Below is a two dimensional illustration of the result of the Pythagorean method. Because it is a function of two variables, Pythagoras actually gives "surface" of values for Pr(Win). That graph is also shown below.



Pythagoras is the first method that completely satisfies three important criteria which by which we would measure a successful method:

- When GF = GA, Pr(Win) is .500
- Pr(Win) becomes asymptotic to and is bounded by 0 and 1.
- Pr(Win) is symmetrical: exchange GF and GA in the formula to get 1 – Pr(Win).



These attributes, and its relative simplicity, have made it the standard for many years. For certain purposes, I prefer the following equivalent formula, which demonstrates that Pr(Win) is driven by the square of the **ratio** of GA to GF:

$$Pr(Win) = 1 / (1 + (GA / GF)^2)$$

James later claimed that 1.83 was a more optimal exponent for baseball. For hockey a better exponent would be 1.86. An exponent under 2 reduces the expected winning percentage for a given ratio of GF to GA in the formula. For hockey this makes sense because of ties. For baseball (I think) it makes sense because there is more variation in scoring. For basketball (hold on tight) the best exponent looks like 16.5 because there is way more scoring. These observations suggest the possibility of a "Generalized Pythagorean" expression:

$$Pr(Win) = GF^E / (GF^E + GA^E)$$

I will explore this formula in greater depth when we allow E to be a variable.

Dynamic Linear Methods

There is considerable evidence that scoring context matters. A team that averages 6 GF and 3 GA ought to win more games than a team that averages 2 GF and 1 GA. Pythagoras would predict the same result in both cases. But the second team will play a lot more tie games and lose some games to bad luck whereas the first team is more likely to distinguish itself from the competition. **A higher scoring context gives better winning "resolution"** (higher scoring increases the expected winning percentage for a given ratio of GF to GA). A win probability model that reflects that is a better model.

Dynamic models are therefore normally based on a team's average scoring per game:

$$Gg = (GF + GA) / GP$$

The linear methods that do this simply use a formula based on Gg to estimate GPW (or slope) before estimating Pr(Win). The linear versions of these formulae are still subject to the same limitations as the static linear methods -- they are not bounded by 0 and 1. But they do add more flexibility, especially within the normal scoring ranges.

Recall that the basic linear form is generally.

$$Pr(Win) = .5 + GDg / GPW$$

and a dynamic method normally makes:

$$GPW = f(Gg)$$

There are a number of these methods. Here are a couple of the more famous ones (translated from baseball to hockey):

GPW	$= 10 \times \sqrt{(Gg/9)}$	*Palmer*
	$= Gg$	*Smyth*

The Smyth version of the formula is dead simple and produces results that are nearly as good as those of Pythagoras (but with a poor fit in the tails). Why should GPW = Gg? Because every game produces a win (or two "half wins" known as ties).

Another generalized form that produces good results is:

$$Pr(Win) = .5 + S \times GDg$$

where

$$S = M \times \log(Gg) + K$$

I have used this formula and found that it is optimized with M = -.169 and K = .285

Of course, the problems inherent in linear methods are not resolved just by using a flexible slope. The fit is still poor in the tails. Although this is only a problem about 5% of the time, it is still a problem.

Dynamic Non-linear Methods

The generalized Pythagoras form for hockey is:

$$Pr(Win) = GF^E / (GF^E + GA^E)$$

There have been several published attempts to base E on Runs per Game for baseball data. The simplest form you could consider (expressed in hockey terms) would be:

$$E = M \times Gg + K$$

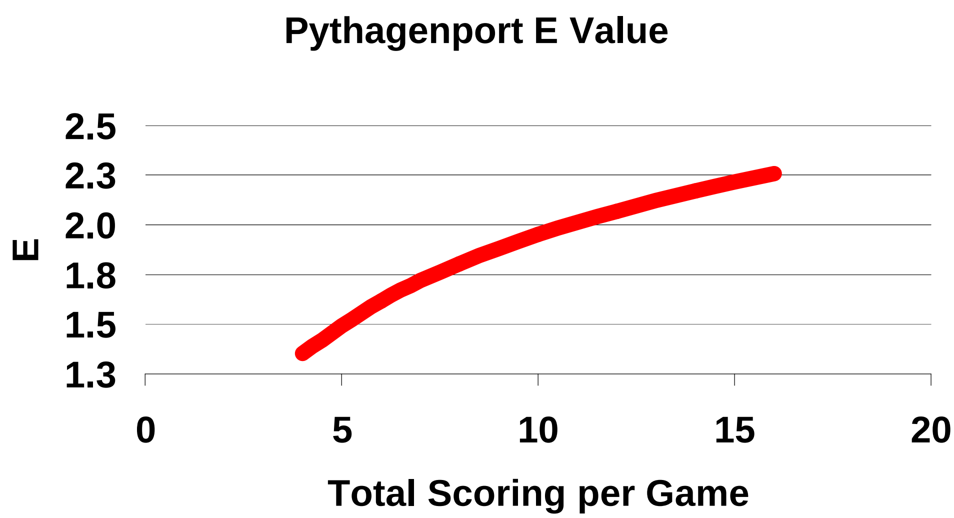
For baseball, Sadowski proposed M = .2174 and K = 0.

Davenport

The most widely known version of the generalized form is attributable to Clay Davenport:

$$E = M \times \log(Gg) + K$$

For baseball, Davenport used M = 1.5 and K = .45 and his resultant version of Pythagoras is known as "Pythagenport". Davenport used some extreme data to find his optimal exponent, which he claims to be accurate for Runs per Game (Rg) ranging between 4 and 40. Below is a graph of Davenport's E values for baseball. I ran Pythagenport through the 1945-46 through 2003-04 seasonal data (not just the tail data) and determined that the



optimal (maximized R-Square) parameters for hockey were K = -.095 and M = 2.74 (R-Square = .941). This is a good result, but a one factor formula would be preferred.

Smyth / Patriot

What about Rg under 4? This is certainly a bigger question in hockey than in baseball. David Smyth observed that the minimum Rg possible in a game is 1, because if neither team scores, the game continues to go on (not so with hockey!). And if a team played all its games at 1 Rg, they would win each game in which they scored a run and lose each time they allowed a run.

In order to make:

$$W / (W + L) = RF^E / (RF^E + RA^E)$$

E must be set equal to 1 when Rg = 1. This is therefore a known point in the domain of the exponent E. The right exponent has to satisfy the following constraints:

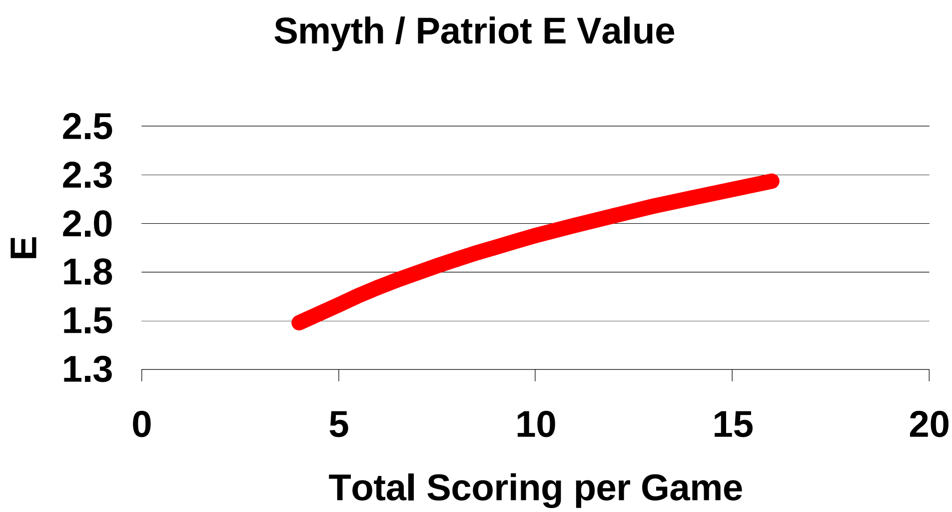
- E = 1 when Rg = 1, and
- be accurate over the normal range of major league teams, and
- E must increase as Rg increases.

An obvious form for E to meet the first condition is:

$$E = Rg^P$$

since 1 raised to any power is 1. In fact, the P of best fit for baseball is .29 and the model using this parameter known as the Smyth / Patriot model (or Pythagenpat).

Smyth / Patriot produces results that look similar to those of Davenport but with a flatter slope for E (see below).



Pythagenport and Pythagenpat are basically equivalent over 5 Gg but the Smyth / Patriot model has the advantages of having:

- application under 4 Gg, and
- only one parameter to estimate.

PythagenPuck: Smyth / Patriot in Hockey

Smyth's logic for E fails us in hockey. Gg = 1 does not imply that E = 1. This is because hockey permits 0-0 ties. Instead, we note that when Gg = 0, Pr(Win) = .500 (every game must be a tie). The problem here is that, for Gg = 0, GF = GA = 0 and the generalized Pythagoras formula:

Pr(Win) $= 0^E / (0^E + 0^E)$
where
E $= 0^P$

is undefined. So we need to come at this from a slightly different angle.

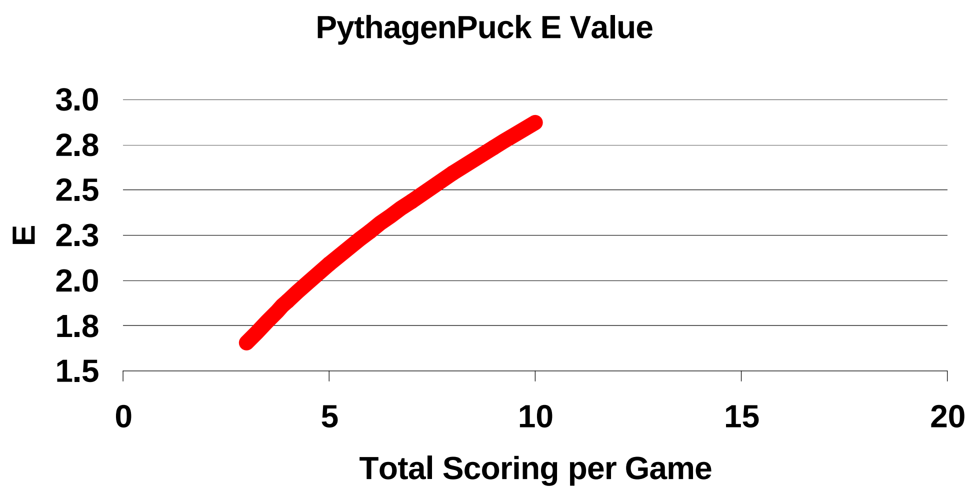
Imagine a team that scores or allows only one goal in its first game of the season and then it ties all of its remaining games 0-0. The more additional 0-0 ties it plays, the more Gg tends towards zero and the closer its record will be to .500. In doing this we have found that the "limit" of Pr(Win), as Gg tends to zero, is .500. This means that, for our formula, we need an E that also causes the "limit" of Pr(Win), as Gg tends to zero, to be .500. It turns out that the Smyth / Patriot model ($E = Gg^P$) will also satisfy this constraint. This can be demonstrated mathematically by taking the limit of the Generalized Pythagorean formula, with $E = Gg^P$, as GF and GA tend to zero. In this case P must be positive.

So PythagenPuck looks like:

Pr(Win) $= GF^E / (GF^E + GA^E)$
where
E $= (GFg + GAg)^P$

The optimal (post WWII) P was .458. This model gives a considerably better fit in the tails than does Pythagoras and produces and R-Square value of .941 (vs .935 for Pythagoras). Over the range of "normal" performance the results are very similar.

Below is a chart of the PythagenPuck E values for P = .458 and in the range of total scoring contexts observed since WWII.



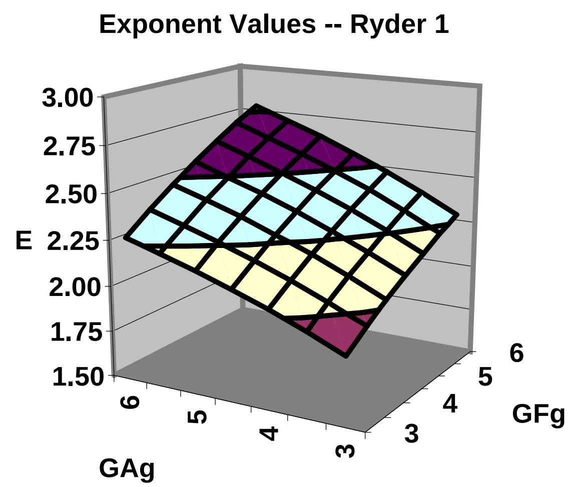
Note that the optimal P and resultant E are considerably larger for hockey than for baseball. But the scoring context for baseball is around 10 runs per game while the average hockey scoring context is more like 6 goals per game. Why is there a big gap for similar levels of scoring? I think it has to do with the differences in the games including the degree of variation in scoring.

Two Ryder Models

One model that I have used produces the same kind of results as PythagenPuck:

E $= (GFg \times GAg)^P$

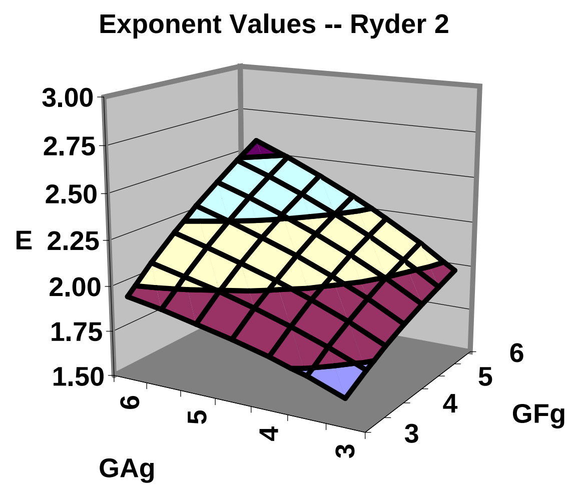
This works because the product of GFg and GAg is another measure of the total scoring context. But, relative to the E for PythagenPuck, this E is larger when GFg and GAg are close and smaller as GFg and GAg diverge (see graph to right). This has the effect, for GFg + GAg = a constant, of increasing the resolution of



the formula when GFg and GAg are close and reducing it when they are not. I don't know if this is a good thing or not, but the results suggest that it is. For this model the optimal P was found to be .285 (R-Square .941).

The second model also uses the product of GFg and GAg and produces a similar result (R-Square .941).

Pr(Win) $= GF^E / (GF^E + GA^E)$
where
E $= M \times (GFg \times GAg) / (GFg + GAg) + K$



When GFg = GAg, the middle part of the formula reduces to GFg / 2. As GFg and GAg diverge, this part of the formula gets smaller. The M and K of best fit are .61 and .74. The resultant exponent values are show in the graph to the right.

You can see that the E values are lower, a little steeper and there is a bit more curvature than with Ryder 1. This method requires the estimation of two parameters and suffers from the potential of a division by zero.

The Law of Cosines

Why does Pythagoras work so well? What is its theoretical underpinning? Why does a dynamic exponent based on total scoring improve the fit? To try to develop answers to these questions, I want to first look at the "Generalized Pythagoras" form:

Pr(Win) $= GF^E / (GF^E + GA^E)$

This can be thought of as the "mass" of goals for expressed as a percentage of the "mass" of total goals.

So let's start by defining a "mass" function M(x,y) to be the "impact" of x and y together on the game and set:

F $= \sqrt{M(GF,0)}$
A $= \sqrt{M(0,GA)}$
C $= \sqrt{M(GF,GA)}$

Furthermore, let's build the same kind of triangle out of F, A and C as we built before when examining Pythagoras in order to graphically determine Pr(Win).

When you build this triangle you will only get a right angled triangle if $M(x,y) = x^2 + y^2$. The Pythagorean relationship works for **right angled triangles**: the square of the hypotenuse is equal to the sum of the squares of the opposite sides. The more general rule for Pr(Win) relies on the graphical determination presented above and the **Law of Cosines**.

The Law of Cosines states that, **for ANY triangle**, the cosine of angle **a** can be expressed in terms of the length of the sides of the triangle:

cos(a) = $(F^2 + C^2 – A^2) / (2 \times C \times F)$

but we also know that, in the right angled triangle with base W:

cos(a) = W / F

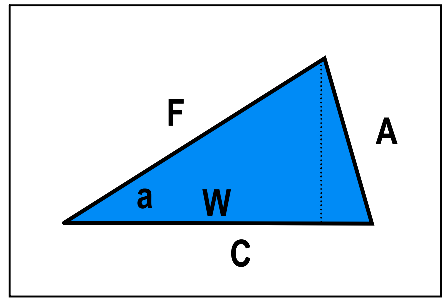
which means that:

W / F = $(F^2 + C^2 – A^2) / (2 \times C \times F)$
W = $(F^2 + C^2 – A^2) / (2 \times C)$

and the general expression for Pr(Win) is:

Pr(Win) = W / C
= $(F^2 + C^2 – A^2) / (2 \times C^2)$
= $.5 + .5 \times (F^2 – A^2) / C^2$
= .5 + .5 x (M(GF,0) – M(0,GA)) / M(GF,GA)

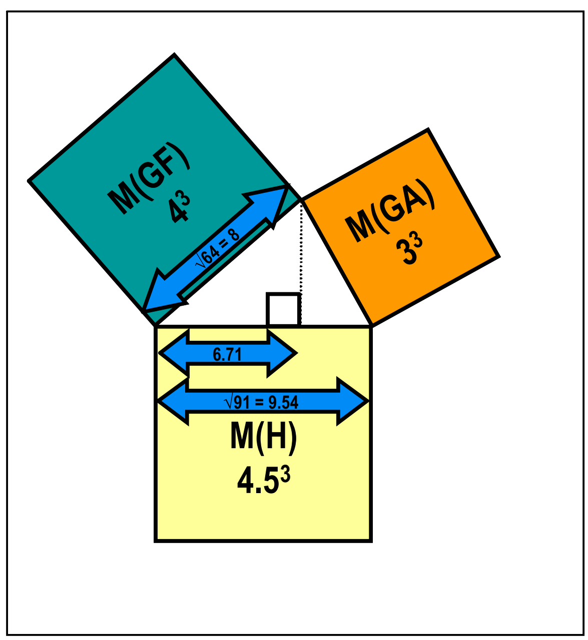
Law of Cosines



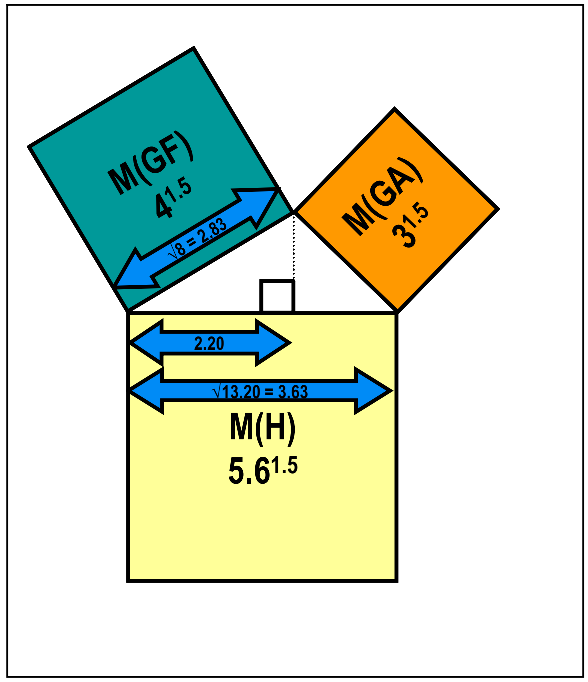
This is true for ANY mass function where M(x,y) ≥ M(x,0) + M(0,y). When M(x,y) = $x^2 + y^2$, this reduces to the Bill James' Pythagorean formula. But, **when E is not 2, Pythagoras won't do**. For any other mass function, including the "Generalized Pythagorean" form M(x,y) = $x^E + y^E$, there are no right angles and it is a discredit to the Law of Cosines to call it a Pythagorean relationship!

Let's look at a couple of examples of the Cosine formula using M(x,y) = $x^E + y^E$ (which I will call a "James" mass function).

First consider the James mass function of M(x,y) = $x^3 + y^3$. In this case the right angle has become an acute (less than 90 degrees) angle. You can see (to the right) that M(GF,GA) < $GF^2 + GA^2$ (the hypotenuse is smaller). The Cosine formula for Pr(Win) gives the following result for GF = 4, GA = 3:



Pr(Win) = .5 + .5 x (M(GF) – M(GA)) / M(GF,GA)]
= $.5 + .5 \times (GF^3 - GA^3) / (GF^3 + GA^3)$
= .5 + .5 x (64 – 27) / (64 + 27)
= .703



Note that, when E > 2, the Pr(Win) predicted for a winning team is higher than it would be if E = 2 (higher resolution).

Next consider the James mass function of $M(x,y) = x^{1.5} + y^{1.5}$. In this case the right angle has become an obtuse angle. You can see (to the right) that $M(GF,GA) > GF^2 + GA^2$ (the hypotenuse is larger). The Cosine formula for Pr(Win) gives the following result for GF = 4, GA = 3:

Pr(Win) = .5 + .5 x (M(GF) – M(GA)) / M(GF,GA)]
= .5 + .5 x $(GF^{1.5} - GA^{1.5}) / (GF^{1.5} + GA^{1.5})$
= .5 + .5 x (8 – 5.2) / (8 + 5.2)
= .606

Note that, when E < 2, the Pr(Win) predicted for a winning team is lower than it would be if E = 2 (lower resolution).

With any sensible mass function the math holds. It does not need to be of the James form. This suggests the possibility of a mass function that recognizes the impact of scoring variation and the covariance between goals scored and allowed. I won't take this further here, but I will leave you with this thought: What is the probability of winning for a team that scores 3 goals in **every** game (no variation) but that allows goals based on a Normal process with mean 2.8 and standard deviation of 0.8? Since there is no variation in goals for, the answer is .599 (the probability of goals against < 3.0). With this example it is easy to see that the variance (and, in fact, the covariance) of GF and GA are considerations in determining the probability of a win.

Ties

A large number of hockey games end up in a tie. Even regular season overtime fails to resolve more than 50% of ties. Because of the history of awarding one point for a tie, it is conventional for hockey analysts to consider ties to be half a win (I define "Win Equivalents" = W + T /2 = WE), but this can distort the results. In baseball a winner is always forced. When a "good" team plays a "bad" team and the game is tied after 9 innings, the teams play on. But, in extra innings, we would agree that there is a better chance for the good team to win than for the bad team. **Extra innings improve the resolution.** "Good" teams have better records than they would if a tie was declared and "half a win" was awarded.

Except in the playoffs, hockey does not require resolution of the game. "Good" teams are frequently stuck with a tie. In the playoffs the game must be resolved. If μ is the intensity of goal scoring by Team A against Team B and ν is the intensity of goal scoring by Team B against Team A, then the probability of Team A winning a playoff overtime game is $\mu / (\mu + \nu)$. If $\mu = 3$ and $\nu = 2$, Team A should win 60% of the time, not half the time as implied by treating a tie as half a win. Because of this treatment of ties, hockey has resolution inferior to that of baseball.

An alternative treatment, ignoring ties, would improve the resolution for hockey. However, we need not lose sleep over this matter. The Pythagorean and Cosine formulae, when applied to hockey, implicitly treat ties as a fraction of a win. In the case of Pythagoras, we have no tool to adjust the resolution of the formulae. But, in the case of the Cosine formula, we can choose a mass function (e.g. a James function with exponent E) to match the observed resolution in the data.

The reduced resolution of winning percentages due to ties is probably the largest single reason that hockey and baseball should not have the same "E" in a James function. We have observed that, in both sports, better resolution occurs in a higher scoring context. I have demonstrated above that better resolution is predicted with a larger "E". Hockey has a different resolution. It needs a different "E".

The Theory: Competing Poisson Probabilities

Goal scoring in hockey is basically a Poisson process. I have written another paper on the applications of the Poisson probability distribution in hockey (see http://www.HockeyAnalytics.com/research_files/Poisson_Toolbox.pdf). In it I derive the probability of a win by z goals:

*Pr(Win by z) = EXP(-(GFg + GAg)) * (GFg / GAg)^(z / 2) * BESSELI(2*SQRT(GFg * GAg),ABS(z))*[2]

This formula holds for all z including 0 and negative values. If you add up these probabilities over all positive z you get the probability of a win. If you add half of the probability of a tie you get the probability of a win equivalent, which is what we have been discussing all along.

Note that Competing Poisson win probabilities predict something that we observe in the data. Resolution improves as total scoring increases. The Competing Poisson model says that this phenomenon is a function of both measures of total scoring: GFg + GAg and GFg x GAg.

This basic Competing Poisson probability model fits the data with an R-Square of .941. This is a result like that of the PythagenPuck and Ryder models.

I also have an enhanced version of the Competing Poisson model that addresses two of hockey's known quirks. Both overtime and what I call the "endgame effect" are non-Poisson aspects of the game. In regular season overtime (i.e. since 1983-84), 5 minutes

[2] *Written in "Excel" for more universal understanding.*

of sudden death hockey is played (4 on 4 since 2000). Because the game ends with a goal, the Poisson probability distribution no longer applies. The endgame effect is the tendency for some one-goal games to turn into ties. When trailing in a close game, coaches pull the goaltender and dramatically change the rate at which goals are scored and allowed. When you adjust for those two things you get the best fit to the data that I have seen (R-Square of .943).

Competing Poisson probabilities are theoretically correct and produce the best results. Does that mean that there is no place for Pythagoras / Cosine models? It turns out that these empirical approaches may not be so empirical after all.

Consider the probability of a win **given** that the game is decided by z goals (i.e. z goal wins or z goal losses):

$$Pr(Win \mid z\ Goal\ Difference) = Pr(Win\ by\ z) / (Pr(Win\ by\ z) + Pr(Win\ by\ -z))$$

If you expand this expression and then simplify the formula you get:

$$= (GFg / GAg)^{(z/2)} / ((GFg / GAg)^{(z/2)} + (GFg / GAg)^{(-z/2)})$$

If you multiply the numerator and the denominator by (GFg / GAg)^(z/2), this becomes:

$$= GFg^{z} / (GFg^{z} + GAg^{z})$$

This has a certain familiarity to it! This is the James form of the mass function to be used in a Cosine model. When z = 2 it is Pythagoras. When we sum across all possible goal differentials we get the probability of a win. But observe that:

$$GFg^{E} / (GFg^{E} + GAg^{E}) = \Sigma_Z\ t_z \times p_z \times GFg^{z} / (GFg^{z} + GAg^{z}),\ z = 0, 1, 2, 3, \dots$$

where p_Z is the probability of a z goal decision and t_Z is .5 for z = 0 (a tie) and 1 otherwise For any given GFg and GAg we can always find an E such that this equation holds. It gives us a rationale for the James form of the mass function (the weighted average, with the t x p factors as weights, of the win probabilities over all goal differentials) and tells us why it is that 2 may not be the best exponent (E is simply the exponent that creates the equality).

Summary

I have taken a tour through many of the simple and more complex win probability models as they apply to hockey. Included in this tour are some approaches that I have developed. In all cases I have fit the models to the data from the NHL from 1945-46 to 2003-04.

Linear methods can work quite well, but fail with extreme data:

$$Pr(Win) = .5 + (GFg - GAg) / GPW$$

For years the state of the art in estimating the probability of a win was the Pythagorean formula developed by Bill James.

$Pr(Win) \quad = GFg^2 / (GFg^2 + GAg^2)$

James acknowledged that the exponent 2 was less than optimal. This gives rise to the Generalized Pythagoras formula:

$Pr(Win) \quad = GFg^E / (GFg^E + GAg^E)$

This is really a special case of the Cosine formula. For any reasonable "mass" function:

$Pr(Win) \quad = .5 + .5 \times (M(GF,0) - M(0,GA)) / M(GF,GA)$

Recently it has been observed that higher levels of total scoring results in improved winning percentage resolution. This suggests that a dynamic E, reflecting total scoring, should fit the data better.

A two factor method is Pythagenport:

$Pr(Win) \quad = GFg^E / (GFg^E + GAg^E)$
where
$E \quad = M \times \log(GFg + GAg) + K$

A one factor method is PythagenPuck:

$Pr(Win) \quad = GFg^E / (GFg^E + GAg^E)$
where
$E \quad = (GFg + GAg)^P$

I have proposed two similar models:

$Pr(Win) \quad = GFg^E / (GFg^E + GAg^E)$
where
$E \quad = (GFg \times GAg)^P$
or
$E \quad = M \times (GFg \times GAg) / (GFg + GAg) + K$

As goal scoring is basically a Poisson process, a Competing Poisson model gives a sound theoretical estimate. Unfortunately it can only be expressed as a sum, making it more difficult to use.

Pr(Win) = Σ_Z EXP(-(GFg + GAg)) * (GFg / GAg)^(z / 2) * BESSELI(2*SQRT(GFg * GAg),ABS(z)), z = 1, 2, 3 …
+ .5 x EXP(-(GFg + GAg)) * BESSELI(2*SQRT(GFg * GAg),0)

Because of hockey's quirks, this basic Competing Poisson model can be enhanced (see http://www.HockeyAnalytics.com/research_files/Poisson_Toolbox.pdf). This approach gives the best results I have yet seen.

Here are the results of some of these methods when compared to NHL results since WWII:

Method	*Parameter*	*R-Square*
Linear		*.934*
Pythagoras		*.935*
Generalized Pythagoras	*1.86*	*.935*
Pythagenport	*-.095, 2.74*	*.941*
PythagenPuck	*.458*	*.941*
Ryder 1	*.285*	*.941*
Basic Competing Poisson		*.941*
Extended Competing Poisson		*.943*

Finally, below is a graph illustrating some of these methods in the tail of their distributions (GA = 3 and GF ranging from 3 to 10). Pythagoras is shown to be weak in the tails. If we consider Competing Poisson to be the truth, then PythagenPuck and Pythagenport seem to bracket reality. My Method 1 looks weaker (and Method 2 would look even weaker). But strengths of my two methods are not well illustrated by this diagram. Pythagenport, PythagenPuck and my methods all have virtually identical R-Square values.