This is a full mathematical explanation of how the Bayesian Resume Rating (BRR) is calculated, including an example for additional clarity.

The goal of the BRR is to rate teams based only on wins, losses, and strength of schedule in an entirely objective manner. The BRR only makes two assumptions. The first is that the talent levels of all teams within a league are normally distributed. The second is that the teams' performances fluctuate in a normally distributed manner from their talent levels. From those two assumptions, all formulas are dictated strictly by mathematical principles.

Other variables, such as margin of victory and venue, are ignored because that would introduce subjectivity. As a result, teams are rated based on what each team has earned, ranking who has the best resume for the season. It doesn't rate who is most likely to win future games. For more conceptual details on the BRR, please read What is the Bayesian Resume Rating?

The BRR is not a simple calculation, but I was intent on finding a mathematically sound, Bayesian solution, even though it meant going beyond basic algebra.

Now, on to the math.

Theory

The BRR has its foundation in statistics, treating teams as weighted random number generators. One of the formulas used is the Bienaymé formula:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

One application of the Bienaymé formula is to calculate how the variances in normally distributed Random Number Generators (RNGs) can be combined. If two or more RNGs have their results added together, the resulting variance will equal the sum of individual RNG variances.

For example, say there is a normally distributed RNG that generates numbers such that $\mu = 2$ and $\sigma^2 = 9$, and another normally distributed RNG that generates numbers such that $\mu = 1$ and $\sigma^2 = 16$. If we sum the outputs of each RNG together, the average and variance of those sums would be $\mu = 3$ and $\sigma^2 = 25$.

If two normally distributed RNGs are compared, it's possible to calculate how frequently one will be greater than the other with the following formula:

$$P_{A>B} = \int_{-\infty}^{\left(\frac{\mu_A - \mu_B}{\sigma_T}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx$$

Where
$P_{A>B}$ = The probability the output from RNG A will be greater than the output from RNG B
μ_A = The average of RNG A
μ_B = The average of RNG B
σ_T = Total standard deviation (calculated using the Bienaymé formula)

So if RNG A has $\mu_A = 2$ and $\sigma_A^2 = 9$, and RNG B has $\mu_B = 1$ and $\sigma_B^2 = 16$, we can first use the Bienaymé formula to find σ_T:

$$\sigma_T = \sqrt{\sigma_A^2 + \sigma_B^2}$$
$$\sigma_T = \sqrt{9 + 16}$$
$$\sigma_T = 5$$

Then we can calculate:

$$P_{A>B} = \int_{-\infty}^{\left(\frac{2-1}{5}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = 58\%$$

Meaning that the probability of RNG A producing a larger number than RNG B is 58%.

This is the basis of the BRR. It assumes that all teams are normally distributed RNGs. A good team will be represented by an RNG with a high average, and a bad team be represented by an RNG with a low average. All teams are assumed to have the same variance (and therefore the same standard deviation). As explained in [What is the Bayesian Resume Rating?](), the argument can be made that these assumptions may not be true, which is certainly a valid argument. Nonetheless, these are logical and fair assumptions, and [evidence exists]() that they are valid.

Lastly, these probabilities can then be fed into Bayes' Theorem, which will give us the most probable rating of each team:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where
P(A) = The prior probability of a given team having a particular level of ability
P(B) = The nominal probability of the team's performance
P(B|A) = The probability of the team's performance, given a certain level of ability
P(A|B) = The updated probability of the team's level of ability, given their performance

Modeling a league's season

As a starting point, for any given league, the average and standard deviation of the league as a whole can be anything. For the BRR, it assumes the league average is 0 and the league standard deviation is 1. From a Bayesian perspective, this is used to establish the prior. Choosing a different league average or standard deviation wouldn't truly change anything. For example, changing the average to 1 would boost every team's rating by 1, but they would all still be in the same order. Choosing the standard deviation to 2 would only double the scale, but it wouldn't change the order either.

It follows that a hypothetical season for a league can be modeled by using random number generators. To simulate an NFL season, we would start by assigning each team a number using an RNG such that μ = 0 and σ = 1.

For example, the talent levels in a league might look like this:

Packers	1.71	Giants	0.66	49ers	0.18	Falcons	0.66
Vikings	-0.41	Cowboys	0.48	Seahawks	-0.32	Saints	0.23
Bears	-0.48	Eagles	0.22	Rams	-0.85	Panthers	-0.44
Lions	-0.71	Redskins	-0.07	Cardinals	-1.05	Bucs	-2.10
Steelers	0.04	Patriots	0.58	Broncos	1.61	Colts	1.77
Browns	-0.15	Dolphins	0.52	Raiders	0.87	Texans	0.94
Ravens	-0.97	Jets	0.08	Chargers	0.46	Titans	0.86
Bengals	-2.83	Bills	-0.13	Chiefs	-0.25	Jaguars	-0.80

Each team has a number assigned to it, and this number will be referred to as a team's Baseline Talent. In this example, the Colts have the highest number, but because games are decided with random number generators, they're not guaranteed to go undefeated, or even necessarily win the most games.

Since this is just an example, I'll choose an arbitrary standard deviation of 1.75 for each individual team's RNG. The number will be referred to as the League Parity. 1.75 is a pretty typical amount of parity for the NFL, though it does vary from year to year. (I'll go into more detail on this number later.)

So, in this example, the Colts will be represented by an RNG such that $\mu = 1.77$ and $\sigma = 1.75$. The Texans will be represented by an RNG such that $\mu = 0.94$ and $\sigma = 1.75$. Using the equations above, we find that Colts have a 63% chance of winning when they face the Texans. The probabilities of all pairings can be found, and RNGs can be used to simulate an entire season. The end of the regular season might look like this…

Packers	12-4	Giants	12-4	49ers	9-7	Falcons	10-6
Vikings	6-10	Cowboys	6-10	Seahawks	6-10	Saints	10-6
Bears	7-9	Eagles	9-7	Rams	8-8	Panthers	6-10
Lions	11-5	Redskins	5-11	Cardinals	9-7	Bucs	4-12
Steelers	9-7	Patriots	8-8	Broncos	11-5	Colts	14-2
Browns	9-7	Dolphins	9-7	Raiders	10-6	Texans	9-7
Ravens	2-14	Jets	5-11	Chargers	10-6	Titans	11-5
Bengals	3-13	Bills	8-8	Chiefs	4-12	Jaguars	4-12

… and result in the Colts having the best record. Or, still using the same inputs, it could look like this…

Packers	13-3	Giants	9-7	49ers	10-6	Falcons	9-7
Vikings	8-8	Cowboys	9-7	Seahawks	9-7	Saints	13-3
Bears	6-10	Eagles	10-6	Rams	8-8	Panthers	4-12
Lions	4-12	Redskins	7-9	Cardinals	5-11	Bucs	2-14
Steelers	9-7	Patriots	8-8	Broncos	12-4	Colts	11-5
Browns	8-8	Dolphins	10-6	Raiders	9-7	Texans	10-6
Ravens	7-9	Jets	7-9	Chargers	11-5	Titans	8-8
Bengals	1-15	Bills	7-9	Chiefs	5-11	Jaguars	7-9

… and a different team could claim the best record. That's the nature of randomness.

Of course, in real life, we don't have the ability to know with full certainty what each team's Baseline Talent is. Instead, we need to analyze it from the other direction: look at who each team beat and who each team lost to, and calculate what their most probable Baseline Talent is. A team's most probable Baseline Talent is its Bayesian Resume Rating.

Calculating the Bayesian Resume Rating

The BRR is calculated by iteration. The ratings are calculated for all the teams, but because each team's rating depends on all of its opponents' ratings, it needs to be re-calculated until equilibrium is found.

A single team's rating is calculated by looking at each of its opponents and the game results. It goes through the possible ratings that the team in question could be, determines the probability of their season happening as it did, and finding the team's most probable Baseline Talent using the equation below. (If the math is starting to get confusing, remember there is an example at the end that's easier to follow along with.)

$$B = \frac{\int_{-\infty}^{\infty} x \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{G} P_i(x)\, dx}{\int_{-\infty}^{\infty} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{G} P_i(x)\, dx}$$

B = Bayesian Resume Rating
G = Number of games
x = Possible BRRs of the team
P(x) = Probability of each game outcome at the given value of x

This answer has a level of uncertainly, because we can't know with 100% confidence what a team's Baseline Talent is. So each team's BRR comes with a standard deviation:

$$S = \sqrt{\frac{\int_{-\infty}^{\infty} (B-x)^2 \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{G} P_i(x)\, dx}{\int_{-\infty}^{\infty} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{G} P_i(x)\, dx}}$$

S = Standard deviation in a team's rating (i.e., the uncertainly of their Baseline Talent)

This level of uncertainty can't be ignored, and again using the Bienaymé formula, must be accounted for when calculating the probabilities of game results. Ignoring this will result in over-fitting errors.

Finding the parity of the league

When I wrote about the simulation of a season, I set the standard deviation of each team's performance equal to 1.75, which represents the parity within the league (i.e., parity = 1.75). But in order to ensure that the BRR is objective, this number cannot simply be arbitrarily chosen. It needs to be calculated based on the actual parity seen in the league.

If we simulate a season in which parity = 0, it means that each team's RNG has a standard deviation of zero. Since no teams will ever fluctuate from their Baseline Talent, the better team will always win and there will never be a single upset. At the other end of the spectrum, if parity = 1000 or more, each game effectively becomes a 50/50 coin flip. The amount of fluctuation is so huge that the Baseline Talent of each team is almost entirely irrelevant.

The amount of parity in a league will change from season to season, so we can't use data from previous seasons to determine the amount of parity in a current season. The parity is measured by looking only at the games in the season being analyzed.

To calculate the parity in a season, the algorithm looks at each game and determines if it was an upset or not, and how close the two teams' ratings were. Of course, the standard deviation of each team's BRR must be accounted for in the calculation, so every matchup is at least partially an upset, and partially not, to account for that uncertainty. The equation below represents finding which level of parity best-fits the game results, and the parity is found by finding the minimum of the equation. (Again, the example is easier to follow than these equations.)

$$f(p)=\sum_{i=1}^{G}\int_{-\infty}^{\infty}\frac{e^{\left(\frac{-\left(y-(B_{Li}-B_{Wi})\right)^2}{2\sqrt{S_{Li}^2+S_{Wi}^2}}\right)}}{\sqrt{2\pi\left(S_{Li}^2+S_{Wi}^2\right)}}\left(\int_{-\infty}^{\frac{y}{p\sqrt{2}}}\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}dx\right)^2 dy$$

G = number of games
B_W = Winner's BRR
B_L = Loser's BRR
S_W = Winner's BRR Standard deviation
S_L = Loser's BRR Standard deviation
p = Parity

Finding the minimum of the curve with respect to p yields the level of parity that the league is playing at.

Again, the BRR is calculated by iteration. So if the new calculated value of p differs from the one used to calculate all teams' BRRs, then all the BRRs need to be re-calculated. And once the teams have new ratings, the level of parity will need to be calculated again. BRRs and parity are both iterated until equilibrium is found. Once equilibrium is found, the algorithm is complete.

An example

Following along with an example is the best way to understand how the BRRs are calculated. For this example, we'll follow the 2009 New Orleans Saints season.

First, a look at their schedule and the results of each game:

Opponent	Result
Detroit Lions	Win
Philadelphia Eagles	Win
Buffalo Bills	Win
New York Jets	Win
New York Giants	Win
Miami Dolphins	Win
Atlanta Falcons	Win
Carolina Panthers	Win
St. Louis Rams	Win
Tampa Bay Buccaneers	Win
New England Patriots	Win
Washington Redskins	Win
Atlanta Falcons	Win
Dallas Cowboys	Loss
Tampa Bay Buccaneers	Loss
Carolina Panthers	Loss
Arizona Cardinals	Win
Minnesota Vikings	Win
Indianapolis Colts	Win

The Saints were undefeated for the first 13 games, then lost the remainder of their regular season games. They then proceeded to win the Super Bowl over Indianapolis.

The BRRs of each opponent is:

Opponent	Rating	σ of rating
Detroit Lions	-1.62	0.65
Philadelphia Eagles	0.62	0.60
Buffalo Bills	-0.38	0.60
New York Jets	0.52	0.56
New York Giants	0.10	0.61
Miami Dolphins	-0.05	0.60
Atlanta Falcons	0.32	0.61
Carolina Panthers	0.16	0.60
St. Louis Rams	-1.93	0.67
Tampa Bay Buccaneers	-1.10	0.63
New England Patriots	0.49	0.59
Washington Redskins	-1.07	0.63
Atlanta Falcons	0.32	0.61
Dallas Cowboys	0.75	0.59
Tampa Bay Buccaneers	-1.10	0.63
Carolina Panthers	0.16	0.60
Arizona Cardinals	0.40	0.59
Minnesota Vikings	0.87	0.60
Indianapolis Colts	1.57	0.61

Note that each team's rating has a level of uncertainty, because there's no way to know exactly what their Baseline Talent is.

As a reminder, to use Bayes Theorem, we need three things:
•P(A), the prior probability, which in this case is the likelihood the Saints have a Baseline Talent of x.
•P(B|A), which is the likelihood of the Saints having the season they had, given a Baseline Talent of x.
•P(B), the marginal likelihood, which in this case is the likelihood of an unknown team having the exact results the Saints had.

The first step is to treat the Saints rating as an unknown, and to try different options to find the most probable Baseline Talent.

First, let's assume that the Saints are a perfectly average team, with Baseline Talent = 0. This is our prior, P(A). The relative probability of a team having a Baseline Talent of zero is calculated with the probability density function (PDF):

$$P_{PDF}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

x = Baseline Talent
$P_{PDF}(x)$ = Likelihood of a team having a Baseline Talent of x
μ = average Baseline Talent of the league, set to 0
σ = standard deviation of the league's Baseline Talent, set to 1

$$P_{PDF}(0) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(0-0)^2}{2}} = 0.399$$

Next, we can start looking at game results. They beat the Lions in week 1. What's the likelihood of that happening?
We start by finding the total standard deviation. This is composed of the performance fluctuations of both teams (i.e., the parity), and the uncertainty in the Lions' rating. At this time, there is no uncertainty in the Saints rating, because we are assuming it to be exactly zero.

$$\sigma_T = \sqrt{p^2 + p^2 + S_{Lions}^2}$$

p = Parity
S_{Lions} = Uncertainty in the Lions BRR
σ_T = Total standard deviation

The level of parity in 2009 was 1.60. (I'll explain how that number was calculated later.)

$$\sigma_T = \sqrt{1.60^2 + 1.60^2 + 0.65^2}$$

$$\sigma_T = 2.35$$

Now that number can be used to determine the probability of a Saints win:

$$P_{Saints\_beat\_Lions} = \int_{-\infty}^{\left(\frac{B_{Saints} - B_{Lions}}{2.35}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx$$

$$P_{Saints\_beat\_Lions} = \int_{-\infty}^{\left(\frac{0-(-1.62)}{2.35}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = 0.755$$

This can be repeated for all the Saints' opponents.

$$P_{Saints\_beat\_Eagles} = \int_{-\infty}^{\left(\frac{0-0.62}{2.34}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = 0.395$$

When it we come to the Saints losses, we need to calculate the probability of a loss.

$$P_{Cowboys\_beat\_Saints} = \int_{-\infty}^{\left(\frac{0.75-0}{2.34}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = 0.626$$

In the end, we have a table that looks like this:

Opponent	Result	Probability of result
Detroit Lions	Win	0.755
Philadelphia Eagles	Win	0.395
Buffalo Bills	Win	0.564
New York Jets	Win	0.411
New York Giants	Win	0.484
Miami Dolphins	Win	0.508
Atlanta Falcons	Win	0.445
Carolina Panthers	Win	0.472
St. Louis Rams	Win	0.793
Tampa Bay Buccaneers	Win	0.680
New England Patriots	Win	0.417
Washington Redskins	Win	0.677
Atlanta Falcons	Win	0.445
Dallas Cowboys	Loss	0.626
Tampa Bay Buccaneers	Loss	0.320
Carolina Panthers	Loss	0.528
Arizona Cardinals	Win	0.431
Minnesota Vikings	Win	0.354
Indianapolis Colts	Win	0.251

Now we take the product of all those probabilities, which comes to 9.81×10^{-7}. This is P(B|A). Multiplying this by the prior of 0.399 comes to 3.91×10^{-7}.

Next, let's repeat the process for another Baseline Talent. Let's try Saints Baseline Talent = 1. Before looking at their game results, the probability density function shows a team having a Baseline Talent of 1 is less likely:

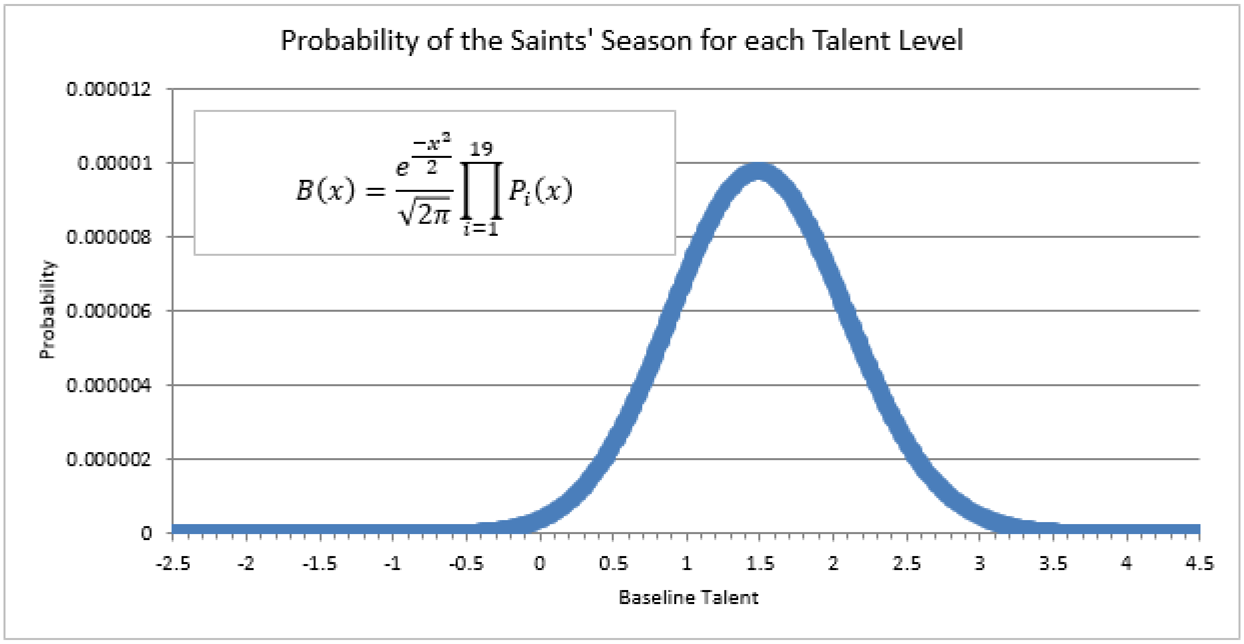
$$P_{PDF} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(1-0)^2}{2}} = 0.242$$

Their win/loss probability table looks like this:

Opponent	Result	Probability of result
Detroit Lions	Win	0.868
Philadelphia Eagles	Win	0.564
Buffalo Bills	Win	0.722
New York Jets	Win	0.581
New York Giants	Win	0.650
Miami Dolphins	Win	0.673
Atlanta Falcons	Win	0.614
Carolina Panthers	Win	0.640
St. Louis Rams	Win	0.893
Tampa Bay Buccaneers	Win	0.814
New England Patriots	Win	0.586
Washington Redskins	Win	0.812
Atlanta Falcons	Win	0.614
Dallas Cowboys	Loss	0.457
Tampa Bay Buccaneers	Loss	0.186
Carolina Panthers	Loss	0.360
Arizona Cardinals	Win	0.601
Minnesota Vikings	Win	0.522
Indianapolis Colts	Win	0.404

Taking the product of all those probabilities, and multiplying by the prior yields 7.01 x 10^{-6}. So, based on the game results, we see that $B_{Saints} = 1$ is much more likely than $B_{Saints} = 0$

This can be repeated for other values of B_{Saints}, and graphed:



The area under this curve is P(B), the likelihood of any unknown team having the season the Saints had.

$$P(B) = \int_{-\infty}^{\infty} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{19} P_i(x)\, dx = 1.47 \times 10^{-5}$$

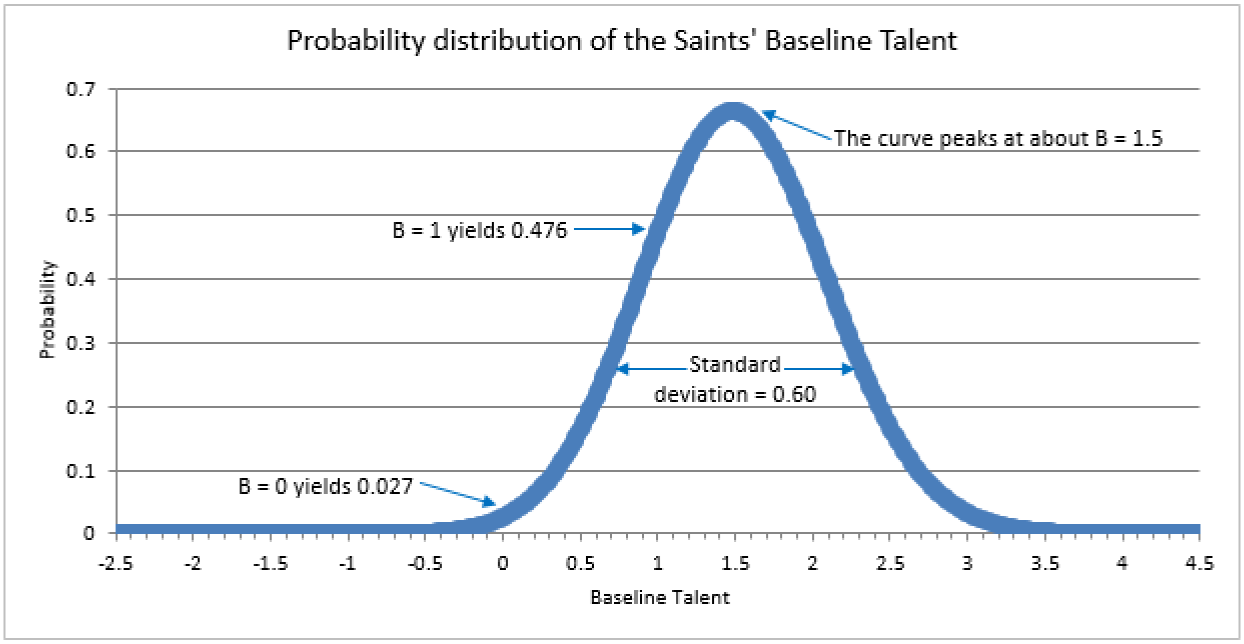
Now we can use Bayes' Theorem to calculate the probability of each level of Baseline Talent. So, going back to the probability the Saints have a Baseline Talent of 0:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{(9.81 \times 10^{-7})(0.399)}{1.47 \times 10^{-5}} = 0.027$$

When we use the prior of Saints Baseline Talent = 1:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{(2.90 \times 10^{-5})(0.242)}{1.47 \times 10^{-5}} = 0.476$$

Charting all these values gives us the probability distribution of the Saints' Baseline Talent:

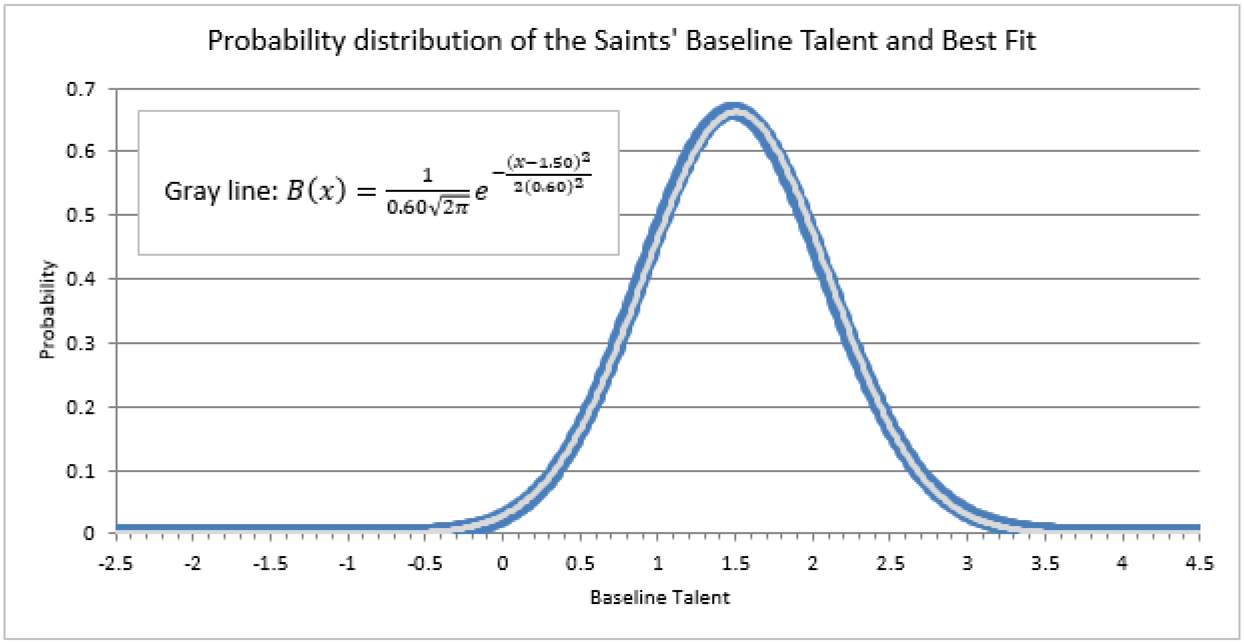


Taking a weighted average and standard deviation of this curve, we find that the most likely Baseline Talent of the Saints is 1.50, which is its BRR. The standard deviation of the rating is 0.60.

$$B_{Saints} = \frac{\int_{-\infty}^{\infty} x \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{19} P_i(x)\, dx}{\int_{-\infty}^{\infty} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{19} P_i(x)\, dx} = 1.50$$

$$S_{Saints} = \sqrt{\frac{\int_{-\infty}^{\infty} (1.50 - x)^2 \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{19} P_i(x)\, dx}{\int_{-\infty}^{\infty} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \prod_{i=1}^{19} P_i(x)\, dx}} = 0.60$$

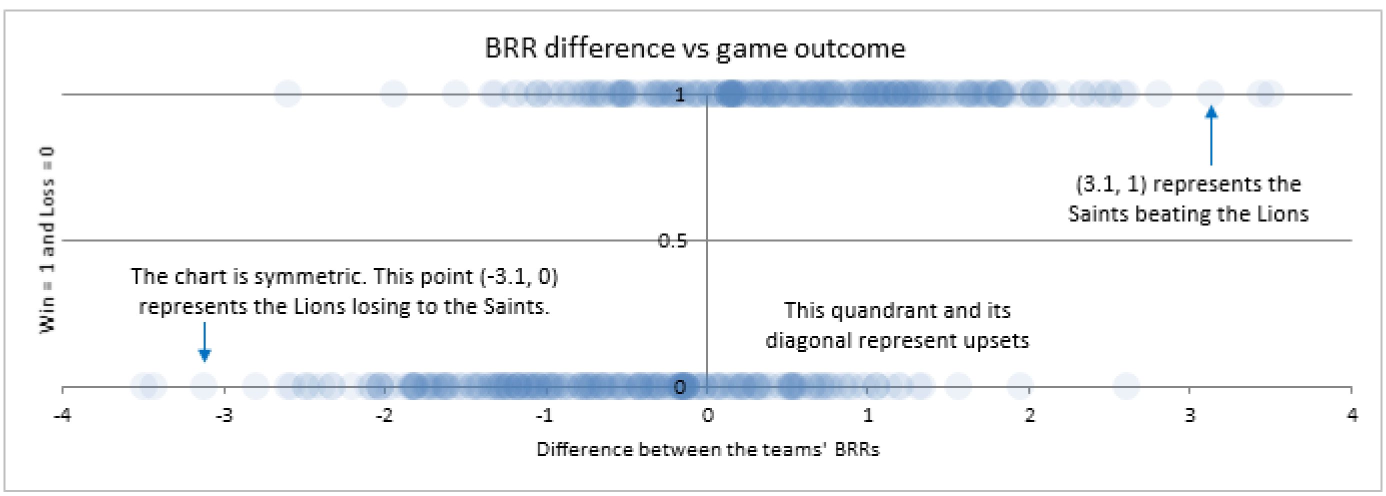
If we overlay this result on the graph it looks like this:



This process is repeated to calculate the BRR for all teams.

Calculating Parity

So now, back to the league parity. To visualize this, we can start by making a binomial regression, in which we plot rating differences versus game results, where a loss is 0 and a win is 1. The 2009 season would look like this:

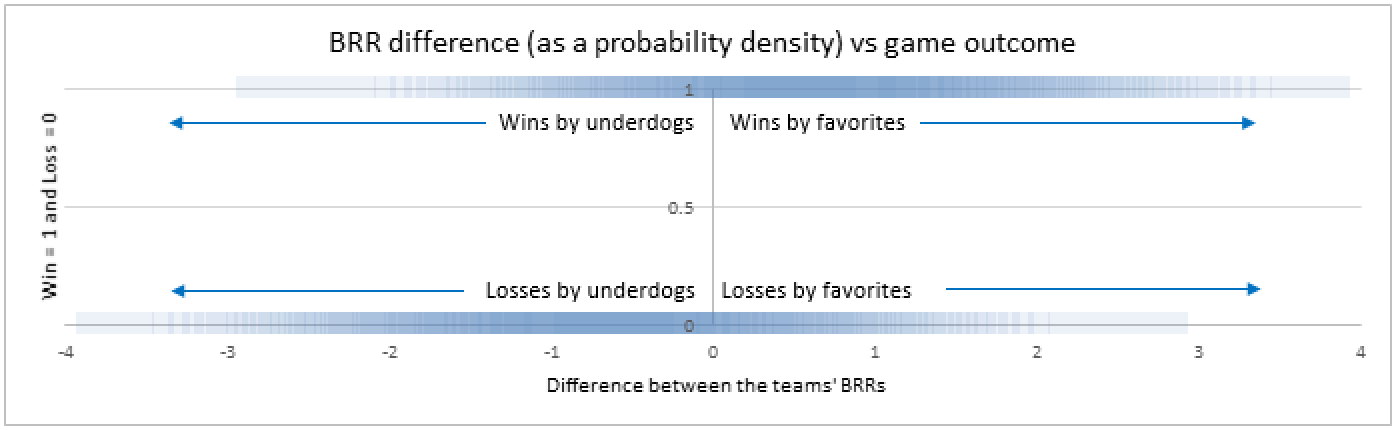


Because the teams are treated as random number generators, a best fit line's form would be based on the Cumulative Distribution Function (CDF):

$$f(Z) = \int_{-\infty}^{\left(\frac{Z}{p\sqrt{2}}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx$$

p = parity
Z = delta between teams in a matchup

So can we simply find the best fit of the data on the graph? Unfortunately, no, because it would result in the data being over-fit. The uncertainty in each team's rating must be accounted for. Effectively, each data point stops being a discrete point, and is instead spread out according to the level of uncertainty. Visually, it looks something like this:



Then, with the data in the graph above, a best-fit curve is found. But how do we spread the data points out according to that uncertainty?

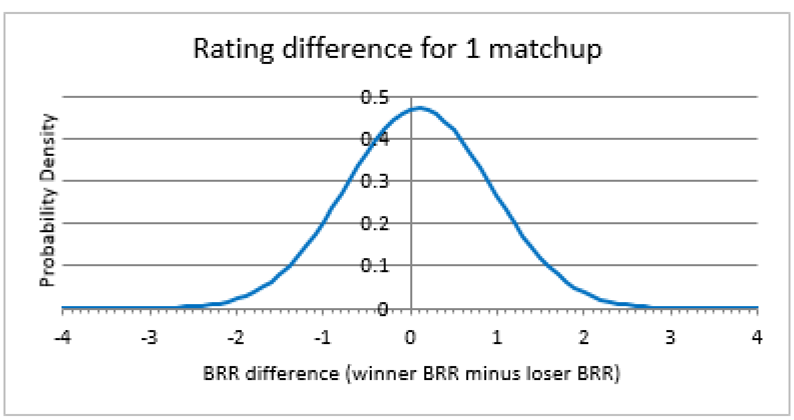
To understand how this is calculated, let's start with the season opener in 2009 which featured the Steelers beating the Titans. The Steelers had a BRR of 0.20 with a standard deviation of 0.59, and the Titans had a BRR of 0.10 with a standard deviation of 0.61.

The presumed favorite won, but due to the uncertainty in their ratings, it's possible that it was an upset. Again, we can't know each teams actual Baseline Talent, and even with a season's worth of data, the best teams may not necessarily always have the better records.

So the Steelers are probably better than the Titans, with their BRR being larger than the Titans' by 0.10, but it's not a certainty they had a better Baseline Talent, so their BRR standard deviation will be accounted for. Combining their standard deviations gives:

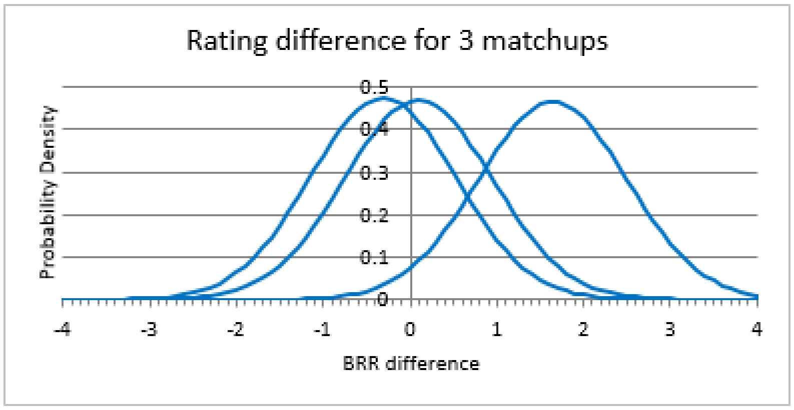
$$\sigma_T = \sqrt{0.59^2 + 0.61^2} = 0.85$$

Graphing the probability density function such that σ = 0.85 and μ = 0.10 looks like this:

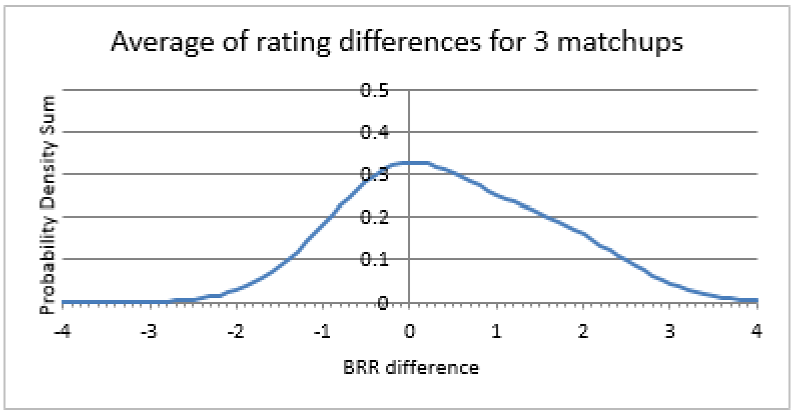


So you can see we are changing each data point from a single point to a spread.

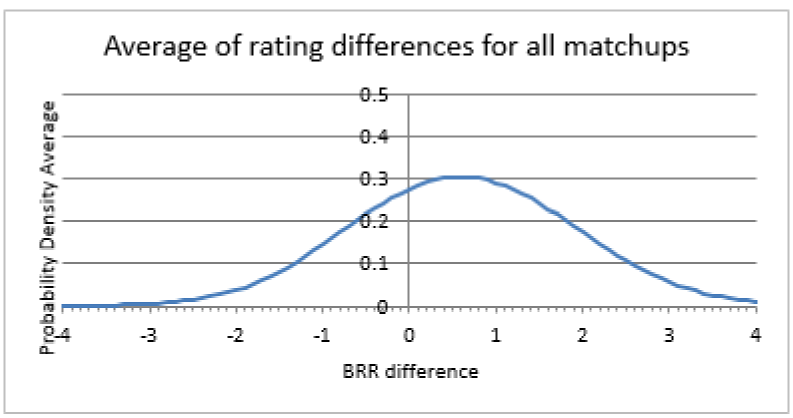
Another apparent favorite that won in week 1 was the Vikings (B = 0.87, S = 0.60) over the Browns (B = -0.77, S = 0.61). One of the week 1 upsets includes the Broncos (B = 0.04, S = 0.60) over the Bengals (B = 0.36, S = 0.59). Adding these to the graph:



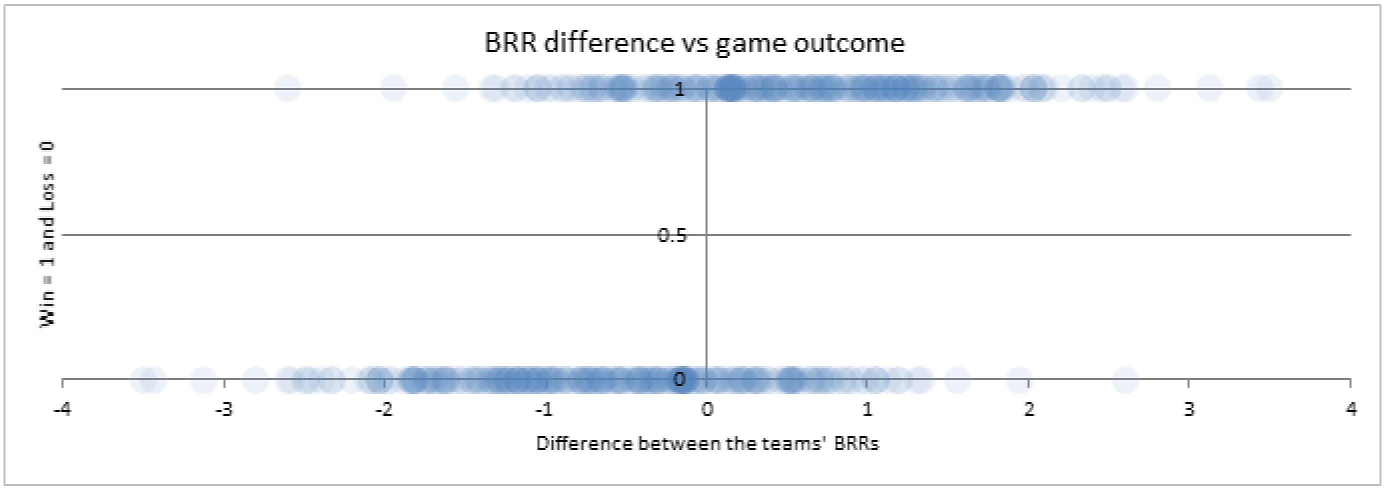
Combining each of these curves into a single curve:



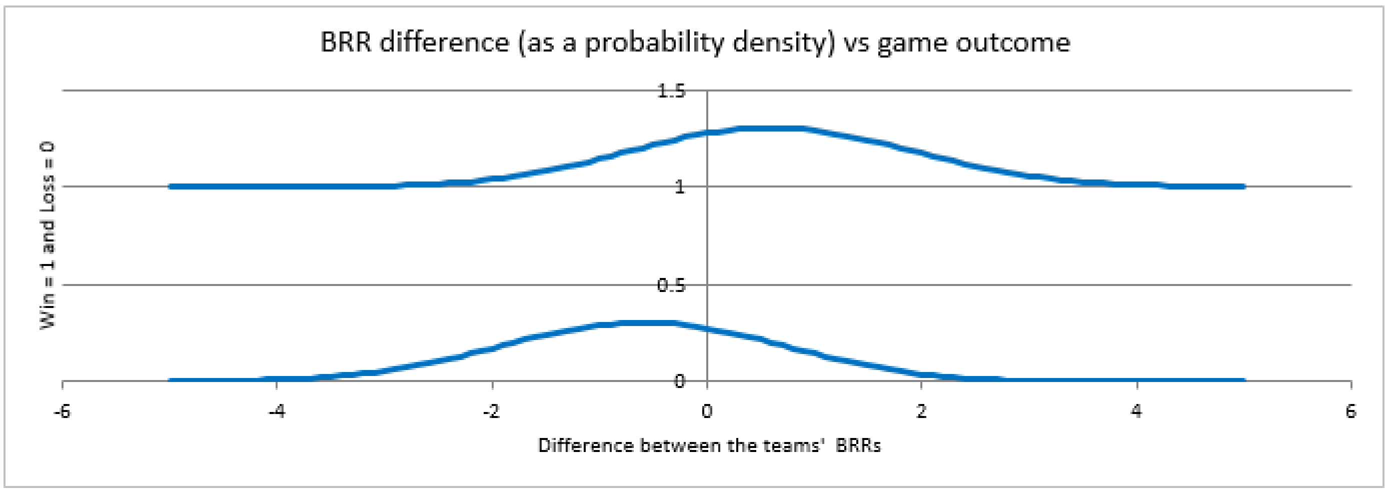
And now combing the curves of all games played the entire 2009 season:



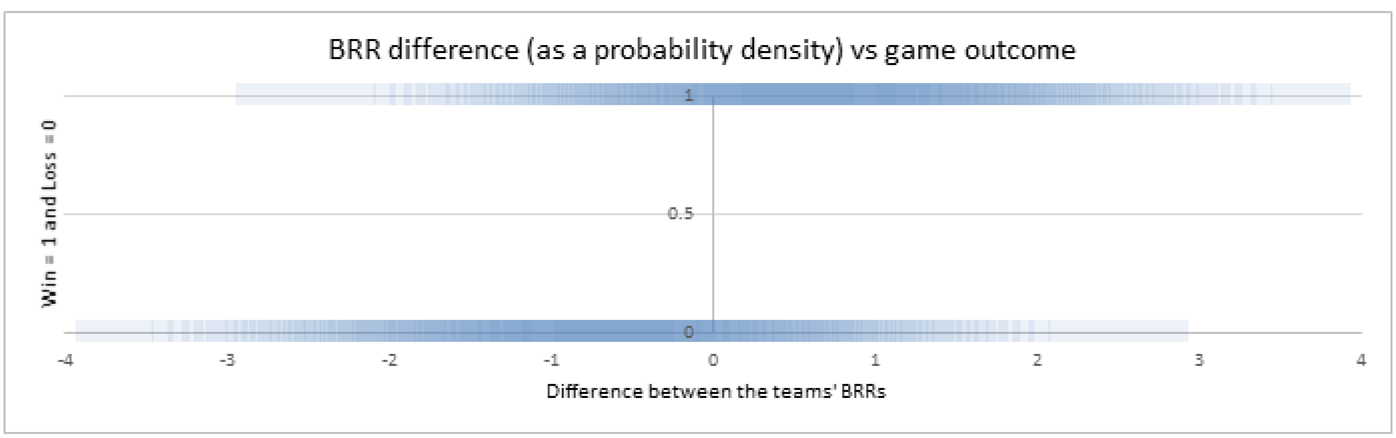
This curve represents each individual game changed from a discrete point to a density. Each discrete point goes from this:



To this:



Which, as shown previously, can also be more intuitively illustrated with a color differential rather than a bell curve:



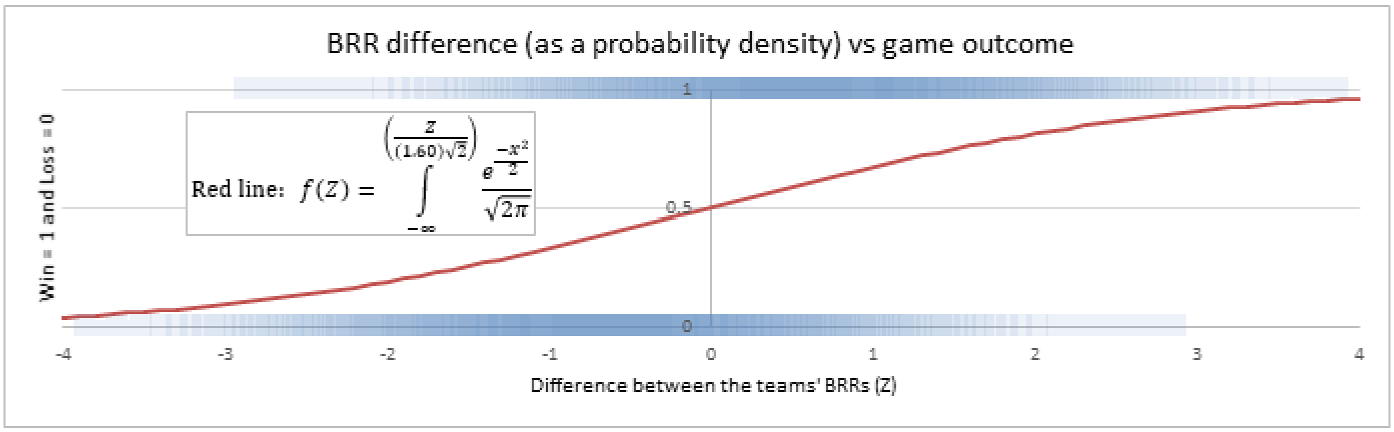
So now all data points are spread out according to the uncertainty in the matchup, and the CDF can be best fit to the data.

Finding the best fit, we find that p = 1.60:

$$\underset{p}{\arg\min}\left(\sum_{i=1}^{G}\int_{-\infty}^{\infty}\frac{e^{\left(\frac{-\left(y-(B_{Li}-B_{Wi})\right)^2}{2\sqrt{S_{Li}^2+S_{Wi}^2}}\right)}}{\sqrt{2\pi\left(S_{Li}^2+S_{Wi}^2\right)}}\left(\int_{-\infty}^{\frac{y}{p\sqrt{2}}}\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}dx\right)^2 dy\right) := \{1.60\}$$

$$\Rightarrow p = 1.60$$

As a visual, once we know p, we can chart the best fit line:



So now parity is calculated, and all teams' BRRs are calculated. Again, if that parity comes out to a different number than what was used to initially calculate all teams' BRRs, the whole process needs to start over with the new value for parity. Once this process is iterated to the point that equilibrium is found, the algorithm is complete.

Addendum A: Convergence

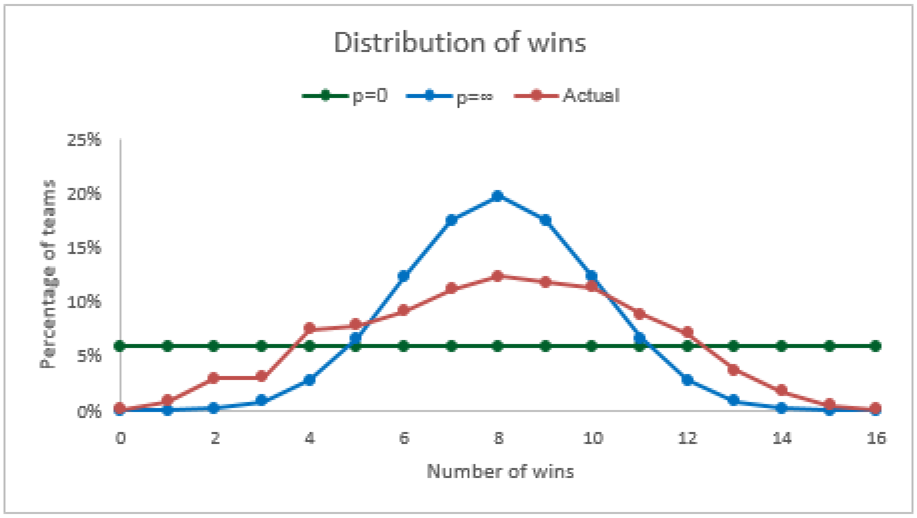
As explained previously, if parity = 0, then there is no fluctuation in teams' performances, and there will never be any upsets. The team with the higher Baseline Talent will win every game. If parity = 1000 or more, the amount of fluctuation is so large that Baseline Talent is nearly irrelevant; every game is effectively a 50/50 tossup.

If parity = 0, what is the probability of a team going undefeated over a randomly chosen 16-game regular season (ignoring that intra-divisional games are each played twice)? It's 1/17 = 0.058, and it's the same probability for each possible record.

As parity approaches infinity, and each game becomes a 50/50 tossup, what is the probability of each record? It's 0.5^{16} multiplied by the number of possible permutations.

If every game is a 50/50 chance of winning for every team			
Wins	Probability	Permutations	Total probability
0	0.5^{16}	1	0.0015%
1	0.5^{16}	16	0.024%
2	0.5^{16}	120	0.18%
3	0.5^{16}	560	0.85%
4	0.5^{16}	1820	2.8%
5	0.5^{16}	4368	6.7%
6	0.5^{16}	8008	12.2%
7	0.5^{16}	11440	17.5%
8	0.5^{16}	12870	19.6%
9	0.5^{16}	11440	17.5%
10	0.5^{16}	8008	12.2%
11	0.5^{16}	4368	6.7%
12	0.5^{16}	1820	2.8%
13	0.5^{16}	560	0.85%
14	0.5^{16}	120	0.18%
15	0.5^{16}	16	0.024%
16	0.5^{16}	1	0.0015%

Of course in real life, the NFL falls somewhere in between zero parity and full parity. Charting the actual season results from the past is unsurprisingly between the two extremes.



So if we want to chart different levels of parity, what's the equation? It's:

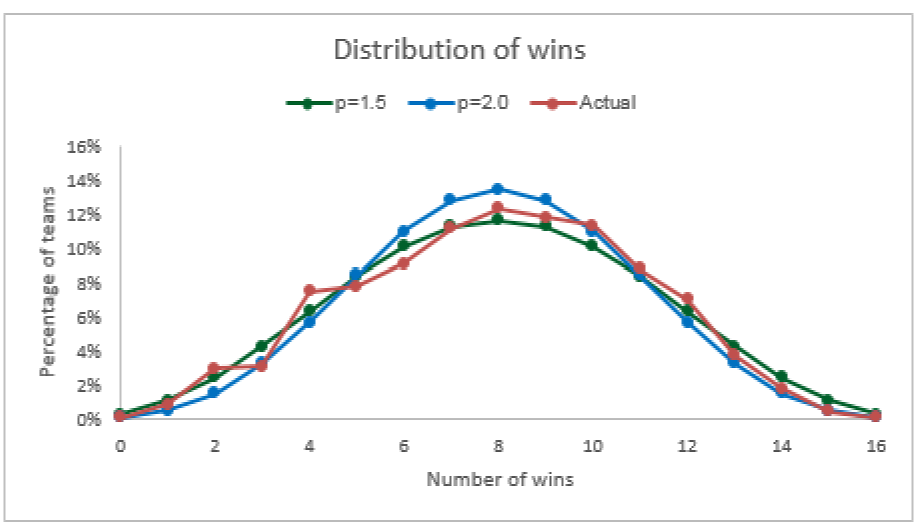
$$P = \int_{-\infty}^{\infty} \frac{G!}{W!\,(G-W)!} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y}{\sqrt{L^2+2p^2}}} e^{\frac{-x^2}{2}} dx \right)^{W} \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y}{\sqrt{L^2+2p^2}}} e^{\frac{-x^2}{2}} dx \right)^{(G-W)} \frac{e^{\frac{-y^2}{2L^2}}}{L\sqrt{2\pi}} dy$$

P = probability
p = parity
W = total wins
G = total games = 16
L = league sigma = 1

For example, in a year when the NFL has a level of parity p=1.75, the likelihood of a randomly chosen team going 12-4 is:

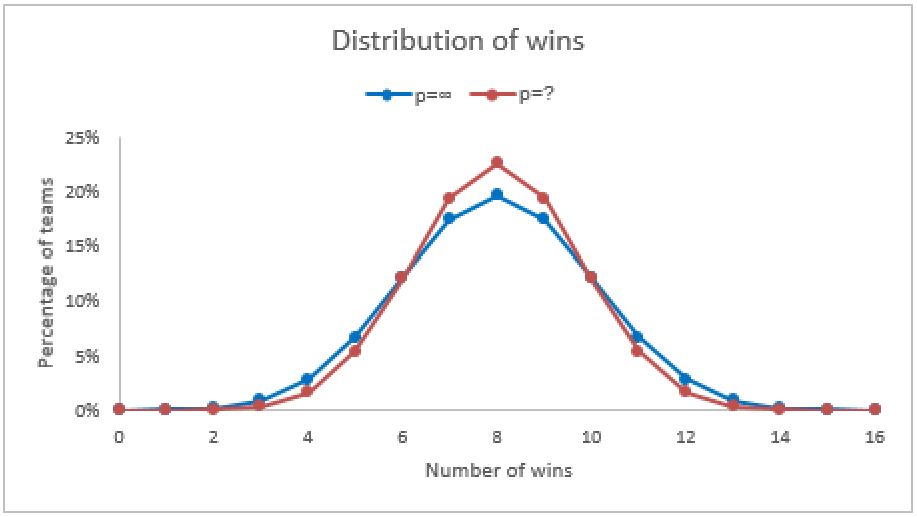
$$\int_{-\infty}^{\infty} \frac{16!}{12!\,(16-12)!} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y}{\sqrt{1+2(1.75)^2}}} e^{\frac{-x^2}{2}} dx \right)^{12} \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y}{\sqrt{1+2(1.75)^2}}} e^{\frac{-x^2}{2}} dx \right)^{(16-12)} \frac{e^{\frac{-y^2}{2(1)^2}}}{1\sqrt{2\pi}} dy = 6.0\%$$

So now we can chart a few more curves with values of p that more closely resemble reality:



The level of parity can vary quite a bit from year to year, but it's typically between 1.5 and 2.0. A couple extremes include 1995 and 1976. In 1995, 15 of the league's 30 teams finished the regular season with records ranging from 7-9 to 9-7. That year the parity was at 3.2. On the other hand, in 1976, Tampa Bay had their winless season, and the Raiders only lost a single game. Only 5 teams finished within a game of .500, and the parity was 1.0.

But what would happen if there were so many upsets that there appeared to be *more* parity than what you'd expect to see from games being 50/50?



In this case, the algorithm would find with each iteration that the level of parity needs to be larger. Which makes sense, because the league is exhibiting infinite parity. Correspondingly, each team's BRR would approach zero. This also makes sense: if the league has so much parity that every game is effectively a coin flip, you can't infer that any team is better than any other. In a coin flipping contest, by pure chance, some will win more than others. Winning records just imply better luck. It doesn't imply superiority.

So the BRR will fail to converge if the amount of parity in the league exceeds what would be expected from each game being 50/50. But the limit as parity approaches infinity still yields the logical result that all teams would be equals.

Addendum B: Ties

There are occasionally ties in the NFL, and there were frequently ties in college football before the overtime rules made it impossible. In the BRR, ties are treated as a half win and a half loss. So a good team will still be downgraded if it ties a bad team; the tie is not simply ignored.

As a consequence of this, if two teams play twice, and tie both times, it would be the same as if the teams had gone 1-1 against each other. For example, way back in 1926, the NFL's Akron Indians and Canton Bulldogs played each other twice, and tied both times. Their BRRs would be the same as if they had each won a game and lost a game.

$$P_{A\_ties\_B} = \sqrt{\left(1 - \int_{-\infty}^{\left(\frac{B_A - B_B}{\sigma_T}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}dx\right)\int_{-\infty}^{\left(\frac{B_A - B_B}{\sigma_T}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}dx}$$

Addendum C: Shortcomings

Because of the Bayesian approach to this ranking system, priors can have a strong influence early in the season when the sample size is small. Each team is assumed to be equal at the start of the season, and ratings get more accurate as the season progresses. If some teams or conferences don't play many games outside their conference, the sample size of non-conference games may be too small to accurately determine a conference's strength. It's debatable whether this is a bug or a feature. Assuming all teams are equal at the start of the year is certainly the most fair assumption, despite not being the most accurate assumption.

Additionally, the level of parity starts high until proven otherwise. It usually takes until all teams have played 6 or 7 games before the level of parity stabilizes.

Addendum D: Making predictions

As described above, the following formula can be used to determine the probability that one team will defeat another:

$$P_{A\_beats\_B} = \int_{-\infty}^{\left(\frac{B_A - B_B}{\sigma_T}\right)} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx$$

Despite this, I maintain that the BRR is for measuring teams' resumes, not for making predictions. Why is that?

If you wanted to make a prediction on a future game, and the only data available was a list of who beat whom for a single season, then I would say, yes, the BRR can make the prediction. In fact, because it's Bayesian, I can confidently say that the BRR will be great as a predictive rating systems under those circumstances.

But of course, we have a lot more data than just who beat whom.

There are many predictive rating systems on the internet, most of which factor in additional variables, such as margin of victory, venue, and more. If your goal is to make the most accurate prediction, using that extra data makes for better predictions (assuming it's used correctly). But at that point, you're no longer objectively ranking resumes, which is the purpose of the BRR. Resumes do not always perfectly correlate to future success, nor should we expect them to. Obviously, there will be a correlation between having a good resume and being a good team, but sometimes teams overachieve or underachieve, and may have a resume that's better or worse than their actual ability.

So, yes, the BRR can predict results, but I don't claim that it's better than other predictive systems. However, when it comes to objectively measuring teams' resumes based only on wins, losses, and strength of schedule, I know of no other rating system that compares to the BRR.

www.jellyjuke.com
jellyjuke@outlook.com