Optimizing Likelihood of MLB Playoff Apperance

Shannon Leiss

12/6/2021

Introduction

Many different attempts have been made to optimize an equation for predicting a team's probability of making the playoffs in all sports.

`Teams` dataset in the `Lahman` package provides team records for the 1871 - 2020 baseball season, along with offensive statistics, defensive statistics, and if the team made the playoffs.

Logistic Model

A simple logistic regression can be used to model the log-odds that a team makes the playoffs, based off of a teams run average and a teams batting average

$$log(\frac{p_i}{1 - p_i}) = \alpha + \beta_1(R_i) + \beta_2(H_i)$$

The terms that are observed are:

$$\begin{aligned} y_i &= \text{ Indicator that a team made the playoffs} \\ R_i &= \text{ Team's run average, found by dividing Runs by At Bats} \\ H_i &= \text{ Team's batting average, found by dividing Hits by At Bats} \end{aligned}$$

Where the unobserved variables are α, β_1, β_2, and p_i, where p_i is defined as:

$$p_i = \frac{e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}{1 + e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}$$

Questions of Interest:

What values of α, β_1, and β_2 optimize the logistic function for the likelihood of a Baseball team making the playoffs?

Which optimization method - Newton-Raphson or Quasi-Newton - gets closest to the optimized values found when the logistic model is fit?

What are the 95% confidence intervals of the $\hat{\alpha}, \hat{\beta}_1$, and $\hat{\beta}_2$ optimized values?

Methods

The following methods will be used to optimize $\theta = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix}$

- Newton-Raphson
- Quasi-Newton
- Bootstrapping for Confidence Intervals

Distribution of Model

Since the response variable of y_i is a binary response, the logistic regression will follow a Bernoulli distribution. The distribution is:

$$f(y_i) = p_i^{y_i}(1 - p_i)^{1-y_i}$$
$$f(y_i) = (\frac{e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}{1 + e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}})^{y_i}(1 - (\frac{e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}{1 + e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}))^{1-y_i}$$
$$f(y_i) = (\frac{e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}{1 + e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}})^{y_i}(\frac{1}{1 + e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}})^{1-y_i}$$

Log-Likelihoods of Model

$$\ell(\theta) = \sum_{i=1}^{n=660} y_i log(\frac{e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}{1+e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}) + (1-y_i)log(\frac{1}{1+e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}})$$

$$\ell(\theta) = \sum_{i=1}^{n=660} y_i(\alpha+\beta_1(R_i)+\beta_2(H_i)) - log(1+e^{\alpha+\beta_1(R_i)+\beta_2(H_i)})$$

$$\ell'(\theta) = \begin{bmatrix} \sum_{i=1}^{n=660} y_i - \frac{e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}{1+e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}} \\ \sum_{i=1}^{n=660} R_i(y_i) - \frac{(R_i)e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}{1+e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}} \\ \sum_{i=1}^{n=660} H_i(y_i) - \frac{(H_i)e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}}{1+e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n=660} y_i - p_i \\ \sum_{i=1}^{n=660} R_i(y_i) - R_i(p_i) \\ \sum_{i=1}^{n=660} H_i(y_i) - H_i(p_i) \end{bmatrix}$$

Letting C $= e^{\alpha+\beta_1(R_i)+\beta_2(H_i)}$

$$\ell''(\theta) = \begin{bmatrix} \sum_{i=1}^{n=660} -\frac{C}{(1+C)^2} & \sum_{i=1}^{n=660} -R_i\frac{C}{(1+C)^2} & \sum_{i=1}^{n=660} -H_i\frac{C}{(1+C)^2} \\ \sum_{i=1}^{n=660} -R_i\frac{C}{(1+C)^2} & \sum_{i=1}^{n=660} -(R_i^2)\frac{C}{(1+C)^2} & \sum_{i=1}^{n=660} -(R_iH_i)\frac{C}{(1+C)^2} \\ \sum_{i=1}^{n=660} -H_i\frac{C}{(1+C)^2} & \sum_{i=1}^{n=660} -(R_iH_i)\frac{C}{(1+C)^2} & \sum_{i=1}^{n=660} -(H_i^2)\frac{C}{(1+C)^2} \end{bmatrix}$$

Methods: Newton-Raphson

The updating rule for Newton-Raphson is:

$$\theta^{(t+1)} = \theta^{(t)} - (\ell''(\theta^{(t)}))^{-1}(\ell'(\theta^{(t)}))$$

The optimization formula was allowed to update until the difference between $\theta^{(t+1)}$ and $\theta^{(t)}$ was less than 0.00001.

Methods: Quasi-Newton

Alternative method to Newton Method when the second derivative of $\ell(\theta)$ is complicated

The updating rule for this method, using the Hessian approximation with the BFGS method:

$$M^{(t+1)} = M^{(t)} - \frac{(M^{(t)}(\theta^{(t+1)} - \theta^{(t)}))(M^{(t)}(\theta^{(t+1)} - \theta^{(t)}))^T}{(\theta^{(t+1)} - \theta^{(t)})^T M^{(t)}(\theta^{(t+1)} - \theta^{(t)})} + \frac{(\ell'(\theta^{(t+1)}) - \ell'(\theta^{(t)}))(\ell'(\theta^{(t+1)}) - \ell'(\theta^{(t)}))^T}{(\theta^{(t+1)} - \theta^{(t)})^T(\ell'(\theta^{(t+1)}) - \ell'(\theta^{(t)}))}$$

Methods: Bootstraping for Confidence Intervals

- Resampling of the Original observed data and then optimizing the resampled data
- 10,000 iterations were run to produce 10,000 optimized values of θ
- Resampled data optimized by Quasi-Newton method

Results

Starting points: $\theta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -8 \\ 80 \\ -20 \end{pmatrix}$

All optimized values, along with iterations and the estimated parameters from the logistic regression

	Alpha	Beta 1	Beta 2	Runs
Newton:(0,0,0)	-7.8077	87.7043	-19.5363	5
Newton:(-1,1,-1)	-7.8077	87.7043	-19.5363	4
Newton:(-8,80,-20)	-7.8077	87.7043	-19.5363	5
Quasi-Newton:(0,0,0)	-7.8015	87.6818	-19.5476	28
Quasi-Newton:(-1,1,-1)	-7.8219	87.6086	-19.4316	25
Quasi-Newton:(-8,80,-20)	-7.8214	87.6017	-19.4302	18
Fisher Scoring(through glm()):	-7.8080	87.7040	-19.5360	4

Results: Confidence Intervals

	Lower	Upper	Estimate	Standard Deviation
Alpha	-12.1018	-3.7367	-7.8636	2.1280
Beta 1	67.6702	109.1632	88.0728	10.5754
Beta 2	-41.5621	2.2570	-19.5275	11.0820

Conclusions

- Newton-Raphson optimization takes the least amount of iterations and gets closest to the estimated values from the logistic regression model.
- Quasi-Newton optimization has shortest computation time and simpler implementation but takes more iterations
- Both methods are fairly close to each other
- Starting value:
  - Doesn't impact number of iterations for Newton-Raphson
  - Impacts number of iterations for Quasi-Newton

Models Created from Optimization

Using Newton-Raphson optimized values:

$$log(\frac{p_i}{1-p_i}) = -7.808 + 87.704(R_i) - 19.536(H_i)$$

Using Quasi-Newton estimate from the bootstrapping method:

$$log(\frac{p_i}{1-p_i}) = -7.864 + 88.073(R_i) - 19.528(H_i)$$

Probability of Making the Playoffs

Team	ARI	ATL	BAL	BOS	CHW	CHC	CIN	CLE
Newton	0.482	0.656	0.232	0.652	0.158	0.583	0.216	0.437
Quasi	0.482	0.657	0.231	0.652	0.157	0.583	0.215	0.437

Team	COL	DET	HOU	KCR	ANA	LAD	FLA	MIL
Newton	0.487	0.035	0.771	0.168	0.387	0.787	0.062	0.389
Quasi	0.487	0.035	0.772	0.167	0.386	0.788	0.061	0.388

Team	MIN	NYY	NYM	OAK	PHI	PIT	SDP	SEA
Newton	0.784	0.856	0.379	0.659	0.396	0.227	0.205	0.412
Quasi	0.785	0.857	0.379	0.659	0.395	0.226	0.204	0.411

Team	SFG	STL	TBD	TEX	TOR	WSN
Newton	0.140	0.425	0.315	0.542	0.302	0.713
Quasi	0.139	0.425	0.314	0.542	0.301	0.714