

# プロ野球の「順位逆転現象」をシミュレーションで紐解く ～2019年のホークス2位は当たり前じゃない!?～

芝浦工業大学 数理科学研究会 BV17057 西脇 友哉

令和元年 11月1日

## 研究背景

今年('19)も福岡ソフトバンクホークスが日本シリーズを制したが、'15や'17ほどの圧倒的な強さはなく、リーグ優勝は逃していた。今年のホークスの様にピタゴラス勝率(式(1))\*<sup>1</sup>が4位のチームがシーズン2位になった例は'11～'19で18チーム中4チームである。その意味ではシーズン2位という結果自体も快挙であると言えそうだが、サンプル数が少ないために「4/18」という割合の珍しさが今ひとつ評価しづらい。そこで、仮想再現した大量のデータを基に判断する事を考えた。\*<sup>2</sup>

## 1 方針

各試合結果や個人成績は無視し、簡易的な手段としてチームの年間成績の再現を図る。プログラムはMATLABで記述する。得点や失点の決定要素\*<sup>3</sup>に対して正規乱数を用いる。準備としてそれらのデータに対し正規性(データが正規分布に従うこと)の検定を行い、その結果から正規性を仮定する。

## 2 ピタゴラス勝率の最適化

プロ野球の順位は勝率の序列で決定される。本研究では勝率をピタゴラス勝率( $Py(x)$ )で近似する。

$$Py(x) := \frac{1}{1 + \rho^x} \left( \rho = \frac{\text{失点数}}{\text{得点数}} \right) \quad (1)$$

一般的に $\rho^x$ の指数部は $x = 2$ として用いられるが、厳密には競技の種類によって異なる指数を要する。そこで、式(2)の平均絶対偏差(MAD\*<sup>4</sup>)を最小化する $x$ をピタゴラス勝率の最適な指数 $x^*$ とする。 $(m$ はサンプル数、 $\omega$ はシーズン中の勝率)

$$MAD(x) := \frac{1}{m} \sum_{i=1}^m \left| \omega_i - \frac{1}{1 + \rho_i^x} \right| \quad (2)$$

計算結果は $x^* = 1.69$ ,  $MAD(x^*) = 0.0230$ となった。

## 3 勝率補正

実際の勝率( $\omega$ )とピタゴラス勝率( $Py(x)$ )との差を勝率補正( $ERR(x)$ )と名付け、以下に定義する。

$$ERR(x) := \omega - Py(x) \quad (3)$$

勝率補正は $\mu = 0.000$ ,  $\sigma = 0.0275$ の正規乱数で再現する。

## 4 主結果

勝率の決定には複数の方法があり、式(4)の場合と式(5)の場合では、同じチームでも最終的な勝率が異なる。

$$(\text{勝率}) \sim Py(1.69) \quad (4)$$

$$(\text{勝率}) \sim Py(1.69) + ERR(1.69) \quad (5)$$

この違いによって順位逆転が起こる例を示したものが表1であり、それを基にピタゴラス勝率の順位とシーズン順位を仮想データ500年分集計したものが表2である。

表1 勝率決定方法の違いによる順位変動の例(仮想データの抜粋)

得点	失点	Py(1.69)		Py(1.69) + ERR(1.69)		順位変動	
		勝率	順位	ERR	勝率		順位
640	547	0.566	2	+0.022	0.588	1	↗
841	652	0.606	1	-0.028	0.578	2	↘
803	814	0.494	3	+0.020	0.514	3	—
533	587	0.459	5	-0.002	0.457	4	↗
703	774	0.459	4	-0.027	0.432	5	↘
511	698	0.371	6	-0.035	0.336	6	—

表2 ピタゴラス勝率の順位とシーズン順位(仮想データ)

		シーズン順位					
		1	2	3	4	5	6
ピタゴラス勝率	1	785	170	33	9	3	0
	2	160	585	199	45	11	0
	3	44	163	504	216	67	6
	4	9	64	195	505	188	39
	5	2	16	63	188	570	161
	6	0	2	6	37	161	794

表2より、仮想データ上ではピタゴラス勝率4位かつシーズン2位となるチームの割合は64/1000であることが分かり、それを達成する難しさがより明確になったと言える。

## 今後の課題

仮想データの精度に関わる項目の数値(勝率の分散など)を改善するために、必要に応じてプログラムを修正する。

## 参考文献

- [1] NPB.jp 日本野球機構, <http://npb.jp/>, 最終アクセス: 2019.4.28.

\*<sup>1</sup> ピタゴラス勝率: 得点と失点からチームの勝率を推定する指標。

\*<sup>2</sup> 以下、シミュレーションによって擬似的に生成したデータを「仮想データ」、過去の試合結果に基づくデータを「実データ」と呼ぶことにする。

\*<sup>3</sup> OPSなど。詳細は資料を参照されたい。

\*<sup>4</sup> MAD: mean absolute deviation