

第 III 部

結論

7 シミュレーション結果

7.1 再現精度の評価

仮想データの質を「データのばらつきをどれ程実データに寄せられているか」によって評価したい。そのために、以下の2つの数値を指標として導入する。

1. 全サンプル（全順位）に対する勝率の標準偏差
2. 各順位の勝率平均における実データとの絶対誤差

1 は仮想データの標準偏差が実データの標準偏差である 0.06913 に近いほど精度が良く、2 については「絶対誤差の合計」は数値が小さいほど良い。それぞれの計算結果を以下に示した。

実データ : 0.06913, $Py(2)$: 0.08779, $Py(1.69)$: 0.07483, $Py(1.69) + ERR$: 0.08008

この結果より、「データのばらつき」に関しては $Py(1.69)$ が最も実データに近いと言える。

表 5 順位別平均勝率

順位	実データ	仮想データ					
		$Py(2)$		$Py(1.69)$		$Py(1.69) + ERR$	
		勝率平均	絶対誤差	勝率平均	絶対誤差	勝率平均	絶対誤差
1	.60102	.62282	.02180	.60471	.00369	.61038	.00936
2	.54556	.56278	.01722	.55322	.00766	.55641	.01085
3	.51294	.51737	.00443	.51470	.00176	.51592	.00298
4	.48208	.48014	.00194	.48320	.00112	.48196	.00012
5	.45065	.43843	.01222	.44783	.00282	.44337	.00728
6	.40841	.38027	.02814	.39798	.01043	.39198	.01643
絶対誤差の合計		.08575		.02748		.04702	

6.5.2 でも述べた様に、ピタゴラス勝率のみで順位を付ける場合と勝率補正を加える場合では異なる順位が出力される事がある。出力された仮想データに基づき、「年度 1・リーグ 1」のデータを「順位変動が起こらない例」、「年度 1・リーグ 2」のデータを「順位変動が起こる例」として表 6 に示す。

表 6 勝率決定方法の違いによる順位変動の例（仮想データ：年度 1, リーグ 2）

No.	OBP	SLG	OPS	得点	失点	$Py(1.69)$		$Py(1.69) + ERR$			順位 変動
						勝率	順位	ERR	勝率	順位	
8	.326	.412	.737	640	547	0.5660	2	+0.0215	0.5875	1	↗
12	.361	.482	.843	841	652	0.6059	1	-0.0281	0.5778	2	↘
9	.369	.455	.823	803	814	0.4943	3	+0.0199	0.5142	3	—
7	.315	.367	.682	533	587	0.4593	5	-0.0021	0.4572	4	↗
10	.358	.413	.771	703	774	0.4594	4	-0.0265	0.4330	5	↘
11	.303	.367	.670	511	698	0.3712	6	-0.0353	0.3359	6	—

ピタゴラス勝率の順位とシーズン順位の対応関係について、全 6000 個のサンプルを集計したものが表 7（勝率補正なし）と表 8（勝率補正あり）である。

「仮想データの中で最も実データに近いのは Py(1.69) である」と述べたが、「ピタゴラス勝率の順位とシーズン順位の逆転が全く起こらない状況」を表す表 7 は冒頭の表 1 と見比べてかなり現実離れしていることがうかがえる。よって、多少標準偏差が大きくなっても勝率補正を加える方が妥当である。

7.2 「順位逆転現象」の再現結果

冒頭で提起した「'19 ソフトバンクの様にピタゴラス勝率 4 位でシーズン 2 位になることは珍しいのか？」という問題について考える。表 8 におけるピタゴラス勝率 4 位の行を参照すると、ピタゴラス勝率 4 位のサンプル全 1000 個のうちシーズン 2 位以上はわずか 73 個に過ぎないことが分かる。

表 7 ピタゴラス勝率の順位とシーズン順位 (勝率補正なし)

		シーズン順位					
		1	2	3	4	5	6
ピタゴラス勝率	1	1000	0	0	0	0	0
	2	0	1000	0	0	0	0
	3	0	0	1000	0	0	0
	4	0	0	0	1000	0	0
	5	0	0	0	0	1000	0
	6	0	0	0	0	0	1000

表 8 ピタゴラス勝率の順位とシーズン順位 (勝率補正あり)

		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

研究の発展性

本研究では勝率をピタゴラス勝率と勝率補正の和で表現した。ピタゴラス勝率は得点と失点のみで決まるため、チームの実力が反映されやすい指標であると言える。

一方、勝率補正のみに影響を与える因子を抽出するのは非常に難しく、*3 勝率補正の決定要素を解明する事は困難である。そこで、勝率という結果に対して、ピタゴラス勝率を実力の要素、勝率補正を運の要素と考えることで、今後は「多少運が悪くてもどれくらいの実力があればカバーできるか」等の基準を探ることに取り組みたい。順位を限定した散布図（図 12～15）を元に勝率とピタゴラス勝率の分布の様子を観察すると、いわゆる「安全圏」と「それ以外」の判別が可能である事が推測出来る。今後は判別関数を具体的に求め、その関数の精度や「安全圏にギリギリ入れる基準」などを考察出来れば、野球ファン目線からも現場目線からも有意義な研究となる事が期待できる。

今後の課題

今回は失点よりも得点の導出過程に重きを置いたプログラムを組んだが、次は失点を重視したモデルも構築したい。また、「野球ゲームにおけるペナントモードのオート進行」を実際に行い、本研究のプログラムと精度を比較したい。

参考文献

- [1] <http://npb.jp/> : NPB.jp 日本野球機構, 最終アクセス日 : 2019.11.25
- [2] <http://baseballdata.jp/> : データで楽しむプロ野球, 最終アクセス日 : 2019.11.25
- [3] <https://ja.wikipedia.org/wiki/> : Wikipedia—野球の各種記録, 最終アクセス日 : 2018.5.15
- [4] 小西貞則, 多変量解析入門—線形から非線形へ, 岩波書店, 2010

*3 例えば、勝率補正から監督の采配の質を測ることが出来るという論調も存在する様だが、采配はピタゴラス勝率にも影響を与えるため、この論調は誤りであるとされている。

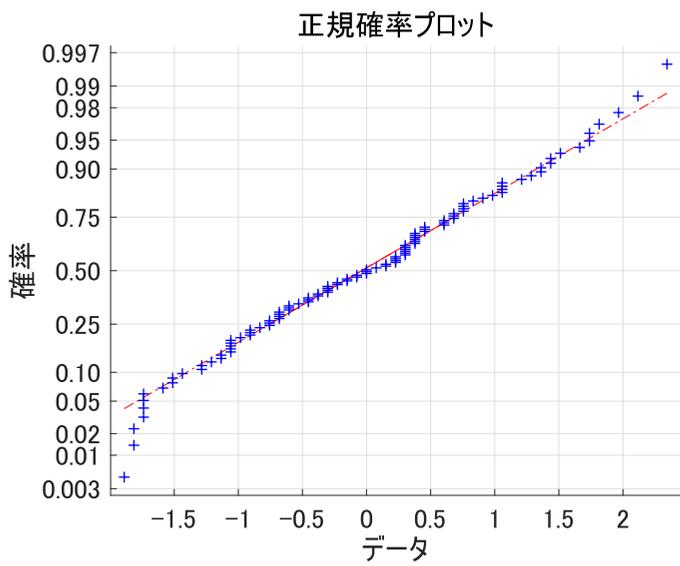


図6 正規確率プロット (OBP)

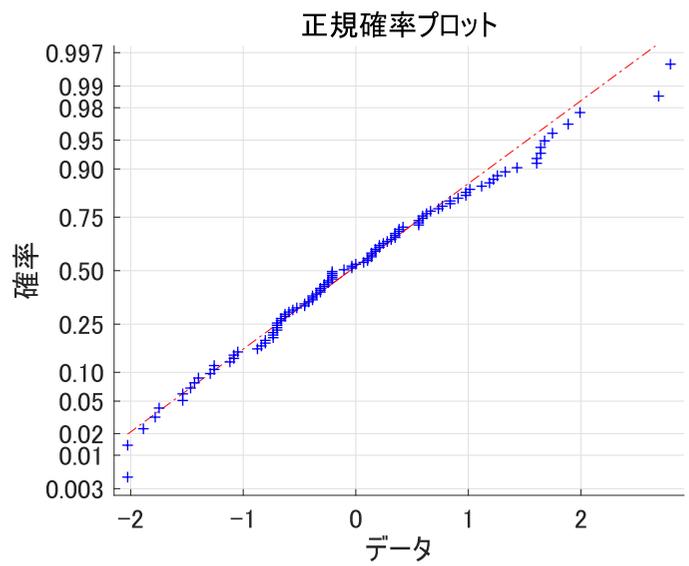


図7 正規確率プロット (SLG)

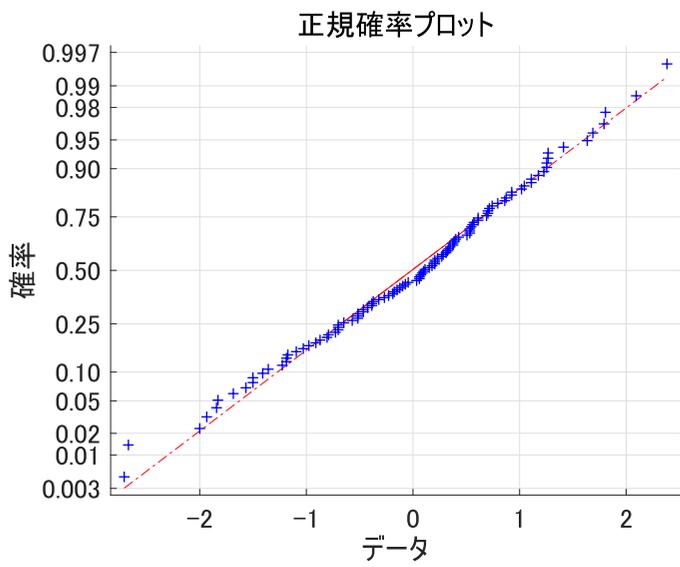


図8 正規確率プロット (年間の失点数)

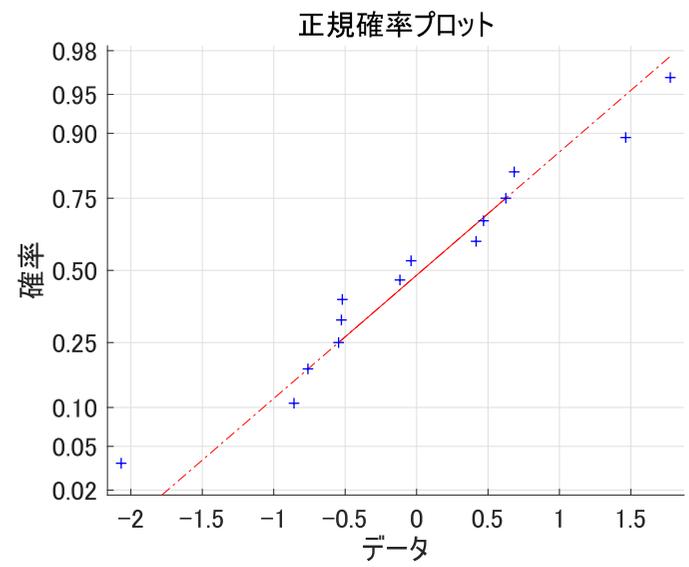


図9 正規確率プロット (交流戦の得失点差)

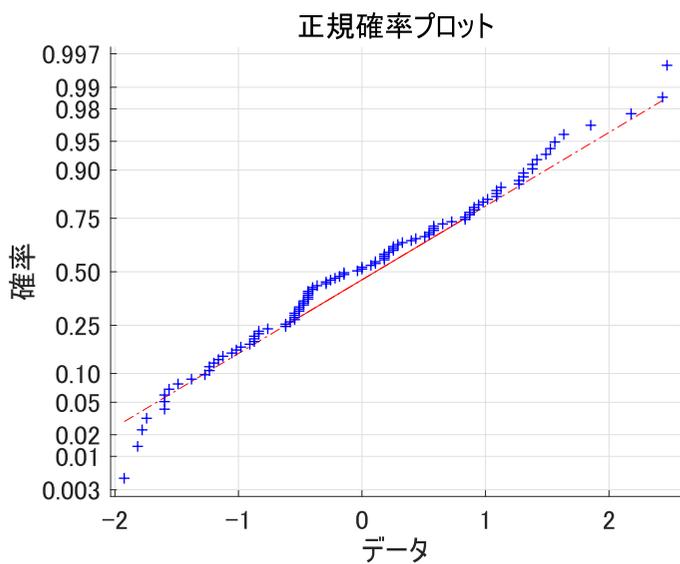


図10 正規確率プロット (勝率 - $P_y(1.69)$)

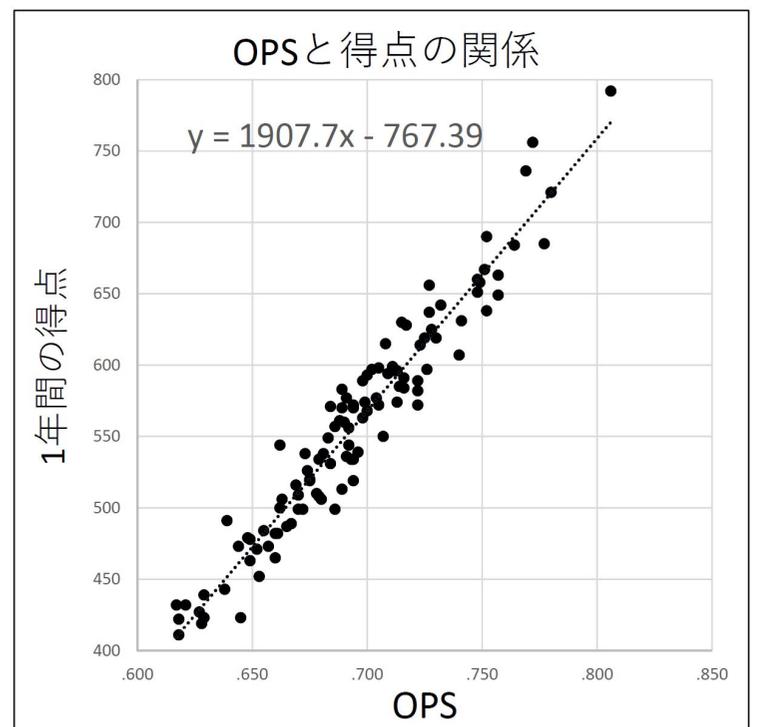


図11 OPS, 総得点の散布図と回帰直線 ('11~'19)

- [5] 下川敏雄, 実践のための基礎統計学, 講談社, 2016
- [6] データスタジアム株式会社, 野球 × 統計は最強のバッテリーである セイバーメトリクスとトラッキングの世界, 中央公論新社, 2015
- [7] <https://bellcurve.jp/statistics/course/12931.html> : 統計 WEB—統計学の時間—正規性の確認, 最終アクセス日 : 2019.4.19
- [8] Amy N. Langville, Carl D. Meyer (訳) 岩野和生, 中村英史, 清水咲里, レイティング・ランキングの数理, 共立出版, 2015
- [9] <https://jp.mathworks.com/help/stats/kstest.html> : MATLAB—ドキュメンテーション—1 標本コルモゴロフ・スミルノフ検定, 最終アクセス日 : 2019.4.19
- [10] <https://bellcurve.jp/statistics/glossary/1360.html> : 統計 WEB—統計用語集—コルモゴロフ=スミルノフ検定, 最終アクセス日 : 2019.4.19
- [11] <https://bellcurve.jp/statistics/glossary/2071.html> : 統計 WEB—統計用語集—正規 Q-Q プロット, 最終アクセス日 : 2019.4.19

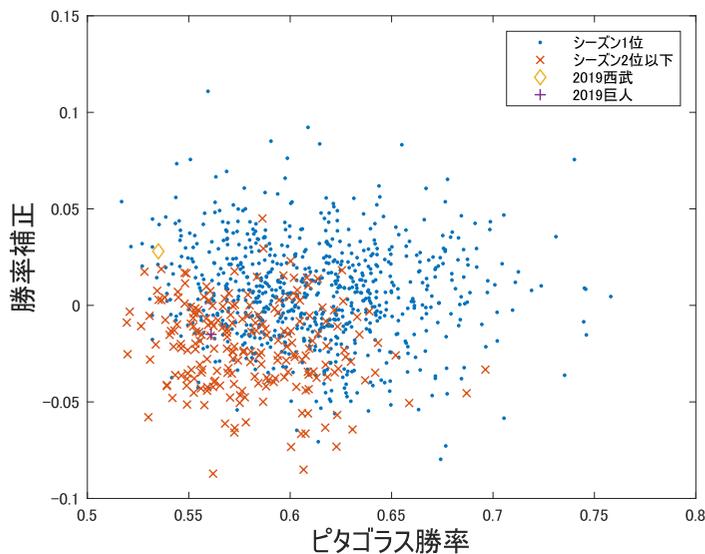


図 12 仮想データ上でピタゴラス勝率 1 位のチームと'19L, '19G

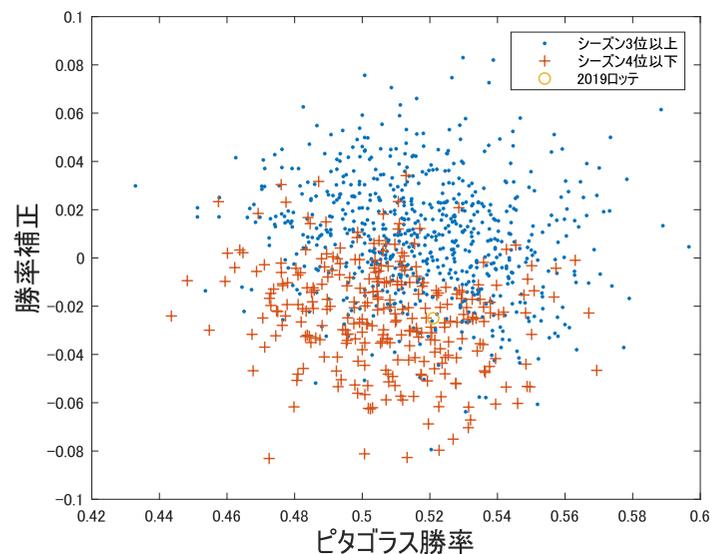


図 13 仮想データ上でピタゴラス勝率 3 位のチームと'19M

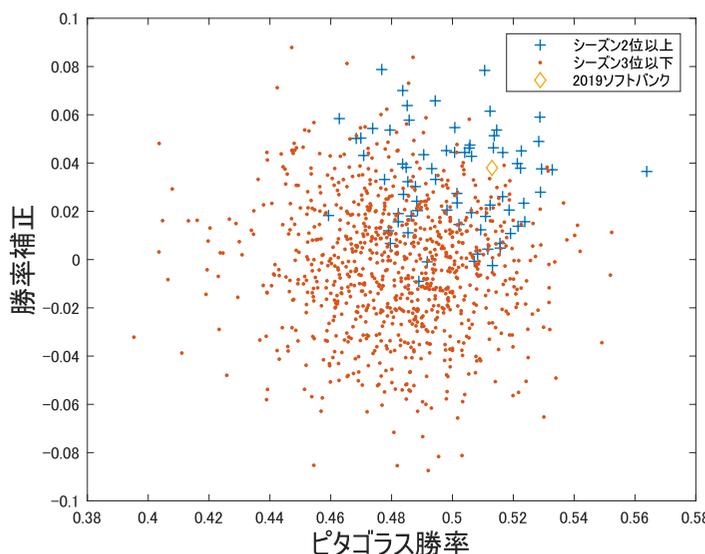


図 14 仮想データ上でピタゴラス勝率 4 位のチームと'19H

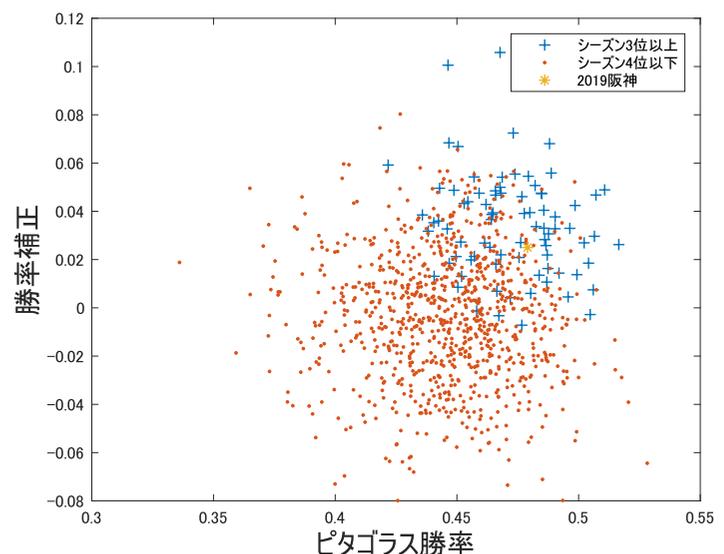


図 15 仮想データ上でピタゴラス勝率 5 位のチームと'19T