

# プロ野球における観客数最大化スケジュール作成法の提案

## Scheduling the Professional Baseball League of Japan

11N8100031C 増子 尚志

中央大学大学院理工学研究科情報工学専攻 松井研究室

2013年 3月

### 1. はじめに

我々の生活の中には様々なスケジュールがある。学校の時間割、電車の運行表、病院などの勤務スケジュール、そして、スポーツにおける対戦表のスケジュール等がある。中でも、アメリカのメジャーリーグ、オリンピック、ヨーロッパのプロサッカーリーグに代表されるように近年のプロスポーツの経済規模の拡大に伴い、スポーツのスケジュールは権利者や選手、ファンのみならず様々な方面で注目されている。

スポーツにおけるスケジュールの作成はスポーツスケジュールリングと呼ばれ、スポーツにおける対戦順序や開催施設、開催日程の割当等を決定する分野である。この分野の研究が活発になったきっかけのひとつとして、アメリカの西海岸大学対抗バスケットボール選手権のスケジュールリングを扱った Nemhauser and Trick による 1998 年の論文 [1] がある。彼らはスポーツスケジュールリングの分類を行い、実際の要求の整理して問題の特徴を明らかにした。さらに、整数計画法と列挙算法を組み合わせた解法を用いてスケジュールを作成した。彼らが作成したスケジュールが実際に使用されたことは、スポーツスケジュールリングの分野の普及に影響を与えた。

日本には、野球、サッカー、バレーボール、バスケットボール等様々なプロスポーツリーグがある。中でも、表 1 が示すように、プロスポーツの中でプロ野球における観客動員数は他のスポーツと比べ、圧倒的に多い事がわかる。

表 1. スポーツ別観客動員数 (2011 年度)

スポーツ	総観客動員数	合計試合数	平均観客動員数
野球	21570196 人	864 試合	24966 人
サッカー	4833782 人	306 試合	15797 人
バスケットボール	764712 人	494 試合	1548 人
バレーボール	249067 人	240 試合	1037 人

現在のプロ野球におけるスケジュール作成は以下

のとおり行なっている。

- 各球団が本拠地 (ホームスタジアム) の使えない日程をプロ野球連盟に提出。
- リーグ戦 120 試合を 4 期に分け (1 期 30 試合)、スケジュールを作成。
- 各球団がホームゲームを正規本拠地で行うか、非正規本拠地で行うかの選択。

この他、雨が降った場合は、シーズンの最後に用意された予備日程で試合と行う。

また、表 2 のとおり各スタジアムの充足率 (平均観客動員数/最大収容人数) は、最も高い福岡ソフトバンクホークスで 0.78 と、改善の余地があることがわかる。

表 2. スタジアムの最大観客動員数と充足率

	北海道日本ハムファイターズ	東北楽天ゴールデンイーグルス
スタジアム名	札幌ドーム	日本製紙クリネックススタジアム
最大収容人数	42002	23036
平均観客動員数	27714	15230
充足率	0.659825723	0.661139087
	千葉ロッテマリーンズ	オリックスバファローズ
スタジアム名	QVC マリンフィールド	京セラドーム大阪
最大収容人数	30082	35000
平均観客動員数	20923	18961
充足率	0.695532212	0.541742857
	埼玉西武ライオンズ	福岡ソフトバンクホークス
スタジアム名	西武ドーム	福岡 Yahoo! Japan ドーム
最大収容人数	33921	38561
平均観客動員数	21469	30225
充足率	0.632911766	0.783823034

本論文では、2010 年度の観客動員数のデータを用い、パ・リーグにおける観客動員数最大化を達成するスケジュール作成と、その評価を行う。日本プロ野球機構に所属する各球団はそれぞれ本拠地を持ち、各試合は対戦するチームのどちらかの本拠地で必ず試合を行うと仮定する。自身の本拠地での試合をホームゲーム、対戦相手の本拠地での試合をアウェイゲームと呼ぶ。実際のスケジュールでは、地方球場等で試合を行うことがあるが、本論文ではそれは考慮しない。表 3 のとおり、一般的に対戦相手の本拠地での試合は自身の本拠地での

試合に比べて不利な点が多い。

表 3. 2010 年度パ・リーグにおけるホーム・アウェイ別勝利数・勝率

チーム名	ホーム勝利数	アウェイ勝利数	ホーム勝率	アウェイ勝率
ファイターズ	39	35	0.542	0.486
イーグルス	37	25	0.514	0.347
ライオンズ	44	34	0.611	0.472
マリナーズ	44	31	0.611	0.431
ホークス	43	33	0.597	0.458
パファローズ	39	30	0.547	0.417

## 2. 用語の説明

### 2.1. 日本プロ野球機構

現在、日本プロ野球機構には、2つのリーグが存在する。それらのリーグはそれぞれセントラル・リーグ、パシフィック・リーグと呼ばれ、それらのリーグには6チームずつ、計12チームが所属している。パシフィック・リーグ所属のチームは、北海道日本ハムファイターズ (F)、埼玉西武ライオンズ (L)、福岡ソフトバンクホークス (H)、千葉ロッテマリーンズ (M)、東北楽天ゴールデンイーグルス (E)、オリックス・パファローズ (B) である。1シーズンは144試合で構成されており、同じリーグのチーム同士での試合をリーグ戦と呼び、各チームは120試合のリーグ戦を行う。また、異なるリーグ同士で行う試合を交流戦と呼び、各チームは24試合の交流戦を行う。各チームは自身の本拠地、すなわちホームスタジアムを持ち、自身のホームスタジアムでの試合をホームゲーム (H)、相手の本拠地での試合をアウェイゲーム (A) とそれぞれ呼ぶ。リーグ戦は4期に分けられ、それぞれを1期、2期、3期、4期と呼ぶ。1期は30試合で構成され、1期で各チームは、同じリーグの他の5チームと6試合ずつ試合をホームで3試合、アウェイで3試合を行う。この連続した3試合のことをカードと呼ぶ。すなわち、1期は10カードで構成され、ホームゲーム、アウェイゲームをそれぞれ5カード、15試合行うことになる。必ず1カードは試合が雨天等で中止になる場合を除き、連続して3試合を行わなければならない。試合が中止になった場合は、4期終了後に用意された予備日程で試合を行う。また、交流戦は通常2期と3期の間、もしくは2期の途中に行われることが多い。

### 3. 観客動員数の予想

本論文では、2010年度のパシフィックリーグの集客データを用いて、それぞれのチームごとに線形回帰分析を行う。その際、日程・対戦相手別に予想観客動員数を推定する。天気、先発投手等を変数に置けば、モデルの当てはまりの良さは良くなる。しかしながら、このような変数は実際にスケジュールを作成する際には不明なので、用いることは出来ない。

本研究では、以下の情報が与えられた問題を扱う。2010年度の日本プロ野球機構のパシフィックリーグの観客動員数データを用いる。1シーズン144試合から交流戦24試合を除いた120試合を4期、各30試合に分割し、各期の予想観客数最大化を行う。各期に必ず3連戦を10回行うこととし、すべての試合は正規本拠地で行うことにし、地方球場などは一切使わないことにする。球団によっては地方球場などの非正規本拠地で試合を行うことがあるが、その場合はデータから除外する。モデルによっては、予想観客動員数が回帰分析の結果から著しく誤差が生じるものが見られたため、このようなデータは外れ値として、4つまで除外を許し、当てはまりの良さが良くなるように務めた。ただし、 $R^2$ の値が0.8以上の場合は、これを行わない。

モデルは、

- 4戦以上連続する場合の変数を置いた休日・平日を同一モデルで扱うもの (モデル1)、
- 4戦以上連続する場合の変数を置かない休日・平日を同一モデルで扱うもの (モデル2)、
- 4戦以上連続する場合の変数を置かない休日・平日を分離した2つのモデルで扱うもの (モデル3)

の3つを扱う。

線形回帰分析を行うにあたり、以下の変数を用いた。

- 対戦相手 (チームの集合 {F, E, L, M, H, B})
- 長期休暇 (お盆, ゴールデンウィーク) であるかどうか
- 開幕戦・閉幕戦かそれ以外
- 月 (月の集合 {3月4月5月6月7月8月9月})

- 8月末かどうか
- 4戦以上の本拠地での連戦かどうか
- 平日か休日か (モデル1, 2のみ)
- 4戦以上連続してホームゲームが続くか (モデル1のみ)

3つのモデルの違いは、平日と休日と同一モデルで扱っているか、4戦以上連続する場合を変数を置くかどうかである。

モデル3の当てはまりの良さは、表4のとおりである。

表4.  $R^2$  結果

チーム	$R^2$	平日		
		観測数	非正規本拠地試合数	外れ値数
パッファローズ	0.68764	31	0	4
ホークス	0.88334	32	0	3
マリーンズ	0.83836	33	0	1
ライオンズ	0.92013	32	4	0
イーグルス	0.94236	26	4	0
ファイターズ	0.80224	25	10	0

チーム	$R^2$	休日		
		観測数	非正規本拠地試合数	外れ値数
パッファローズ	0.90059	25	0	0
ホークス	0.80166	25	0	0
マリーンズ	0.84817	22	0	4
ライオンズ	0.9031	24	0	0
イーグルス	0.87089	26	0	3
ファイターズ	0.84688	21	4	0

ほとんどの回帰分析において、外れ値を除外すると  $R^2$  の値は0.8を超えたが、ドーム球場を持たないために天候の影響を受けやすいチーム、本拠地の最大収容人数が少ないチーム、また充足率の低いチームでは  $R^2$  が0.8を超えないこともあった。

#### 4. 0-1 整数計画問題

本節では、予想観客動員数最大化問題を0-1整数計画問題として定式化する。定式化にあたり、下記の集合、添字、定数、変数を用意する。チームの集合を  $T$ 、カードの集合を  $C$  と表す。  $u, v$  を  $T$  の要素、  $c$  を  $C$  の要素とする。予想観客動員数を、行が  $T$ 、列が  $C$  でインデックスされた観客動員数行列を  $W$  で表す。すなわち、任意の2チーム  $u, v \in T, u \neq v$ 、カード  $c \in C$  について、  $W$  の要素  $w_{u,v,c}$  は、チーム  $u$  が本拠地でチーム  $v$  とカード  $c$  で戦った場合の予想観客動員数を示す。  $W_{u,v,c}$  は、3章で求めた定数である。カード  $c$  でチーム  $u$  が自身の本拠地でチーム  $v$  と戦う場合1、それ以外の場合には0とする0-1整数変数を  $x_{u,v,c}$  とする。なお、本論文ではリーグ戦の観客動員数を最大化することを目

的としているので、交流戦は除外する。交流戦以外の120試合は、4期に分かれているため、1期ごとに予想観客動員数の最大化を目指す。

予想観客動員数を最大化する0-1整数計画問題は、

$$\begin{aligned}
 & \text{maximize } \sum_{u \in T} \sum_{v \in T} \sum_{c \in C} w_{u,v,c} \cdot x_{u,v,c} \\
 & \text{subject to } \sum_{c=1}^C x_{u,v,c} = 1, \\
 & \quad (u \neq v, u \in \{1, \dots, T\}, v \in \{1, \dots, T\}), \\
 & \quad \sum_{v=1}^T x_{u,v,c} + \sum_{v=1}^T x_{v,u,c} = 1 \\
 & \quad (u \in \{1, \dots, T\}, c \in \{1, \dots, C\}), \\
 & \quad x_{u,u,c} = 0 \\
 & \quad (u \in \{1, \dots, T\}, c \in \{1, \dots, C\}), \\
 & \quad \sum_{i=0}^3 \sum_{v=1}^T x_{u,v,c+i} \leq 3 \\
 & \quad (u \in \{1, \dots, T\}, c \in \{1, \dots, c-3\}), \\
 & \quad x_{u,v,c} \in \{0, 1\} \\
 & \quad (u \in \{1, \dots, T\}, v \in \{1, \dots, T\}, c \in \{1, \dots, C\}),
 \end{aligned}$$

と定式化される。

解を求めるために汎用ソルバであるNUOPT12.1.0を使用した。3つのモデルの総観客動員数、計算時間は表5に表す。

表5. モデル比較

	モデル1	モデル2	モデル3
総観客動員数 (人)	8897227	8796114	8899512
総計算時間 (秒)	0.19	0.2	0.27

#### 5. 評価

4章で得られたスケジュールと、2010年のスケジュールの観客動員数の比較を行う。ただし、実際に行われたスケジュールでは、雨天中止などで延期になり、4期の後に用意された予備日程で試合を行なっているが、すべて予定通りに行うとする。そのスケジュールに、3章で求めた予想観客動員数を当てはめる。実際に行われたスケジュールによる観客動員数のデータを用いないのは全ての試合を正規本拠地で行うと仮定して予想観客動員数を求めているため、実際のデータに比べ観客動員数が大きくなる可能性が高いためである。

その結果、実際のスケジュールと提案したスケジュールの観客動員数は以下ようになった。

表 6. 期別観客動員数比較

	実際のスケジュール	モデル 3	比率
1 期	1763932	2041668	1.15745
2 期	1879687	2178140	1.15878
3 期	2042729	2293085	1.12256
4 期	2061652	2386619	1.15762
合計	7748000	8899512	1.14862

提案したスケジュールの観客動員数は 8899512 人、実際のスケジュールの観客動員数は 7748000 人となり、14.9% 増加することが分かった。期別にみると、もともと観客動員が見込める 7, 8 月に試合を行う 3 期はあまり予想観客動員数が増加しない結果となった。

表 7. 球団別比較

	F	E	L
実際のスケジュール	1596353	903338	1271546
モデル 3	1904564	913452	1401020
比率	1.19307	1.0112	1.10182
	M	H	B
実際のスケジュール	1191343	1796965	988455
モデル 3	1350892	1848249	1481335
比率	1.13392	1.02854	1.49864

表 7 が示すように、全てのチームにおいて、実際のスケジュールより観客動員数が増加することがわかった。しかしながら、球団によって増加量に差があり、最も増加したオリックスバファローズと東北楽天ゴールデンイーグルスの間には比率で約 0.48 もの差が見られた。表 2 のとおりオリックスバファローズの平均充足率が他球団に比べて低いこと、本拠地の最大観客動員数が多いことがあげられる。

## 6. まとめと今後の課題

### 6.1. 予想観客動員数のまとめ

本論文では、2010 年度の日本野球機構パシフィックリーグの観客動員数を用いて線形回帰分析を行い、その結果を用いて 0-1 整数計画問題を解くことによって、観客動員数が約 14.9% 増加するスケ

ジュールを求めるモデルを作成することが出来た。線形回帰分析の当てはまりの良さ  $R^2$  の値は、ほとんどの球団で 0.8 以上となった。計算時間は 1 つのモデルで 0.19 秒から 0.27 秒と非常に短くなった。正規本拠地の最大観客動員数や、平均充足率によっても予想観客動員数の増加率は大きく左右されることがわかった。また、全てのチームにとっても観客動員数が増加する結果となった。

### 6.2. 今後の課題

最大の課題であるのは、精度のいいモデルの作成である。予想観客動員数は、入力データに大きく依存するので、集客数予測に必要であると考えられるなるべく少ない説明変数で、どこまで相関性を高められるかが課題になってくる。先発投手や天気等を変数に入ればもっと良くなると言われているが、スケジュール作成のためにはこれらの変数は使えない。また、2010 年度以外、またセリーグの観客動員数予想を同じモデルで行うことによってより精度を高めていく必要もある。

本論文のモデルでの制約に加えて、

- 正規本拠地以外での試合もモデルに含む、
- 平日と休日の割り振りを均等にする、
- 各球団の観客動員数の増加率の平均化、
- そもそも 120 試合のゲームを行う必要があるのか

などを考慮したモデルについて研究することが今後の課題となる。

### 参考文献

- [1] G. Nemhauser and M. Trick, "Scheduling a Major College Basketball Conference," *Operations Research*, **46** (1998), pp. 1-8.
- [2] 大木 竜也, Jリーグ集客予測とスケジュールの構築, 卒業論文, 中央大学, 2009 年 3 月
- [3] 間瀬茂, 神保雅一, 鎌倉稔成, 金藤浩司, 「工学のためのデータサイエンス入門」, 数理工学社, 2004.
- [4] 日本野球機構, 日本野球機構オフィシャルサイト, Web Page. (available at <http://www.npb.or.jp/>)