

# 在来線優等列車の座席予約状況に基づく旅客需要の分析\*

Passenger demand analysis based on the situation of reserved seats for limited train\*

奥田大樹\*\*・柴田宗典\*\*\*・田村一軌\*\*\*\*・武藤雅威\*\*\*\*\*

By Daiki OKUDA\*\* Munenori SHIBATA\*\*\* Kazuki TAMURA\*\*\*\* Masai MUTOH\*\*\*\*\*

## 1. はじめに

優等列車の指定席は、購入した旅客に着席を保障するものであり、その需要は大きく十分な座席数を確保することが望ましい。しかし、在来線においては優等列車の増発余地は少なく、需要が集中した列車では、需要を満たすことなく指定席が満席になってしまい、一部の旅客が予定の変更や取りやめなど、不利益を被る事象が往々にして発生する。優等列車の指定席は、乗車日の30日前から予約が可能であり、この需要は連休や週末など暦日の並び（以下、暦日情報）と深い相関があると考えられる。この相関性を定量的に評価できれば、指定席の混雑予測情報を指定席の発売に先立って提供することが可能となり、指定席不足により旅客が被る不利益を低減させることに繋がると考えられる。

本研究の目的は、在来線優等列車の予約状況データから、指定席の混雑発生と暦日情報の相関性を分析して、混雑予測に資する情報を見出すことであり、後述する実際の予約状況データを用いて検討をおこなう。ここで、混雑は旅客の指定席需要が一定以上になったときに発生するものであり、この需要は様々な目的（帰省や旅行など）の需要が複雑に混ざりあったものである。よって、精度の高い情報を得るためには、指定席需要を旅客の目的ごとの指定席需要に細分化する必要があると考えられる。また、指定席需要の発生は正規分布に従わないと考えられる。そこで、抽出する因子や成分が統計的に独立、かつ非正規性という特徴を持つ独立成分分析を用いて、混雑発生と暦日情報の相関性を分析することとする。

## 2. 在来線優等列車の予約状況データの分析

### (1) 独立成分分析

独立成分分析とは、いくつかの独立な原信号 $s_j(t)$ が混ざ

\*キーワード：鉄道計画，交通情報，公共交通運用

\*\*正員，修(工)，(財)鉄道総合技術研究所

(東京都国分寺市光町2-8-38, TEL042-573-7309, FAX042-573-7305)

\*\*\*正員，修(工)，(財)鉄道総合技術研究所

\*\*\*\*非会員，博(都市・地域計画)，(財)福岡アジア都市研究所

\*\*\*\*\*正員，博(工)，(財)鉄道総合技術研究所

り合い構成された観測信号 $x_i(t)$ から、元の互いに独立な信号を復元する手法である。観測信号 $x_i(t)$ は式(1)で表される。また、前処理として、観測信号の平均を0、分散を1とする正規化をおこなう。

$$x_i(t) = \sum_j a_{ij} s_j(t) \quad (1)$$

$a_{ij}$ は原信号 $s_j(t)$ が観測信号 $x_i(t)$ に占める割合の大きさ、つまり重みを表す。ただし、推計される成分の大きさと符号、そして順序には任意性が残る。

### (2) 在来線優等列車の予約状況データの分析手順

本研究では、インターネット上で公開されている在来線優等列車の予約状況データを分析対象とし、分析対象路線の特定区間で運行される全優等列車の、乗車日30日前から前日までの予約状況データを、順行・逆行それぞれ取得期間中毎日収集した。ここで、大都市から地方都市へ向かう列車を順行とし、地方都市から大都市へ向かう列車を逆行とする。また、予約状況データは、残席割合に応じて「○」「△」「×」といった形式で公開されており、残席割合の正確な情報は公表されていない。そこで、「×」を満席、「△」を80%以上の座席が予約された状態とし、「○」はそれ以外の状態とした。ここで、予約状況データが「×」の状態になっている列車の指定席は既に予約不可であり、「△」の状態になっている列車の指定席についても、希望する座席の予約ができない場合があるなど、旅客が指定席を購入するに際して不利益を被る可能性があり、その大きさを表す指標として混雑得点を導入した。それぞれ混雑得点として「×」は1点、「△」は0.8点を付与した。また、「○」は旅客が不利益を被ることなく自由に指定席を購入できる状態とし、混雑得点は0点とした。この混雑得点は、列車の混雑状況を比較する間隔尺度の性質を有した量的データであり、乗車日30日前から前日までの混雑得点の合計値（総混雑得点）によって、列車ごとの指定席の混雑状況を表す。

7:00~22:00を1時間ごと15の時間帯 $t$  ( $t=1, 2, \dots, 15$ )に分割し、暦日 $i$ において時間帯 $t$ 内に分析対象区間の基準点を通る全列車の総混雑得点の平均値を、時間帯 $t$ の混雑指数 $x_i(t)$ とし、順行・逆行ごとに整備する。こ

の混雑指数の値が0の時間帯は、混雑が発生していないことを意味する。ここで、混雑指数の値が0.8点を下回る可能性が考えられるが、これは「△」の混雑得点である0.8点以下ということになり、意味合い的には「○」の状態となる。よって混雑指数 $x_i(t)$ が0.8点以下となった場合は、一律で0点とする。

データ取得期間を $m$ 日間とし、混雑指数データ $\mathbf{X}(x_1(t), x_2(t), \dots, x_m(t))$ に独立成分分析を適用し、 $n$ 個の原系列 $\mathbf{S}(s_1(t), s_2(t), \dots, s_n(t))$ と、原系列の重み $a_{mn}$ を順行・逆行ごとに推計する。原系列 $\mathbf{S}$ は時間帯による旅客の目的別需要度合いを表すものであり、重み $a_{mn}$ は各日の混雑指数に対する原系列それぞれの寄与度を表すものである。

抽出する原系列の数が多すぎると、各原系列の解釈が困難となるので、主成分分析による次元縮約をおこなう。本研究では、累積寄与率が90%を超えるまでの合成変量を基準とし、適切な数の原系列となるよう調整する。

原系列の重み $a_{mn}$ を目的変数として重回帰分析をおこない、各原系列と暦日情報の相関性を分析し、得られた情報が混雑予測に資するかどうかを検討する。

表3-1 暦日情報の一覧

暦日情報		期間
大型連休	ゴールデンウィーク	2007年5月3日～6日 2008年5月3日～7日
	お盆	2007年8月11日～15日 2007年8月13日～17日
	年末年始	2007年12月29日～1月3日
	通常	連休 週末
夏季	夏休み	2007年7月20日～9月2日 2008年7月19日～8月31日

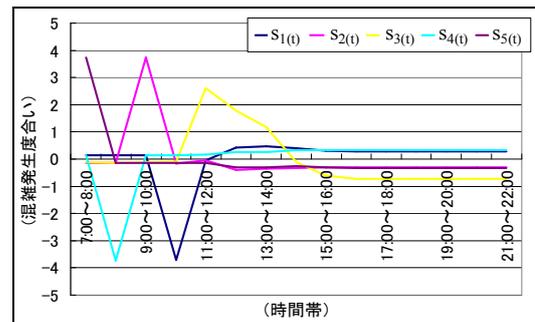


図4-1 Z本線順行の原系列の時間帯分布

### 3. Z本線における予約状況データの概要

本研究の分析対象路線はZ本線であり、分析対象区間は同線上の $\alpha$ 駅～ $\beta$ 駅間である。順行は $\alpha$ 駅から $\beta$ 駅方面へ向かう列車であり、基準点を順行は $\alpha$ 駅の発時刻、逆行は $\alpha$ 駅の着時刻とした。データ取得期間は、順行が2007年4月26日～2008年9月30日までの524日間、逆行は2008年7月20日～2008年9月30日までの73日間であり、期間中の暦日情報は表3-1のように分類した。

前章の定義に従い取得データを加工し、順行の混雑指標 $\mathbf{X}(x_1(t), x_2(t), \dots, x_{524}(t))$ と、逆行の混雑指標 $\mathbf{X}'(x_1'(t), x_2'(t), \dots, x_{73}'(t))$ のデータを得た。

### 4. Z本線予約状況データの分析結果と考察

次元縮約の結果、抽出する原系列は、順行で5系列、逆行で3系列となり、独立成分分析を適用することで、原系列 $\mathbf{S}(s_1(t), s_2(t), s_3(t), s_4(t), s_5(t))$ 、 $\mathbf{S}'(s_6(t), s_7(t), s_8(t))$ と、それぞれに対応する重み $\mathbf{A}(a_1, a_2, a_3, a_4, a_5)$ 、 $\mathbf{A}'(a_6, a_7, a_8)$ を得た。図4-1と図4-2は、原系列 $\mathbf{S}$ と $\mathbf{S}'$ を示す。

以下では、分析結果とそれに基づく考察を示す。

#### (1) 予約状況データの分析

順行はデータ取得期間の約33% (171日) で混雑指標の値が0以外となる時間帯が存在し、その内約70% (118日) が休日であった。また、午前中から午後の早い時間帯に混雑の発生が集中していた。

逆行はデータ取得期間の約49% (36日) で混雑指標の値

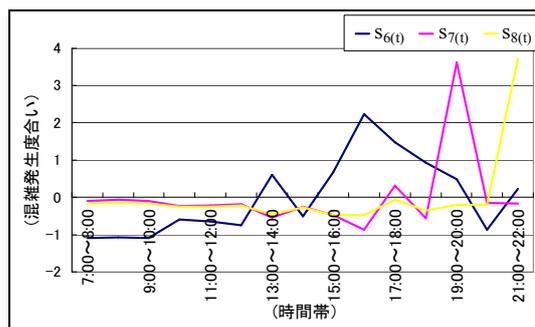


図4-2 Z本線逆行の原系列の時間帯分布

が0以外となる時間帯が存在し、その内約61% (22日) が休日であった。また、午後の比較的遅い時間帯を中心に混雑の発生が集中していた。

順行では大型連休の前半に混雑指数の値が非常に高くなり、逆行では大型連休(お盆のみ)の後半に混雑指数の値が非常に高くなった。混雑の主要因は帰省や旅行に関する需要と考えられ、順行では、それぞれの帰省先や旅先に向かうものが主であり、逆行ではそのUターンと考えられる。ただし、大型連休の中でもゴールデンウィークは、旅行に関する需要の割合が高いと考えられ、宿泊日数などの旅程も旅客によって様々であり、順行・逆行共に需要が集中する日時にはばらつきがあると考えられる。一方で、年末は帰省に関する需要の割合が高いと考えられ、旅客の帰省期間は概ね同程度であることから、旅程は比較的同じようなものであり、順行・逆行共に需要が集中する日時のばらつきは小さいと考えられる。お盆はこれらの中間と考えられる。

通常期間では、連休や週末などを中心に混雑が発生しており、順行では土曜日の約66%(49日)、日曜日の約43%(31日)で混雑指標の値が0以外となる時間帯が存在し、それらの1日の混雑指数の合計値の平均は土曜日が11.05、日曜日が5.76であった。また、逆行では土曜日の約67%(6日)、日曜日の約78%(7日)で混雑指標の値が0以外となる時間帯が存在し、それらの1日の混雑指数の合計値の平均は土曜日が5.57、日曜日が14.89であった。よって、通常期間においても週末や連休を中心として、旅行や帰省に関する需要が発生していると考えられ、順行における旅客は主に土曜日にそれぞれの旅先や帰省先に向かい、逆行は主に日曜日にUターン需要が発生していると考えられる。ただし、大型連休期間と比較して時間的な余裕は少なく、日帰りや1泊2日の旅行など、短期のものが中心であると考えられる。また、前後に大型連休や連休がある週末は、混雑指数の合計値が比較的低くなっており、出控が発生しているものと考えられる。

夏季については、順行・逆行共に混雑指数の値が高く、平日でも混雑が発生していた。これは、夏季が最も旅行需要の大きい季節であることに加え、夏休み期間中の学生による帰省や旅行に関する需要や、それに合わせた家族での帰省や旅行に関する需要などが発生しているものと考えられる。

(2) 各原系列と暦日情報の相関性に関する分析

各原系列と暦日の相関性を分析するため、原系列ごとの重みを目的変数として重回帰分析をおこなった。説明変数は表4-1に示すもので、全てダミー変数である。ここで、逆行において、2008年9月2日(火)に特異な混雑が発生していることから、これも説明変数に加えた。

以下では、原系列ごとの分析結果と考察を示す。

a) 順行の各原系列と暦日情報の相関性に関する分析

原系列 $s_1$ に対する重み $a_1$ について、表4-2より決定係数は0.574であり、中程度の当てはまりである。説明変数の

表4-1 説明変数の一覧

説明変数		対応する暦日情報
大型連休	GW初日	ゴールデンウィークの初日
	GW2日目	ゴールデンウィークの2日目
	お盆前半	お盆の前半期間
	お盆後半 年末	お盆の後半期間 12月29日～12月31日
通常	週末初日	週末の初日(土曜日)
	週末終日	週末の終日(日曜日)
	連休初日	連休の初日
	連休終日	連休の終日
	前後連休初日	前後に連休がある週末の初日
	前後連休終日	前後に連休がある週末の終日
夏季	夏休み平日	夏休み期間の平日
	夏休み休日	夏休み期間の休日
2008年9月2日		2008年9月2日の1日

うち、GW初日とお盆前半の影響が大きく、同じ大型連休の年末は比較的影響は小さい。その他では、連休初日の影響も大きい。また、図4-1より、 $s_1$ は10:00～11:00の比較的遅い時間帯に混雑発生度合いが大きく、出発時間に比較的余裕があると考えられる。よって原系列 $s_1$ は、長期の休暇中(GW, お盆, 連休など)などを中心に発生する、比較的時間に余裕のある旅行(宿泊を伴う旅行など)をおこなった旅客の混雑発生度合いを中心に表す原系列と考えられる。

原系列 $s_2$ に対する重み $a_2$ について、表4-2より決定係数は0.513であり、中程度の当てはまりである。原系列 $s_1$ と同様の傾向ではあるが、GW初日とお盆前半の影響が相対的に低くなっており、逆に週末初日の影響が大きくなっており、連休初日の影響も大きい。更に前後連休初日の負の影響も大きく、出控を示している。その他では、夏休み休日や2008年9月2日の影響も見られるが、夏休み休日については、夏季ということを除けば通常の週末と同義である。また、図4-1より、 $s_2$ は9:00～10:00の時間帯に混雑発生度合いが大きくなっており、 $s_1$ よりも早い時間帯である。よって原系列 $s_2$ は、週末などに発生する比較

表4-2 重回帰分析の推計結果

説明変数	順行								逆行							
	$S_1$		$S_2$		$S_3$		$S_4$		$S_5$		$S_6$		$S_7$		$S_8$	
	推定値	t値	推定値	t値	推定値	t値	推定値	t値	推定値	t値	推定値	t値	推定値	t値	推定値	t値
GW初日	3.313	18.09 **	3.002	11.76 **	5.652	61.93 **	3.875	12.93 **	0.914	4.23 **	—	—	—	—	—	—
GW2日目	0.080	0.62	0.535	2.95 **	-0.035	-0.54	0.261	1.23	-0.080	-0.52	—	—	—	—	—	—
お盆前半	1.588	13.36 **	1.767	10.66 **	1.491	25.17 **	1.976	10.16 **	-0.232	-1.65	-0.066	-0.29	-0.220	-1.93	-0.074	-0.08
お盆後半	0.197	1.40	0.106	0.54	0.273	3.91 **	-0.300	-1.30	-0.276	-1.67	3.109	12.08 **	0.462	3.64 **	-1.090	-1.11
年末	0.781	5.22 **	-0.253	-1.21	1.354	18.15 **	-0.165	-0.67	-0.274	-1.55	—	—	—	—	—	—
週末初日	0.125	2.58 *	0.546	8.09 **	-0.029	-1.21	0.364	4.60 **	0.109	1.90	-0.636	-2.47 *	-0.096	-0.75	0.031	0.03
週末終日	0.021	0.43	-0.022	-0.33	-0.017	-0.71	0.142	1.79	0.050	0.88	-0.148	-0.57	-0.051	-0.40 **	3.188	3.25 **
連休初日	0.770	10.50 **	1.019	9.97 **	0.157	4.29 **	1.110	9.25 **	0.452	5.23 **	0.055	0.27	-0.077	-0.77	1.381	1.78
連休終日	-0.025	-0.35	-0.025	-0.25	-0.003	-0.09	0.133	1.12	-0.019	-0.22	0.191	1.07	-0.111	-1.27	0.462	0.68
前後連休初日	-0.111	-1.64	-0.483	-5.12 **	0.020	0.60	-0.133	-1.20	-0.043	-0.54	0.236	0.99	0.016	0.14	0.364	0.40
前後連休終日	-0.064	-0.95	-0.020	-0.21	0.019	0.57	-0.065	-0.59	-0.076	-0.96	0.832	3.49 **	0.156	1.33	-2.918	-3.20 **
夏休み平日	0.049	1.26	-0.001	-0.01	-0.015	-0.79	0.450	7.07 **	-0.006	-0.13	-0.004	-0.05	0.004	0.10	-0.318	-1.03
夏休み休日	0.130	2.42 *	0.399	5.31 **	0.012	0.44	0.761	8.63 **	0.381	5.99 **	0.601	3.82 **	0.110	1.41	-0.578	-0.96
2008年9月2日	0.011	0.04	1.816	5.04 **	-0.057	-0.44	-0.105	-0.25	-0.102	-0.33	0.794	2.82 **	3.370	24.24 **	-0.745	-0.69
定数項	0.013	0.93	0.017	0.85	-0.004	-0.53	0.030	1.27	0.027	1.57	0.035	0.58	-0.014	-0.47	0.547	2.38 *
自由度調整済み決定係数	0.574		0.513		0.905		0.535		0.142		0.855		0.900		0.174	

\*\* 1%有意 \* 5%有意

的時間に余裕のない旅行（日帰りや1泊2日の旅行など）をおこなった旅客の混雑発生度合いを中心に表す原系列と考えられる。

原系列 $s_3$ に対する重み $a_3$ について、表4-2より決定係数は0.905であり、当てはまりは非常に良い。説明変数のうち、GW初日、お盆前半、年末の影響が非常に大きいですが、その他の原系列と比較すると、年末の影響が相対的に大きい。また、図4-1より、 $s_3$ は11:00~14:00の広い時間帯で混雑発生度合いが大きいことがわかる。これは、旅客の時間選択に比較的余裕があること、若しくは特定の日に需要が集中し、希望の時間帯の列車に乗車できなかった旅客が、前後の時間帯の列車に流れることで混雑が広がったものと考えられる。よって原系列 $s_3$ は、大型連休に帰省をおこなった旅客の混雑発生度合いを中心に表す原系列と考えられる。

原系列 $s_4$ に対する重み $a_4$ について、表4-2より決定係数は0.535であり、中程度の当てはまりである。原系列 $s_2$ と同様の傾向ではあるが、原系列 $s_2$ と比較して夏休み平日、夏休み休日の影響が大きい。また、図4-1より、 $s_4$ は8:00~9:00の早い時間帯に混雑発生度合いが大きいことが分かる。よって原系列 $s_4$ は、原系列 $s_2$ と同様の目的を持った旅客の混雑発生度合いを中心に表す原系列と考えられるが、ゴールデンウィークから夏季にかけては日の出時刻が早く、それに伴い旅客の出発時刻が早くなっているものと考えられる。また、夏休み平日の影響も大きいですが、平日には業務を中心とした旅客の需要が一定量あると考えられ、そこに夏休み期間中の休暇を利用した学生などの旅行需要が加わり、混雑が発生したものと考えられる。

原系列 $s_5$ に対する重み $a_5$ について、表4-2より決定係数は0.142であり、当てはまりは悪い。よって、原系列 $s_5$ は暦日との相関性は認められない。

#### b) 逆行の各原系列と暦日情報の相関性に関する分析

原系列 $s_6$ に対する重み $a_6$ について、表4-2より決定係数は0.855であり、当てはまりは非常に良い。説明変数のうち、お盆後半の影響が非常に大きい。また、図4-2より、15:00~20:00の広い時間帯で混雑発生度合いが大きいことが分かる。これは原系列 $s_3$ と同様の理由が考えられる。よって原系列 $s_6$ は、お盆に帰省をおこなった旅客のUターンの混雑発生度合いを中心に表す原系列と考えられる。

原系列 $s_7$ に対する重み $a_7$ について、表4-2より決定係数は0.900であり、当てはまりは非常に良い。ただし、説明変数のうち2008年9月2日以外の影響はほとんど無く、原系列 $s_8$ は、2008年9月2日に突発的に発生した混雑発生度合いを表し、暦日との相関性は少ないと考えられる。

原系列 $s_8$ に対する重み $a_8$ について、表4-2より決定係数は0.174であり、当てはまりは悪い。よって原系列 $s_8$ は、暦日との相関性は認められない。

#### (3) 各原系列と暦日情報の相関性に関する考察

以上の分析より、以下の知見を得ることができた。

- 指定席の混雑は、主に休日の帰省や旅行に関する需要の影響が中心であり、順行では主に期間の前半、逆行では主に期間の後半に混雑が発生する。
- 大型連休で発生した帰省需要に関する混雑発生度合いを表す原系列は、順行・逆行共に暦日情報と高い相関性があり、順行・逆行間の混雑発生にも相互関係がある。
- 大型連休や通常の連休、週末などに発生した旅行需要の混雑発生度合いを表す原系列は、順行・逆行間の混雑発生に相互関係は認められるものの、暦日情報との相関性が認められるのは順行のみであり、それも、帰省需要による混雑ほどの相関性は無い。

以上より、大型連休の帰省需要に関する混雑発生度合いについては、順行・逆行ともに指定席の混雑予測に資する情報に成りえると考えられる。

大型連休や連休・週末の旅行需要に関する混雑発生度合いについては、帰省需要ほどではないが順行においては指定席の混雑予測に資する情報に成りえると考えられる。一方で、逆行においては暦日情報との相関性が認められず、指定席の混雑予測に資する情報としては不適切であると考えられる。帰省と比較して、旅行の宿泊日数などの旅程は旅客それぞれの意思が強く反映されるためばらつきが大きく、特にUターン需要のばらつきが大きいため、逆行では混雑の発生はあるものの、暦日情報との相関性が認められなかったと考えられる。ただし、逆行は順行と比較してデータ取得期間が短く、データに偏りがある可能性も考慮する必要がある。

#### 5. おわりに

今後の課題としては、以下が挙げられる。

- より精度の高い予約状況データ（残席の実数データなど）を用いて、より詳細な分析をおこなうとともに、主に平日の発生が予想される業務目的の指定席需要に関しても分析をおこなう必要がある。
- 同程度の規模の都市間を結ぶ在来線優等列車の予約状況データなど、今回取り上げたものとは性格の異なるデータを利用し分析をおこなうこと。

#### 参考文献

- 1) Aapo Hyvarinen : 詳解独立成分分析, 東京電気大学出版局, 2005
- 2) 塚井誠人, 井上真一, 奥村誠 : データマイニングによる広島空港流入車両の時間分布パターンの分析, 交通工学研究発表会論文報告集, No. 28. pp. 269-272, 2008.