

# 国内旅行行動の地域性:EM アルゴリズムによる需要分類

佐藤 彰洋 \*

京都大学 大学院情報学研究科 数理工学専攻

## 概要

本研究では、国内ホテル予約サイトから毎日収集した利用可能な宿泊プランに関する網羅的なデータを用いることにより、予約可能なプラン数から推計される需要の地域依存性を調べた。国内 4 つの地域における日次需要代理変数に対して有限混合ポワソン分布のパラメータを EM アルゴリズムにより推定する。EM アルゴリズムを計算するためのパラメータ初期値は Finch らの方法を用い、最適な混合数の決定に赤池情報量基準 (AIC) を用いた。計算の結果、需要の多いカテゴリに分類される宿泊日 (繁忙期) は、需要の少ないカテゴリに分類される宿泊日 (閑散期) と比較して、利用可能な宿泊プランの平均価格の差異が大きかった。

## 1 はじめに

近年インターネットを通じて商品やサービスを購入する仕組みが普及してきている。1990 年代後半に始まったインターネットの商用利用は商品やサービスの購入と販売に対して大きな変化をもたらした。Web ページを介したインターネットにおける商品販売では、自動的に購入決済を行うことができるため、商品やサービスの購入履歴に関する多くのデータが蓄積できるようになってきている。これらのデータを用いることで、間接的に我々の社会に関する経済状況を分析できるようになりつつある。

このようなインパクトは旅行観光において、宿泊予約の電子化という形で普及してきている [1]。近年、インターネットを通じた宿泊予約の網羅性は高まりつつある。国土交通省観光庁の統計データによると 2010 年現在、日本国内にはホテル、旅館、民宿、保養所を含め約 5 万の宿泊施設が存在している [2]。本研究で用いるホテル予約サイトのひとつリクルート社じゃらんの場合、インターネットを通じて予約可能な国内の宿泊施設総数は約 16,000 にのぼる。

インターネットを用いた宿泊予約の個人的な経験から、我々は予約できる宿泊プランが時間的に変化することを知っている。この宿泊プランの利用可能性はランダムに変化しているようにさえ見える。更に、宿泊プランを探すとき、日時と場所は重要な因子である。予約できるかできないかはカレンダー日 (平日、週末、祝日) と地域に強く依存している。

この宿泊施設における予約可能性は将来の旅行者の移入を示しているかもしれない。すなわち、網羅的なデータを用いて宿泊可能性を調べることにより、国内における旅行者の移動を理解する手がかりを得ることができる。

実際社会的な移入動学はこれまで社会力学 (socio-dynamics) の文脈において活発に研究されてきた [3, 4]。Weidlich と Haag は地域と時間に依存した効用と嗜好を含む遷移確率を含むマスター方程式を提案している。このモデルによりエージェントの移入に関する意思決定の傾向を記述することが可能である [5]。

\* 〒 606-8501, 京都市左京区吉田本町, TEL:075-753-5515, FAX:075-753-4919, E-mail:aki@i.kyoto-u.ac.jp

移入の動機は心理的かつ物理的因子から来ているように思われる。観光学ではこれを push 効果と pull 効果として理解している [6]。push 効果とは旅行者をその場所から押し出そうとする因子であり、旅行者の側の理由 (需要) である。一方 pull 効果とは地域それぞれが持つ魅力であり、旅行者を引き付ける観光地の側の持つ理由 (供給) である。この旅行における動力学を理解することは、社会全体におけるエージェントの内部状態と集団行動に関する有益な洞察を与えることになる。

本稿では、約 1 年間にわたる日本国内ホテル予約サイトから収集した約 16,000 の宿泊施設が提供する宿泊プランに関するデータから各地域における需要代理変数を計算し、この需要代理変数の時系列に対して、有限混合ポワソン分布の EM アルゴリズムを用いることで、宿泊日を分類する方法を提案する [7]。更に、各地域における宿泊日を需要の大小にわけて分類し、利用可能プランの平均価格と宿泊日が属するカテゴリとの関係を調べることで、地域の宿泊需要と価格選好の特徴付けをおこなう。

本稿は以下の構成からなっている。2 節では、分析に用いたデータと収集方法について概説する。3 節では、宿泊日を需要代理変数に基づき分類するアルゴリズムを説明する。4 節では提案手法を用いて、強度の異なる複数のポワソン分布から抽出した人工的な混合乱数列を用いてパラメータの推定を行ない、推定アルゴリズムの有効性を確認する。5 節では提案手法を用いて 4 つの地域における需要代理変数のパラメータ推定と宿泊日の分類をおこなう。6 節ではまとめと今後の課題を記す。

## 2 データ

この節では、利用可能なプランを収集する上で用いたホテル予約サイトの仕組みを簡単に説明する。本研究で用いたデータは、じゃらん Web サービス上で利用可能なプランを Web API 経由で毎日収集し、csv ファイルとして保存したものである。

じゃらん (<http://www.jal-an.net>) は日本国内でのホテル予約をインターネット経由で行うことができるサービスを提供しているホテル予約サイトのひとつである。国内約 16,000 箇所の宿泊施設の予約が可能である。

ホテル予約サイトは次のような仕組みで成り立っている。(i) 宿泊施設のマネージャーはサーバー上に提供可能な部屋を入力する。(ii) 提供可能な部屋やプラン情報はじゃらん Web ページ上で表示される。(iii) 顧客はじゃらん Web ページで表示されている宿から予約したいものを選び、予約を行う。(iv) 利用者の情報は予約した宿泊プランを提供するホテルフロントへと送られる。

本分析では、じゃらん Web ページ上で大人 2 名が 1 泊宿泊可能なプランの情報を集めた。表. 1 は収集したプラン属性を表わしている。データには地域情報が含まれているため、地域ごとの利用可能な宿泊プラン数を計測することが可能である。この分析では、利用可能な宿泊プラン数をその地域における利用可能な部屋数の代理変数とみなし、地域ごとの需要が期間でどのように変化するかに着目して分類をおこなう。分析対象期間は 2009 年 12 月 24 日から 2010 年 11 月 4 日までとした。図. 1 はデータ期間中の代表的な 3 日間における 1,000 円ごとのプラン価格のヒストグラムである。プラン価格のヒストグラムは、2010 年 1 月 1 日では 8,000 円、2010 年 3 月 4 日では 9,000 円、2010 年 8 月 31 日では 9,000 円を最頻値とする形状をしており、日本全体において、滞在日に応じ宿泊日前日において利用可能な価格帯が異なることが確認される。

図. 2 は日本国内全体で利用可能なプラン数を表わす。サンプル期間を通じて週末 (土曜日)、祝日前、連休日に予約可能な宿泊プラン数が減少することが確認できる。特に年末年始と春の行楽シーズンであるゴールデンウィーク、および秋の行楽シーズンにおけるシルバーウィークにおいて利用可能な宿泊プラン数の減少が顕著である。更に、7 月後半から 8 月終りにかけての夏季休みの期間

表 1: プラン属性

データ収集日時
宿泊日
宿泊施設識別番号
宿泊施設名称
宿泊施設名称 (カナ)
郵便番号
住所
URL
緯度
経度
プラン名称
食事
直近の最安値
一泊当たりの価格

は期間全体を通じて利用できる宿泊プランが減少する傾向がある。また、大型連休直後の土曜日は利用可能な宿泊プラン数はほとんど減少しないことが確認される。一方、1月から2月および6月の週末は他の月の週末と比べて宿泊プラン数の減少は顕著でない。このように、日本国内の旅行にともなう集団行動は気象条件や暦に強く影響を受けることが確認される。

本研究では各地域で前日に利用可能なプランの数がその地域における需要と関係すると考え、観測期間内での最大プラン数からどの程度プラン数が減少したかを需要のめやすとする。そして、需要の代理変数として  $z_m(t) = \max Y_m(t) + 10 - Y_m(t)$  を定義する。ここで、 $Y_m(t)$  は宿泊日  $t$  のいける地域  $m$  で確認される宿泊プラン数である。小樽 (北海道)、会津高原、湯野上、南会津 (福島県)、白金 (東京都)、湯沢 (新潟県) における需要代理変数の時間変化を図. 3 に示す。

### 3 分類アルゴリズム

#### 3.1 有限混合ポワソン分布

宿泊日時  $t$ 、地域  $m$  における日次需要代理変数  $z_m(t)$  を有限混合ポワソン分布からの抽出とみなす。

$$\Pr_{Z_m}(\cdot)$$

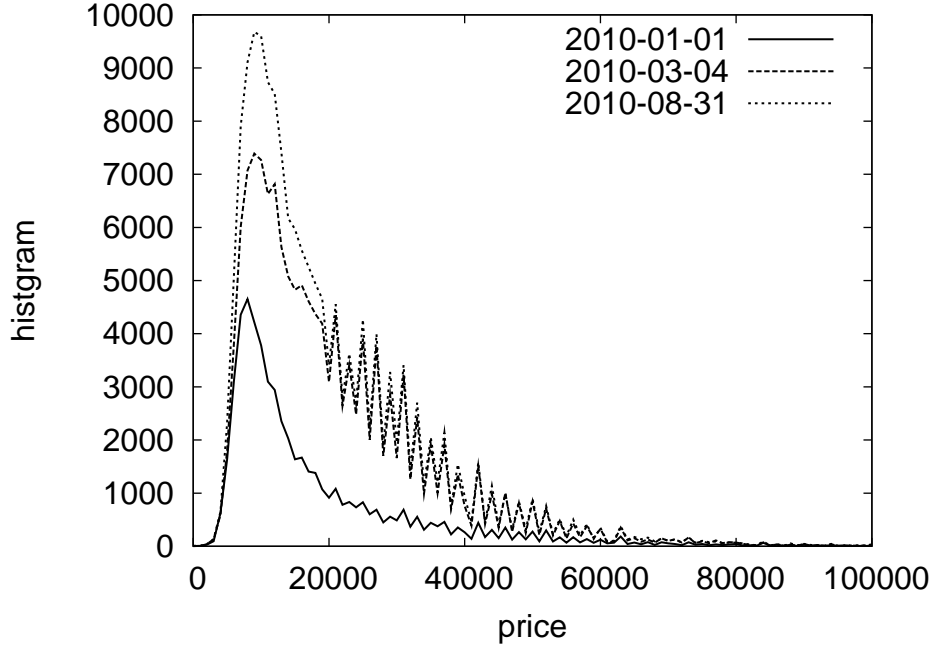


図 1: 2010 年 1 月 1 日, 2010 年 3 月 4 日, 2010 年 8 月 31 日における日本全体でのプラン価格のヒストグラム.

$z_m(t)$  ( $t = 1, \dots, T$ ) を地域  $m$  における日時  $t$  での需要代表変量であるとする. このとき, (1) 式に対する対数尤度関数は

$$L_m(a_{m1}, \dots, a_{mK_m}, r_{m1}, \dots, r_{mK_m}) = \sum_{s=1}^T \log \left( \sum_{i=1}^{K_m} a_{mi} \frac{(Mr_{mi})^{z_m(s)}}{z_m(s)!} e^{-Mr_{mi}} \right), \quad (2)$$

と書かれる. パラメータ  $r_{mi}, a_{mi}$  ( $i = 1, \dots, K_m$ ) は最尤法により, 条件  $\sum_{i=1}^{K_m} a_{mi} = 1$  のもとでの対数尤度関数の最大化

$$\{\hat{a}_{m1}, \dots, \hat{a}_{mK}, \hat{r}_{m1}, \dots, \hat{r}_{mK}\} = \arg \max_{\{a_{mi}\}, \{r_{mi}\}} L_m(a_{m1}, \dots, a_{mK_m}, r_{m1}, \dots, r_{mK}) \quad (3)$$

からを推定することが可能である. A 節に示すように, (3) 式からパラメータに対する漸化式を得ることができる.

$$a_{mi}^{(\nu+1)} = \frac{1}{T} \sum_{t=1}^T \frac{a_{mi}^{(\nu)} F_{mi}^{(\nu)}(z_m(t))}{G_m^{(\nu)}(z_m(t))} \quad (i = 1, \dots, K_m), \quad (4)$$

$$r_{mi}^{(\nu+1)} = \frac{1}{M} \frac{\sum_{t=1}^T z_m(t) \frac{F_{mi}^{(\nu)}(z_m(t))}{G_m^{(\nu)}(z_m(t))}}{\sum_{t=1}^T \frac{F_{mi}^{(\nu)}(z_m(t))}{G_m^{(\nu)}(z_m(t))}} \quad (i = 1, \dots, K_m). \quad (5)$$

ここで,

$$F_{mi}^{(\nu)}(x) = \frac{(Mr_{mi}^{(\nu)})^x e^{-Mr_{mi}^{(\nu)}}}{x!}, \quad (6)$$

$$G_m^{(\nu)}(x) = \sum_{i=1}^{K_m} a_{mi}^{(\nu)} F_{mi}^{(\nu)}(x), \quad (7)$$

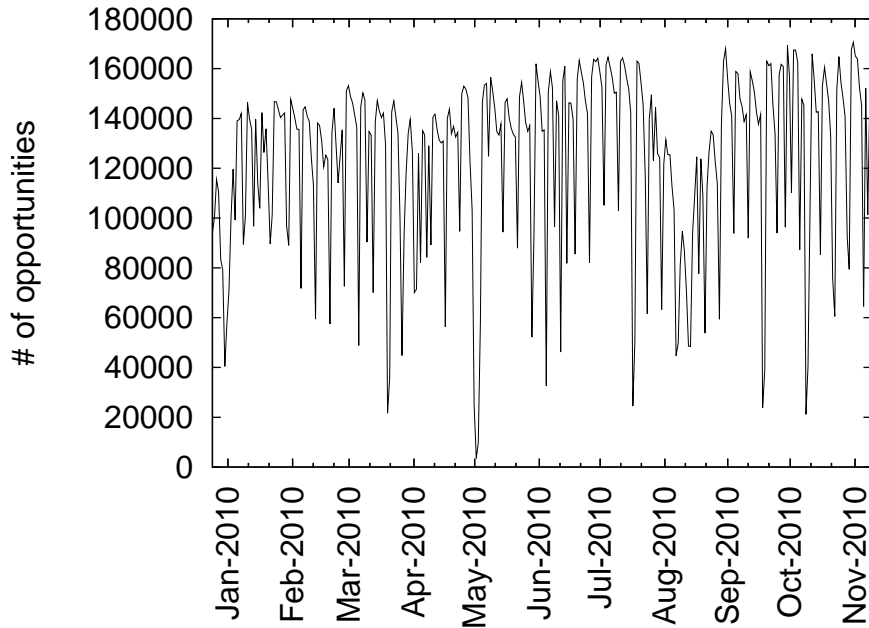


図 2: 2009 年 12 月 24 日から 2010 年 11 月 4 日までの日本国内における日次総プラン数の推移.

と置く. これらの式は有限混合ポワソン分布に対する EM アルゴリズムとして知られている [9, 10].

### 3.3 Finch らのアルゴリズム

EM アルゴリズムは局所最適性しか満足しないため, そのまま EM アルゴリズムを計算しても, 有限混合ポワソン分布の対数尤度関数の多峰性により計算を始めるパラメータ初期値依存して異なる収束値が得られる. これを回避するためには, パラメータ初期値の選択が重要である. ここでは, 確率的に初期値を選択し, 繰り返し計算したパラメータ推定値のなかで尤度が最大となるパラメータを推定値として与える方法を採用する. さらに, パラメータ初期値の選択は Finch らのアルゴリズムを用いる [11].

Finch らのアルゴリズムのアイデアは以下のとおりである.  $a_{mi}^{(0)}$  が  $\sum_{i=1}^{K_m} a_{mi}^{(0)} = 1$  満足するように与えられていると仮定する. そして, この既知の  $a_{mi}^{(0)}$  を用いて観測量の順位  $s$  番目の順位統計量  $z_m(t_s)$  ( $s = 1, \dots, T$ ) を  $K_m$  個に  $a_{mi}^{(0)}$  の割合で分割する. そして,  $i$  番目の区画に分類される観測量の列から  $r_{im}^{(0)}$  を時系列平均値  $\mu_{mi}$  を求め, この値を用いて  $r_{im}^{(0)} = \mu_{mi}/M$  としてパラメータ初期値とする.

モンテカルロステップにおいて,  $a_{m1}^{(0)}, \dots, a_{mK_m}^{(0)}$  は区間  $[0, 1]$  を満足する一様乱数から抽出した独立な乱数列を総和が 1 となるように規格化して与える. そして,  $r_{m1}^{(0)}, \dots, r_{mK_m}^{(0)}$  を Finch らのアルゴリズムで設定する. これらのパラメータ値を用いて対数尤度を (18) 式から計算する. この値が有限であった場合には, これをパラメータ初期値として EM アルゴリズムによってパラメータの計算を対数尤度の値が変化しなくなるまで続ける.

このとき計算される対数尤度の値がこれまでモンテカルロステップにおいて計算されたパラメータ値の対数尤度より大きい場合はこのパラメータ値をパラメータ推定値の候補とする. この操作を

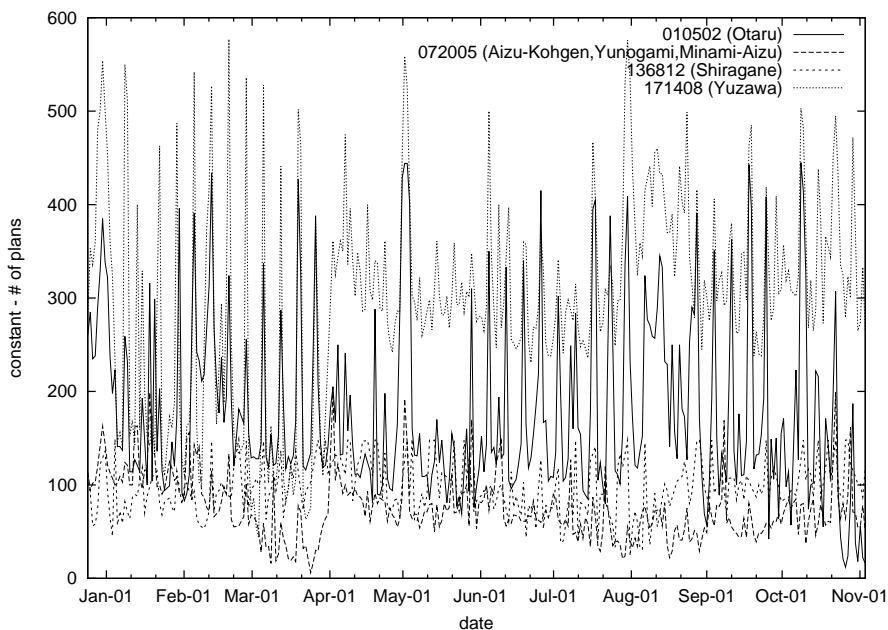


図 3: 010502 小樽 (北海道), 072005 会津高原, 湯野上, 南会津 (福島県), 136812 白金 (東京都), 171408 湯沢 (新潟県) における需要代理変数の変化.

パラメータ推定値の候補が更新されなくまるまで繰り返すことにより, 適切なパラメータ推定値を得る. このアルゴリズムをまとめると以下ようになる.

- (0)  $maxobj = 0$  かつ  $counter = 0$  と設定する.
- (1)  $[0, 1]$  の一様乱数  $b'_{mi}$  を用いて,  $a'_{mi}$  を  $a'_{mi} = b'_{mi} / \sqrt{\sum_{i=1}^{K_m} b'_{mi}}$  として与える.
- (2)  $r_{mi}$  を Finch らのアルゴリズムによって計算する. もし  $counter > MAXCOUNT$  ならばステップ (6) へ移動する.
- (3) もし対数尤度  $L_m(a'_{m1}, \dots, a'_{mK_m}, r'_{m1}, \dots, r'_{mK_m})$  が  $maxobj$  より大きいならば,  $maxobj$  をその値に更新し,  $r_{mi} := r'_{mi}$  として  $a_{mi} := a'_{mi}$  ステップ (4) に移動する. (4). もしそうでないならばステップ (1) に移動する.
- (4) パラメータ初期値  $(a_{m1}, \dots, a_{mK_m}, r_{m1}, \dots, r_{mK_m})$  から, (4) 式と (5) 式を対数尤度の変化が閾値以下になるまで繰り返し計算する.
- (5) もし, 対数尤度関数の最大値が  $maxobj$  以上ならば, そのときの収束値を推定値の候補として記録する.
- (6)  $counter = counter + 1$  とする. もし  $counter < MAXCOUNT$  ならば, ステップ (1) へ移動. そうでなければ, ステップ (7) へ移動する.
- (7)  $maxobj$  とパラメータ推定値候補を表示してプログラムを終了する.

### 3.4 カテゴリ数と観測量が属するカテゴリの決定方法

最適なカテゴリ数は赤池情報量基準 (AIC)

$$AIC(K_m) = 4K_m - 2\hat{L}_m, \quad (8)$$

が最小となる  $K_m$  として与える.

$$\hat{K}_m = \arg \min_{K_m} AIC(K_m). \quad (9)$$

ここで,

$$\hat{L}_m = \sum_{s=1}^T \log \left( \sum_{i=1}^{K_m} \hat{a}_{mi} \frac{(M\hat{r}_{mi})^{z_m(s)}}{z_m(s)!} e^{-M\hat{r}_{mi}} \right). \quad (10)$$

は  $2K_m$  パラメータに関する対数尤度の最大値である.

観測量  $z_m(s)$  がどのカテゴリに属するべきかは, 推定されたパラメータをもつ個別のポワソン分布の対数尤度

$$R_{mi}(z) = z \log M + \log \hat{r}_{mi} - M\hat{r}_{mi} \log(z!). \quad (11)$$

を基準として, これを最大化するカテゴリとして決定する.

$$\hat{i}_s = \arg \max_i R_{mi}(z_m(s)). \quad (12)$$

## 4 数値シミュレーション

実証分析を行う前に人工的に複数の強度の異なるポワソン分布から抽出した混合時系列に対して提案するパラメータ推定方法が有効であることを確認する.  $z(s)$  ( $s = 1, \dots, T$ ) を以下で生成される有限混合ポワソン分布からの抽出であるとする.

$$\begin{cases} r(t) = r_i \text{ w.p. } a_i \\ z(t) \sim \Pr(l = Z(t)|r(t)) = \frac{(Mr(t))^l}{l!} e^{-Mr(t)} \end{cases} \quad (13)$$

ここで,  $K$  はカテゴリ数,  $a_i$  は  $i$  番目のカテゴリを表わすポワソン分布が選択される確率である ( $i = 1, \dots, K; \sum_{i=1}^K a_i = 1$ ).

$K = 12$ ,  $M = 100,000,000$  とし, 表. 2(左) に示すパラメータを用いて  $T = 200$  の時系列を生成した. 次にパラメータに関する事前知識を用いず時系列のみからパラメータの推定をおこなう. 図. 4(左) に示すように  $K$  に関する AIC の値は  $\hat{K} = 12$  において最小値をとっている. 推定されたパラメータによる有限混合ポワソン分布が時系列をよく表わすことを確認するために Kolmogorov-Smirnov 検定をおこなった. 図. 4(右) は KS 統計量の値を示している.  $\hat{K} = 12$  における KS 統計量は  $0.327 (< 1.36)$  と計算されるので, 5% の有意水準において同じ分布からの抽出であることは棄却されない. 表. 2(右) はこのときのパラメータ推定値を示している. さらに, 各観測量がどのカテゴリに属するポワソン分布からの抽出であったかを推定した. 図. 5 に示すように, 推定誤差は  $8.0 \times 10^{-6}$  以下であり, 相対誤差  $|\hat{r}_i(t) - r_i(t)|/r_i(t)$  は 0.3 % 以下となった.

表 2: 人工的な時系列を生成するために用いた有限混合ポワソン分布のパラメータ値 (左). カテゴリー数  $K = 12$  とした. EM アルゴリズムを用いて時系列から計算した有限混合ポワソン分布のパラメータ推定値 (右). AIC を最小とするカテゴリー数は  $\hat{K} = 12$  であり, このとき  $AIC = 3803.20$  を得た.

$r_1$	0.000025	$a_1$	0.109726	$r_1$	0.000024	$a_1$	0.09000
$r_2$	0.000223	$a_2$	0.070612	$r_2$	0.000222	$a_2$	0.07000
$r_3$	0.000280	$a_3$	0.073355	$r_3$	0.000280	$a_3$	0.05500
$r_4$	0.000479	$a_4$	0.077612	$r_4$	0.000479	$a_4$	0.06500
$r_5$	0.000613	$a_5$	0.094848	$r_5$	0.000613	$a_5$	0.08000
$r_6$	0.000652	$a_6$	0.073841	$r_6$	0.000651	$a_6$	0.09500
$r_7$	0.001219	$a_7$	0.090867	$r_7$	0.001218	$a_7$	0.10417
$r_8$	0.001233	$a_8$	0.062191	$r_8$	0.001232	$a_8$	0.08082
$r_9$	0.001295	$a_9$	0.077662	$r_9$	0.001294	$a_9$	0.08500
$r_{10}$	0.001341	$a_{10}$	0.102573	$r_{10}$	0.001341	$a_{10}$	0.10500
$r_{11}$	0.001412	$a_{11}$	0.085892	$r_{11}$	0.001412	$a_{11}$	0.08000
$r_{12}$	0.001570	$a_{12}$	0.080821	$r_{12}$	0.001568	$a_{12}$	0.09000

## 5 実証分析

ここでは, 実際のデータを用い提案手法によりパラメータの推定をおこなう.  $M = 1,000,000,000$  とし需要に対し, 4つの地域におけるパラメータの推定をおこなった. 図. 3で示した4つの地域を分析対象とした. AIC が最小となる値から, 最適なカテゴリー数は表. 3示すように決定された. 更に, 各地域における AIC と KS 統計量の示す. この表から, 最小となる AIC を与えるカテゴリー数において KS 統計量はどれも 5% 有意水準で同一の分布であることは棄却されないことがわかる.

地域  $m$  での宿泊日  $t$  における利用可能なプランの価格  $p_{mk}$  ( $k = 1, \dots, n_m(t)$ ) から平均価格を

$$P_m(t) = \frac{1}{n_m(t)} \sum_{k=1}^{n_m(t)} p_{mk}(t) \quad (14)$$

によって計算する. この平均価格と宿泊日  $t$  における予約確率  $r_m(t)$  との関係は宿泊日  $t$  の需要の

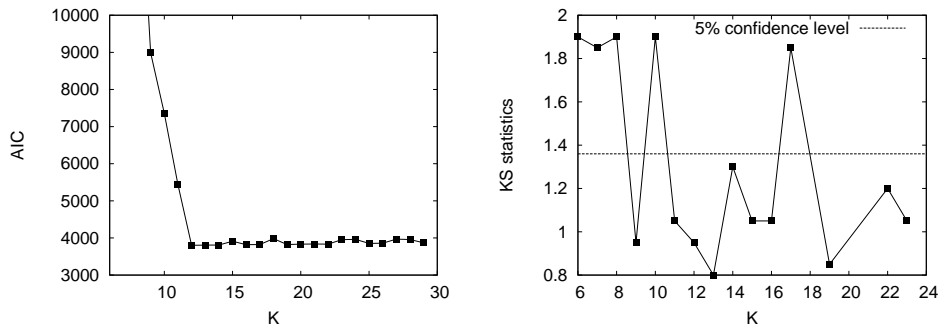


図 4: 人工的に生成したデータの  $K$  に対する AIC の変化 (左). 最小の AIC の値は  $K = 12$  において 3803.20 を得た.  $K$  に対する KS 統計量の変化 (右). 破線は 5% 信頼値.



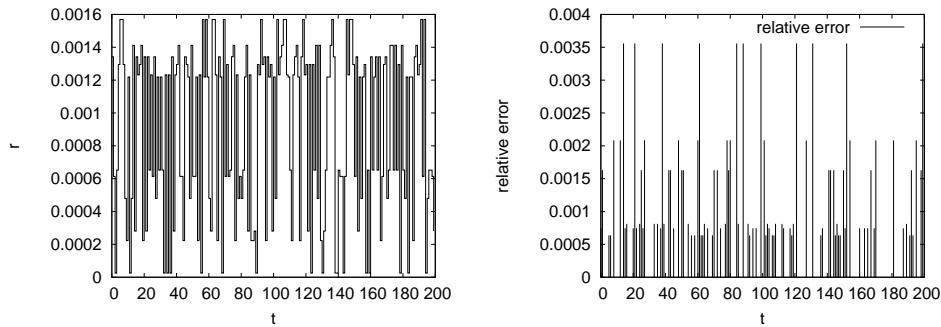


図 5: 各観測量に対する推定された  $r$  の値 (左). 各観測値に対する相対誤差 (右).

表 3: 4 つの地域における AIC, 最大対数尤度 ( $ll$ ),  $p$  値, および KS 統計量

地域名	$K$	AIC	$ll$	$p$ -value	KS value
小樽 (北海道)	12	3558.31	1756.15	0.807	0.532
会津高原, 湯野上, 南会津 (福島県)	10	3009.10	1485.55	0.187	1.088
白金 (東京都)	5	2572.33	1277.17	0.107	1.245
湯沢 (新潟県)	11	3695.25	1826.62	0.465	0.850

様子を特徴を付ける量となりえる. 図. 6 に示す. この図から需要が小さい場合, 各地域における平均価格のずれは小さいが需要が大きい場合には, 各地域における平均価格のずれが大きくなることが確認された. これは需要が大きい場合, 消費者が嗜好するプラン価格帯に大きな差異が存在していることを意味する.

## 6 まとめ

日本国内のホテル予約サイトから収集したデータを用い, 2009 年 12 月 25 日から 2010 年 11 月 4 日までの各宿泊日次に対する需要を分類する方法を議論した. 提案手法では EM アルゴリズムに Finch らによるパラメータ初期値選択方法を用い, ランダムに選択したパラメータ初期値から尤度関数を最大化する方法を用いた. 複数のポワソン分布分析から抽出した乱数のある混合比で混ぜた時系列に対してカテゴリ分類をおこなった. その結果, 時系列データのみからパラメータを推定しカテゴリ数を決定できることがわかった.

更に, 提案手法を用いることにより, 4 つの地域における宿泊需要の代理変数のパラメータを推定し, 宿泊日を需要の違いによって分類した. 推定パラメータを用いた Kolmogorov-Smirnov 検定により 5% の有意水準で時系列は推定分布からの抽出であることを棄却されないことを確認した. 利用可能なプランの平均価格と需要確率との関係を確認し, 需要の大きなカテゴリに分類される宿泊日には利用可能な宿泊プラン平均価格に大きな差異が存在することを確認した. これは, 宿泊日の需要が大きい場合, 宿泊プランの価格帯に対する顧客の価格選好が大きく異なることを意味する. すなわち, 需要の大きな宿泊日には, 宿泊者の客層が大きく異なっていると考えられる. このことから, 宿泊施設は単純に需要の大小によってプラン価格帯を考慮するのではなく, 需要を生み出す客層を考慮してプラン価格帯を決定するのがのぞましいことが示唆される.

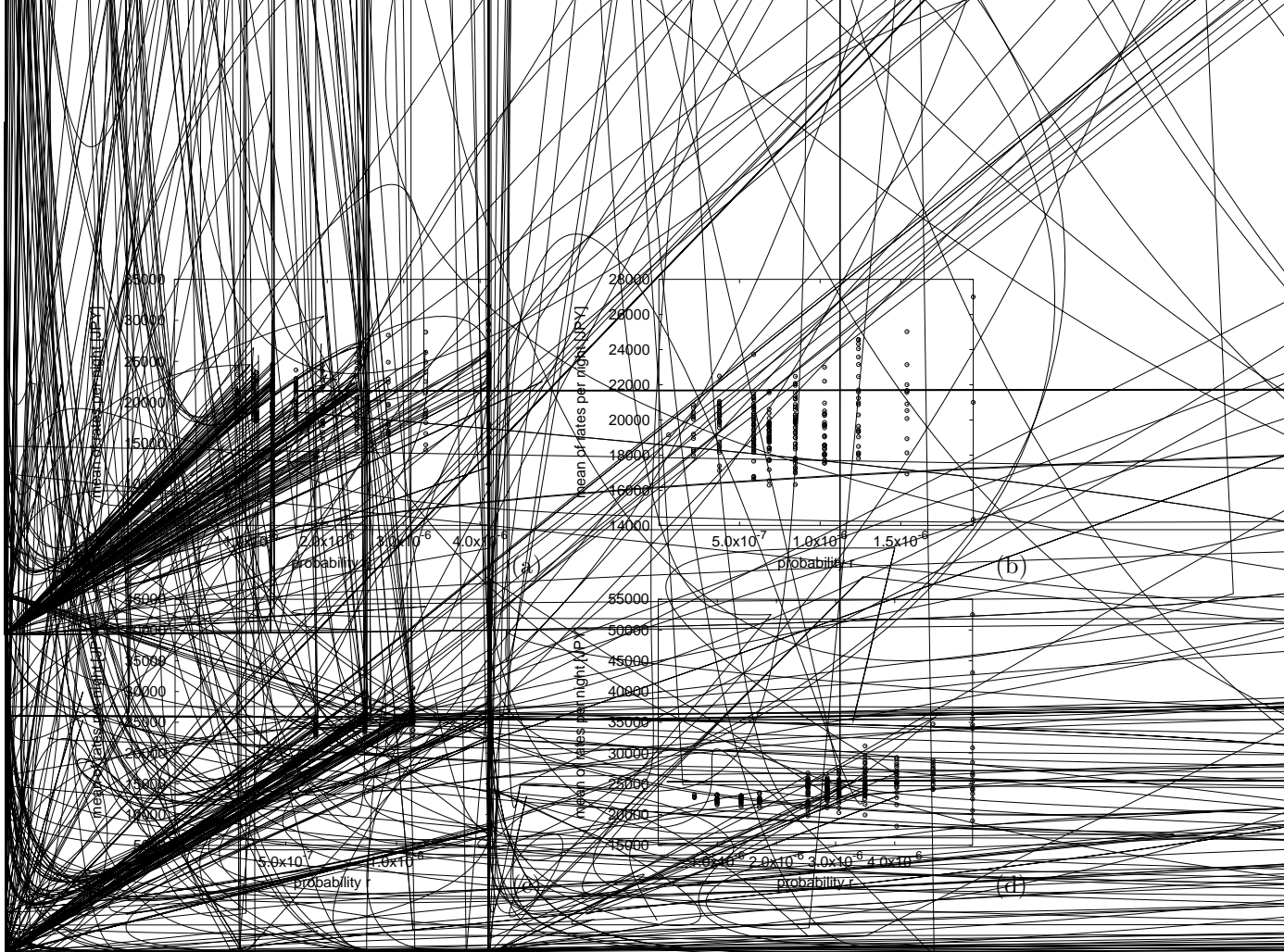


図 6: 各カテゴリーにおける平均価格と予約確率  $r_{mi}$  との関係. (a)010502 小樽 (北海道), (b)072005 会津高原, 湯野上, 南会津 (福島県), (c)136812 白金 (東京都), (d)171408 湯沢 (新潟県).

今後の課題として, 地域の需要特性の網羅的な特徴付けおよび各地域において利用可能な部屋数の推定があげられる. 各地域における需要と供給の特徴をイベントに応じて分析することにより, 各地域で起った出来事や気象条件を理解することが可能になると推察される.

## 謝辞

本研究は, 日本学術振興会優秀若手研究者海外派遣 (#21-5341) の助成を受けて行われた. ここに深く感謝の意を表す. 本研究をおこなうに当たり, キール大学 Thomas Lux 教授に感謝の意を表す. また, スイス連邦工科大学 Dirk Helbing 教授との議論が有益であった.

## A EM アルゴリズム

有限混合 Poisson 過程に対する EM アルゴリズムを最尤法から導出する方法を示す.  $G_m(z)$  を  $K_m$  個のポワソン分布  $F_{mi}(z)$  ( $m = 1, \dots, K_m$ ) の混合分布であるとする.

$$F_{mi}(z) = \frac{(Mr_{mi})^z}{z!} e^{-Mr_{mi}}, (i = 1, \dots, K_m) \quad (15)$$

$$G_m(z) = \sum_{i=1}^{K_m} a_{mi} F_{mi}(z). \quad (16)$$

ここで  $a_{mi}$  は混合比率を意味し, 規格化されているとする.

$$\sum_{i=1}^{K_m} a_{mi} = 1. \quad (17)$$

$T$  個の観測量  $\{z_m(t)\}$  からパラメータ  $a_{mi}, r_{mi}$  ( $i = 1, \dots, K_m$ ) を最尤法で求めることを考える. 有限混合ポワソン分布に対する対数尤度は

$$L_m(a_{m1}, \dots, a_{mK_m}, r_{m1}, \dots, r_{mK_m}) = \sum_{t=1}^T \log G_m(z_m(t)). \quad (18)$$

と書ける. (15) 式と (17) 式を (18) 式に代入すると, 以下を得る.

$$L_m(a_{m1}, \dots, a_{mK_m}, r_{m1}, \dots, r_{mK_m}) = \sum_{t=1}^T \log \left( \sum_{i=1}^{K_m-1} a_{mi} \frac{(Mr_{mi})^{z_m(t)}}{z_m(t)!} e^{-Mr_{mi}} + (1 - \sum_{i=1}^{K_m-1} a_{mi}) \frac{(Mr_{mK_m})^{z_m(t)}}{z_m(t)!} e^{-Mr_{mK_m}} \right). \quad (19)$$

(19) 式を  $a_{mi}$  と  $r_{mi}$  に対してそれぞれ偏微分することで

$$\frac{\partial L_m}{\partial a_{mi}} = \sum_{t=1}^T \frac{F_{mi}(z_m(t)) - F_{mK}(z_m(t))}{G_m(z_m(t))} \quad (i = 1, \dots, K_m - 1), \quad (20)$$

$$\frac{\partial L_m}{\partial r_{mi}} = \sum_{t=1}^T \frac{a_{mi} F_{mi}(z_m(t))}{G_m(z_m(t))} \left( \frac{z_m(t)}{r_{mi}} - M \right) \quad (i = 1, \dots, K_m - 1), \quad (21)$$

$$\frac{\partial L_m}{\partial r_{mK_m}} = \sum_{t=1}^T \frac{(1 - \sum_{i=1}^{K_m-1} a_{mi}) F_{mK}(z_m(t))}{G_m(z_m(t))} \left( \frac{z_m(t)}{r_{mK_m}} - M \right). \quad (22)$$

を得る.  $a_{mi}$  を (20) 式に代入し  $i$  に対して和を取ると

$$\sum_{t=1}^T \frac{F_{mi}(z_m(t))}{G_m(z_m(t))} = T \quad (i = 1, \dots, K_m), \quad (23)$$

となる. そして,  $a_{mi}/T$  を (23) 式に掛けると

$$a_{mi} = \frac{1}{T} \sum_{t=1}^T \frac{a_{mi} F_{mi}(z_m(t))}{G_m(z_m(t))} \quad (i = 1, \dots, K_m), \quad (24)$$

を得る. (21) 式と (22) 式より,

$$r_{mi} = \frac{1}{M} \frac{\sum_{t=1}^T z_m(t) \frac{F_{mi}(z_m(t))}{G_m(z_m(t))}}{\sum_{t=1}^T \frac{F_{mi}(z_m(t))}{G_m(z_m(t))}} \quad (i = 1, \dots, K_m). \quad (25)$$

を得る. この式は, パラメータに対する漸化式となっており, もし適切なパラメータの初期値  $\{a_{mi}^{(0)}\}$ ,  $\{r_{mi}^{(0)}\}$  を設定できるとするとパラメータ  $\{a_{mi}^{(\nu)}\}$  and  $\{r_{mi}^{(\nu)}\}$  は漸化式の更新

$$a_{mi}^{(\nu+1)} = \frac{1}{T} \sum_{t=1}^T \frac{a_{mi}^{(\nu)} F_{mi}^{(\nu)}(z_m(t))}{G_m^{(\nu)}(z_m(t))} \quad (i = 1, \dots, K_m), \quad (26)$$

$$r_{mi}^{(\nu+1)} = \frac{1}{M} \frac{\sum_{t=1}^T z_m(t) \frac{F_{mi}^{(\nu)}(z_m(t))}{G_m^{(\nu)}(z_m(t))}}{\sum_{t=1}^T \frac{F_{mi}^{(\nu)}(z_m(t))}{G_m^{(\nu)}(z_m(t))}} \quad (i = 1, \dots, K_m), \quad (27)$$

によって得ることができる。ここで,

$$F_{mi}^{(\nu)}(z) = \frac{(Mr_{mi}^{(\nu)})^z}{z!} e^{-Mr_{mi}^{(\nu)}}, \quad (28)$$

$$G_m^{(\nu)}(z) = \sum_{i=1}^{K_m} a_{mi}^{(\nu)} F_{mi}^{(\nu)}(z), \quad (29)$$

と置いた。

## 参考文献

- [1] R. Law, “Disintermediation of reservations”, *International Journal of Contemporary Hospitality Management*, **21** (2009) 766-772.
- [2] 国土交通省観光庁統計資料: <http://www.mlit.go.jp/kankocho/siryou/toukei/index.html>.
- [3] W. Weidlich, *Sociodynamics: A Systematic Approach to Mathematical Modelling in the Social Sciences*, Taylor and Francis, London (2002).
- [4] S. Alfrano and T. Lux, “A noise trader model as a generator of apparent financial power laws and long memory”, *Macroeconomic Dynamics*, **11** (2007) 80–101.
- [5] G. Haag, and W. Weidlich, “A stochastic theory of interregional migration”, *Geographical Analysis*, **16** (1984) 331-357.
- [6] Sukbin Cha, Ken W. McCleary, and Muzaffer Uysal, “Travel Motivations of Japanese Overseas Travelers: A Factor-Cluster Segmentation Approach”, *Journal of Travel Research*, **34**, (1995) 33–39.
- [7] A.-H. Sato, “Patterns of Regional Travel Behavior: An Analysis of Japanese Hotel Reservation Data”, *International Review of Financial Analysis*, in press.
- [8] V. Hasselblad, “Estimation of Finite Mixtures of Distributions from the Exponential Family”, *Journal of the American Statistical Association*, **64** (1969) 1459–1471.
- [9] A.P. Dempster, N.M. Laird, and D.B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm”, *Journal of the Royal Statistical Society, Series B*, **39** (1977) 1–38.
- [10] Z. Liu, J. Almhana, V. Choulakian, R. McGorman, “Online EM algorithm for mixture with application to internet traffic modeling”, *Computational Statistics and Data Analysis*, **50** (2006) 1052–1071.
- [11] S.J. Finch, N.R. Mendell, H.C. Thode, “Probabilistic measures of adequacy of a numerical search for a global maximum”, *Journal of the American Statistical Association*, **84** (1989) 1020–1023.