Patterns of Regional Travel Behavior: An Analysis of Japanese Hotel Reservation Data

Aki-Hiro SATO[1,2] [1], Thomas LUX[2]
[1] Department of Applied Mathematics and Physics, Kyoto University, Japan
[2] Department of Economics, University of Kiel, Germany

Abstract

This study considers availability of room opportunities collected from a Japanese hotel booking site. The status of room opportunities is empirically analyzed from a comprehensive point of view. To determine migration tendency of travel and tourism a mixture of Poisson processes is discussed. In order to estimate time-dependent room existence ratio we employ the EM-algorithm for the mixture of Poisson distributions. We characterize states of travel and tourism by means of relationship between mean of room prices and room existence ratio.

Keywords: Japanese Hotel Reservation, Mixture of Poisson distributions, EM-algorithm, Parameter Estimation
Classification codes: C80, L81, L83

1. Introduction

Recent technological development enables us to purchase various kinds of items and services via E-commerce systems. The emergence of Internet applications has had an unprecedented impact on our life style to purchase goods and services. From availability of items and services at such E-commerce platforms utilities of agents in socio-economic systems are expected to be estimated.

Such an impact on travel and tourism, specifically, on hotel room reservations, is significantly considered [1]. According to Pilia [2], 40 per cent of hotel reservations will be made via Internet in 2008, up from 33 per cent in 2007 and 29 per cent in 2006. Therefore the coverage of Internet room opportunities may be sufficient to provide us with statistically significant results. Furthermore it is possible for us to conduct comprehensive investigation based on hotel booking data collected from Internet site.

From our personal experience, it is found that it is becoming more popular to make reservations of hotels via Internet. As we have some experiences at the

[1] Contact information for the corresponding author: Department of Applied Mathematics and Physics, Graduate School of Informatics, Yoshida-Homcho, Sakyo-ku, 606-8501 Kyoto, JAPAN, E-mail: aki@i.kyoto-u.ac.jp

hotel reservation site, we sometimes can make a reservation of several hotels or not. Namely the hotel accessibility seems to be random. Moreover availability of the hotels depends on calendars (weekdays, weekends, and holidays) and the hotel. It is found that the number of the booked rooms is eventually increasing as the date of stay reaches. In our experience the stay of date is an important factor to determine the availability of room opportunities.

This availability of the hotel bookings may indicate the future migration by travel and tourism. Therefore it will be able possible for us to detect migration in countries from comprehensive data if we have data on availability of all the rooms.

Migration processes have been intensively studied in the context of socioeconomic dynamics with particular interest for quantitative research. Weidlich and Haag proposed the Master equation with transition probabilities depending on regional-dependent and time-dependent utility and mobility in order to describe collective tendency of agent decision in migration chance [3]. Since the motivation of migration seems to come from both psychological and physical factors, the understanding of the dynamics of the migration is expected to lead to inner states of agents and insights on collective behavior of agents.

In the present article we will consider available opportunities of hotels and investigate statistics of the number of such available opportunities from several perspectives.

2. Data description

In this section we will give a brief explanation of a method to collect data on hotel availability. In this study we used a Web API (Application Programing Interface) in order to collect the data. The API is an interface code set which is designed for a purpose to simplify development of application programs.

The Jalan Web API [2] is provided by the Jalan hotel reservation site which is one of the most popular hotel reservation sites in Japan. The Jalan web site serves interfaces for both hotel managers and customers (See Fig. 1). On the one hand the hotel managers can input information on rooms served by the hotel via Internet. On the other hand consumers can book available rooms via the Jalan Web page. The third parties can even built their web services with the Jalan data by using the Web API.

We are collecting all plans in the Jalan web page under the condition that two adults will be able to stay at the room per night and is accumulating data as csv files. In the dataset there exist over 100,000 room opportunities from over 14,000 hotels. Tab. 1 represents contents included in the data set. Since the data contain regional information, it is possible for us to analyze regional

[2] The data is provided by the Jalan Web service.

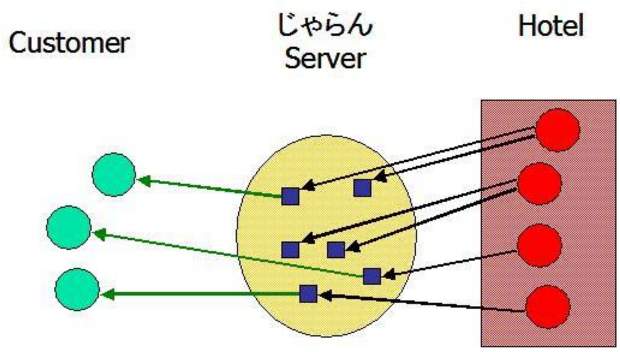


Figure 1: A conceptual illustration of the Jalan web service. The hotel managers enter information on rooms (plans) which will be served at the hotels. Customers can search and book rooms from all the available rooms (plans) via the Jalan web page.

dependence of hotel rates. Throughout the investigation we regard the number of available root opportunities as a proxy variable of the rest stocks of rooms.

Table 1: An example of opportunities

Sampledate	Date of collection	from 2009-12-24
Targetdate	Date of Stay	Up to 6 days before
Hotelid	Hotel identification number	300000
Hotelname	Hotel name	Hotel ABC
Hotelnamekana	Hotel name (Kana)	ほてるえーびーしー
Postcode	Postal code	066-8520
Hoteladdress	Address	千歳市本町
Hoteldetailurl	URL	http://www.jalan...
X	Latitude	509943536
Y	Longtitude	154132695
Planname	Opportunity name	旅の達人！
Meal	Meal availability	no meal
Sample rate	the latest best rate	3500 (JPY)
Rate	Rate per night	7000 (JPY)

Throughout the investigation we regard the number of available hotels (plans) as a proxy variable of the rest stocks of rooms.

The data are sampled from the Jalan net web site (*http://www.jalan.net*) every day. This data set contains plans that two adults will be able to stay at the hotel for a night. Each plan also contains sampled date when we sampled the data, target date when we will stay at the hotel, regional sequential number, hotel identification number, hotel name, postal address, URL of the hotel web page, geographical position, plan name, rate. The data period is from 25th Dec 2009 to 15th July 2010.

Fig. 2 shows an example of hotel distribution and representative rates. Yellow (black) filled squares represent hotel plans cost 50,000 JPY (1,000) JPY per night. Red filled squares represent hotel plans cost over 50,000 JPY per night. As you can see it is found there is strong regional dependence on concentration. Specifically we find that many hotels are located around several centralized cities such as Tokyo, Osaka, Nagoya, Fukuoka and so on. Fig. 2 (bottom) shows a probability density distribution on 15th April 2010 all over the Japan. It is found that there are two maximal values around 10,000 JPY and 20,000 JPY.

3. Empirical Analysis

The total number of room opportunities at which two adults have possibility to stay was counted up from the data throughout the whole sampled period. Fig. 3 shows the total number of room opportunities per day.

From this graph we found three points:

(1)There exists weekly seasonality for the total number of hotels.

(2)There is a strong dependence of the rest number of hotels on Japanese calendar. Namely holidays influence reservation activities of consumers. For example during new year days (around 12/30-1/1) and a holidays in spring seasons (around 3/20) the total number of the rest hotels show steep decreases.

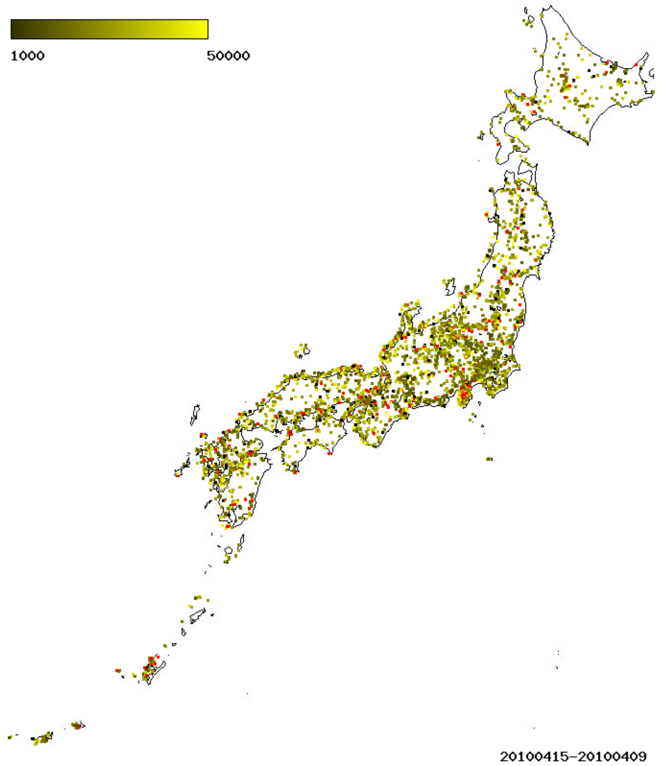
(3)As D decreases the rest number of hotels decreases. Specifically two days before drastic decreases of the number of hotels are observed.

Fig. 4 shows the number of available room opportunities for several regions for a period from 25th Dec 2009 to 15th July 2010. It is found that there are regional dependences of their temporal development.

Furthermore we show that dependence of average rates all over the Japan on calendar dates in Fig. 5. On the new year holidays in 2010 it is confirmed that the mean rates rapidly decrease, meanwhile, on the spring holidays in 2010 the mean rates rapidly increase. This difference seems to come from the difference of consumers motivation structure and preference on price levels between these holiday seasons.

4. Model

Let M and N be the number of consumers and one of opportunities, respectively. The number of opportunities N may be assumed to be constant since the Internet booking style has been sufficiently speeding, and almost hotels offer their rooms via Internet.



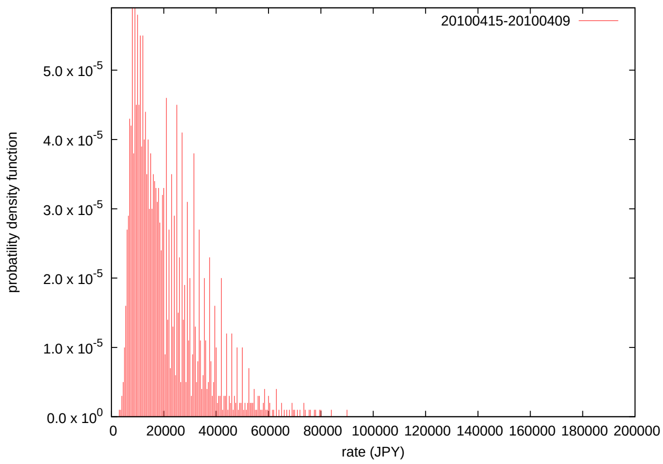


Figure 2: An example of rates distributions of which two adults will be able to stay at the hotel for a night at 15th April 2010 (Top). A probability density distribution of rates at 15th April 2010 (Bottom). This data have been sampled on 9th April 2010. Yellow (black) filled squares represent hotel plans cost 50,000 JPY (1,000) JPY per night. Red filled squares represent hotel plans cost over 50,000 JPY per night.

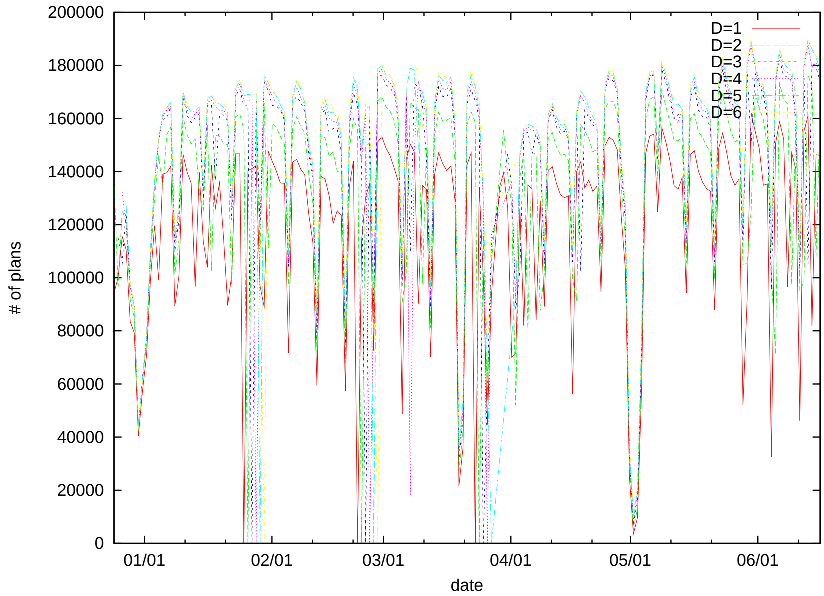


Figure 3: The temporal development for the number of hotels where two adults would be able to stay per night for a period from 25th Dec 2009 to 15th July 2010. D represents a daily duration between target date and sample date.

We assume that a Bernoulli random variable represents booking decision of a consumer from N kinds of room opportunities. We consider M Bernoulli random variables with success probability p,

$$y_i = \begin{cases} 1 & w.p. \quad p \\ 0 & w.p. \quad 1-p \end{cases} . \tag{1}$$

If $y_1, \ldots, y_M$ are independently, identically, distributed then

$$Y = \sum_{i=1}^{M} y_i, \tag{2}$$

$$\tag{3}$$

follows a binomial distribution $\mathrm{B}(M, p)$. Furthermore assuming $p \ll 1$, $M \gg 1$, $Mp \gg 1$, we can approximate Y as a Poissonian random variable, which follows

$$P_Y(m|p) = \frac{(Mp)^m}{m!} e^{-Mp}. \tag{4}$$

Therefore $Z = N - Y$ can also approximate a Poissonian random variable with intensity $N - Mp$,

$$P_Z(l|r) \quad = \quad \frac{N(1-M/Np)^l}{l!} e^{-N(1-M/Np)} \tag{5}$$

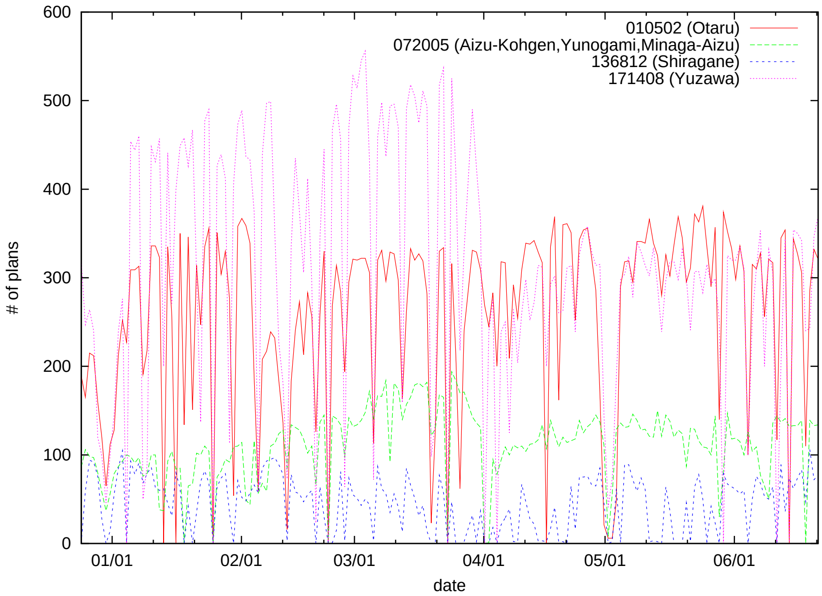


Figure 4: The number of available room opportunities for each region per day. It is found that there exists regional dependence of their fluctuations.

$$= \frac{Nr^l}{l!} e^{-Nr}, \tag{6}$$

$$\tag{7}$$

where we assume $r = 1 - M/Np$.

Since the agents have some interactions with one another, the psychological atmosphere (mood), which is collectively created by agents, influences their decision. Such a psychological effect may be expressed as probability fluctuations for success probability p in the Bernoulli random variable.

Let us introduce the probability fluctuations $r \quad (0 \leq r \leq 1)$ which are sampled from probability density $F(r)$. Then the marginal distribution for the Poisson distribution conditional on r with probability fluctuation $F(r)$ gives the distribution of Z,

$$P_Z(l) = \int_0^1 F(r) \frac{(Nr)^l}{l!} e^{-Nr} dr. \tag{8}$$

Since we can observe the number of available opportunities $Z = N - Y$, we may estimate parameters of the distribution $F(r)$ from the successive observations. For the sake of simplicity we assume that r can take discrete states such as r_i with probability a_i ($0 \leq r_i \leq 1$, $i = 1, \ldots, K$ and $\sum_{i=1}^{K} a_i = 1$). These parameters are expected to describe motivation structure of consumers on calendar days (weekdays/weekends and special holidays, business purpose/recreation and

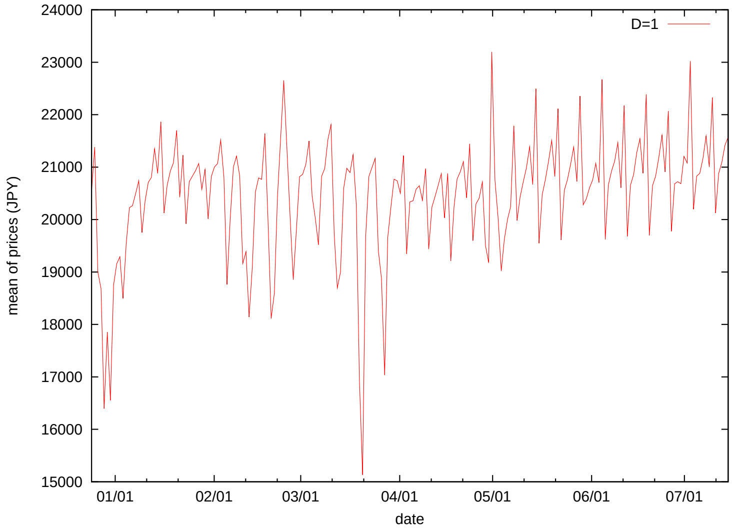


Figure 5: Time series of average rates of room opportunities on stay dates. The mean value of rates is calculated from all the available room opportunities which are observed on each stay date.

so forth). Then since $F(r)$ is given by

$$F(r) = \sum_{i=1}^{K} a_i \delta(r - r_i), \tag{9}$$

$P_Z(Z)$ is calculated as

$$\begin{aligned} P_Z(l) &= \int_0^1 F(r) \frac{(Nr)^l}{l!} e^{-Nr}, \\ &= \sum_{i=1}^{K} a_i \frac{(Nr_i)^l}{l!} e^{-Nr_i}. \end{aligned} \tag{10}$$

Therefore Eq. (10) is concerned with a finite mixture of Poisson distributions.

5. Estimation procedure by means of the EM algorithm

The construction of estimators for finite mixtures of distributions has been considered in the literature of estimation. Estimation procedures for Poissonian mixture model have been successively studied by several researchers. Specifically moment estimators and maximum likelihood estimators are intensively studied.

Moment estimators were tried on a mixture of two normal distributions by Karl Peason as early as 1894. Graphical solutions have been given by Cassie [6], Harding [7] and Bhattacharya [8]. Rider discusses mixtures of binomial and mixtures of Poisson distributions in the case of two distributions [9].

Hasselblad proposed the maximum likelihood estimator and derived recursive equations for parameters [10]. The effectiveness of the maximum likelihood estimator for the mixtures of Poissonian distributions is widely recognized. Dempster discusses the EM algorithm for mixtures of distributions in the several cases [11]. By using the EM algorithm we can estimate parameters from mixing data.

Let $z(1), \ldots, z(T)$ be the number of available room opportunity observed at each observation date. Let us consider that the way to estimate parameters given in Eq. (8) based on a maximum likelihood estimator.

In this case the log likelihood function can be described as

$$L(a_1, \ldots, a_K, r_1, \ldots, r_K) = \sum_{t=1}^{T} \log\Big(\sum_{i=1}^{K} a_i \frac{(Nr_i)^l}{l!} e^{-Nr_i}\Big), \tag{11}$$

and an adequate set of parameters is estimated by maximization of the log likelihood function $L(a_1, \ldots, a_K, r_1 \ldots, r_K)$,

$$\{\hat{a_1}, \ldots, \hat{a_K}, \hat{r_1}, \ldots, \hat{r_K}\} = \mathrm{argmax} L(a_1, \ldots, a_K, r_1 \ldots, r_K) \tag{12}$$

with a constraint $\sum_{i=1}^{K} a_i = 1$.

As the maximum log likelihood estimator for the Poissonian mixture model given by Eq. (12), by setting partial differentiations of Eq. (11) with respect to each parameter as zero,

$$\frac{\partial L}{a_i} = 0, \quad \frac{\partial L}{r_i} = 0, \quad (i = 1, \ldots, K) \tag{13}$$

the following recursive equations for parameters can be derived,

$$\begin{aligned} a_i^{(\nu+1)} &= \frac{1}{T} \sum_{t=1}^{T} \frac{F_i^{(\nu)}(z(t))}{G^{(\nu)}(z(t))} \quad (i = 1, \ldots, K), & (14) \\ r_i^{(\nu+1)} &= \frac{1}{N} \frac{\sum_{t=1}^{T} z(t) \frac{F_i^{(\nu)}(z(t))}{G^{(\nu)}(z(t))}}{\sum_{t=1}^{T} \frac{F_i^{(\nu)}(z(t))}{G^{(\nu)}(z(t))}} \quad (i = 1, \ldots, K), & (15) \end{aligned}$$

where

$$\begin{aligned} F_i^{(\nu)}(x) &= a_i^{(\nu)} \frac{(Nr_i^{(\nu)})^x e^{-Nr_i^{(\nu)}}}{x!}, & (16) \\ G^{(\nu)}(x) &= \sum_{i=1}^{K} F_i^{(\nu)}(x). & (17) \end{aligned}$$

These recursive equations give us a way to estimate parameters by starting from an adequate set of initial values. These recursive equations are also referred as to the EM-algorithm of mixtures of distributions [11, 12].

In order to determine an adequate number of parameters we introduce Akaike Information Criteria (AIC) defined as

$$AIC = 4K - 2\hat{L}, \tag{18}$$

where $\hat{L}$ is the maximum value of log likelihood estimator where the estimated parameters are given at the model with $2K$ parameters. It is known that the preferred model should be the one with the lowest AIC value.

Furthermore we consider to determine an underlying Poissonian distribution adequate to the sample $z(s)$ from the Poissonian distributions which are described by estimated parameters. We proposed a method based on the maximum log likelihood procedure for each distribution. Let $R_i(z) \quad (i = 1, \ldots, K)$ be Poisson distribution with estimated intensity $N\hat{r}_i$ defined as

$$R_i(z) = \frac{(N\hat{r}_i)^z}{z!} e^{-N\hat{r}_i}. \tag{19}$$

Then it is possible for us to select the adequate distribution where $z(s)$ is sampled, by finding a log likelihood function $\log R_i(z(s))$ which is maximized for $i = 1, \ldots, K$. Namely the adequate Poissonian distribution for the sample $z(s)$ should be given as $R_{\hat{i}_s}$, where

$$\hat{i}_s = \mathrm{argmax} \log R_i(z(s)). \tag{20}$$

6. Numerical simulation

Before starting empirical analysis on room opportunities by using our proposal method, we performed the estimation procedure with artificial data.

We generate the time series $z_t \quad (t = 1, \ldots, T)$ from a Poissonian mixture model given by

$$\begin{cases} Pr(Z = l) &= \frac{(Nr)^l}{l!} e^{-Nr} dr, \\ Pr(r = r_i) &= a_i \quad (i = 1, \ldots, K) \end{cases}, \tag{21}$$

where K is the number of states, and $a_i \quad (i = 1, \ldots, K)$, which hold $\sum_{i=1}^{N} a_i = 1$, represent probabilities for the i-th state to be.

Fig. 6 shows a sample path of this data at $K = 12$ and $T = 200$, and Tab. 2 shows parameters to generate this time series. From this time series we estimate parameters without any knowledge on parameters. As shown in Fig. 7 (top) the AIC value for each K has a minimal value at $K = 10$. The estimated parameters are seen in Tab. 2. Furthermore we select values of r_i for each sample by means of the proposed method mentioned in Sec. 5. The estimated parameters are shown as a function of samples as shown in Fig. 7 (bottom). It is confirmed that the proposed parameters estimation method based on the mixture of Poissonian model is well workable for the artificial data set.

We confirmed that estimation error between estimated probability $\hat{r}$ and given probability r defined as $|\hat{r} - r|$ at each data point is less than 8.0×10^{-6}, and that its relative error, defined as $|\hat{r} - r|/r$, is less than 0.7 %.

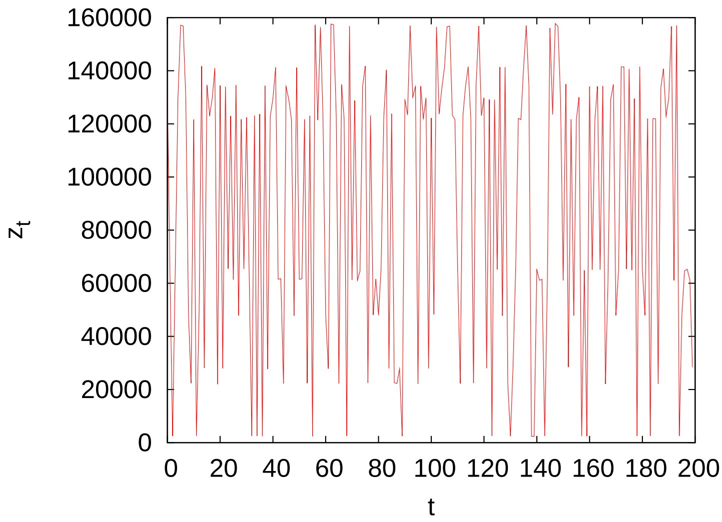


Figure 6: Examples of time series generated from the Poissonian mixture model for $K = 12$.

7. Empirical results and discussion

We compute log likelihood functions following Eq. (11) for each observations on the total number of rooms opportunities for a period from 25th December 2009 till 15th July 2010 at $D = 1$ (See Fig. 3). We assume that N is almost equivalent to the total population of Japan, so that $N = 1,000,000,000$.

In order to determine the adequate number of parameters we calculated the value of AIC at each K. As shown in Fig. 8 the AIC values show dependence on the number of parameters. In this case $K = 23$ is the lowest case.

The estimated parameters at $K = 23$ are shown in Tab. 3 and Fig. 9. The parameters for weekdays (from Monday to Friday) seem to be estimated as higher values of $\hat{r}$, and for special holidays as lower values of $\hat{r}$.

Furthermore by means of procedure to select the underlying distribution from a set of distributions with estimated parameters for each sample, we compute probabilities r_i for each sample date. As shown in Fig. 9 (bottom) probabilities show strong dependence on calendar dates. Specifically on holidays and Saturdays they take small values, on weekdays they are large.

Next we estimated probabilities r_i of the number of opportunities for four regions. According to values of AIC the adequate number of parameters is estimated as $K = 2$ or 3. Fig. 4 shows probabilities on four regions (See Fig. 11). It is confirmed that there is both regional and temporal dependence of probabilities. The correlation among probabilities on regions is one of the important factors of determining availability of rooms all over the Japan. It is obviously confirmed that synchronous behavior of probabilities on each regions exists. Synchronization of probability fluctuations is deeply related to synchronization

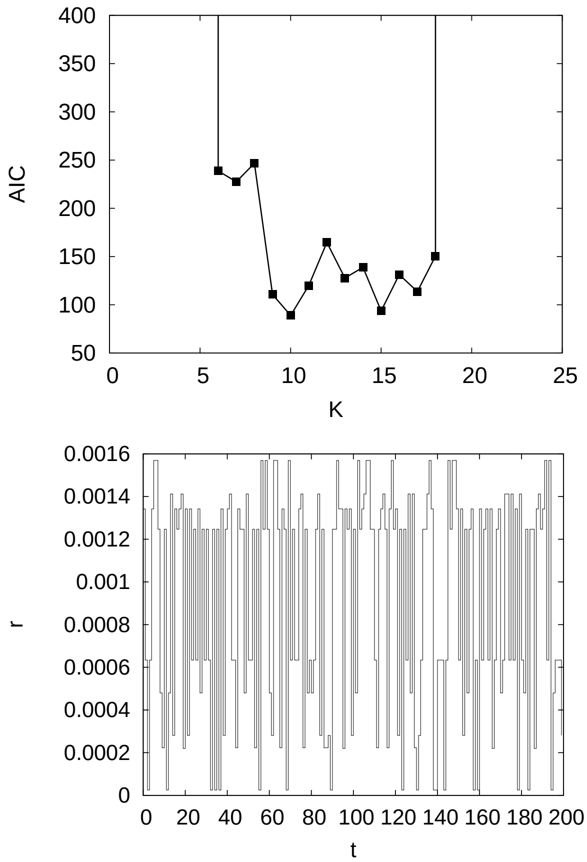


Figure 7: The value of AIC for artificial data is shown as a function in terms of the number of parameters K (Top). The lowest value of AIC is found as 3908.41 at $K = 10$. Estimated parameters are shown as a function in terms of samples (Bottom).

Table 2: We estimated parameters for the artificial data set generated from a Poissonian mixture model. The number of model parameters is set as to $K = 12$. Parameters set of the Poissonian mixture model which generate time series (top) and ones estimated by using the EM estimator (bottom). The number of estimated parameters is $K = 10$ and its AIC value is obtained as $AIC = 88.9952152$.

r(1)	0.000025	a(1)	0.109726
r(2)	0.000223	a(2)	0.070612
r(3)	0.000280	a(3)	0.073355
r(4)	0.000479	a(4)	0.077612
r(5)	0.000613	a(5)	0.094848
r(6)	0.000652	a(6)	0.073841
r(7)	0.001219	a(7)	0.090867
r(8)	0.001233	a(8)	0.062191
r(9)	0.001295	a(9)	0.077662
r(10)	0.001341	a(10)	0.102573
r(11)	0.001412	a(11)	0.085892
r(12)	0.001570	a(12)	0.080821

r(1)	0.0000247783	a(1)	0.0900000000
r(2)	0.0002202500	a(2)	0.0000000000
r(3)	0.0002229207	a(3)	0.0700000000
r(4)	0.0002806173	a(4)	0.0550000000
r(5)	0.0004798446	a(5)	0.0650000000
r(6)	0.0006343063	a(6)	0.1750000000
r(7)	0.0012466507	a(7)	0.2699999815
r(8)	0.0013420367	a(8)	0.1050000185
r(9)	0.0014120537	a(9)	0.0800000000
r(10)	0.0015688622	a(10)	0.0900000000

among consumer behavior on travel and tourism in Japan.

Thirdly we confirmed that the relationship between mean of room prices and probability r at and each observation. Fig. 10 shows scatter plots during January and May 2010. Each point represents the relation at each observation day. In the New Year holidays cheaper opportunities remained in comparison with expensive ones. In the Golden week holidays expensive opportunities remained in comparison with cheaper ones. This tendency of price dependence on the probability r is found on Statuary. The Japanese consumer preference of hotel booking on weekends and holidays after April 2010 seems to be different from one before it.

From the dependence of mean prices on the probability r we can understand preference and motivation structure of consumers for travel and tourism.

8. Conclusion

We analyze data on a Japanese hotel reservation site and found that there is strong dependence of the number of available opportunities on calendar date.

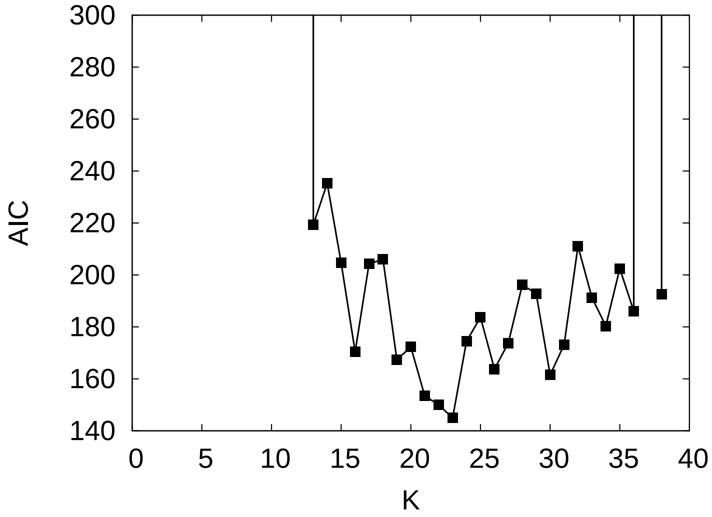


Figure 8: The value of AIC is shown as a function in terms of the number of parameters K. The lowest value of AIC is found as 144.95 at $K = 23$.

We proposed a model of hotel booking activities based on Poissonian mixture model with intensity depending on calendar dates. From binomial model with a time-varying successful probability we derived the Poissonian mixture model.

From the Poissonian mixture model we can consider that the numbers of room population at each date with different parameters shows motivation structure of consumers dependent on calendar dates. Multiple time series of the numbers of room opportunities at distinct constructed from this data will prove the difference among motivation structure of consumers to distinct.

It is found that this large-scale data on hotel opportunities provide us with invisible properties on human activities related to travels and tourism in Japan. A future emerging technology will make it possible to see or foresee something which we can not see at this moment.

Acknowledgement

This study was financially supported by the Excellent Young Researcher Overseas Visiting Program (# 21-5341) of the Japan Society for the Promotion of Science (JSPS). One of the authors (A.H. Sato) expresses his sincere gratitude to Prof. Dr. Dirk Helbing for stimulative discussions. This is a research study, which has been started in collaboration with Dirk Helbing.

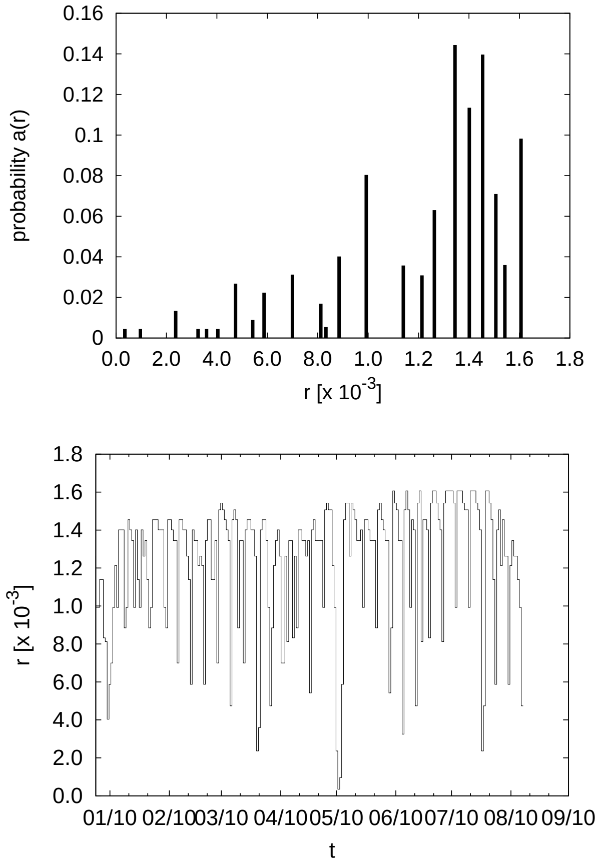


Figure 9: The illustrative representation of estimated parameters and their probabilities (Top). Each bar position represents the estimated parameters and its height corresponds to its probability. The estimated intensity are shown as a function of days (Bottom).

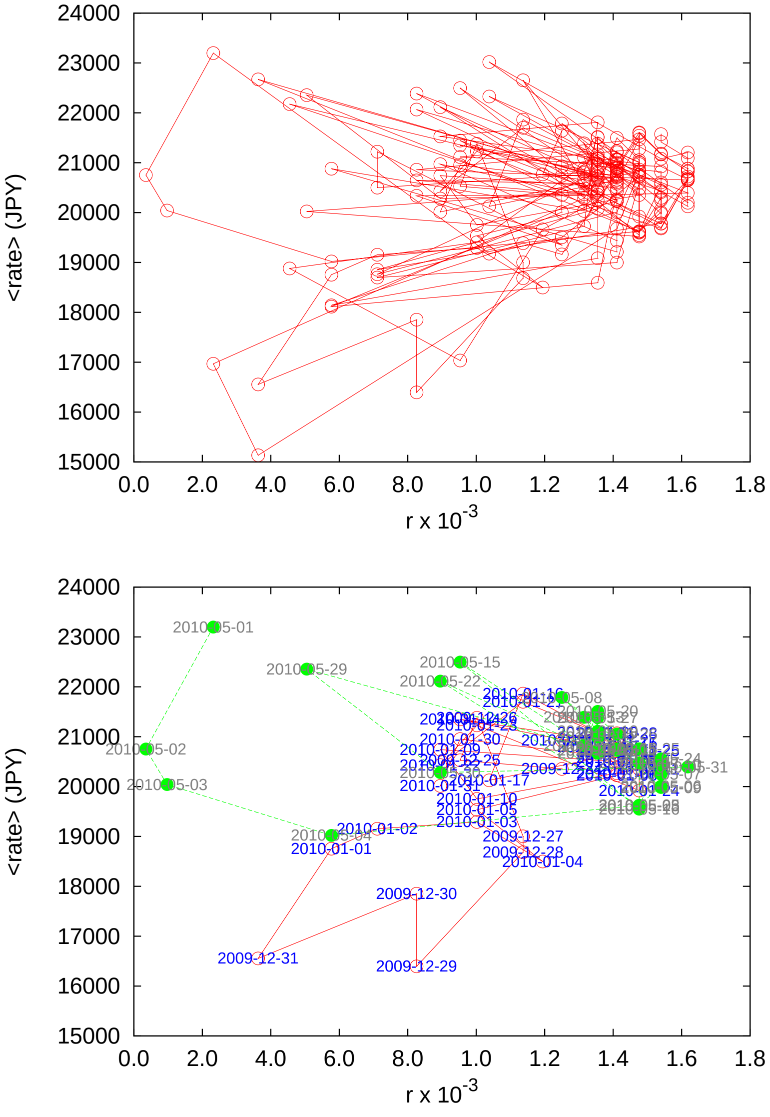


Figure 10: The relationship between a mean of room opportunities and existence probability of room opportunities for a period from 25th December 2009 to 15th July 2010 (top) and January and May 2010 (bottom). Each point represents the relation on each observation day.

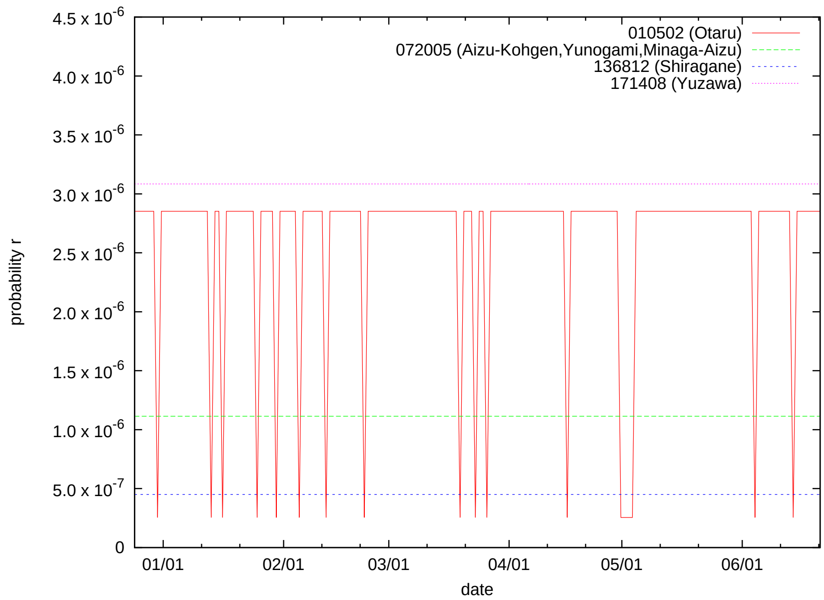


Figure 11: The parameters estimated from the number of available opportunities for each region on observation date. It is found that there exist regional dependences of probability fluctuations.

Appendix A. Procedure to estimate parameters

In this section we describe a numerical method to estimate parameters by using the EM algorithm. The update rule of parameters in the EM algorithm is given as

$$a_i^{(\nu+1)} = \frac{1}{T}\sum_{t=1}^{T}\frac{F_i^{(\nu)}(z(t))}{G^{(\nu)}(z(t))} \quad (i=1,\ldots,K), \tag{A.1}$$

$$r_i^{(\nu+1)} = \frac{1}{N}\frac{\sum_{t=1}^{T} z(t)\frac{F_i^{(\nu)}(z(t))}{G^{(\nu)}(z(t))}}{\sum_{t=1}^{T}\frac{F_i^{(\nu)}(z(t))}{G^{(\nu)}(z(t))}} \quad (i=1,\ldots,K), \tag{A.2}$$

where

$$F_i^{(\nu)}(x) = a_i^{(\nu)}\frac{(Nr_i^{(\nu)})^x e^{-Nr_i^{(\nu)}}}{x!}, \tag{A.3}$$

$$G^{(\nu)}(x) = \sum_{i=1}^{K} F_i^{(\nu)}(x). \tag{A.4}$$

We compute these recursive equations by setting an arbitrary set of parameters. Some of them are convergent and the others divergent. Therefore it is important

Table 3: Parameters estimated by means of the maximum likelihood estimator. In order to estimate them we used the data set for a period from 24th December 2009 till 15th July 2010. The number of model parameters is set as to $K = 23$. The AIC value was obtained as $AIC = 144.95115$.

r(1)	0.0000966000	a(1)	0.0044642857
r(2)	0.0002365800	a(2)	0.0133928571
r(3)	0.0000350600	a(3)	0.0044642857
r(4)	0.0008848885	a(4)	0.0401747715
r(5)	0.0003253200	a(5)	0.0044642382
r(6)	0.0003592096	a(6)	0.0044643332
r(7)	0.0008325793	a(7)	0.0053903109
r(8)	0.0004040700	a(8)	0.0044642858
r(9)	0.0004740300	a(9)	0.0267857143
r(10)	0.0005874029	a(10)	0.0223599584
r(11)	0.0005421173	a(11)	0.0088900416
r(12)	0.0006998500	a(12)	0.0312500000
r(13)	0.0008123711	a(13)	0.0169349176
r(14)	0.0009926461	a(14)	0.0803571429
r(15)	0.0011395580	a(15)	0.0357147066
r(16)	0.0012133569	a(16)	0.0308473690
r(17)	0.0012627651	a(17)	0.0630041846
r(18)	0.0013446027	a(18)	0.1443472488
r(19)	0.0014543992	a(19)	0.1396581557
r(20)	0.0014011806	a(20)	0.1134539949
r(21)	0.0015066568	a(21)	0.0709306351
r(22)	0.0015423484	a(22)	0.0359238138
r(23)	0.0016064234	a(23)	0.0982627486

for us to find an adequate set of initial values when we use the EM-algorithm given by Eqs. (A.1) and (A.2) for estimating parameters. To do so we propose a way to find a candidate of parameters in a stochastic manner.

In the Monte Carlo step a set of parameters is randomly selected and the log likelihood function is evaluated at the point. If the log likelihood value at this point is finite, we choose this set of parameters as the new starting point for the recursive equation. Further setting the set of parameters as an initial condition, we recursively calculate the EM algorithm until the log likelihood value converges. If the log likelihood value is greater than the maximum value of log likelihood function obtained in the Monte Carlo step, then this set of parameters as a candidate of estimators.

Repeating this procedure until we can not find any points which improve the log likelihood value in the Monte Carlo step, we estimate an adequate set of parameters. This algorithm is described as follows.

(0) We set $maxobj = 0$ and $counter = 0$.

(1) From independent and identical uniform distributions ranging from 0 to $r0$,

r'_i are randomly selected. b'_i are randomly selected from independent and identical uniform distribution with arbitrary range and a'_i are normalized as $a'_i = b'_i / \sqrt{\sum_{i=1}^{K} b'_i}$. If *counter* $> MAXCOUNT$, then we go to (6).

(2) If $L(a'_1, \ldots, a'_K, r'_1, \ldots, r'_K)$ is greater than *maxobj*, then we set *maxobj* as the value, $r_i := r'_i$ and $a_i := a'_i$, and go to (3). Otherwise go to (1).

(3) From the starting point $(a_1 \ldots, a_K, r_1, \ldots, r_K)$ we compute Eqs. (14) and (15) recursively until the value of log likelihood function converges.

(4) If the maximum value of log likelihood function in terms of the converged set of parameters is larger than *maxobj*, then we set the value as *maxobj* and record the solution as a candidate of estimation.

(5) *counter* = *counter* + 1 and if *counter* $< MAXCOUNT$, then we go to (1). Otherwise we go to (6).

(6) We stop this computer program and display *maxobj* and the recorded candidate as estimated parameters.

References

[1] R. Law, "Disintermediation of researvations", International Journal of Contemporary Hospitality Management, **21** (2009) 766-772.

[2] R. Pilia, "Disintermediation of the hotel industry: the new perspects of online booking" available at : `www.mymarking.net/agora/editoriali/contributi/dettaglio?articolo.asp?a=29\&s=134\&i=2584` (2008).

[3] G. Haag, and W. Weidlich, "A stochastic theory of interregional migration", Geographical Analysis, **16** (1984) 331-357.

[4] W. Weidlich, Sociodynamics: A Systematic Approach to Mathematical Modelling in the Social Sciences, Taylor and Francis, London (2002).

[5] 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.

[6] R.M. Cassie, "Some uses of probability paper in the analysis of size frequency distributions", Australian Journal of Marine and Freshwater Research, **5** (1954) 513–522.

[7] J.P. Harding, "The use of probability paper of the graphical analysis of polymodal frequency distributions", Journal of the Marine Biological Association, **28** (1949) 141–153.

[8] C.G. Bhattacharya, "A simple method of resolution of a distribution into Gaussian components", Biometrics, **23** (1967) 115-137.

[9] R.P. Rider, "Estimating the parameters of mixed Poisson, binomial, and Weibull distributions by the method of moments", Bulletin of the International Statistical Institute, **39** (1961) 225–232.

[10] V. Hasselblad, "Estimation of Finite Mixtures of Distributions from the Exponential Family", Journal of the American Statistical Association, **64** (1969) 1459–1471.

[11] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", **39** (1977) 1–38.

[12] 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.