Analysis of the Characteristics of Customer Defection on a Hair Salon Considering Individual Differences

Mana Iwata[1](✉), Kohei Otake[2], and Takashi Namatame[3]

[1] Graduate School of Engineering, Tokyo Institute of Technology,
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
manachan.kororo@gmail.com
[2] School of Information and Telecommunication Engineering,
Tokai University, 2-3-23, Takanawa, Minato-ku, Tokyo 108-8619, Japan
[3] Faculty of Science and Engineering, Chuo University,
1-13-27, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Abstract. In recent years, as the number of hair salons increases, and the scale of the hair salon market has declined. Hence, competition for hair salon acquisition will be intensified. It is important for a hair salon to specify the cause of customer defection and plan to settle that. In this study, we analyzed the POS data with customer distinguishing information of a hair salon chain. Concretely, we performed logistic regression and hierarchical Bayes logit model to identify the cause of customer defection on the hair salon. After that, we categorized and considered the customers using hierarchical cluster analysis. Using these models, we extract the characteristics of customer defection on a hair salon and propose marketing measures to prevent defection.

Keywords: Customer defection · Logistic regression analysis · Hierarchical Bayes logit model

1 Introduction

The number of hair salon reached over 240,000 stores according to a research of the Ministry of Health, Labour and Welfare in fiscal year 2017, and new opening keeps increasing [1]. An emporium is increased newly, but also the scale of the hair salon market decreases in recent years, and it seems it is becoming severe in customer acquisition competition of a hair salon [1]. It is an important to reduce customer defection. It seems that the problem of compatibility with the hair salon is great for new customers, and it is difficult to completely prevent this. However, for customers who visit two or more times, we can think about marketing measure to prevent defection.

In this research, we aim to clarify factors of customer defection by analyzing customers' behavior at the last visit and one time ago of the last visit to the last visit, and their changes. Moreover, we aim to propose marketing measures to prevent customer defection.

G. Meiselwitz (Ed.): HCII 2019, LNCS 11579, pp. 378–391, 2019.
https://doi.org/10.1007/978-3-030-21905-5_29

2 Data Sets

In this study, we target on 10 stores of a hair salon chain in Japan, all stores are located in urban area and near railroad stations. In this study, we used following data.

- Member registration information data: Information on customer attributes such as gender, date of birth, member ID
- Purchasing history data with customer information: Information on purchasing behavior such as visit date/time, sales amount, purchased item (menu), staff rank, etc. (data period: July 1, 2015 to June 30, 2017)

We analyzed 11,683 customers with their information and visited at least twice during the data period of ID-POS data.

First, we defined the defect situation. Because each customer's visit interval was not same, we defined each customer's defect that was over 30 days from the visit interval maximum of former.

In addition, it is thought that the situation at the time of the last visit and change from the previous visit will lead to customer defection. Therefore, we made variables about the accounting contents which are at the last visit, the contents one time ago of the last visit and the amount which changed to the last visit.

3 Analysis of Customer Defection Factors

In this section, we describe our analysis procedure.

3.1 Flow of Analysis

We show the outline of analysis in Fig. 1.

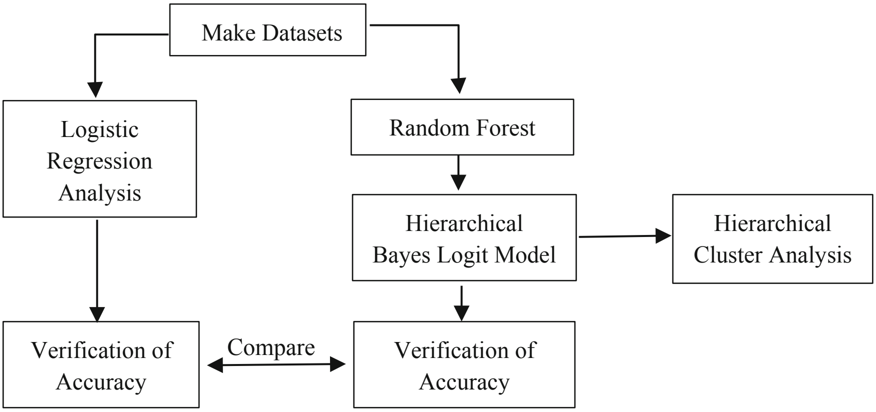
Make Datasets
Logistic Regression Analysis
Random Forest
Hierarchical Bayes Logit Model
Hierarchical Cluster Analysis
Verification of Accuracy
Compare
Verification of Accuracy

Fig. 1. Outline of analysis

In this research, first, we perform logistic regression analysis and then a hierarchical Bayes logit model. Then, we compare the accuracy of these two analyzes.

In the hierarchical Bayes logit model, variables cannot be selected sequentially like a general regression model, and it is necessary to select explanatory variables in advance. Since there are many explanatory variables created in this study, we select explanatory variables beforehand by using random forest. After that, we create a hierarchical Bayes logit model and obtain accuracy. Then, we perform hierarchical cluster analysis in order to cluster customers based on the determined individual parameters.

3.2 Logistic Regression Analysis

In order to identify factors of customer defection, we try to perform logistic regression analysis. When regression coefficient is defined as β_k and explanatory variable is defined as x_k, probability of occurring defection p of event y is shown as Eq. (1) [2].

$$p_y = \frac{\exp\{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\}} \tag{1}$$

where y of (1) defines below, and we estimate parameter β_k.

$$y = \begin{cases} 1 \cdots \text{Defect} \\ 0 \cdots \text{Continue} \end{cases}$$

As explanatory variables used in the model construction, we created twelve variables from customers' behavior at the last visit, thirteen variables from changes from one time ago of the last visit to the last visit, four variables from personal attributes of customers and hair salon's staffs. Moreover, we normalized sales, discount amount, point balance, diff of sales, diff of discount amount, and diff of point balance

Details of the explanatory variables are shown in Table 1.

Table 1. Demographic variables and customers' behavior variables used in the model construction

Type of variable			Variable name	Data type
Objective variable			Defection or not	0 or 1
Explanatory variable	Behavior at the last visit	No interaction	Sales	Decimal
			Discount amount	Decimal
			Point balance	Decimal
			Store ID	Factor
		Interaction	Cut	Integer
			Treatment	Integer
			Color	Integer
			Blow shampoo hair set	Integer
			Private brand item	Integer
			Perm	Integer

(*continued*)

Table 1. (*continued*)

Type of variable			Variable name	Data type
	Change in behavior from one time ago of the last visit to the last visit	No inter action	Sales	Decimal
			Discount amount	Decimal
			Point balance	Decimal
			Store ID	Factor
			Staff ID	Factor
		Interaction	Cut	Integer
			Treatment	Integer
			Color	Integer
			Blow shampoo hair set	Integer
			Private Brand item	Integer
			Perm	Integer
	Interaction in a day of the week		A day of week at the last visit	Factor
			A day of week at one time ago of the last visit	Factor
	Interaction in time		Time at the last visit	Factor
			Time at one time ago of the last visit	Factor
	Interaction in rank		Staff's rank at the last visit	Factor
			Staff's rank at one time ago of the last visit	Factor
	Interaction in demography		Customers' sex	Factor
			Customers' age	Integer

3.3 Hierarchical Bayes Logit Model

Conventional statistical methods such as logistic regression analysis common obtaining point-estimate parameters by most likelihood method. On the other hand, the hierarchical Bayes logit model assumes a prior distribution for each parameter and performs distribution convergence and individual parameter estimation for each case by repeatedly generating random numbers based on prior distribution in simulation and update parameter values using Bayesian theory. As a result, it is possible to flexibly express the parameters of each individual and each group. Based on the premise that customer's desire for service varies from individual to individual, we think that it is appropriate to estimate parameters for each customer rather than uniquely estimate parameters, and we apply a hierarchical Bayes logit model.

3.3.1 Formulation of Defection Discrimination Model

In this section, we show a model that discriminates whether or not it is a defection using a hierarchical Bayes logit model. The proposed model is expressed in the framework of a logistic regression model. The defection probability p_i is shown below as a proposed model in Eq. (2).

$$p_i = \Pr\{y_i = 1\} = \tfrac{e^{u_i}}{1 + e^{u_i}} \tag{2}$$

Equation (3) concretely shows the utility u_i for the defection of the customer *i.*

$$u_i = X_i^T B_i + x_i[StoreID[i]] \tag{3}$$

X_i is an explanatory variable vector containing intercept terms related to the defection of customer *i*, B_i is a parameter vector containing intercept term of each customer, and x_i is a store specific term. Also, in order to take account of heterogeneity of customers, parameters are assumed to be different for each customer [2].

3.3.2 Parameter Hierarchy

In this study, we used Markov Chain Monte Carlo methods, (MCMC) to estimate parameters and No-U-Turn-sampler for sampling [6]. Sampling was done 5000 times. Then the first 500 times was the burn-in period that the samples were discarded.

To consider customer heterogeneity, we constructed the hierarchical Bayes logit model that was assumed that individual parameters for each customer. To estimate the parameters of a hierarchical Bayes logit model, simulations are performed in which the prior distribution is assumed for each parameter and the generation of random variables is repeated from the distribution [3, 4].

In this study, it is assumed that the parameter vector β_i follows the multivariate normal distribution as Eq. (4).

$$\beta_i \sim MVN(\Delta, \Sigma_B) \tag{4}$$

In addition, as a hyper prior distribution for expressing the variation of parameters for each customer, for Δ and $\sum_B$ in Eq. (4), a uniform wide distribution (no information prior distribution) and the inverse Wishart distribution as shown in Eqs. (5) and (6).

In Eq. (6), M represents the number of variables, v_0 represents a constant M, and V_0 represents a square matrix of $M \times M$ with a diagonal component $M + 3$.

$$\Delta \sim U(0, 100) \tag{5}$$

$$\textstyle\sum_B \sim IW(v_0, V_0), v_0 = M, V_0 = \begin{bmatrix} M+3 & 0 & \cdots & 0 \\ 0 & M+3 & & \\ & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M+3 \end{bmatrix} \tag{6}$$

In addition, we use $\hat{R}$ to confirm convergence of parameters. $\hat{R}$ is a convergence determination index proposed by Gelman and Rubin [5]. If $\hat{R}$ is close to 1, convergence to a steady distribution is suggested, and when it is larger than 1 it is considered not converging. In [5], if $\hat{R}$ should be less than 1.1 or 1.2, we use the criterion that it can be judged that it converged.

Random Forest

In the hierarchical Bayes logit model, when the more variables are, it takes more time to calculate the simulation. In addition, since all 29 explanatory variables used in this analysis do not necessarily work significantly, so we perform random forest as a preliminary analysis before creating a model. Using the significance due to the average decrease in impurity of the Gini coefficient calculated from the result, the top eight variables with the highest importance are selected in advance. The top eight variables with high importance obtained as a result of performing the random forest using explanatory variables shown in Table 1 are shown in descending order in Table 2.

Table 2. Variable selected by random forest

Point balance at the last visit
Difference of point balance
Store ID at the last visit
Sales at the last visit
Difference of sales
Day of the week at last visit
Day of the week at one time ago of the last visit
Age

In the hierarchical Bayes logit model, if we use categorical variables, it takes very long time to calculate. In order to think about marketing measures, we thought that it is not necessary to obtain a specific day of the week, so the day of the week was converted into two values, that is, the weekend or not. Furthermore, since we thought that the store ID at the last visit could be explained by the store specific that introduced this time, it is excluded from the explanatory variables.

3.4 Hierarchical Cluster Analysis

This analysis attempts to identify relatively homogeneous groups of cases (or variables) based on selected characteristics, using an algorithm that starts with each case (or variable) in a separate cluster and combines clusters until only one is left. We use the Ward method.

3.5 Dataset and Evaluation Indicator

Although the number of target customers in this research is 11,683, when constructing, we randomly sample the number of continuing customers by setting the number equal to the number of defective customers.

Furthermore, in order to verify the prediction accuracy of the model, we set 75% of the data for training data and 25% for the test data, for each continuing customer and each defective customer. As a result, the datasets used in the model construction was divided as follows (Table 3).

Table 3. Datasets used in the model construction

	Training data	Test data	Total
Defective customers	2,493	831	3,324
Continuing customers	2,493	831	3,324
Total	4,986	1,662	6,648

In order to confirm the prediction accuracy of the constructed model, we performed hold-out validation by using the train data and test data. Specifically, we created a confusion matrix like a following table and we calculated prediction accuracy of the constructed model by using following equations (Table 4).

Table 4. Confusion matrix

		Predicted class	
		Positive	Negative
Actual class	Positive	True Positive (TP)	True Negative (TN)
	Negative	False Negative (FP)	False Negative (FN)

Accuracy (ACC): Percentage of the total number correctly predicted among the total number predicted.

$$\mathrm{ACC} = \frac{TP + TN}{FP + FN + TP + TN}$$

Precision (PRE): Percentage of the total number that is a positive class actually among the total number predicted positive class.

$$\mathrm{PRE} = \frac{TP}{TP + FP}$$

Recall (REC): Percentage of the total number predicted positive class among the total number that is a positive class actually

$$\text{REC} = \frac{TP}{FN + TP}$$

F-measure: harmonic mean of PRE and REC

$$\text{F - measure} = 2 \times \frac{PRE \times REC}{PRE + REC}$$

4 Results

In this section, we summarize our results.

4.1 Logistic Regression

We built a model that predicts defection for the entire customer using binomial logistic regression analysis with stepwise selection method. Then, we selected explanatory variables of coefficient of significant probability less than 0.05.

Table 5. Estimated value of selected partial regression coefficient

Explanatory variables	Partial regression coefficient
Intercept	1.232
Sales	−0.137
Discount amount	−0.135
Point balance	−0.607
Store ID at the last visit was B	1.301
Store ID at the last visit was C	0.510
Store ID at the last visit was D	0.606
Store ID at the last visit was E	0.401
Store ID at the last visit was F	0.402
Store ID at the last visit was G	0.720
Store ID at the last visit was H	0.751
Store ID at the last visit was K	0.786
Last visit was on Friday	−0.316
Last visit was on Sunday	−0.287
Menu at the last visit was cut	−0.840

(*continued*)

Table 5. (*continued*)

Explanatory variables	Partial regression coefficient
Menu at the last visit was blow shampoo hair set	−0.771
Difference of point balance	0.319
Not changed staffs	−0.204
Add cut	0.384
Add private brand goods	0.123
Menu at the last visit was cut and perm	0.750
Menu at the last visit was cut and blow shampoo hair set	0.585
Age	−0.010

From Table 5, we found that sales, point balance, store ID, the day, menu and staff's sex at the last visit become significant.

Tables 6, 7, 8 and 9 show the confusion matrix and the evaluation indicator for the train data and the test data.

Table 6. Confusion matrix of model for the train data

		Predicted class	
		Positive	Negative
Actual class	Positive	836	1657
	Negative	339	2154

Table 7. Evaluation indicator of model for the train data (%)

ACC	PRE	REC	F-measure
60.0	71.1	33.5	45.6

Table 8. Confusion matrix of model for the test data

		Predicted class	
		Positive	Negative
Actual class	Positive	266	565
	Negative	114	717

Table 9. Evaluation indicator of model for the test data (%)

ACC	PRE	REC	F-measure
59.1	70.0	32.0	43.9

From Tables 6, 7, 8 and 9, it turns out that it is almost resulted in Negative

4.2 Hierarchical Bayes Logit Model

Table 10 shows the average value and $\hat{R}$ of the individual parameters of the explanatory variable obtained as a result of the hierarchical Bayes logit model.

Table 10. Variable selected by random forest

Explanatory variables	Average of parameters	$\hat{R}$
Intercept	−0.052	1.001
Point balance at the last visit	−1.123	1.001
Difference of point balance	0.753	1.011
Sales at the last visit	−0.144	1.008
Difference of sales	−0.029	1.001
Day of the week at the last visit	−0.148	1.001
Day of the week at one time ago of the last visit	−0.137	1.001
Age	−0.245	1.001

From the Table 10, we also found that $\hat{R}$, which determines the convergence of each parameter, also falls below 1.1, which is a measure of convergence. It is understood that point balance and age parameter are negative, and difference of point balance is positive. Also, we can see that difference of sales and sales parameter may be negative or positive. In addition, the confusion matrix for the train data and the evaluation indicator are as shown in Tables 11 and 12 below.

Table 11. Confusion matrix of model for the train data in hierarchical bayes logit model

		Predicted class	
		Positive	Negative
Actual class	Positive	2057	436
	Negative	11	2482

Table 12. Evaluation indicator of model for the train data (%)

ACC	PRE	REC	F-measure
91.0	99.5	82.5	90.2

Comparing Tables 6, 7, 8 and 9 with Tables 11 and 12, it is understood that the hierarchical Bayes logit model was better than logistic regression in performance.

As a result of hierarchical cluster analysis, it was divided into four clusters. The average values of the parameters of each cluster are shown in Table 13.

Table 13. The average value of each parameter

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Intercept	0.154	0.169	−0.277	−0.237
Point balance at the last visit	−1.179	−1.105	−1.052	−1.158
Difference of point balance	0.728	0.828	0.768	0.695
Sales at the last visit	−0.270	0.005	−0.021	−0.279
Difference of sales	−0.103	0.076	0.032	−0.106
Day of the week at the last visit	−0.080	−0.085	−0.217	−0.210
Day of the week at one time ago of the last visit	−0.069	−0.075	−0.214	−0.183
Age	−0.181	−0.170	−0.315	−0.311

From Table 13, the positive and negative of the average value of most parameters did not change depending on the cluster. However, there is a case that sales and the difference of sales differed, and it turned out that a difference arises depending on the cluster.

5 Discussions

In this section, we discuss about the results of our analysis and propose some efficient marketing strategies.

5.1 Discussions for Logistic Regression

As the result of logistic regression analysis, we found that customers with low sales and small discount amount at the last visit are easy to defect. The sales are considered to represent money sense for beauty. Therefore, it can be said that customers who think that they do not want to pay so much for beauty and customers who are dissatisfied with less discount amount are easy to defect. Therefore, because coupons generated at a certain amount of money or more cannot prevent such customer defection, measures that distribute coupons that arise only by visiting stores are considered to be effective.

Also, when the point balance at the final visit was small, and customer with a difference in point balance know that it is easy to defect. The point balance is considered to represent the loyalty to the store of the customer. Also, the difference of the point balance is the point award amount one time ago of the last visit - the point usage amount one time ago of the last visit, so the difference of the point balance is considered to represent the intention to accumulate points. In other words, it can be said that customer with low royalties for the store and having little intention to accumulate points tend to defect. Considering that customers who do not intend to accumulate points generally are hard to become good customers, it is thought that firstly it is necessary to raise the royalty for the store by urging to accumulate points. Therefore, measures such as preparing some benefits when saving points more than a certain amount are considered to be effective.

In addition, variables for staff and menu are also selected. Regarding the staff, we found that customers who are changing staff are more likely to defect. Therefore, it can be said that it is necessary to take careful handover when the customer changes staff. Regarding the menu, we found that it was easy to continue if the customer selects blow shampoo hair set at the last visit. This is because menus are less likely to cause mistakes than perms and colors.

Also, we found customers choosing a cut at the last visit are more likely to continue, but adding a cut tends to lead to defect. The staff are familiar with the preferences of customers who continue to select a cut, but they do not know the preferences very much about customers who add a cut. It is thought that as the hearing about the preferences does not go properly, it may lead to defect. Hence, it is considered that it is necessary to carefully hear the preferences for customers who have added a new menu, even if they are not visiting for the first time.

Also, we found that customers who visit on Friday and Sunday are harder to defect. However, because it is difficult to identify the cause of such a result, we found that it was necessary to compare it between weekdays and weekends. The results compared on weekdays and weekends are stated in Sect. 5.2. It also turned out that there was a difference in departure rate for each store. Therefore, rather than doing a campaign common to all stores, it is considered more effective to conduct different campaigns for each store. It is necessary to analyze each salon, which is also a future work.

5.2 Discussions for Hierarchical Bayes Logit Model

As the result of the analysis using the hierarchical Bayes logit model, the overall tendency is point balance at the last visit is small, and customer with a difference in point balance know that it is easy to defect. Hence, measures for points as stated in Sect. 5.1 are considered to be effective.

Also, customers with low sales and little increase from the last visit in sales are easy to defect. Sales represent the sense of money for beauty, and the difference in sales represents how good the menu compared to last time. In other words, customers not only do not want to pay so much for beauty, but also want to cheaper than raising the rank of the menu are easy to defect. Therefore, it is considered good practice to advertise to customers like this, appealing cheaper than appealing the quality of treatment.

Furthermore, it turned out that customers who come to the store on weekdays are more likely to defect. Hence, it seems that establishing benefits such as restricted visits on weekdays is a measure to prevent defection. In addition, it is found that the lower the age, the easier it is to defect, so it is considered that setting privileges for young people is a measure to prevent defection.

As a result of hierarchical cluster analysis, the average value of parameters for most clusters did not change much, but there was a difference in the difference between sales and difference of sales. In other words, unlike the overall trend, it turned out that there are clusters that can be defected even if sales are high or there is a difference in sales. Hence, there is a group that tends to defect, despite spending money on beauty or changing to a good menu. In such a group, it is better to think about improving the quality of service rather than devising measures to lower the price. Hence, it is thought that there is a group that should appeal affordable sales rather than the quality of treatment and there is a group that should appeal quality and satisfy rather than sales. In order to identify the characteristics of attributes of such a group, further analysis is necessary and it can be said that this is also our future work.

6 Conclusion

In this research, for a hair salon chain using member registration information data, purchase history data at the last visit, purchase history data one time ago of the last visit, purchase history data changed from one time ago of the last visit to the last visit, we constructed a model that predicts customer defection. As a result, we are able to grasp actions specific to customers who defected.

In addition, in order to consider individual differences, we construct a hierarchical Bayes logit model to predict customer defection. As a result, accuracy has improved and individual differences among customers can be considered. Furthermore, by using hierarchical cluster analysis, we divided the clusters into four, and analyzed the characteristics of each cluster. Hence, we can propose some marketing measurements to prevent defection.

In the defection prediction model constructed in this research, although it is possible to clarify the characteristics of customers by performing logistic regression analysis. However, the prediction accuracy of the constructed model is not satisfactory and it was found that REC was especially low, and it tended to be subject to negative expectations, so there is room for improvement. In addition, in the hierarchical Bayes logit model, only a part of the result variable considering calculation time was selected, and categorical variables such as age were also treated as numerical values. As a result, the variables were not sufficient and we could not fully grasp the characteristics of the customers that were divided into clusters. In order to solve the problem, it is necessary to increase the number of variables, and accordingly it is necessary to consider the reduction of calculation time.

Furthermore, we could confirm the accuracy in the hierarchical Bayes logit model only for the train data. Since there is no general method for checking the accuracy with respect to the test data, it is necessary to construct an algorithm in consideration of this point. In addition, it can be said that more concrete measures can be devised by constructing a more detailed model using variables not used in this research.

Acknowledgment. We thank a hair salon company for permission to use valuable datasets. This work was supported by JSPS KAKENHI Grant Number 16K03944 and 17K13809.

References

1. Ministry of Health, Labor and Welfare: Statistical Information White Paper “Overview of Health Administration Administrative Report in Heisei 29” (2018). https://www.mhlw.go.jp/toukei/saikin/hw/eisei_houkoku/17/. (in Japanese)
2. Yano Economic Research Institute: 2018 Edition beauty care marketing general press release (2018). https://www.yano.co.jp/press-release/show/press_id/1884. (in Japanese)

3. Sato, Y., Otake, K., Namatame, T.: Analysis of the characteristics of repeat customer in a golf EC site. In: Meiselwitz, G. (ed.) SCSM 2017. LNCS, vol. 10282, pp. 223–233. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58559-8_19
4. Matsura, K.: R2 Bayesian statistics modeling with Stan and R. Kyouritu (2016). (in Japanese)
5. Sato, S., Asahi, Y.: The model of purchasing and visiting behavior of customers in an e-commerce site for consumers. Commun. Oper. Res. Soc. Jpn. **58**(2), 16–22 (2013)
6. Oikawa, Y., Otake, K., Namatame, T.: Purchasing behaviors considering various search actions of customers at EC sites. In: Abstracts of Spring Conference for 2018 of Operational Research Society of Japan, pp. 270–271 (2017). (in Japanese)
7. Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**, 457–472 (1992)
8. Gilks, R.W., Richardson, S., Spiegelhalter, J.D.: Markov Chain Monte Carlo in Practice, pp. 131–143. Chapman & Hall, London (1996)