

應用層級貝氏法於線上消費行為管制 之探索研究

HIERARCHICAL BAYES FOR INTERPURCHASE TIMES WITH APPLICATION TO QUALITY CONTROL OF ONLINE CUSTOMERS

羅淑娟 蔡佳璋

台北科技大學工業工程與管理系

Shu-chuan Lo Chia-chang Tsai

Dept. of Industrial Engineering and Management

National Taipei University of Technology

摘 要

在本研究中我們使用消費間隔模式分析顧客間的異質性，並使用混合模型將顧客分成理想狀態、合理狀態和危險狀態。這兩種模式都是使用層級貝式的觀念，並透過馬可夫蒙地卡羅法的方法來估計所需的參數。最後，我們選擇以合理狀態的機率密度函數，運用管制圖的方式來監控顧客消費行為。管制圖通常是被企業用在產品品質的管制上，它提供一個合理的時間點去修正產品品質的偏差。本研究的管制圖包含一個顧客的消費間隔時間和最近消費時間。

關鍵字：層級貝氏、混合模式、顧客異質性、管制圖、消費者行為

ABSTRACT

In this paper, we follow the model of interpurchase to get the heterogeneity across customers. Then we use mixture model to segment customers into three groups as super-active, active and inactive states. Both models are solved by hierarchical Bayes via Markov Chain Monte Carlo method. At last, we employ X chart based on the density of active component to monitor consumer behavior. Control charts have been widely used in industry to control the quality of products. They provide an economical timing to launch corrective actions. The information on X chart includes interpurchase times and recency for an individual customer.

Keywords: Hierarchical Bayes, Finite Mixture Model, Heterogeneity Across Customers,

壹、緒論

近幾年來網際網路的盛行創造了許多的商業機會，像是以網際網路為主體的線上公司來說，由於交易環境的快速成長，創造了一個全新的商業管道，使得網路上的商業行為日漸普及。根據經濟部技術處委託資策會 ECRC-FIND 進行的「我國網際網路用戶數調查統計」，截至 2002 年 6 月底為止，我國上網人口達 808 萬人，連網普及率為 36%，比前一年同期成長了 10%。亞洲資訊資源公司在 2000 年針對網路人口進行一項調查，調查中指出在 2000 年中，中國大陸線上消費者平均每個人花費 887 元。對於以網際網路為主體的線上公司來說，由於交易環境的快速成長，創造了一個全新的商業管道，使得網路上的商業行為日漸普及。

線上公司的行銷手法不同於傳統企業，傳統企業的封閉式行銷手法有區域性的顧客，顧客的流動性較低，會有穩定的長期顧客。線上公司雖然有較廣泛的顧客群，但技術門檻低，當第一家線上公司成立時，相同性質的線上公司會很快的成立，顧客的選擇增加，相對的流動性較高，因此如何透過顧客關係管理 (Customer Relationship Management, CRM)，使顧客從短期消費者成為長期消費的主要顧客，就成為線上公司最重要的工作了。很多研究都試著預測出消費者行為和下次的交易時間，而我們企圖

從管理面來提供另外一個觀點。本研究中我們以某 KTV 網站所收集的顧客人口特徵和消費記錄資料為分析對象，這網站每天有超過 1 千個人上站，經營者比較注意顧客狀態的變動，尤其是從合理狀態轉變為危險狀態的顧客。當顧客從合理狀態轉變為危險狀態時，線上公司應該推出一些促銷方案或新的服務來刺激顧客繼續上站。Allenby, G. M., Leone, R. P., and Jen, L. (1999) 以層級貝氏 (Hierarchical Bayes) 推估的交易間隔模式用來表示顧客消費行為是相當有效的。另外，使用管制圖來監控消費者的狀態是相當有效和經濟的。管制圖不只能夠顯示出歷史資料，還能發現當顧客轉變到較差狀態前的警示現象。在本研究中，我們結合了這兩種方法在一起，而且提出一個有效和經濟的程序來判斷顧客的狀態。

Allenby et al. (1999) 所提出的交易間隔時間模式是以一般化伽瑪 (Generalized Gamma, GG) 分配來模擬顧客的消費間隔時間 (Interpurchase time)。

$$t_{ij} \sim GG(\alpha, \lambda_i, \gamma) = \frac{\gamma}{\Gamma(\alpha)\lambda_i^{\alpha\gamma}} t_{ij}^{\alpha\gamma-1} e^{-\left(\frac{t_{ij}}{\lambda_i}\right)^{\gamma}} \quad (1)$$

上式的 t_{ij} 表示第 i 個顧客在第 j 次的消費間隔時間，參數 γ 和 α 建立分配的弧形 (Shape)，參數 λ_i 為分配的衡量 (Scale) 參數，為了模擬顧客之間的異質性，指定 λ_i 的先驗分配為反一般化伽

瑪 (Inverse generalized gamma, IGG) 分配

$$\lambda_i \sim IGG(\nu, \theta, \gamma) = \frac{\gamma}{\Gamma(\nu)\theta^{\nu\gamma}} \lambda_i^{-(\nu\gamma+1)} e^{-\left(\frac{1}{\theta\lambda_i}\right)^\gamma} \quad (2)$$

使用一般化伽瑪分配來推估顧客的消費行為是相當有效的，因為它是一個相當具有彈性的分配，它所包含的特殊分配有指數分配 $GG(1, \cdot, 1)$ 、gamma 分配 $GG(\alpha, \cdot, 1)$ 、Erlang 分配 $GG(2, \cdot, 1)$ 和 Weibull 分配 $GG(1, \cdot, \cdot)$ 。在 Allenby et al. (1999) 的研究中，一般化伽瑪分配對於資料的估計較伽瑪分配更為合適。一般化伽瑪分配的期望值為

$$E(t_{ij}) = \frac{\Gamma(\alpha + 1/\gamma)}{\Gamma(\alpha)} \lambda_i \quad (3)$$

本研究的目的是了解每位顧客所處的狀態以及是否有狀態的變動。異質性分析是我們的第一個步驟，參數 λ_i 代表顧客間的異質性，我們假設 λ_i 為反一般化伽瑪分配，期望值為 $E(\lambda_i) = \frac{\Gamma(\nu - 1/\gamma)}{\Gamma(\nu)\theta}$ 。

第二個步驟將顧客分成合理的群數，由於針對顧客的異質性很難使用一個分配將顧客分群，因此我們透過混合模式將顧客分群。在混合模式的結合上，我們透過貝氏方法並使用 Dirichlet 決定各個分配所佔的比例。

$$\phi(x_i) \sim \sum_{j=1}^c \theta_j f_j(x_i | \alpha_j), \quad i = 1, 2, \dots, n \quad (4)$$

上式的 $f_j(x_i | \alpha_j)$ 是一個參數已知的分配，如常態分配、指數分配和伽瑪分配。 $\theta_j, j=1, \dots, C$ 為各個分配所佔的比

例，其總合為 1。在本研究中，我們假定所有顧客的交易間隔時間被分為三個群組，理想狀態 (super-activeness)，合理狀態 (activeness) 和危險狀態 (inactiveness) (Allenby et al., 1999)。這樣的顧客分類在直接行銷 (direct marketing) 是有意義的。最後我們用管制圖來監控顧客消費行為的變動，管制圖通常被用來管制作業流程或產品的品質，因為它是容易被了解和使用的。顧客的消費行為也能夠在低成本的情況下受到有效的控制。

馬可夫蒙地卡法 (Markov chain Monte Carlo, MCMC) 是一種以模擬參數的先驗分配 (posterior distributions) 的方法來估計未知參數 (Gelfand & Smith, 1990)，在本研究中，我們使用 Winbugs (Windows Bayesian inference Using Gibbs Sampling) 這個套裝軟體來估計平均交易間隔時間和混合模式的參數，這套軟體是由英國醫學學院 (Imperial College School of Medicine) 所發表的，它是使用 MCMC 中的吉氏抽樣 (Gibbs Sampling) 來估計參數。

在下一節我們將討論資料特性和一些敘述統計。在第三節中，我們將使用 Allenby et al. (1999) 的交易間隔時間的模式來估計顧客異質性。在第四節中，我們將利用混合模式來判斷的顧客所處的狀態。在第五節將討論顧客狀態和最近時間的結合。在第六節中，我們將應用管制圖來監控顧客交易間隔時間的消費行為。第七節為我們的結論。

貳、背景

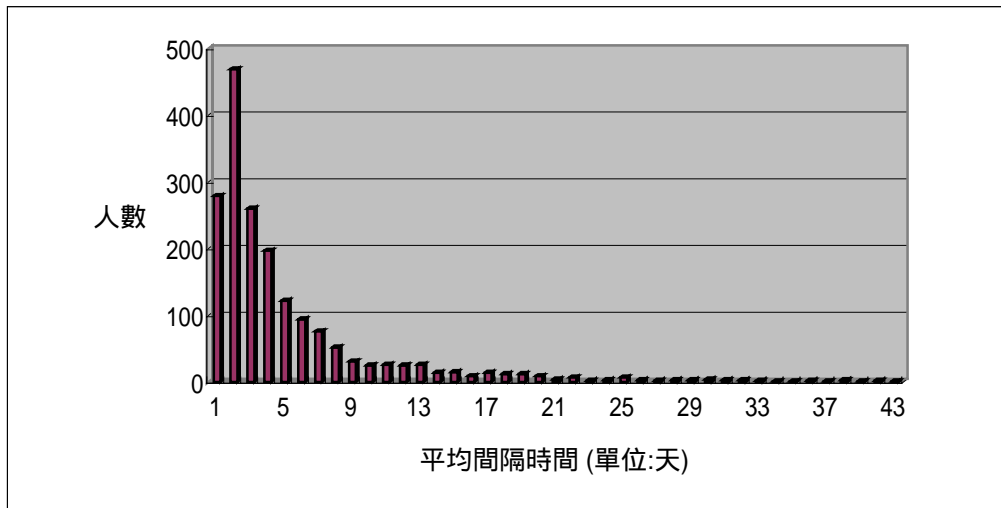


圖 1 顧客平均間隔時間的條狀圖

本研究資料是來自一家線上卡拉OK的公司，他們提供顧客點歌、伴唱和錄製歌曲的服務，固定每個月跟會員收取月費。當顧客登錄會員時，記錄顧客的基本資料，並記錄顧客每次上網的時間和點歌的歌曲。在與這家公司的經營者和分析師的討論下，由於顧客消費週期短，所以決定以天為單位來計算顧客交易間隔時間，即無論顧客在同一天上網點歌多少次都當作同一筆交易記錄。

我們取得三個月的資料，自 2002 年 8 月 16 日到 11 月 15 日的顧客交易資料，我們捨棄了交易記錄少於三次的顧客，實際之樣本數為 1825 位顧客，總點歌次數為 10,548 次。

圖 1 為 1825 位顧客的平均間隔時間。平均間隔時間的範圍從 1 天到 42.5 天。由於有些顧客幾乎是每天上站點歌，有些顧客是十幾天或幾十天才上站一次，這樣的資料顯示出顧客彼此之間的異質性很大。

圖 2 為 1825 位顧客的點歌次數。在圖中我們可以發現有許多顧客的點歌次數相當的少，在傳統統計中的參數估計會產生偏差，而層級貝氏可以提升少量樣本下估計參數的準確性 (Congdon, 2001)。

圖 3 為一位顧客的交易間隔時間的時間序列。圖中有些間隔時間點特別突出，這些點代表該顧客於那段時間是較不活躍的，這些點也是網站經營者所希望避免掉的。會發生這些點的原因很多，如果有一個服務人員能夠接觸這些顧客，提供一些他們所需要的服務或推出特別的促銷來刺激顧客，將能夠減少顧客轉變成危險狀態，這也是本研究的主要目的。

參、層級貝氏推估顧客異質性

所謂的層級貝氏 (Hierarchical Bayes) 是使用兩層的先驗分配假設來估

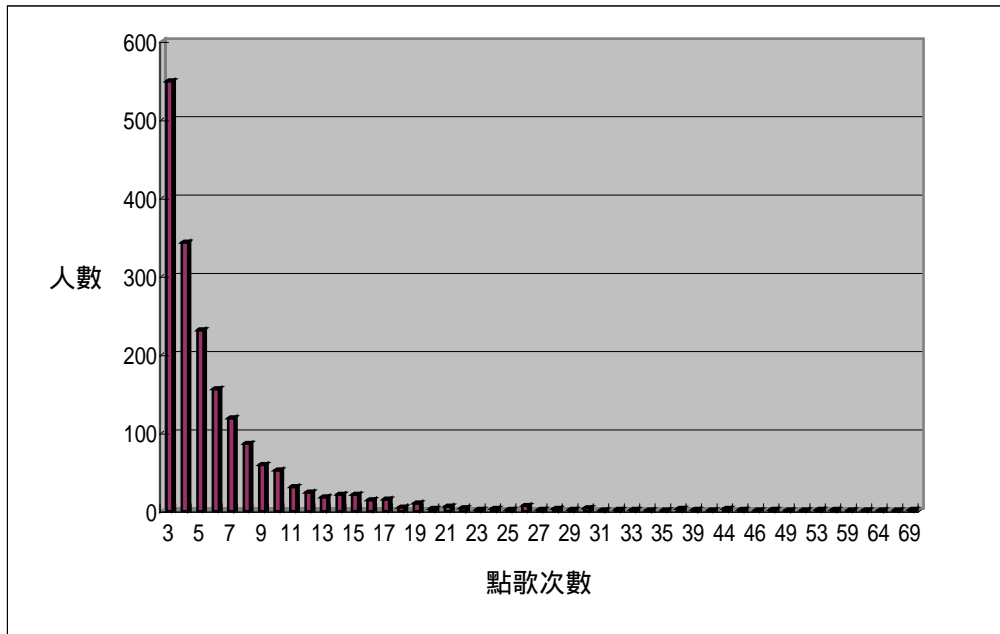


圖 2 顧客上站次數的條狀圖

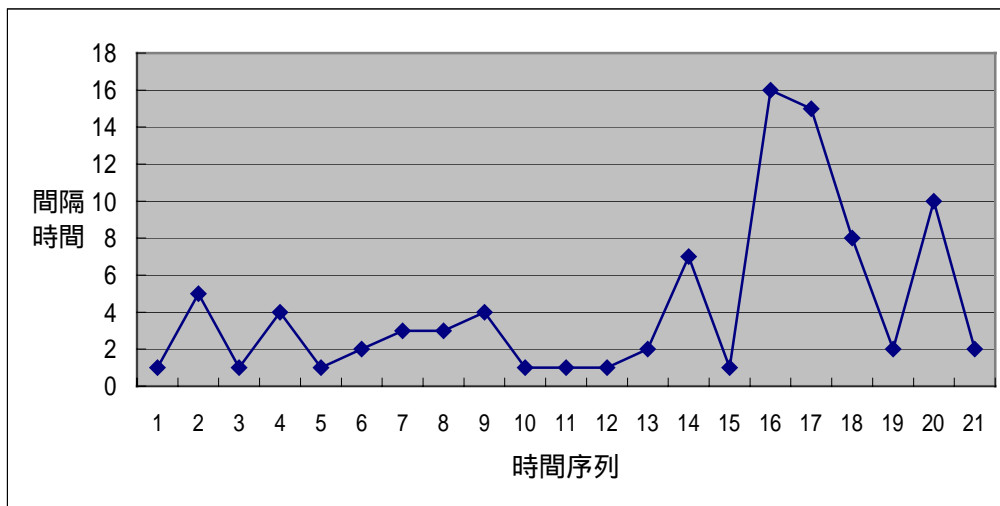


圖 3 顧客間隔時間的時間序列圖

計參數。舉例來說，估計常態群體的平均值，在主要抽樣單位 (The Primary Sampling Unit, PSU) 層，我們假設每位個體的成分效用 (part-worth) 即個體實際觀測值所形成群體之參數來自常態分配，表示為

$$\beta_i \sim N(\mu, \Lambda) \tag{5}$$

其中 i 是一個個體的成分效用 (part-worth), 平均值向量 μ 代表個別個體參數所形成群體的平均值，常態分配的共變異矩陣 包括個別個體的異值性

和個別個體間在成分效用 (part-worth) 上的相關性。在主抽樣的第二階段的抽樣 (The Secondary Sampling Unit, SSU) 上, 我們假設個別個體選定案例的機率分配是邏輯複常態 (multinomial logit), 其第 i 個個體選擇第 k 個事例的機率為

$$p_k = \exp(x_k \beta_i) / \sum_j \exp(x_j \beta_i) \quad (6)$$

其中 p_k 是在給定參數 β_i 下抽出第 k 個實際事例的機率, 變數 x_j 是描述第 j 個事例所得到的值向量。於此例中, 所需估計的參數包括每個個體的成分效用 (part-worth) 的平均參數向量 μ 和 σ 。

我們使用層級貝氏的程序並以 Allenby et al. (1999) 所提的模型來處理顧客間的異質性, 每個顧客的交易間隔時間是服從 Generalized Gamma 分配, 如下:

$$t_{ij} \sim GG(\alpha, \lambda_i, \gamma) = \frac{\gamma}{\Gamma(\alpha)\lambda_i^{\alpha\gamma}} t_{ij}^{\alpha\gamma-1} e^{-\left(\frac{t_{ij}}{\lambda_i}\right)^\gamma} \quad (7)$$

上式的 t_{ij} 表示第 i 個顧客在第 j 次的消費間隔時間, 為了模擬顧客之間的異質性, 指定 λ_i 為反一般化伽瑪 (Inverse Generalized Gamma, IGG) 分配,

$$\lambda_i \sim IGG(\nu, \theta, \gamma) = \frac{\gamma}{\Gamma(\nu)\theta^{\nu\gamma}} \lambda_i^{-(\nu\gamma+1)} e^{-\left(\frac{1}{\theta\lambda_i}\right)^\gamma} \quad (8)$$

然後採用層級貝氏模式進行模型推估時, 將 α 、 ν 和 θ 參數的先驗分配指定如下:

$$\alpha \sim U(0, +\infty) \quad (9)$$

$$\nu \sim U(0, +\infty) \quad (10)$$

$$\theta \sim IGG(a, b, r) \quad (11)$$

上式中的 a 和 b 可自行決定 (Allenby et al., 1999), 參數 γ 設定為 1.0。在我們的研究中指定 a 和 b 都為 10。我們以使用 Winbugs 執行了 2000 次的吉氏抽樣, 取最後的 1000 次的平均做為參數估計值, 其中參數的估計值 $\alpha=1.355$, $\nu=2.354$ 和 $\theta=5.112$ 。在表 1 中列出部份顧客層級貝氏估計的平均間隔時間的期望值

$$E(t_{ij}) = \frac{\Gamma(\alpha+1/\gamma)}{\Gamma(\alpha)} \lambda_i, \text{ 在表 1 中我們可得知}$$

層級貝氏的調整機制, 當樣本數不夠時, 如帳號 15536 其點歌次數只有 2 筆, 其樣本平均的代表性不足, 因此層級貝氏的估計值會偏向母體平均數。當樣本數足夠時, 如帳號 36268 其點歌數有 43 筆資料, 樣本平均數的代表性較高, 則層級貝氏的估計值偏向樣本平均數。

肆 透過MCMC方法的混合模式

Evans, Guttan, and Olkin (1992) 使用層級貝氏法在無先前資訊下分析雙元素常態分配的混合模型。Carlin and Chib (1995) 在元素個數不確定下配置合適的混合模型。Allenby et al. (1999) 的研究中, 使用 3 個對數型態的延遲平均間隔時間 (lagged interpurchase time) 於 3 元素的混合機率函數中, 來區隔理想狀態、合理狀態和危險狀態的消費行為。在本研究中我們也是使用混合模式來區分顧客的消費行為。我們將所有顧客的交易間隔時間混合在一起, $t = \{t_{ij}\}$ 是所有顧客的交易間隔時間的集合, q_k 表示混合模式中每一個分配所佔的比例, 並且

$$\sum_k q_k = 1。$$

表 1 層級貝氏估計的平均點歌間隔時間

帳號	點歌間隔次數	平均點歌間隔時間	層級貝氏模型估計的平均點歌間隔時間
15536	2	41.5	29.4
49825	2	37.5	26.5
12657	3	28.33333	22.31
32775	6	10.33333	9.637
25745	11	7	6.858
86395	14	2.357142	2.5
85397	16	1.5625	1.757
80240	21	1.857142	2.014
61302	34	2.441176	2.497
36268	43	2	2.084

$$t \sim \sum_k q_k GG(\alpha, \lambda_k, \gamma) \quad (12)$$

各個參數的先驗分配分別為

$$\lambda_k \sim IGG(\nu, \theta, \gamma) = \frac{\gamma}{\Gamma(\nu)\theta^{\nu\gamma}} \lambda_k^{-(\nu\gamma+1)} e^{-\frac{1}{\theta\lambda_k}^\gamma} \quad (13)$$

$$\alpha \sim U(0, +\infty) \quad (14)$$

$$\nu \sim U(0, +\infty) \quad (15)$$

$$\theta \sim IGG(a, b, r) \quad (16)$$

$$q_k \sim Dirichlet(p_k) \quad (17)$$

在本研究中我們將顧客分成三個群組，所以公式(12)會變成

$$t \sim \sum_{k=1}^3 q_k GG(\alpha, \lambda_k, \gamma) \quad (18)$$

我們同樣設定 $\gamma = 1.0$ ，經過 Winbugs2000 次的抽樣，在表 2 中列有參數估計的結果。一般化伽瑪在 $\gamma = 1.0$ 的

期望值為 $\frac{1}{\lambda_k}$ 。三個分配的期望值分別為 1.244，5.141 和 25.532 天。我們假設危險狀態的期望值大於合理狀態的期望值，合理狀態的期望值大於理想狀態的期望值。所以理想狀態的期望值為 1.244 天，合理狀態的期望值為 5.141 天，危險狀態的期望值為 25.532 天。因此在本研究的資料中理想狀態的顧客幾乎每天上網，合理狀態的顧客大約 1 個禮拜上網一次，而危險狀態的顧客接近 1 個月才上網一次。

圖 4 為顧客交易間隔時間的直方圖，圖 5 為三個分配（實線表示）和四個分配（虛線表示）的配適機率分配函數。由圖 4 和圖 5 的對照中，我們可以發現在本研究中，增加分配的數量並不會讓分配的更符合原始資料。

我們計算每一位顧客的平均交易間隔時間，代入這三種分配的機率密度函數，求出每一位顧客在三種分配的機率

表 2 三個分配的混合模式參數

參數	估計值	標準差
α	6.564	0.1198
λ_1	0.19	0.004035
λ_2	0.7832	0.01881
λ_3	3.904	0.1067
q_1	0.6531	0.005105
q_2	0.2725	0.004765
q_3	0.07439	0.00289
ν	1.77	0.7063
θ	0.9422	0.2573

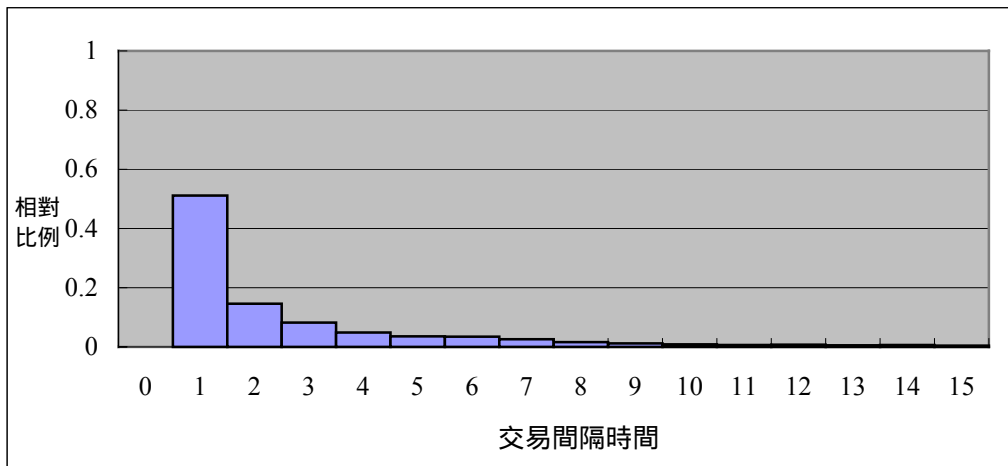


圖 4 交易間隔時間圖

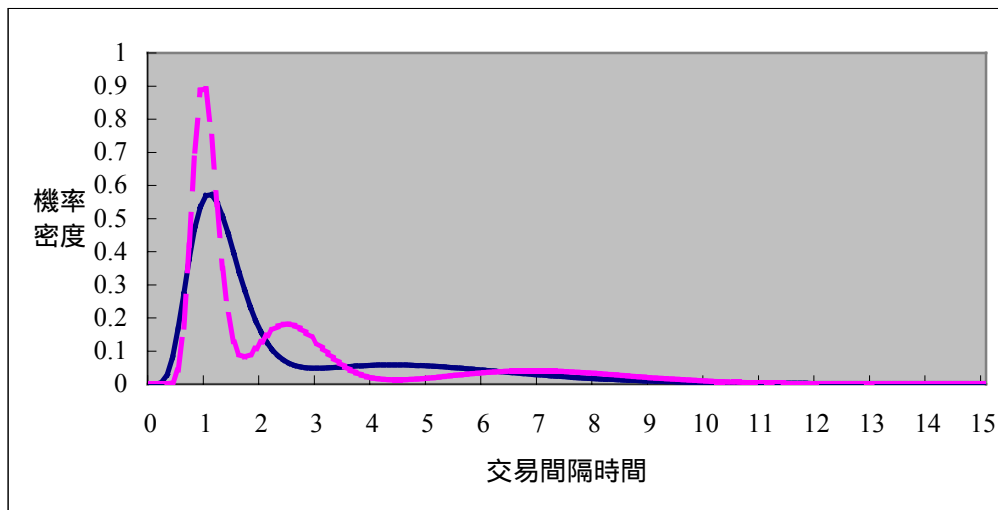


圖 5 機率分配圖 (實線為三個分配、虛線為四個分配)

表 3 依平均間隔時間判斷的顧客狀態

帳號	層級貝氏平均間隔時間($\alpha\lambda_i$)	狀態一的機率密度	狀態二的機率密度	狀態三的機率密度	屬於何種狀態
81203	3.985	0.000286476	0.208194193	0.000320337	2
96261	1.614	0.492741045	0.028132652	3.84908E-06	1
35883	1.804	0.336704628	0.040999452	6.80998E-06	1
89403	3.895	0.000405142	0.205671959	0.000288682	2
67157	2.35	0.082821941	0.088906915	2.57827E-05	2
59583	3.855	0.000472165	0.204367459	0.000275375	2
62701	22.88	3.09653E-43	1.15925E-07	0.042353756	3
72310	5.31	1.32464E-06	0.189400782	0.001126814	2
72026	6.865	1.54306E-09	0.108581618	0.003158717	2
71438	2.524	0.049320407	0.10593851	3.66924E-05	2

表 4 依平均間隔時間判斷各個狀態的顧客人數

狀態	理想狀態	合理狀態	危險狀態	總數
人數	324(17.75%)	1316(72.1%)	185(10.15%)	1825

密度，並以機率密度較大的代表顧客所處的狀態。在表 3 中我們以 1、2 和 3 分別代表顧客的三種狀態，其中帳號 96261 的顧客在狀態一的機率密度最大，所以他是屬於理想狀態的顧客。依此同理，帳號 81203 和 62701 則分別屬於合理狀態和危險狀態。

在表 4 中，我們可以發現在 1825 位顧客中有 17.75% 的顧客為理想狀態，72.1% 的顧客屬於合理狀態，最後 10.15% 的顧客被判定為危險狀態。這些比例所代表的是顧客過去行為所呈現的狀態，並不呈現顧客現在所處的狀態。在消費行為的觀點上，顧客最後一次消費距離現在的時間的間隔時間是很重要的，最近交易時間 (recency) 的長短與顧客消

費行為有強烈的相關性 (Hughes, 1994)。在下一節中，我們考慮最近交易時間來重新判斷顧客的所處的狀態。

伍、考慮最近交易時間的顧客狀態

顧客的價值在於顧客的關係強度，雖然在關係強度的研究中，行為和心理層面的關聯是可以分開研究的，但在市場策略上，這兩種層面都是值得研究。因為行為資料較容易獲得且成本較低，本研究將採取行為資料作為顧客行為研究的初期探索。最近交易時間 (Receny)

表 5 考慮最近交易時間的顧客狀態判斷

帳號	層級貝氏平均 間隔時間 ($\alpha\lambda_i$)	點歌 間隔次 數(n)	最後點歌時 間間隔(R)	R'	狀態一 的機率密度	狀態二 的機率密度	狀態三 的機率密度	狀態三 的累積機率	屬於 何種狀態
81203	3.985	11	10	6.84929	1.65E-09	0.109379	0.003131	0.995785	2
96261	1.614	11	0	1.614	0.492741	0.028133	3.85E-06	0.999999	1
35883	1.804	45	3	1.87875	0.284771	0.046713	8.37E-06	0.999997	1
89403	3.895	7	0	3.895	0.000405	0.205672	0.000289	0.999803	2
67157	2.35	2	74	72.1145	5.3E-153	3.44E-32	8.39E-05	0.000456	4
59583	3.855	3	76	73.2603	1.4E-155	8.7E-33	6.83E-05	0.000369	4
62701	22.88	2	26	25.7771	1.43E-49	5.57E-09	0.039147	0.442126	3
72310	5.31	3	60	57.3957	6.6E-120	1.4E-24	0.001022	0.006107	3
72026	6.865	2	73	71.2364	5.1E-151	9.87E-32	9.82E-05	0.000536	4
71438	2.524	5	79	74.4479	3E-158	2.09E-33	5.51E-05	0.000296	4

是一個相當重要的行為指標，其他還有交易次數 (Frequency) 和交易金額 (Monetary)，這三個指標合稱為 RFM 模型。最近交易時間也是一個時間序列的資料，因此我們將每一個顧客的平均交易間隔時間 (i) 和最近交易時間 (R) 結合成 R'

$$R' = w_1 * \alpha\lambda_i + w_2 * R \quad (19)$$

在我們的資料中，我們設定平均間隔時間的權重為 $w_1 = \frac{n}{n+R}$ ，最近交易時間的權重為 $w_2 = \frac{R}{n+R}$ ，這權重是隨著交易次數 n 的大小和最近交易時間的長短來變動的 (此權重的設定是依試誤實驗後所得)，使用者可以依據自己的資料型態來改變權重公式。在表 5 中，我們利用 R' 再次計算三個分配個別的機率密度，倒數第二個欄位是在危險狀態中 R' 的累積機率。在這裡有兩個步驟來判斷顧客的狀態，首先比較每一位顧客的三種狀態

機率密度的大小，並以機率密度較大的代表顧客所處的狀態。當顧客的狀態以狀態三的機率密度最大時，就進入第二階段的判斷，在危險狀態 (即狀態三) 的機率分配下，計算大於 R' 的累積機率，若大於臨界值 (0.001) 則判斷顧客處於狀態三，若小於臨界值 (0.001) 則判斷顧客處於流失狀態，在表 5 中以數字 4 來表示。

表 6 為全部顧客在四種狀態中所佔的數量和比例。在表 6 中我們有 55% 的顧客是處於流失狀態，在與網站公司的管理者討論過後，發現這些被判定為流失的顧客，其點歌時間大多是 2002/8/16 至 2002/9/15 這段時間，由於這個卡拉 OK 網站的顧客大多都是學生，這段時間還在暑假期間，當暑假結束之後，這些顧客也停止上站。這是一個合理又有趣的現象，也顯示出如何在淡季提高顧客上網消費是這個卡拉 OK 網站所要注意的

表 6 考慮最近交易時間在四種狀態的人數

狀態	理想狀態	合理狀態	危險狀態	流失狀態	總數
人數	61(3.34%)	284(15.56%)	464(25.43%)	1016(55.67%)	1825(100%)

表 7 顧客狀態轉換

狀態轉變	1→1	1→2	1→3	1→4	總數
數量	55(3%)	16(0.88%)	18(1%)	235(13.9%)	324
狀態轉變	2→1	2→2	2→3	2→4	
數量	6(0.3%)	237(13%)	296(16.2%)	777(42.6%)	1316
狀態轉變	3→1	3→2	3→3	3→4	
數量	0(0%)	31(1.7%)	150(8.2%)	4(0.2%)	185
總數	61(3.3%)	284(15.6%)	464(25.4%)	1016(55.7%)	1825

事項。

在表 7 中顯示出顧客從表 3 (未考慮最近交易時間的狀態判定) 的狀態轉變到表 6 (考慮最近交易時間的狀態判定) 的差異性的人數統計, 符號 1→1 表示顧客在考慮最近交易時間下依然維持在理想狀態。而從低階狀態轉變為高階狀態是不可靠的, 因為資料擷取時間是取決於使用者, 而不是自然發生的。從高階狀態轉變為低階狀態是可信的, 因為這真實的狀態是等於或更糟的狀態變動。在括弧中的數字是該狀態變動在 1825 位顧客中所佔的百分比。在表 7 中, 我們能夠發現一些有趣的現象, 第一, 有 24.2% 的顧客維持在原本的狀態, 第二, 在理想狀態和合理狀態有高達 73.7% 的顧客轉變為危險狀態或流失狀態, 而原本處於危險狀態的顧客轉變為流失顧客的卻只有 0.2%。因此如何即時的監控顧客消費行為變成該網站公司的重要目標。在本研究中我們提供了簡單有效的

管制圖來及時監控, 當顧客要轉變為低階狀態時, 能夠提早發現。我們將在下一節來討論。

陸、顧客狀態的管制圖

以歷史觀點來看, 統計管制的技術多半用在製造程序中, 然而近年來, 其他程序如服務業已開始用這些簡易有效的統計管制技術。指數加權移動平均管制圖 (exponentially weighted moving average control chart, EWMA 圖) 最早於 1959 年被 Roberts 提議出來, 後來 Wortham and Ringer (1971) 也實際應用在金融和管理上的控制系統上。累計總和管制圖 (cumulative sum control chart, CUSUM 圖) 對於查覺在 0.5σ 和 2.5σ 之間的小變化特別有用。對於大於 2.5 的變動, Shewhart-type 圖 (這一系列的管制圖被 Shewhart 博士在 1924 年發明) 則

是較佳的選擇。

Shewhart (1931) 建議利用線圖來觀察資料，他指出有二個變異的原因，第一個稱為機遇原因 (chance causes)，這些事件會導致資料有相當小的變化，這些小變化的發生並不重要，甚至可以不用去更正它；第二個變異的原因稱為非機遇原因 (assignable causes)，這些原因可以從資料的變動中觀察出來，而且它所導致的變動也較機遇原因大。Shewhart 博士所開發的管制圖要假設機遇原因的變異來自於常態分配其在分配上的設定管制界線，任何觀察值落在這個管制界線之外，代表該觀察值發生了非機遇原因所產生的變異。因為是以觀察值的發生順序來觀察，所以可以容易的觀察出趨勢和其他不自然的現象。

在我們的研究中，交易間隔時間是一個顧客品質的指標。對每一位顧客，我們每次只能取得一筆觀測資料，一次只有一筆資料的管制圖稱為 X 圖 (X chart)，而 X 圖通常搭配移動全距圖 (Moving-Range chart) 來使用。在本研究中，每位顧客的資料一天發生一次，觀察值的計算為交易間隔時間，我們以最近交易時間為變異值，因此我們將交易間隔時間和最近交易時間的資料一併繪製在 X 圖上，並不搭配移動全距圖來使用。一般而言，經營者比較在意顧客的消費頻率降低，尤其是當顧客降為危險狀態時。因此，我們選擇合理狀態為管制圖的分配，中心線是合理狀態分配的期望值 5.139，而我們只設定上管制界線為 $\mu + 3\sigma = 11.16$ 在圖 6 中有三位消費者的 X 圖，橫軸為時間序列，縱軸為間隔時間。當顧客在第二次上站時，會在 X 圖上畫上第一個交易間隔時間點，隔天

若該顧客沒上站，將累積為最近交易時間。在圖中交易間隔時間是以三角形來表示，最近交易時間以線條來表示，一旦超出管制界線，將是個嚴重的警訊。另外，即使不超出管制界線有時也會呈現一些警訊，如顧客二在的管制圖在連續出現數個大幅躍動的現象，之後開始超出管制界線。因此我們找出一些特殊的現象，一旦發生這些現象，可以提供經營者應變的方法。在這裡我們整理一些除了超出管制界線以外的警示現象。

1. 幅度的變動— 在圖 6(a)和(b)中顯示了間隔時間的幅度變動。在圖 6(b)中，該顧客原本的交易間隔時間的很短，之後交易間隔時間大幅的增加。
2. 不穩定的循環— 在圖 6(c)中顯示了不穩定的循環。在圖 6(c)中，顧客在連續的上網交易後會有一段長時間不在上網，可是這些集中的天數並不一致。
3. 上升的趨勢— 在圖 6(c)的後半段，顧客的交易間隔時間有越來越長的趨勢，並出現鋸齒狀的現象。

通常在兩種情況下我們會修正及重新計算管制界線，第一個是改善作業程序已穩定步上常規，且無出現超出管制界線的現象。在這種情形下，重新修定會提供群體標準差較佳的估計。第二種發生重新修訂管制界線是當某些型態如上升或循環型態被辨識出來並予以去除時，一旦非機遇原因被發現且修正後，並確定不再影響未來程序時，管制界線即可重新計算。此兩種重新計算管制界線的法則都將作為顧客品質，管制圖重新計算管制界線的法則。除此之外，網路交易的週期非常短，很多型態可能已經發生但卻沒有即時發覺。

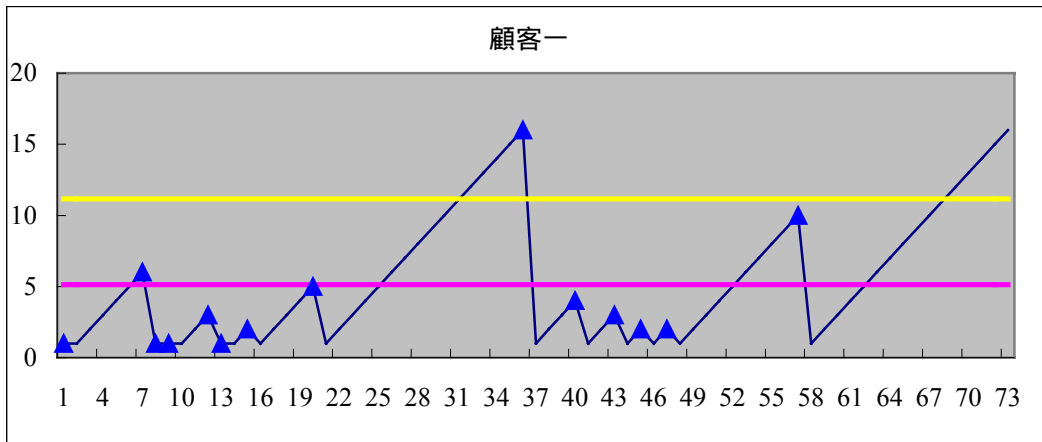


圖 6(a) 顧客一的管制圖

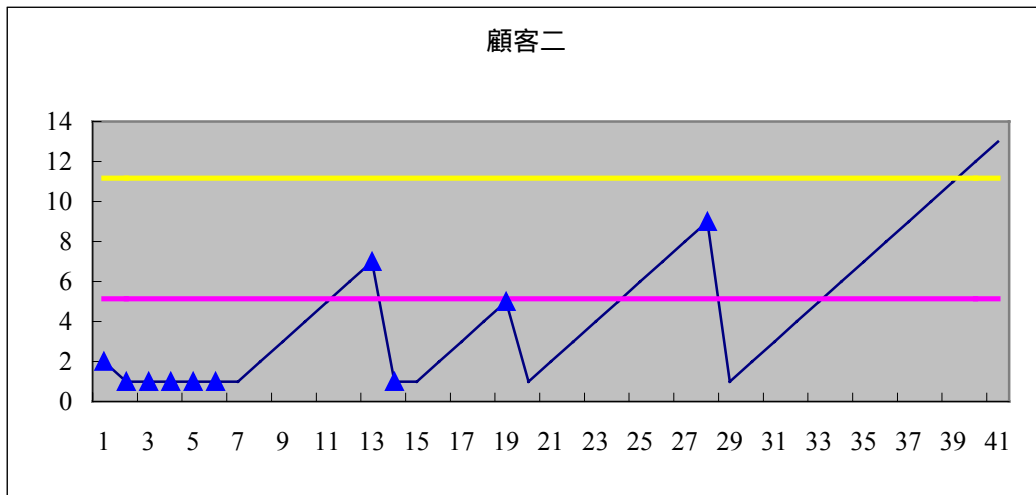


圖 6(b) 顧客二的管制圖

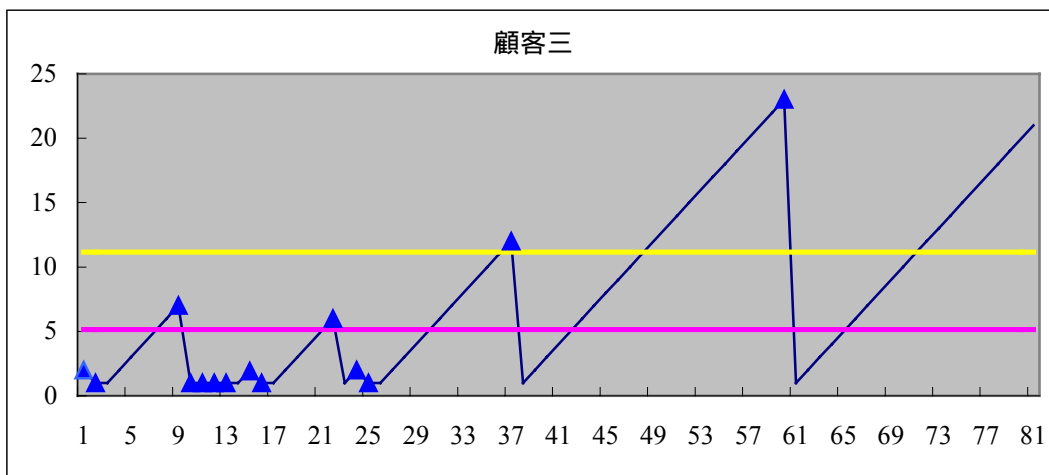


圖 6(c) 顧客三的管制圖

在我們的分析中，顧客從合理狀態轉變為危險狀態的平均時間大約為一個月，所以我們建議，即使沒有新的操作或非機遇因素的刪除，還是得重新計算第四節所提到的混合模式，重新取得最近合理狀態的分配參數值，以修正管制界線和中心線。

柒、結論

由於個別顧客的交易資料相當的少（如圖 2 大部份顧客僅上站 3 次），很難使用傳統的統計方法來估計個別顧客交易間隔時間的平均值，層級貝氏方法可以藉由母體平均值來修正樣本平均值。當樣本數多時會偏向樣本平均值，反之，當樣本數少時會偏向母體平均值。本研究使用層級貝氏方法來分辨顧客的異質性，網站每天有超過 1000 位顧客上站，相對於交易間隔時間，管理者比較關心顧客的消費狀態。我們使用混合模式將顧客分成理想狀態、合理狀態和危險狀態等三種。最近交易時間對於顧客現在的狀態是一個重要的指標，我們結合交易間隔時間和最近交易時間，並給定適當的權重來判斷顧客現在所處的狀態。最後意外的發現有 50% 的顧客會在淡季的時候流失掉。因此我們需要一個便宜簡單的方法來監控顧客的行為，當顧客出現從合理狀態轉變成危險狀態時，能夠即時的顯現出來。本研究使用管制圖的方式來監控顧客行為，我們也提出了一些除超出管制界線的警示型態，幫助判定顧客行為是否有脫離合理狀態進入危險狀態，至於管制圖應用之成效，將於未來研究中作更進一步的探

討。

參考文獻

1. Allenby, G. M., Leone, R. P., & Jen, L. (1999). A Dynamic Model of Purchase Timing With Application to Direct Marketing, Journal of the American Statistical Association, 94(446), 365-373.
2. Carlin, B. P., & Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Method, Journal of the statistical Society, Ser. B, Vol.57, 473-484.
3. Congdon, P. (2001). Bayesian Statistical Modelling, pp. 1-8, John Wiley & Sons, West Sussex, England.
4. Evans, M., Guttan, I., & Olkin, I. (1992). Numerical aspects in estimating the parameters of a mixture of normal distributions, J. Comput. Graph. Statist., 1, 351-365.
5. Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-Based Approach to Calculating Marginal Densities, Journal of the American Statistical Association, 85, 398-409.
6. Hughes, Arthur M. (1994). Strategic Database Marketing. Chicago : Probus Publishing.
7. Roberts, S. W. (1959). Control Chart Tests Based on Geometric Moving

Averages, Technometrics, 1(3),
239-250.

8. Shewhart, W. A. (1931). Economic control of quality of Manufactured Product, 1931, Van Nostrand-Reinhold, A print of this classic work has been published and is available from the American Society for Quality, Milwaukee, New York.
9. Wortham, A. W., & Ringer, L. J. (1971). Control Via Exponential Smoothing. The Logistic Review, 7(3), 33-40.

2004 年 02 月 13 日收稿

2004 年 03 月 04 日初審

2004 年 08 月 20 日複審

2004 年 09 月 03 日接受