Life-span distributions of supermarket products

Hiromichi UENO, Tsutomu WATANABE and Misako TAKAYASU

Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259-G3-52 Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan and Institute of Economic Research, Hitotsubashi University 2-1 Naka, Kunitachi, Tokyo 186-8603, Japan

E-mail: ueno@smp.dis.titech.ac.jp

Abstract. We have analyzed the lifetime distributions of more than 0.7 million products sold across approximately 400 Japanese supermarkets. The distributions are well approximated by an exponential function for products with lifetimes longer than 1000 days, implying that the manufacturers’ decisions about whether to continue production are purely random. However, the distributions tend to deviate from an exponential distribution for products with lifetimes shorter than 1000 days. Specifically, the distributions for food products exhibit a quicker decay in a short time scale, suggesting the existence of competing products during the initial stages of the product lifecycle. On the other hand, the distributions for toiletry products exhibit a slower decay in a short time scale.

1. Introduction

Developments in information technology have enabled the storage of large volumes of real-time economic data on computer systems. In the field of econophysics, such data are analyzed using statistical physics methods, and important results have already been established for financial markets and company’s financial data[1][2]. Recently, data related to various types of human activities have been stored, and these have become the latest focus of attention in the field of econophysics. For example, Sornette et al. conducted an extensive study of the time series of books sales by using a unique database obtained from Amazon.com[3][4].

Recently, point of sale (POS) data has also been analyzed in this field. Most retailers employ a POS system for the purpose of efficient inventory management, accounting, and marketing. Such systems maintain a real-time record of the number of sales of products and the price of each. By using this system, retailers can monitor all purchasing activities in a retail store. Since POS data contains detailed information, it can be used to analyze economic activities from various points of view. Mizuno et al. studied the repeat purchase behavior of consumers by using POS data collected from Japanese convenience stores[5][6]. Groot analyzed the fluctuations in product sales using sales data collected from Dutch supermarkets and observed that these fluctuations exhibit properties similar to those of a stock market[7]. Around 30 years ago, Montroll analyzed the product pricing in the catalogue of Sears, Roebuck and Company and observed that the price distribution follows a log-normal distribution[8].

In this study, we analyze the product lifetime distributions using POS data collected from Japanese supermarkets. It is thought that life span characterizes the object’s property. The lifetime distributions of various phenomena are typically represented by an exponential distribution or a Weibull distribution. For example, radioactive decay follows an exponential...
distribution; here, an element independently decays at a constant probability. On the other hand, the probability of fault occurrence in a machine, say an electron tube, follows the Weibull distribution[9].

We define the lifetime of a retail product as the period between the day when it first becomes available at a store (entry date) and the final day when it is sold (exit date). Manufacturers determine the entry and exit dates for a product according to their merchandising strategy. Therefore, the product lifetime influences the merchandising strategy and retail markets, and therefore, it is useful to analyze the product lifetime distribution. In the field of economics, the turnover of a product has been considered to be related to firms’ technological innovation, in particular product innovation, which is a key element of firm and industry dynamics. The results of this paper will be useful in the given field[10][11].

2. Data
In this study, we analyze the POS data (provided by Nikkei Digital Media Inc.) collected from Japanese supermarkets. This data comprises many records, each of which consist of 6 numerical codes, namely, the JAN (Japanese Article Number) code (a unique product identification code), genocode (a supplementary code of JAN code), store code, date of a sale, gross sales price of the day, and number of sales in a day.

This set of codes is unique to each product. The POS data analyzed by us contains information about approximately 1.4 million types of products available across approximately 400 Japanese supermarkets between March 1988 and April 2008. We analyzed approximately 3.7 billion records.

This POS data includes information about all commercial products with a JAN+genocode that are sold at supermarkets; this enables a comprehensive analysis of product lifetimes. It should be noted that this data does not contain information abut products without JAN codes, including fresh foods such as vegetables, meat, and fish.

3. Definition of product lifetime
We define the product lifetime as follows. First, we define the entry date as the day when a product first becomes available in a store, and the exit date as the day when the product is sold. The sales lifetime is then defined as the number of days between the entry and exit dates. Fig.1 shows a schematic diagram that explains this definition. Here, the entry date is the date when the product first becomes available at store 2 (circle), and the exit date is the date when the product is sold in store 5 (square).

To define the product lifetime more explicitly, we excluded some products from the data analysis; these either had entry dates earlier than the beginning of our observation period or had exit dates later than the end of our observation period. Fig.2 shows a schematic diagram of the case in which the entry date of a product was earlier than the beginning of our observation period. The entry date of product B is represented by the circle; however, this date is the first day of our observation period, and the real entry date is earlier than this date. To avoid such estimation errors for both the entry and the exit dates, we apply the following rules. Products whose entry and exit dates were within 364 days of our first and last observation dates, respectively, were excluded from the analysis.

The threshold period was selected as 364 days because in some cases, a store or a manufacturer stopped selling a product for a certain period before resuming its sale. We observed the distribution of such intermissions and found that the probability of them lasting longer than a year was practically zero.

After filtering the products according to the above rules, approximately 0.7 million products remained, and we analyzed their lifetimes.
4. Distribution of sales lifetime

We observe the cumulative distribution of the sales lifetime of all products that were analyzed, \( P(\geq t) \), that is, the probability of finding a product having a lifetime greater than \( t \) days. Fig.3 shows a semilog plot of the probability distribution of products having \( P(\geq t) \). It is evident that the curve can be approximated by a straight line except for the very short days and very long days. The dotted line represents the best-fit exponential function with an exponent of 0.001206, implying that the characteristic life span is 829 days (2.3 years).

The sharp decay in the distribution curve after around 6500 days (Fig.3) is caused by the limitation of the observation period to around 6500 days. If no such limitation is enforced, the exponential distribution is expected to extend longer.

To investigate the deviation from the exponential distribution in very short time scales, we plot the probability density function, that is, the derivative of \(-P(\geq t)\), for a short time scale, as shown in Fig.4. We can confirm the presence of several peaks around the 31-, 62-, 92-, 123-, 154-, 183-, and 369-day periods. This implies that there is a tendency of making decision on determining the period of sales in the unit of month or year.

![Figure 1. Schematic diagram explaining the definition of product lifetime on the basis of our sales data. The entry and exit dates are denoted by a circle and a square, respectively.](image1)

![Figure 2. Schematic diagram explaining exclusion of a product having an entry date earlier than our observation period. The real entry date of product B (circle 1) does not coincide with the first day of the observation period.](image2)

![Figure 3. Semilog plot of cumulative distribution of sales lifetimes of all products in Japanese supermarkets. The dotted line represents an exponential function with a characteristic time of 829 days.](image3)
5. Dependence of distribution on different product categories

Products are classified into two broad categories, i.e., food products and toiletries, and these are further divided into 36 subcategories such as beverages, alcoholic drinks, detergents, etc. To investigate the heterogeneity of lifetimes in each category, we analyze the lifetime distributions of each category. Fig.5 shows the cumulative lifetime distribution for detergents (dotted line) and bread (solid line) on a semilog scale. The former roughly follows an exponential function over the entire time scale until the observation threshold, while the latter does not follow an exponential function over a time scale shorter than approximately 1000 days.

A short time scale is one that has probability values $P(\geq t) > 0.5$, while a long time scale is one that has probability values in the range $0.1 \geq P(\geq t) > 0.01$. Fig.6 shows a graph of the combination of characteristic lifetimes for short and long time scales for all product categories. The horizontal (T1) and vertical (T2) axes represent the characteristic lifetimes for the short and long time scales, respectively. The plots on the guide line represent the case in which the two exponents coincide, i.e., when the distribution is approximated by one exponential function for the entire time scale.

The values for the categories of toiletries and food products are plotted on the lower-right and upper-left sides of the graph, respectively (Fig.6). The distributions are observed to be different for each category.

The characteristic lifetime of an exponential function depends on the product category. In
Figure 6. Combination of values of characteristic lifetimes for short and long time scales. The dotted line represents the case in which these values coincide. Products are divided into two broad categories, i.e., food products and toiletries.

the short time scale (T1), the average lifetimes lie in the range of 200-1300 days, as shown in Fig.6. For food products, the lifetimes are shorter than 800 days, while for toiletries, they are comparatively longer. In the long time scale, the characteristic life spans are generally around 800 days (2.2 years).

Fig.7 shows the lifetime distributions for 4 food products, namely, bread, chilled desert, terrine, and sweets. The distribution for each product is well approximated by the dotted straight line in the range of 2000-5000 days. In this time scale, the lifetime distributions for all food products roughly follow a universal exponential function. However, in the short time scale, the distribution exhibits a rapid decay.

Figure 7. Cumulative lifetime distributions for 4 food products, namely, bread, chilled desert, terrine, and sweets.

Fig.8 shows the lifetime distributions for 4 toiletry products, namely, bath and body-care goods, sanitary goods, food and sanitary products for pets, and detergents. The distribution for each item is well approximated by the dotted line in the range of 2000-5000 days. However, in the short time scale, the distribution exhibits a slower decay.
6. Summary
In this study, we observed that the probability density function of the sales lifetimes of supermarket products roughly follows an exponential function accompanied by several peaks around the unit of months or year. The distributions follow an exponential function rather than a Weibull distribution, implying that the manufacturers’ decisions about whether to continue to produce or not can be well approximated by a random process such as a Poisson process; however, there is a tendency of making decision on determining the period of sales in the unit of month or year.

In a long time scale, the lifetime distributions of all product categories are well approximated by a common exponential function. This implies that in a large time scale, a product of each category exits nearly randomly at a constant rate.

However, in a short time scale, the distribution sales lifetimes is found to be dependent on the product category. In this time scale, the characteristic lifetimes of food products are generally short. This is because food products tend to exit at a rapid rate in a short time scale, possibly due to the existence of competing products. On the other hand, the characteristic lifetimes of toiletry products are generally long. This is because toiletry products tend to exit at a slow rate in a short time scale. The analysis of POS data revealed that in a short time scale the lifetime distributions for different product categories are different. It is believed that manufacturers implement different merchandising strategies for different product categories.

The POS data contains information about the number of sales of products and the price of each. By using a POS dataset, we intend to analyze the relation between a manufacturer’s merchandising strategy and fluctuations in sales.

Acknowledgement
The authors appreciate the corporation of ”Nikkei Digital Media Inc.” for providing the POS data. This work is partly supported by the Research Fellowships of the Japan Society for the Promotion of Science.

References
[1] R.N.Mantegna and H.E.Stanly 1995 Nature 376 46-49.
[2] M.H.R Stanly 1996 Nature 379 804-806.
[3] D. Sornette, F. Deschatres, T. Gilbert, and Y. Ageon 2004 *PRL* 93 228701.
[4] F. Deschatres and D. Sornette 2005 *PRE* 72 016112.
[5] Takayuki Mizuno, Masahiro Toriyama, Takao Terano and Misako Takayasu 2008 *Physica A* 387 3931-3935.
[6] Takayuki MIZUNO and Misako TAKAYASU 2009 *Progress of Theoretical Physics Supplement* 179 71-79.
[7] Robert D. Groot 2005 *Physica A* 353 501-514.
[8] ELLIOT W. MONTROLL 1981 *Proc. Natl. Acad. Sci. USA* 78-12 7839-7843.
[9] John H. K. Kao 1959 *Technometrics* 1-4 389-407.
[10] Christian Broda and David E. Weinstein 2007 *NBER Working paper* 13041.
[11] Bils Mark 2004 *NBER Working paper* 10606.