Physica A 390 (2011) 499–504

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Collective purchase behavior toward retail price changes

Hiromichi Ueno [a,*], Tsutomu Watanabe [b], Hideki Takayasu [c], Misako Takayasu [a]

[a] *Department of Computational Intelligence & Systems Science, Interdisciplinary Graduate School of Science & Engineering, Tokyo Institute of Technology, 4259-G3-52 Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan*

[b] *Institute of Economic Research, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8603, Japan*

[c] *Sony Computer Science Laboratories Inc., 3-14-13 Higashigotanda, Shinagawa-ku, Tokyo 141-0022, Japan*

ARTICLE INFO

Article history:
Received 5 September 2010
Available online 7 October 2010

Keywords:
Collective behavior
Power law
POS data
Log-normal distribution

ABSTRACT

By analyzing a huge amount of point-of-sale data collected from Japanese supermarkets, we find power law relationships between price and sales numbers. The estimated values of the exponents of these power laws depend on the category of products; however, they are independent of the stores, thereby implying the existence of universal human purchase behavior. The rate of sales numbers around these power laws are generally approximated by log-normal distributions implying that there are hidden random parameters, which might proportionally affect the purchase activity.

1. Introduction

Developments in information technology have enabled the storage of large volumes of high-frequency data of human activities, and soon, scientists began paying attention to such data [1]. People act intentionally based on their own will; in this sense, human behavior should be very different from the motion of materials. It will be very difficult to find a universal law for individual behavior, which may be based on private preferences or habits; however, there is a possibility that universal statistical laws can be found in collective human behavior.

Several pioneering studies have reported possible universal laws in the mass of human activity. At the end of the nineteenth century, Pareto investigated individual income distribution in many countries and found that power laws were dominant in the case of the high-income group of people [2]. In 1949, Zipf listed power law distributions in various types of human behavior from word-frequency to city population [3]. Shockley pointed out that the distribution of the productivity of scientists followed a log-normal law in 1957 [4]. In 1963, Mandelbrot found scale-invariance and a power law distribution in the market prices of cotton [5], and in 1981, Montroll analyzed the price distribution of products, and found that the distribution follows a log-normal law [6].

Electronic databases became available from the end of the last century, and the quality of data analysis rose considerably. In 1995, Mantegna and Stanley confirmed power law distributions of market price changes [7]. M.H.R. Stanley et al. surveyed business firm databases and discovered that the variance of the growth rate of annual sales of a firm decreases following an inverse power law of its sale in 1997 [8], and Redner found a power law distribution in scientific citations in 1998 [9].

Recently, sales data such as point-of-sale (in short POS) data are studied from the viewpoint of physics. Sornette et al. analyzed a time series data of book sales obtained from Amazon.com and found that a functional form of increase and decrease in bestsellers can be approximated by power laws [10,11]. Groot observed fluctuations in sales using sales data collected from Dutch supermarkets and observed that these fluctuations exhibit properties similar to those of the stock market [12]. Fu et al. reported a universal growth rate distribution through an exhaustive investigation of various economic activity data such as the POS of products, business firm's sales, and even GDP [13]. Mizuno et al. focused on the amount

* Corresponding author.
E-mail address: ueno@smp.dis.titech.ac.jp (H. Ueno).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2010.09.032

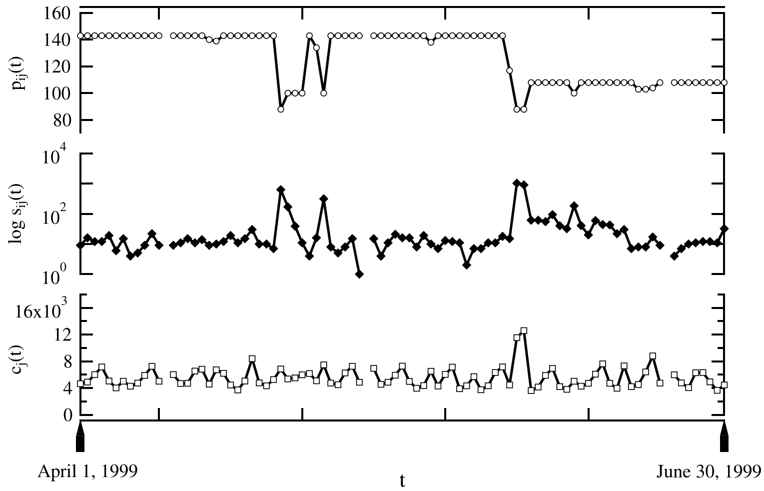

Fig. 1. Time series of $p_{ij}(t)$ (circle), $s_{ij}(t)$ (black-lozenge, in log scale) and $c_j(t)$ (square), $i = 1$; the product is a popular instant cup noodles, for store number $j = 1$ from April 1, 1999 to July 30, 1999.

of payment at one purchase and found a power law distribution [14]; moreover, repeaters' characteristic behaviors were detected by analyzing the POS data with IDs collected from Japanese convenience stores [15]. As a basic property derived from the POS data, Ueno et al. demonstrated that life span distributions of products are generally approximated by an exponential function for a time scale longer than about 4 years, thereby implying that a long seller's life span may follow a Poisson process [16].

In this paper, we analyze a huge amount of POS data from Japanese supermarkets to clarify consumers' responses to price changes. Qualitatively, it is apparent that people will rush to purchase discounted products; here, we quantitatively observe the functional relationship between the rate of price discount and the rate of sales number increase.

2. The relationship between price and the number of sales

We analyze the POS data provided by Nikkei Digital Media Inc. collected from 384 Japanese supermarkets from March 1988 to May 2009. This data comprises approximately 4.2 billion records, each consisting of 5 numbers; the JAN (Japanese Article Number) code identifies the product name, the store code, the date, the total of gross sales, and the number of sales for this product sold in this store on this day. The total number of products specified by the JAN code is approximately 1.6 million. In addition to this POS data, an additional data set comprising the number of customers per day in each store together with the store's name and address was analyzed. This POS data covers comprehensive information about all commercial products having JAN codes, which were sold at one of the 384 supermarkets. It should be noted that this data does not contain information about products without JAN codes, for example, fresh foods such as vegetables, meat, and fish are excluded. All the products are classified into two broad categories—food and toiletries. There are 213 sub-categories such as milk, instant cup noodles, and shampoo.

From this data set, we define the variables as follows: the number of sales of product i at store j on the tth day, $s_{ij}(t)$; the gross sales of product i at store j on the tth day, $g_{ij}(t)$; and the number of customers at store j on the tth day, $c_j(t)$. As there is no direct information about the price of each product in the data, we define the price of product i at store j on the tth day by $p_{ij}(t) = g_{ij}(t)/s_{ij}(t)$.

Fig. 1 shows an example of a set of time series of $p_{ij}(t)$, $s_{ij}(t)$, and $c_j(t)$ for product $i = 1$ (this is a popular instant cup noodles product) at store $j = 1$ for 3 months from April 1, 1999 onwards. As is evident in this figure, store 1 sold product 1 at approximately 143 yen with the number of sales around 10. Sometimes, there were bargain sales at a price of around 88 yen, when the number of sales peaked to about 1000, that is, about 100 times the regular price. There are cases when the store was closed, and there is no record of the store on those days; the plots of such days are missing.

In order to estimate a quantitative relationship between the price and sales numbers, we introduce two quantities—the rate of sales numbers of product i at store j on the tth day, $S_{ij}(t)$, and the price rate, $P_{ij}(t)$.

$$S_{ij}(t) = \frac{s_{ij}(t+1)}{s_{ij}(t)} \tag{1}$$

$$P_{ij}(t) = \frac{p_{ij}(t+1)}{p_{ij}(t)}. \tag{2}$$