

RESEARCH ARTICLE

Property of Fluctuations of Sales Quantities by Product Category in Convenience Stores

Gaku Fukunaga¹, Hideki Takayasu^{2,3}, Misako Takayasu^{1,2*}

1 Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama, Japan, **2** Advanced Data Analysis and Modeling Unit, Institute of Innovative Research, Tokyo Institute of Technology, Midori-ku, Yokohama, Japan, **3** Sony Computer Science Laboratories, Shinagawa-ku, Tokyo, Japan

* takayasu.m.aa@m.titech.ac.jp



OPEN ACCESS

Citation: Fukunaga G, Takayasu H, Takayasu M (2016) Property of Fluctuations of Sales Quantities by Product Category in Convenience Stores. PLoS ONE 11(6): e0157653. doi:10.1371/journal.pone.0157653

Editor: Wei-Xing Zhou, East China University of Science and Technology, CHINA

Received: February 6, 2016

Accepted: June 2, 2016

Published: June 16, 2016

Copyright: © 2016 Fukunaga et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The raw data include some confidential information and they cannot be copied freely without permission of the data provider, Seven-Eleven Japan Co., Ltd. How to access the data: Similar POS data in general can be purchased from data provider companies or be obtained directly from individual companies which are using POS data in their business. Our data was provided by Seven-Eleven Japan Co., Ltd. for academic study with the authors. The data was provided free of charge from Mr. Toshifumi Suzuki, a former CEO and president of Seven-Eleven Japan Co., Ltd., limited to use for academic purpose in 2010. Those researchers who is interested in analyzing the same dataset can ask

Abstract

The ability to ascertain the extent of product sale fluctuations for each store and locality is indispensable to inventory management. This study analyzed POS data from 158 convenience stores in Kawasaki City, Kanagawa Prefecture, Japan and found a power scaling law between the mean and standard deviation of product sales quantities for several product categories. For the statistical domains of low sales quantities, the power index was 1/2; for large sales quantities, the power index was 1, so called Taylor's law holds. The value of sales quantities with changing power indexes differed according to product category. We derived a Poissonian compound distribution model taking into account fluctuations in customer numbers to show that the scaling law could be explained theoretically for most of items. We also examined why the scaling law did not hold in some exceptional cases.

Introduction

In recent years with the progress and spread of information technologies and the Internet, data superior in both quality and quantity have been accumulating. Many are striving to utilize the laws and knowledge obtained from detailed analyses of these data in practical economics and to improve their daily lives. Human purchasing behavior around products and services is no exception. Numerous companies are keeping point of sales (POS) data on minute and second time-scales for all products at their stores, as well as on what and how many products are sold to each customer and where. Analyses of these data had practical and academic dimensions—such as the distribution of monetary amounts paid by customers for single-time purchases [1], product life [2], price elasticity [3]—that led to discussions of important characteristics based on quantitative observations.

The POS data accumulated over a specific time period enable the study of dynamic characteristics such as changes in the relationships between competing products, time series of sales quantities [4] and its fluctuations [5, 6]. Companies engage in advertising, price cuts, and other promotions, as well as sales on certain weekdays or days of the month, all of which have major

Seven-Eleven Japan Co., Ltd. for academic collaboration.

Funding: This work was partly supported by JSPS KAKENHI Grant Number 26310207 (MT) (<https://kaken.nii.ac.jp/d/p/26310207.en.html>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Coauthor Hideki Takayasu is employed by Sony Computer Science Laboratories. Sony Computer Science Laboratories provided support in the form of salaries for author HT, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of manuscript. The specific role of this author is articulated in the "author contributions" section.

Competing Interests: Coauthor HT is employed by Sony Computer Science Laboratories. There are no patents, products in development, or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

impacts on increases and decreases in sales. These items are also being studied, and methods using models such as state space model have been proposed [7–10].

Meanwhile, Robert [5] studied the distribution of sales quantities time series and autocorrelation while excluding sales data from price reduction and sales promotion periods. These time series fluctuations are inherent in goods and services markets and are difficult for companies to predict and control. Companies require knowledge of not only the means of sales amount time series but also their distributions and fluctuations. For example, to determine the effectiveness of a sales promotion, a company needs information about statistical distributions when there is no promotion [11]; even when mean sales amounts are the same, considering inventory or opportunity loss, companies have greater difficulty deciding delivery amounts for products with large fluctuations. It is essential, especially for convenience stores and other small retailers that handle thousands of products every day, to efficiently and accurately follow all processes from production to product delivery. Production lines at factories produce many products for many different stores; they must know not only the size of sales-quantity time series fluctuations for smaller domains such as single products for single stores but also the fluctuation sizes for major domains (i.e., the totals for their customers over many stores and many products).

The proportionality relationship of a quantity associated with a variable representing scale is generally called a scaling law. Scaling law for fluctuations as a function of the mean value began with Taylor's studies of the means and fluctuations in the numbers of individuals in natural ecology [12]; these were followed by extensions into monetary markets [13–16], networks [17, 18] and other various fields [19–22]. Sano et al. [23, 24] studied the frequencies of word appearances on Internet blogs and showed that they could be explained by combining fluctuations in an internal system factor (e.g., written quantities of individual bloggers) with those in an external factor (e.g., total blogger numbers). An internal factor for product sales quantities would be fluctuations in the number of items purchased by a single person, while an external factor would be fluctuations in customer quantities. Thus, considering these two fluctuations simultaneously should enable the construction of a fluctuation scaling law for individual products' sales-quantities time series.

Accordingly, this paper uses POS data on convenience stores to analyze a scaling law for sales-quantities time series fluctuations when customer numbers are changed via changes in time scales (of one day or one week) and space scales (of one store or multiple stores). We also derive a fluctuation scaling model by constructing a model that simultaneously considers fluctuations due to number of products purchased by a single customer and customer-quantity fluctuations, and we compare between the scaling laws introduced by the model and actual scaling laws for each product category. Our study observes and models scaling laws concerning not only limited numbers of products within the limited parameter of products in stores but also varieties of products, including foodstuffs and non-foodstuffs, at multiple stores.

Analysis of actual data

Analyzed POS data

The POS data used for analysis were taken from 158 chain stores of a major convenience store company in Kawasaki City, Kanagawa Prefecture, Japan. The data cover the five month period from June 1 to October 31, 2010. These 158 stores had continuous operations during the study period (with operations commencing before June 1 and without any store closing during the period). The data were divided into two physical localities, one with 77 stores, and the other with 81. Two types of data were used for analysis: receipt data and daily data. Receipt data are recorded every time a customer completes a purchase at the cash register; these data comprise sales date, sales time, store number, store location, the sequential customer number for that

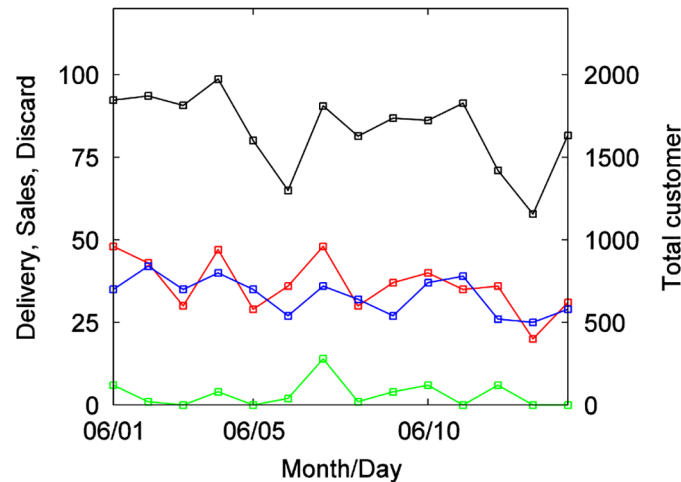


Fig 1. Time series of delivery quantities for a single product, rice ball at a store (red), sales quantities (blue), discard quantities (green), total customer numbers (black).

doi:10.1371/journal.pone.0157653.g001

day at that store, product category, product code, product name, sales quantity, and sales monetary amount. Sales time is recorded in minutes. Product categories are divisions by product type, with a total of 60 categories. When a customer submits a point card at the time of purchase, the card’s ID is recorded. However, our study computed sales quantities regardless of whether such a card was submitted. These data also include data on returned items; in such cases, the items recorded are subtracted from the total sales quantity amounts. However, as returns are extremely rare, returned items have virtually no effect on the calculations in this paper. Our second data type, daily data, indicates number of products delivered, sales quantities, and discard quantities; items recorded are date of sale, store number, store location, product category, product code, product name, delivered quantities, sales quantities, and discard quantities. Using these data types, one can recreate various time series, such as in Fig 1. This shows time series for delivery quantities, sales quantities, and discards for individual days for a single product, rice ball at a specific store, as well as a time series for total customer quantities (i.e., anyone who purchased any item) at that store for each day.

Analyzed products

The analyzed products were the top 10 products in terms of sales quantities for all product groups (within each product category) for which at least one item was sold during one day at any one store. In some cases, there were fewer than 10 products within a single category that fulfilled these conditions. In some categories, no product fulfilled these conditions; these categories were excluded from analysis. Of all 60 categories, 45 were used for analysis (30 foodstuffs and 15 non-foodstuffs; see Table 1). For products that included any discount period, sales data within the period were excluded from the time series to exclude the effects of monetary discounts. By changing time scales (i.e., one day or one week) and space scales (i.e., one store or multiple stores) for the time series, we investigated the means and standard deviations for the sales-quantity time series for various customer numbers.

Table 1. Product categories and the coefficients of the variation (σ_N/μ_N) for each week's totals for all stores.

Foodstuff categories				Non-foodstuff categories	
Beer	0.059	Candies1	0.080	Magazines	0.093
Japanese Sake	0.101	Candies2	0.104	Newspapers	0.041
Chuhai alcohol	0.052	Snack candies	0.085	Menswear	0.097
Rice balls	0.060	Instant noodles	0.094	Health care items	0.096
Sandwiches	0.055	Spices	0.030	Tissues	0.052
Noodles products	0.182	Rice	0.069	Shampoos and oral care products	0.056
Salads and soups	0.055	Gelled products	0.104	Detergents, cleaning	0.104
Desserts	0.047	Snack items1	0.041	Cleaning material	0.087
Fried items	0.120	Frozen foods	0.213	Hair-styling	0.194
Packaged beverages	0.072	Ice creams	0.291	Goods featuring popular character	0.155
Cheese	0.057	Soft drinks	0.098	Office supplies	0.063
Eggs	0.070	Nutrition drinks	0.137	Electrical items and umbrellas	0.098
Snack items1	0.168	Nutrition solids	0.129	Stamps	0.078
Breads1	0.051	Candies3	0.765	Coffee	0.043
Breads2	0.044	Other foodstuffs	0.166	Oversize garbage sticker	0.093

doi:10.1371/journal.pone.0157653.t001

Analysis results for sales-quantity time series for one product, one store, per one day

First, we analyze the smallest scale, sales-quantity time series for one product, one store, per one day. Fig 2(a) shows the semi-log plot of cumulative distribution of sales time intervals for two time frames—weekdays and non-weekdays (e.g., weekends, holidays)—for one product, soft drinks, at a single store. In this paper, a time interval is the time from the sale of one product until the next time an identical product is sold again. For sales-quantity time series, the periodicity of weekdays is usually considered; thus, our analysis divided the days of the week into “weekdays” and “non-weekdays”. In our results, the “daytime” timeframe includes data from 11:00 a.m. to 2:59 p.m., and the “evening” timeframe includes data from 6 p.m. to 9:59 p.m. The distributions are displayed by individual (approximately) straight lines, showing that sales-quantity time intervals can be approximated with exponential distributions with differing statistical means.

Fig 2(b) shows the autocorrelation of the time series of sales-quantity time intervals. (CDFs and autocorrelations of some other categories are shown in S1 Fig.) The result for each is near 0. The Ljung-Box test [25] was performed for these. Under the null hypothesis that the autocorrelation coefficient is 0, the test statistic Q as defined in Eq (1) follows the chi-square distribution of degree n :

$$Q = T(T + 2) \sum_{k=1}^n \frac{\hat{\rho}_k^2}{T - k} \tag{1}$$

Here, T and ρ_k are the number of sample and autocorrelations of time series data quantity with lag k , respectively. In the results of tests that changed n to 1 through 10, there was no rejection of the null hypothesis at a 5% significance level (see S1 Table).

For this store, the mean μ and standard deviation σ of the product's sales quantity time series for every Friday are 40.9 and 7.0, respectively, showing the following relationship:

$$\sigma \approx \mu^{\frac{1}{2}} \tag{2}$$

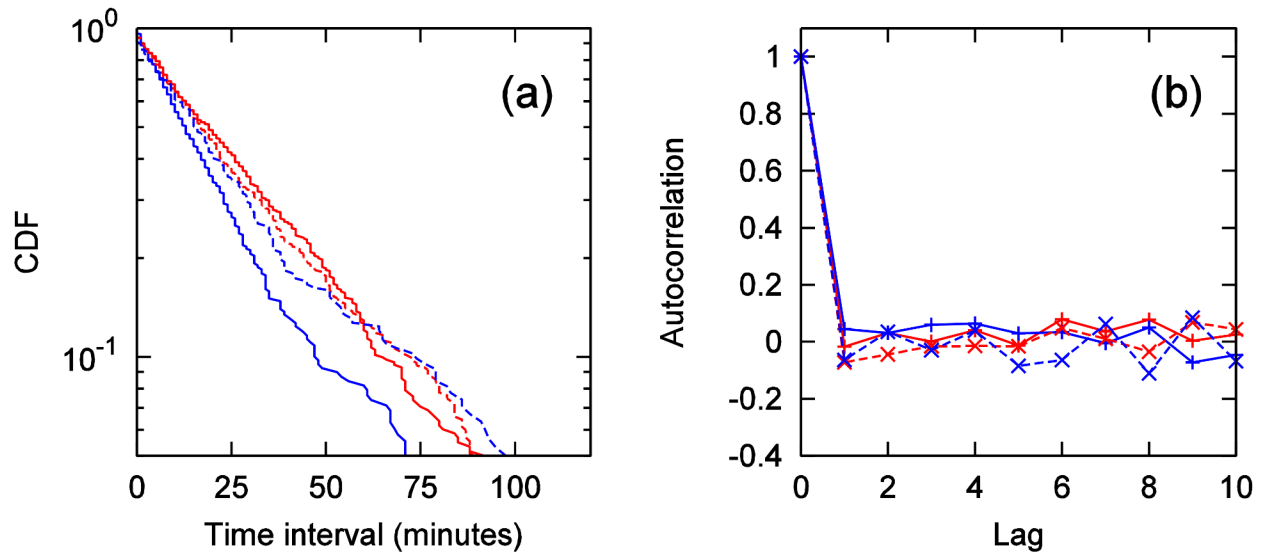


Fig 2. (a) Cumulative distribution of time interval (minutes) in semi-log plot: daytime on weekday (red solid line), daytime on non-weekday (blue solid line), evening on weekday (red dashed line), evening on non-weekday (blue dashed line). (b) Autocorrelation of time interval.

doi:10.1371/journal.pone.0157653.g002

Some studies (e.g., [26]) assume that sales quantities for single products (i.e., one item) follow a Poisson process; for sales of one item at one convenience store per one day, this assumption is reasonable.

Analysis results for a variety of scales

By changing the scale of customer quantities observing with different time and space scales, we investigated the relationship between the means and standard deviations of sales and customer quantities.

Sales-quantity time series. For each product category, the following 12 combinations of sales-quantity time series were analyzed:

1. 1product, 1store, each day of the week (See Fig 3(a) and 3(b))
2. 1product, 1store, weekly totals
3. 1product, all stores within a locality, each day of the week
4. 1product, all stores within a locality, weekly totals
5. 1product, all stores within the city, each day of the week (See Fig 3(c) and 3(d))
6. 1product, all stores within the city, weekly totals
7. all 10 products, 1store, each day of the week
8. all 10 products, 1store, weekly totals
9. all 10 products, all stores within a locality, each day of the week
10. all 10 products, all stores within a locality, weekly totals
11. all 10 products, all stores within the city, each day of the week

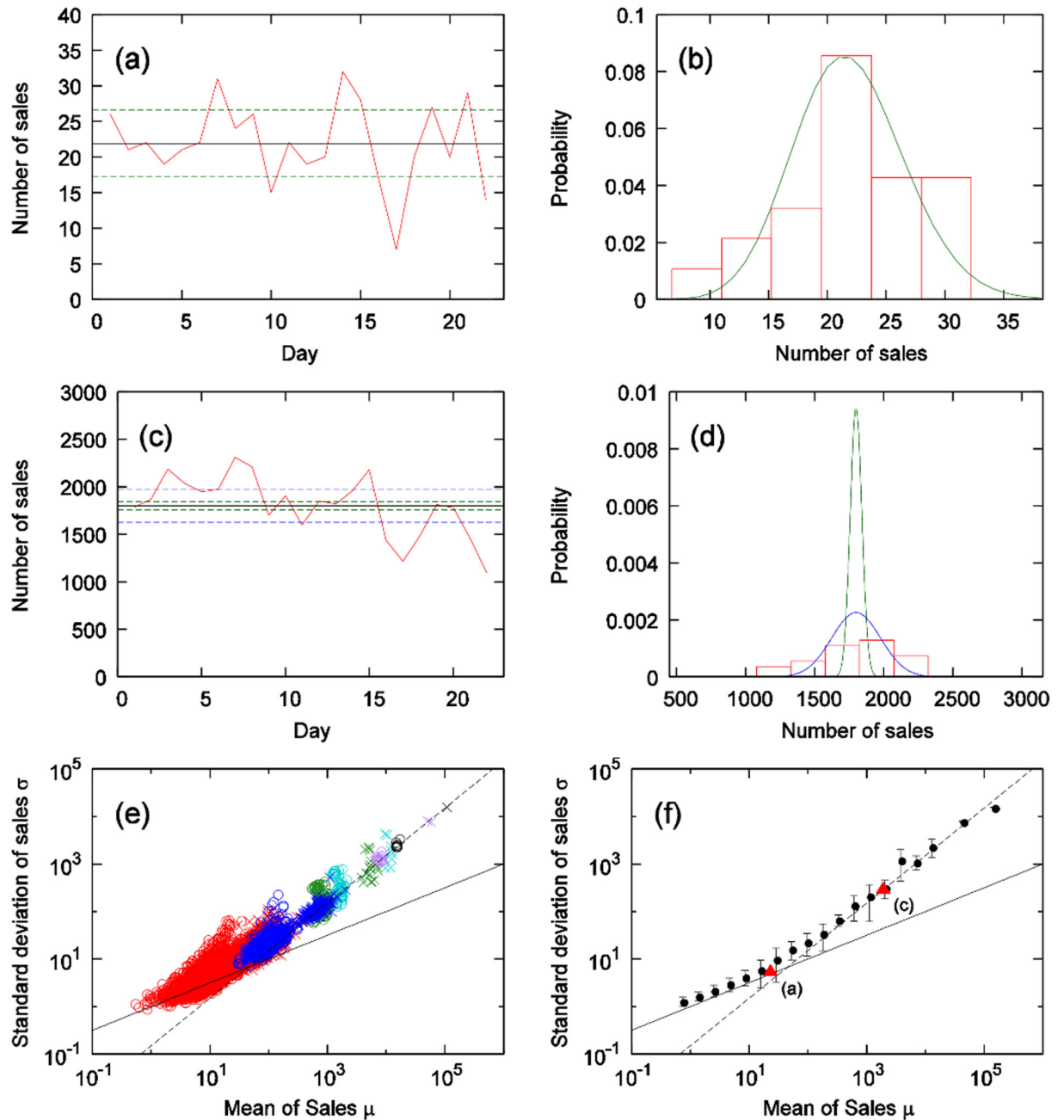


Fig 3. (a),(b) Sales-quantity time series and distribution of 1 product, 1 store, each day of the week. (c),(d) Sales-quantity time series and distribution of 1 product, all stores within the city, each day of the week. In (b),(d), green line shows a Poisson distribution whose mean is the same as the mean of real data. In (d), blue line is a Gaussian distribution whose mean is the same as the mean of real data and the standard deviation is 0.1 times of the mean of real data. In (a),(c), green and blue dashed lines represent the standard deviations of (b),(d), respectively, and black line is the mean of real data. **(e) Relationship between mean and standard deviation of sales quantities in the soft drinks category.** Regarding plot points, for each day of week: circle, weekly totals; cross, 1 product on 1 store; red, all products on 1 store; blue, 1 product within locality; magenta, 1 product within city; cyan, all products within city; black. Black solid line shows $\sigma = \sqrt{\mu}$, and black dashed line shows $\sigma = 0.15\mu$. **(f) Means of (e) with error bars given by the standard deviation.** Two red triangles indicate mean and standard deviation of (a) and (c).

doi:10.1371/journal.pone.0157653.g003

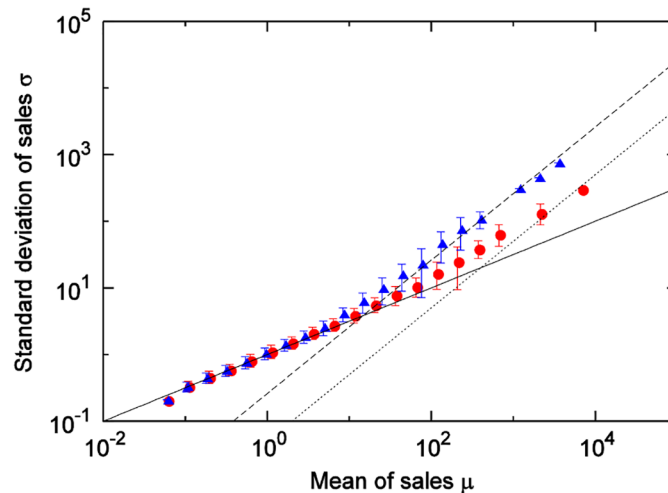


Fig 4. Relationship between mean and standard deviation of sales quantities by category: cheese (red) and hair-styling (blue). Solid, dashed and dotted line show $\sigma = \sqrt{\mu}$, $\sigma = 0.26\mu$, and $\sigma = 0.05\mu$ respectively.

doi:10.1371/journal.pone.0157653.g004

12. all 10 products, all stores within the city, weekly totals

Fig 3(e) and 3(f) shows the means μ and standard deviations σ of product sales-quantity time series for the soft drinks category. We see that, when μ is in a region sufficiently smaller than 10, Eq (2) holds; when μ is in a region sufficiently larger than 10, this relationship does not hold; rather, the following holds:

$$\sigma \propto \mu \tag{3}$$

This type of transition of power law is called Taylor’s law, and in our study, the region (statistical domain) where Eq (2) holds is called the “Poisson area”, and the region where Eq (3) holds is called the “non-Poisson area”.

We next investigate the differences according to product category. Fig 4 shows the results of a comparison between the cheese category and the hair-styling products category. When the sales quantity means of both categories are in the small region, Eq (2) holds; when both are in the large region, Eq (3) holds. However, the boundary values where the movements of the means and the standard deviations change from Eqs (2) and (3) differ.

customer-quantity time series. Before presenting the results of the customer-quantity time series analysis, we first comment on the analysis method for customer quantities used in this paper. In this paper, individual customers are not distinguished; rather, the numbers of purchases completed at the register are denoted as “customer quantities”. For example, suppose the same person makes three purchases in one day at the same convenience store. If individuals were distinguished, this person would be counted as one customer; since individuals cannot be distinguished in POS data, however, this person would be counted as three persons in our study. Customers are also analyzed by classifying these for each product category. If a customer were to purchase any product within a category targeted for analysis, regardless of the number of products bought, this would be counted as one person. Conversely, even if a customer were to purchase several items of products not within a targeted category, they would not be counted within the customer quantities for the targeted categories. If a customer were to purchase two

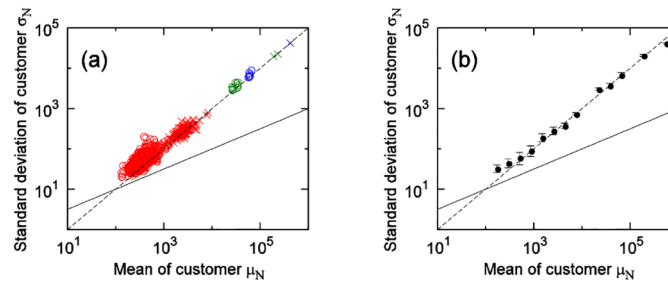


Fig 5. (a) Relationship between mean and standard deviation of customer-quantities time series for the soft drink category. Regarding plot points, for each day of week: circle, weekly totals: cross, on 1 store: red, within 1 locality: green, within city: blue. Black solid line shows $\sigma_N = \sqrt{\mu_N}$, and black dashed line shows $\sigma_N = 0.10\mu_N$. **(b) Mean of (a).**

doi:10.1371/journal.pone.0157653.g005

items from different targeted categories, more customers would be counted than the actual customer quantities (number of purchases completed at the register) since these would be counted as individual purchases within each category. For each product category, the following six combinations of customer-quantity time series were analyzed:

1. 1store, each day of the week
2. 1store, weekly totals
3. all stores within a locality, each day of the week
4. all stores within a locality, weekly totals
5. all stores within the city, each day of the week
6. all stores within the city, weekly totals

[Fig 5\(a\) and 5\(b\)](#) shows the results of the customer-quantity time series for the soft drinks category. Unlike in the results for sales-quantity time series, [Eq \(3\)](#) holds in all regions (statistical domains).

We next investigate differences according to product category. [Fig 6](#) shows the results of a comparison between the cheese category and the hair-styling products category. As with the results for the sales-quantity time series, when the customer quantity means of both categories are in the small region, [Eq \(2\)](#) holds; when both are in the large region, [Eq \(3\)](#) holds. Moreover, the boundary value points where the movements of the means and the standard deviations change are different. As [Fig 5](#) shows, the customer-quantity time series for one store, one product, and one day—the smallest scale—fell within the non-Poisson area for the soft drinks category because it is larger in terms of customer scale than of the cheese or hair-styling products categories, which are the same scale. If one were to observe the soft drinks category at an even smaller time scale (for example, every hour), it would likely fall within the Poisson area.

Summary of analysis results

The results of the analysis of actual data are summarized below:

1. For the statistical mean μ and standard deviation σ of the sales-quantity time series, when $\mu \ll A$, then $\sigma \approx \mu^{1/2}$, and when $\mu \gg A$, then $\sigma \propto \mu$. Here, A is a constant determined by the product category.

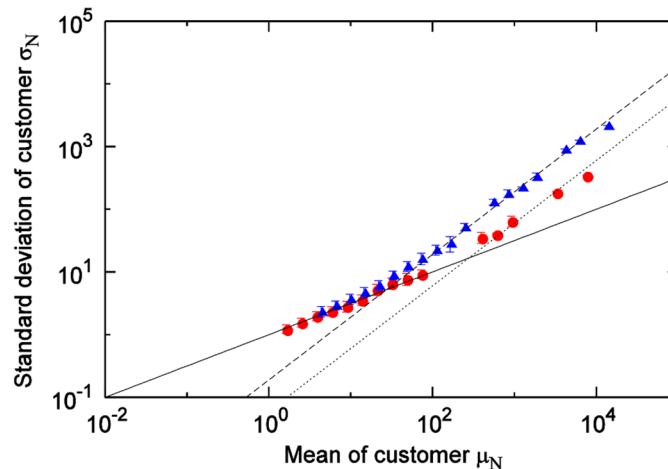


Fig 6. Relationship between mean and standard deviation of customer-numbers time series by category: cheese (red) and hair-styling (blue). Solid, dashed and dotted line show $\sigma_N = \sqrt{\mu_N}$, $\sigma_N = 0.19\mu_N$, and $\sigma_N = 0.06\mu_N$ respectively.

doi:10.1371/journal.pone.0157653.g006

2. For customer-quantity time series classified by category, the above holds.
3. In the Poisson area, the sales process can be approximated by the Poisson process.
4. Even though the scaling of customer quantities is non-Poisson-like ($\sigma \propto \mu$) as shown in Fig 5, the scaling of sales quantities can be Poisson-like ($\sigma = \sqrt{\mu}$) from result of soft drink category as shown in Fig 3(f) in the range $\mu < 10^1$.

Hereafter, compound distribution will be constructed, and the reproducibility of the characteristics described above will be discussed.

Derivation of scaling laws for fluctuations using compound distributions

In this section, we derive scaling laws for fluctuations in sales quantities by using a model that simultaneously considers fluctuations in the number of items purchased at one time and fluctuations in customer quantities (number of purchases completed at the register). The next section compares the results of this section with the actual data.

Here, X_i is a random variable showing the number of items of a single product purchased at a single shopping event by customer i . N is the customer quantity (number of purchases completed at the register) of the relevant product category. We then consider that X summed N times equals the total sales quantity S of one product. That is,

$$S = X_1 + X_2 + \dots + X_i + \dots + X_{N-1} + X_N \tag{4}$$

Here, each X_i ($i = 1, 2, \dots, N - 1, N$) independently follows the same distribution and becomes a series of random variables with no time (temporal) or space (spatial) dependency. The characteristics of customers are considered to be identical and not dependent on the day of the week or the store. Customer quantity N is also considered a random variable not dependent on time. In general, the distribution followed by S as defined in Eq (4) is called a “compound distribution” wherein the distribution followed by X is compounded with the distribution followed by N (for compound distribution, see [27]). The probability density functions of $X, N,$

and S are, respectively, $f_X, f_N,$ and $f_S,$ and the expected values for each of these are $E_X, E_N,$ and $E_S.$ The expected values of arbitrary function g can thus be shown, respectively, as

$$E_X[g(X)] = \sum_x g(x)f_X(x) \tag{5}$$

$$E_N[g(N)] = \sum_n g(n)f_N(n) \tag{6}$$

$$E_S[g(S)] = \sum_s g(s)f_S(s) \tag{7}$$

Using these, the respective means μ and standard deviations σ of $X, N,$ and S are defined as follows:

$$\begin{aligned} \mu_X &= E_X[X] & \sigma_X &= \sqrt{E_X[X^2] - E_X[X]^2} \\ \mu_N &= E_N[N] & \sigma_N &= \sqrt{E_N[N^2] - E_N[N]^2} \\ \mu_S &= E_S[S] & \sigma_S &= \sqrt{E_S[S^2] - E_S[S]^2} \end{aligned}$$

The scaling relationship of the mean μ_S and the standard deviation σ_S of sales quantity S can thus be strictly derived (see [S1 Appendix](#)):

$$\sigma_S = \sqrt{\frac{\sigma_X^2}{\mu_X} \mu_S + CV(N)^2 \mu_S^2} \tag{8}$$

Here, $CV(N)$ expresses the coefficient of the variation in customer quantity and is $CV(N) = \sigma_N/\mu_N.$ From our hypothesis, the square root σ_X^2/μ_X of [Eq \(8\)](#) is not dependent on time or space scales, and $CV(N)^2,$ from the actual data, is $1/\mu_N$ in the Poisson area, and, in the non-Poisson area, a constant whose value differs for each product category. Thus, from [Eq \(8\),](#)

$$\sigma_S = \sqrt{\frac{\sigma_X^2}{\mu_X} + \mu_X \cdot \mu_S^{1/2}} \quad (\mu_S \ll A_m) \tag{9}$$

$$\sigma_S \approx CV(N) \cdot \mu_S \quad (\mu_S \gg A_m) \tag{10}$$

are derived. A_m is the crossover value where the scaling law changes, and is

$$A_m = \frac{\sigma_X^2}{\mu_X} CV(N)^{-2} \tag{11}$$

Thus, by considering the fluctuations in numbers of items purchased and in customer quantities for each time point, one can express the difference in the scaling laws for sales quantity fluctuations in the small and large (statistical) regions.

The scaling law in the region where $\mu_S \gg A_m$ is satisfied does not contradict the scaling law in the Poisson area observed in the actual data, as shown in the previous section. Conversely, in the region where $\mu_S \ll A_m$ is satisfied, although σ_S is proportional to $\mu_S^{1/2},$ the proportionality coefficient is dependent on the distribution of $X.$ Moreover, when one assumes that, as with the soft drinks category, the scaling of customer quantities is non-Poisson-like in all regions, in the scaling law for the region where $\mu_S \ll A_m$ is satisfied, one can ignore the second term of the

square root of Eq (8); this thus becomes $\sigma_s \approx \sqrt{\sigma_x^2/\mu_x} \cdot \mu_s^{1/2}$. Likewise, in this case, although σ_s is proportional to $\mu_s^{1/2}$, the proportionality coefficient is dependent on the distribution of X .

The statistical models for the number of items purchased by a single individual are the negative binomial distribution (NBD) [28–30] model, which postulates that Poisson distribution whose mean follows gamma distribution, and the beta-binomial distribution (BBD) [31–33] model, which postulates that binomial distribution whose purchase probability follows beta distribution (considerations regarding the application of the NBD and BBD models for the purchase of X number of items at one time are provided S2 and S3 Appendix.) One must note, however, that these models are not concerned with the number of items purchased at one time but rather the number of items purchased within a specific time period. More simply than in these models, this paper assumes that the purchase of X number of items at one time is a simple Poisson distribution of the mean λ . Thus, from $\mu_x = \sigma_x^2 = \lambda$, the scaling law Eq (8) becomes

$$\sigma_s = \sqrt{\mu_s + CV(N)^2 \mu_s^2} \tag{12}$$

and Eqs (9) and (10) become

$$\sigma_s = \sqrt{1 + \mu_x} \cdot \mu_s^{1/2} \quad (\mu_s \ll A_m = CV(N)^{-2}) \tag{13}$$

$$\sigma_s \approx CV(N) \cdot \mu_s \quad (\mu_s \gg A_m = CV(N)^{-2}) \tag{14}$$

Furthermore, when it is assumed that $\mu_x \ll 1$ (for example, for the product with the highest sales in the soft drink category, $\mu_x = 0.030$.), then, when $\mu_s \ll A_m$, this becomes $\sigma_s \approx \mu_s^{1/2}$. The same equation holds when the scaling of customer quantities is non-Poisson-like in all ranges. The scaling of customer quantities is then in both Poisson and non-Poisson areas, and the scaling of sales quantities is expressed as

$$\sigma_s \approx \sqrt{\mu_s + CV(N)^2 \mu_s^2} \tag{15}$$

and the boundary value A_m where there is a change from $\sigma_s \propto \mu_x^{1/2}$ to $\sigma_s \propto \mu_x$ becomes

$$A_m = CV(N)^{-2} \tag{16}$$

Here, $CV(N)$ is the coefficient of the variation in customer quantities within the non-Poisson area. Note that, when it is postulated that X follows a binomial distribution (using the BBD model), the same results can be derived (see S3 Appendix). A noteworthy characteristic here is that, since, with Eq (15), parameters due to X do not appear, a scaling law can be shown only for the coefficient of variation (a constant) of customer quantities. The next section compares the scaling laws from the POS data to the scaling law of Eq (15).

Comparison of actual data to the theory

Here, we compare Eq (15), where the number of items X purchased at one time follows a Poisson distribution, with the results of the scaling laws derived from actual data. For each category, we calculate the coefficient of the variation in customer-quantity time series for weekly totals at all stores in city (actual measured values for each category are found Table 1), and use the results as $CV(N)$ in Eq (15). Fig 7 shows the comparison results for 6 of the 45 categories. In each of these figures, Solid lines show Eq (15) and dashed lines show the boundary value A_m as calculated by Eq (16). From Fig 7, it is clear that, for many categories, scaling laws in the Poisson and non-Poisson areas could be expressed with Eq (15). Fig 8 shows the results for the

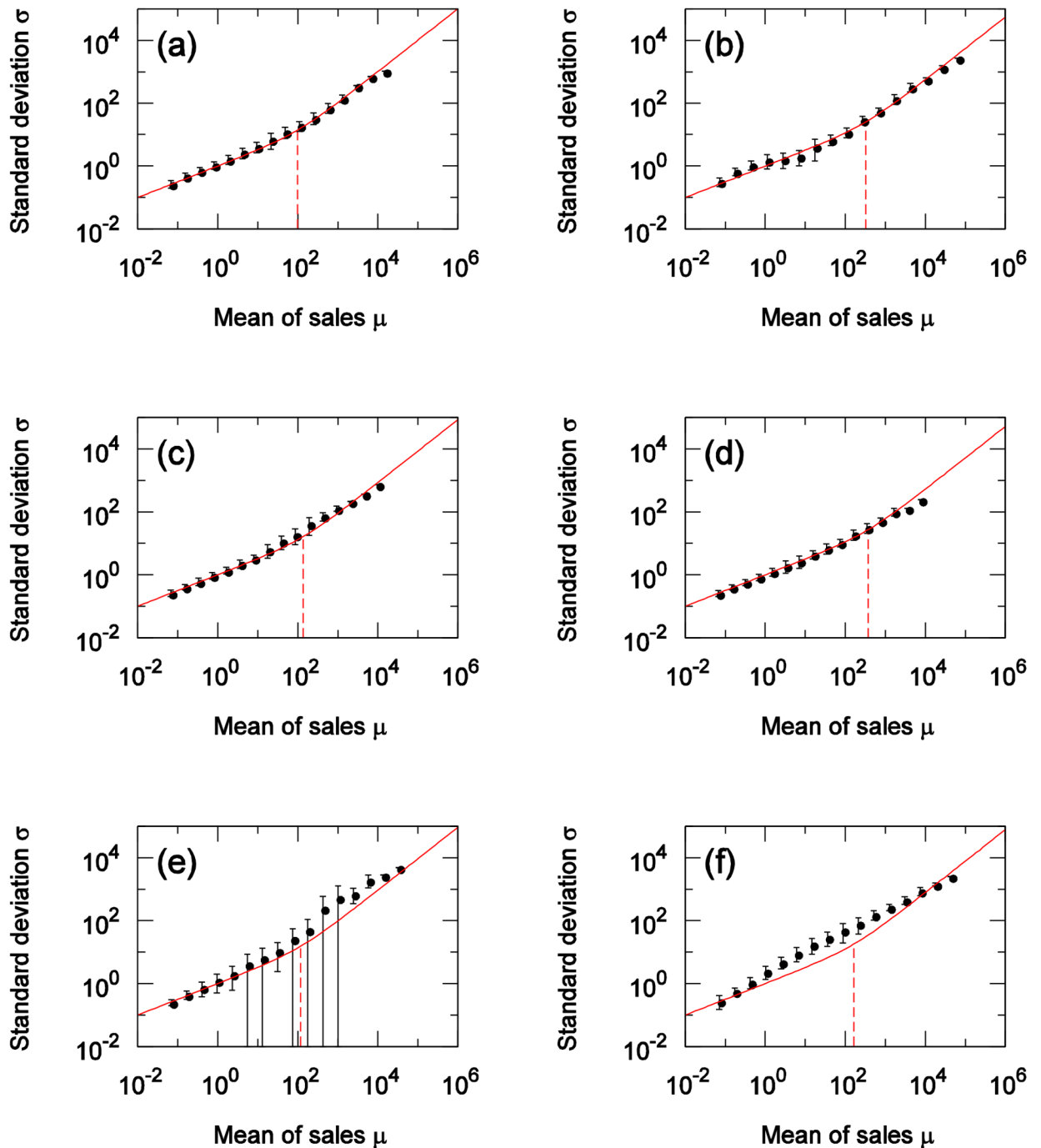


Fig 7. Comparison of actual data and theory. (a) Japanese sake (rice wine), (b) Soft drinks, (c) Eggs, (d) Tissues, (e) Magazines, (f) Stamps. Solid and dashed line show Eq (15), the crossover value A_m , respectively.

doi:10.1371/journal.pone.0157653.g007

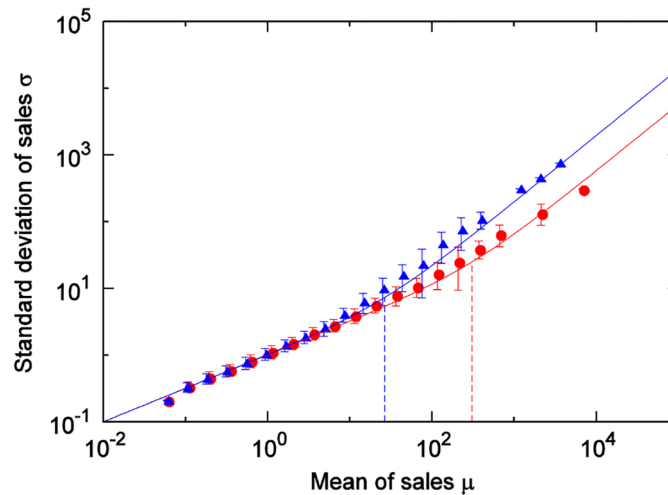


Fig 8. Comparison of reproducibility for two item categories: cheese (red) and hair-styling (blue). Two solid lines show Eq (15) and dashed lines indicate crossover value A_m respectively. Plots are same as in Fig 4.

doi:10.1371/journal.pone.0157653.g008

cheese category and the hair-styling products category. Fig 8 also shows how, with the categorization of customers, the use of the coefficients of the variation in these categories enables the expression of the differences due to these categories. The stamp and magazine categories show relatively large discrepancies between the results of the actual data and those of Eq (15). Considerations are presented for these two items below, category by category.

Regarding the stamps category

From Eq (9), on the Poisson area, $\sigma_s \propto \mu_s^{1/2}$ holds. However, as Fig 7(f) shows, within the Poisson area, $\sigma_s \propto \mu_s^{1/2}$ does not hold. One possible reason is that the distribution of X follows a power-law distribution, whose variance cannot be defined, and thus $\sigma_s \propto \mu_s^{1/2}$ did not hold. Fig 9 shows the cumulative distribution of X from actual data for customers who purchased one or

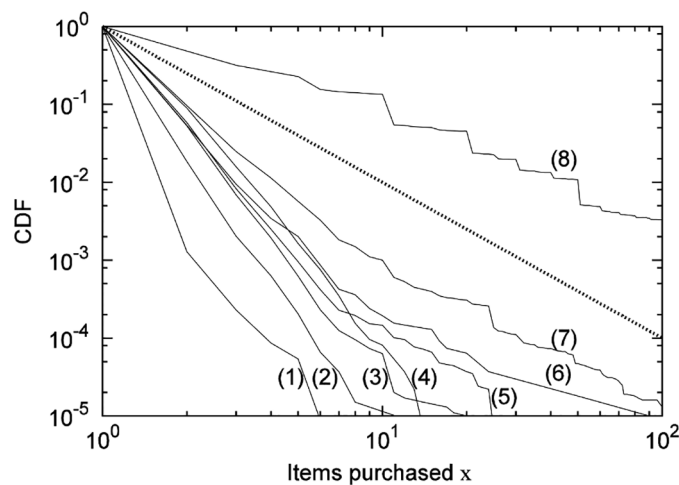


Fig 9. Cumulative distribution of quantities by a purchase. (1)Magazines, (2)Eggs, (3)Salads and soup, (4)Gelled items, (5)Instant noodles, (6)Tissues, (7)Soft drinks, (8)Stamps. Dashed line shows x^{-2}

doi:10.1371/journal.pone.0157653.g009

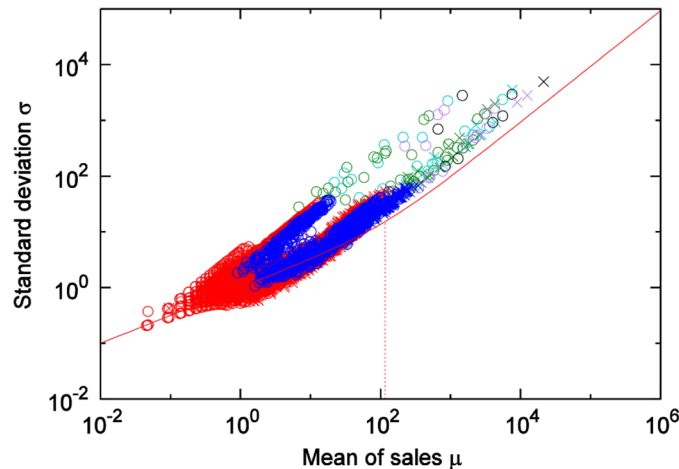


Fig 10. Scaling law in the magazines category. Solid and dashed lines show Eq (15), crossover value A_m respectively.

doi:10.1371/journal.pone.0157653.g010

more products, namely $\Pr(X \geq x|X \geq 1)$ for each category. When the power index α is less than 2 within the power-law distribution $f(x) = \alpha\beta^\alpha/x^{\alpha+1}$, ($\beta \leq x$), variance cannot be defined (it diverges). If we assume that X follows a power-law distribution, Fig 9 suggests that the power index for the stamps category was less than 2, while the power index for other categories larger than 2. For a given finite dataset, a power-law distribution, with a low power index α gives a high mean value. This suggests that the stamps category involves an item that is easy to buy in bulk (i.e., many at one time); and this is the reason of its discrepancy with Eq (15). However, even if the mean of X is large, if the mean and the variance are definable, the discrepancy with the theory is not so great—at the very least, from the consideration that $\sigma_s \propto \mu_s^{1/2}$ holds in that case.

Regarding the magazines category

Fig 10 shows scatter plots of the scaling laws from the POS data for the magazine category. The actual data have wider fluctuations than do those of the theory, and the scattering of each plot point of the actual data is larger than for other categories. In Fig 9, unlike the case with the stamps category, for magazines, the mean of X is small, and, even under the assumption that X is a power-law distribution, its index is still greater than 2. Thus, the divergence between the actual data and Eq (15) is not due to the effects of bulk purchases. In the magazines category, the divergence occurs because all of the nine products targeted for analysis are weekly magazines. In general, the greatest sales of weekly magazines occur on the date of sale, and this quantity diminishes thereafter. Moreover, when the date of sale is a holiday, the magazine will often be sold on another day. For example, if the date of sale is every Monday, when the Monday is a holiday, the magazine may be sold on the immediately preceding Saturday. Studies have considered the periodicity of the days of the week. Analyses of sales quantities per day have examined for each day of the week. If the date of sale is a holiday, however, the periodicity is interrupted, making it necessary to consider periodicity as starting with the date of sale, followed by the next, and the next, and so on. Analysis results of a scaling law for date of sale are shown in Fig 11(a) and 11(b), while Fig 11(c) and 11(d) show the results of a scaling law for the day after the date of sale. As the date of sale was considered as the periodicity standard, overall scattering is reduced, and the divergence with Eq (15) becomes smaller.

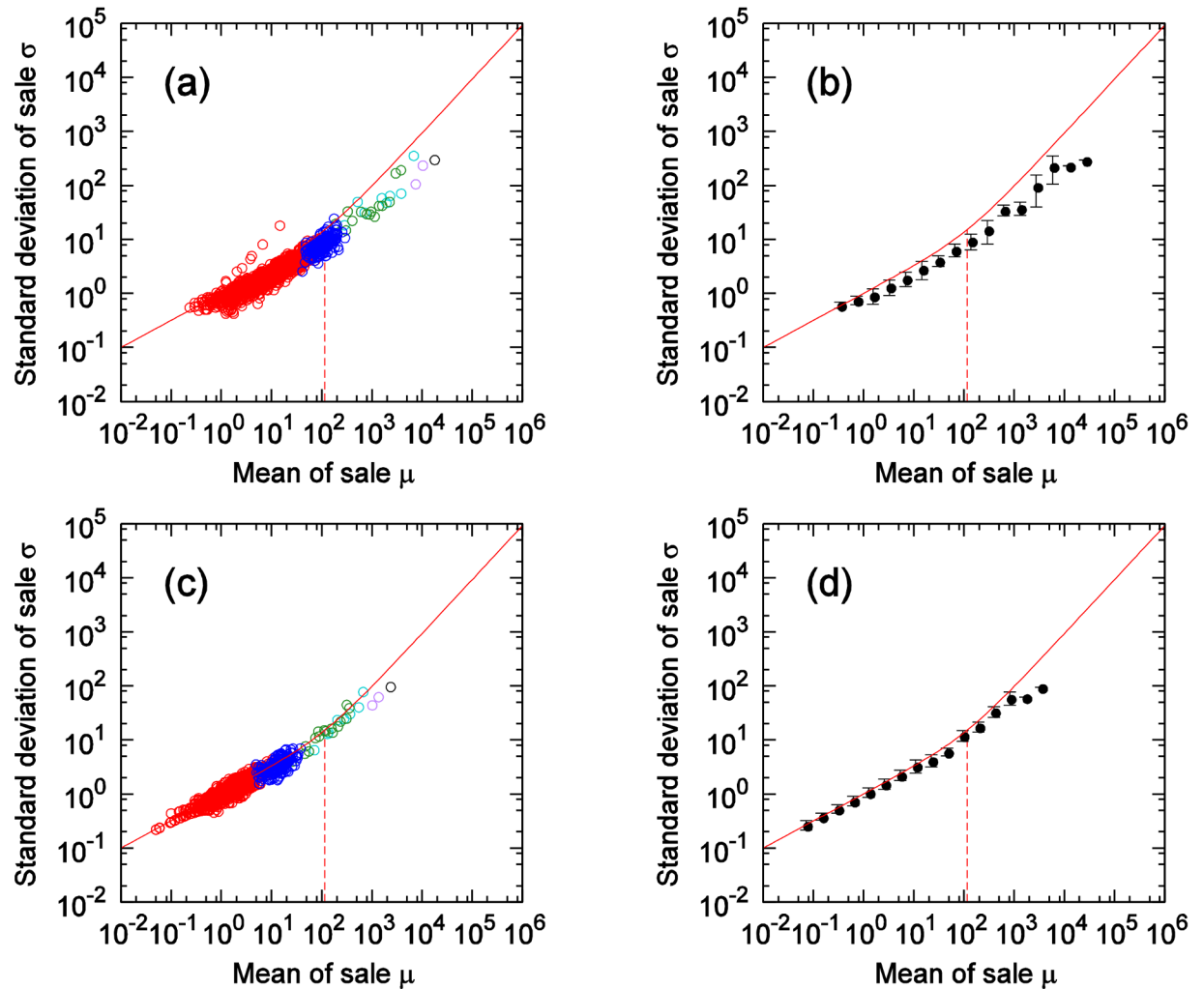


Fig 11. Scaling laws in the magazines category considering the date of sale of each magazine. (a),(b) Results on date of sale. (c),(d) Results on a day after the date of sale. Solid and dashed line show Eq (15), crossover value A_m respectively. (b) is the mean of (a), (d) is the mean of (c).

doi:10.1371/journal.pone.0157653.g011

Standard deviation σ on a Poisson area for date of sale is smaller than that derived from Eq (15), perhaps because of the existence of fixed customers: that some customers always purchase a certain magazine every week may have made this fluctuation smaller than for other categories. Danaher [32] has researched the existence of fixed customers (i.e., periodical fixed purchasers) using monthly magazine purchase data. Danaher [32] modified the BBD model described in S3 Appendix to consider fixed customers, showing that such considerations make for a better fit with actual data. The modified BBD model is called the “MBBD model”; here, with n indicating the number of times a magazine is published, the probability density function $f(X)$ of the random variable X showing purchasing quantities within n times is expressed as

$$f(X) = (1 - \omega) \binom{n}{X} \frac{B(p + X, q + n - X)}{B(p, q)} + \omega I_{\{X=n\}}, \quad X = 0, 1, 2, \dots, n \quad (17)$$

$I_{\{X = n\}}$ is an indicator function defined as, when $X = n$, then $I_{\{X = n\}} = 1$, when $X \neq n$, then $I_{\{X = n\}} = 0$. $\omega (0 \leq \omega \leq 1)$ shows the percentage of fixed customers (those who with probability 1 will buy a magazine); when $\omega = 0$, then this reverts to the BBD model where there is no consideration of fixed customers. Here, $B(p, q)$ is a beta function with positive parameters, p and q . The mean of X and the variance σ_X^2 respectively become [32]

$$\mu_X = n \frac{p + \omega q}{p + q} \tag{18}$$

$$\sigma_X^2 = (1 - \omega) \frac{npq(p + q + n) + n^2\omega q^2(p + q + 1)}{(p + q)^2(p + q + 1)} \tag{19}$$

To compare this with the scaling law for the date of sale of weekly magazines, $n = 1$, and the mean μ_X becomes

$$\mu_X = \frac{p + \omega q}{p + q} \tag{20}$$

When this mean is used, the variance becomes

$$\begin{aligned} \sigma_X^2 &= (1 - \omega) \frac{pq(p + q + 1) + \omega q^2(p + q + 1)}{(p + q)^2(p + q + 1)} \\ &= (1 - \omega) q \frac{p + \omega q}{(p + q)^2} \\ &= (1 - \omega) \frac{q}{p + q} \mu_X \end{aligned}$$

hence,

$$\frac{\sigma_X^2}{\mu_X} = (1 - \omega) \frac{q}{p + q} \tag{21}$$

When this is inserted into Eq (9), a scaling law considered fixed customers within the Poisson area of sales quantities can be derived:

$$\sigma_s = \sqrt{1 - \omega \frac{q}{p + q}} \cdot \mu_s^{1/2} \tag{22}$$

Thus, since $0 \leq \omega q(p+q)^{-1} \leq 1$, by considering fixed customers, the MBBBD model can show that, even if the means within the Poisson area is the same, the standard deviation becomes smaller than $\mu_s^{1/2}$. Therefore, the analysis results of this paper confirm the existence of fixed customers (subscribers) of weekly magazines at convenience stores from the scaling law of fluctuation.

Conclusion

By analyzing POS data from convenience stores, this paper showed that Taylor’s law holds between the means and standard deviations of most of actual quantities and constructed a model to fit with the data. The constructed model enabled the simultaneous consideration of fluctuations in individual item purchase quantities and customer quantities, allowing us to reproduce the characteristics of actual scaling laws such that, for sales quantities in areas (statistical domains) with low sales quantities, the power index was 1/2, and, in areas with large sales quantities, the power index was 1. The fact that the boundary value where the power

index was changed would differ for different product categories was reflected by classifying customers into categories.

We also separately analyzed the magazines and stamps categories, which have special characteristics in terms of their scaling laws. For magazines, considering the standard periodicity of weekly magazines' dates of sale, the scaling law was shown to hold, and we also confirmed the existence of fixed customers from the scaling law, as described by Danaher [32].

The stamps category was a clear example of violation of Taylor's law. Since the number of items purchased each time follows a power-law distribution and since the power index is less than 2, variance cannot be well defined. We showed from the distribution of actual item purchase quantities that this category showed behavior different from that of the scaling laws derived from our model.

Summarizing the results we can say that the reason why the number of sales of some categories are Non-Poisson is classified in two cases. One is the case when fluctuation of the number of customer is bigger than fluctuation of sales quantities by a purchase. The other is the case when distribution of quantities by a purchase follows a power-law distribution whose variance cannot be well-defined (it tends to diverge for larger purchase quantity). The former case follows Taylor's law, however, the latter case deviates from the law as we confirmed in the stamps category. The basic Poisson behavior, that is, the standard deviation of sales quantities is proportional to the square root of mean value, is realized if the following two conditions are fulfilled: Condition 1; the variance of sales quantities by a purchase is finite. Condition 2; the number fluctuation of customers is smaller than fluctuation of sales quantities.

Future possible research tasks include specifying the factors determining the coefficient of customer quantity variation in the non—Poisson area. Since the boundary value between the Poisson and non—Poisson areas of sales quantities is determined by the coefficient of customer quantity variation, specifying the factors that determine this value is a key issue. This paper analyzed POS data for Kawasaki City, Japan covering summer period (June through October); it will also thus be necessary to make comparisons with scaling laws with wider time and space scales. Such time scales should include scaling laws for winter periods and over one year, and the space scales should include not just a city but an entire prefecture or an even broader region. It will also be necessary to ascertain power indices and changes of boundary values where the power index changes for each category.

As a practical application, consideration of the characteristics of sales fluctuations not only for the sales of a single store but also for the sales totals of multiple stores will help make production quantity management and production delivery timing more accurate throughout the entire retailer chain.

Supporting Information

S1 Fig. CDF and Autocorrelation of some categories.

(PDF)

S1 Table. The p values of the Ljung-Box test.

(PDF)

S1 Appendix. Derivation of [Eq \(8\)](#).

(PDF)

S2 Appendix. NBD model.

(PDF)

S3 Appendix. BBD model. (PDF)

Acknowledgments

The POS data used in this research were provided by Seven-Eleven Japan Co., Ltd. This work was supported by a Grant-in-Aid for Scientific Research (B) 26310207.

Author Contributions

Conceived and designed the experiments: GF MT HT. Performed the experiments: GF MT HT. Analyzed the data: GF MT HT. Wrote the paper: GF MT HT.

References

- Mizuno T, Toriyama M, Terano T, Takayasu M. Pareto law of the expenditure of a person in convenience stores. *Physica A*. 2008; 387:3931–3935. doi: [10.1016/j.physa.2008.01.059](https://doi.org/10.1016/j.physa.2008.01.059)
- Ueno H, Watanabe T, Takayasu M. Life-span distributions of supermarket products. *Journal of Physics Conference Series*. 2010; 221. doi: [10.1088/1742-6596/221/1/012018](https://doi.org/10.1088/1742-6596/221/1/012018)
- Ueno H, Watanabe T, Takayasu H, Takayasu M. Collective purchase behavior toward retail price changes. *Physica A*. 2011; 390:499–504. doi: [10.1016/j.physa.2010.09.032](https://doi.org/10.1016/j.physa.2010.09.032)
- Ma Q, Zhang W. Public Mood and Consumption Choices: Evidence from Sales of Sony Cameras on Taobao. *PLoS ONE*. 2015; 10(4):e0123129 doi: [10.1371/journal.pone.0123129](https://doi.org/10.1371/journal.pone.0123129) PMID: [25902358](https://pubmed.ncbi.nlm.nih.gov/25902358/)
- Groot R.D. Lévy distribution and long correlation times in supermarket sales. *Physica A*. 2005; 353:501–514. doi: [10.1016/j.physa.2004.12.064](https://doi.org/10.1016/j.physa.2004.12.064)
- Groot R.D. Consumers don't play dice, influence of social networks and advertisements. *Physica A*. 2006; 363:446–458. doi: [10.1016/j.physa.2005.08.023](https://doi.org/10.1016/j.physa.2005.08.023)
- Bruce N.I. Pooling and dynamic forgetting effects in multitheme advertising: tracking the advertising sales relationship with particle filters. *Marketing Science*. 2008; 27:659–673. doi: [10.1287/mksc.1070.0317](https://doi.org/10.1287/mksc.1070.0317)
- Bass F.M, Bruce N, Majumdar S, Murthi B.P.S. Wearout effects of different advertising themes: a dynamic bayesian model of the advertising-sales relationship. *Marketing Science*. 2007; 26:179–195. doi: [10.1287/mksc.1060.0208](https://doi.org/10.1287/mksc.1060.0208)
- Lachaab M, Ansari A, Jedidi K, Trabelsi A. Modeling preference evolution in discrete choice models: a bayesian state-space approach. *Quantitative Marketing and Economics*. 2006; 4:57–81. doi: [10.1007/s11129-006-6559-x](https://doi.org/10.1007/s11129-006-6559-x)
- Higuchi T, Kitagawa G. Knowledge discovery and self-organizing state space model. *IEICE Transaction on Information and Systems*. 2000; E83-D:36–43.
- Goodhardt G.J, Ehrenberg A.S.C. Conditional trend analysis: a breakdown by initial purchasing level. *Journal of Marketing Research*. 1967; 4:155–161. doi: [10.2307/3149360](https://doi.org/10.2307/3149360)
- Taylor L.R. Aggregation, variance and the mean. *Nature*. 1961; 189:732–735. doi: [10.1038/189732a0](https://doi.org/10.1038/189732a0)
- Sato A, Sakai H, Nishimura M, Holyst J.A. Similarity, clustering, and scaling analyses for the foreign exchange market. *Progress of Theoretical Physics Supplement*. 2009; 179:38–50. doi: [10.1143/PTPS.179.38](https://doi.org/10.1143/PTPS.179.38)
- Eisler Z, Kertész J. Scaling theory of temporal correlations and size-dependent fluctuations in the traded value of stocks. *Phys Rev E*. 2006; 73:046109. doi: [10.1103/PhysRevE.73.046109](https://doi.org/10.1103/PhysRevE.73.046109)
- Sato A, Hayashi T, Holyst J.A. Comprehensive analysis of market conditions in the foreign exchange market. *J. Econ. Interact. Coordinat*. 7(2), 167–179 (2012). doi: [10.1007/s11403-012-0089-2](https://doi.org/10.1007/s11403-012-0089-2)
- Jiang Z.-Q. Guo L. Zhou W.-X. Endogenous and exogenous dynamics in the fluctuations of capital fluxes: An empirical analysis of the Chinese stock market. *Eur. Phys. J. B* 57, 347–355 (2007). doi: [10.1140/epjb/e2007-00174-7](https://doi.org/10.1140/epjb/e2007-00174-7)
- De Menezes M.A, Barabási A.-L. Fluctuations in network dynamics. *Phys Rev Lett*. 2004; 92:028701. doi: [10.1103/PhysRevLett.92.028701](https://doi.org/10.1103/PhysRevLett.92.028701) PMID: [14753972](https://pubmed.ncbi.nlm.nih.gov/14753972/)
- De Menezes M.A, Barabási A.-L. Separating internal and external dynamics of complex systems. *Phys Rev Lett*. 2004; 93:068701. doi: [10.1103/PhysRevLett.93.068701](https://doi.org/10.1103/PhysRevLett.93.068701)
- Eisler Z, Bortos I, Kertész J. Fluctuation scaling in complex systems: Taylor's law and beyond. *Advances in Physics*. 2008; 57:89. doi: [10.1080/00018730801893043](https://doi.org/10.1080/00018730801893043)

20. Azevedo R.B.R, Leroi A.M. A power law for cells. *Proc Natl Acad Sci USA*. 2001; 98:5699. doi: [10.1073/pnas.091485998](https://doi.org/10.1073/pnas.091485998) PMID: [11331756](https://pubmed.ncbi.nlm.nih.gov/11331756/)
21. Kendal W.S. A frequency distribution for the number of hematogenous organ metastases. *Journal of Theoretical Biology*. 2002; 217:203–218. doi: [10.1006/jtbi.2002.3021](https://doi.org/10.1006/jtbi.2002.3021) PMID: [12202114](https://pubmed.ncbi.nlm.nih.gov/12202114/)
22. Reed D.H, Hobbs G.R. The relationship between population size and temporal variability in population size. *Animal Conservation*. 2004; 7:1–8. doi: [10.1017/S1367943004003476](https://doi.org/10.1017/S1367943004003476)
23. Sano Y, Kaski K, Takayasu M. Statistics of collective human behaviors observed in blog entries. *Proceedings of 9 Asia-Pacific Complex Systems Conference*. 2009;195–198. arXiv:0912.1911.
24. Sano Y, Yamada K, Watanabe H, Takayasu H, Takayasu M. Empirical analysis of collective human behavior for extraordinary events in the blogosphere. *Phys Rev E*. 2013; 87:012805. doi: [10.1103/PhysRevE.87.012805](https://doi.org/10.1103/PhysRevE.87.012805)
25. Ljung G.M, Box G.E.P. On a measure of a lack of fit in time series models. *Biometrika*. 1978; 65:297–303. doi: [10.2307/2335207](https://doi.org/10.2307/2335207)
26. Seki Y, Kamekura Y. Gross margin maximization based on price elasticity estimated using the generalized linear model. *Journal of Japan Industrial Management Association*. 2012; 63:161–172.
27. Winkelmann R. *Econometric analysis of count data*. Springer. 2008
28. Ehrenberg A.S.C. The pattern of consumer purchases. *Journal of the Royal Statistical Society. Series C*. 1988; 8:26–41. doi: [10.2307/2985810](https://doi.org/10.2307/2985810)
29. Morrison D.G, Schmittlein D.C. Generalizing the NBD model for customer purchases: what are the implications and is it worth the effort? *Journal of Business & Economic Statistics*. 1988; 6:145–159. doi: [10.2307/1391551](https://doi.org/10.2307/1391551)
30. Wadsworth R.N. The experience of a user of a consumer panel. *Journal of the Royal Statistical Society. Series C*. 1952; 1:169–178. doi: [10.2307/2985620](https://doi.org/10.2307/2985620)
31. Morrison D.G. Purchase intentions and purchase behavior. *Journal of Marketing*. 1979; 43:65–74. doi: [10.2307/1250742](https://doi.org/10.2307/1250742)
32. Danaher P.J. Parameter estimation and applications for a generalisation of the Beta-Binomial distribution. *Austral. J. Statist*. 1988; 30:263–275. doi: [10.1111/j.1467-842X.1988.tb00621.x](https://doi.org/10.1111/j.1467-842X.1988.tb00621.x)
33. Danaher P.J. A Markov chain model for magazine exposure. *Austral J. Statist*. 1990; 32:163–176. doi: [10.1111/j.1467-842X.1990.tb01010.x](https://doi.org/10.1111/j.1467-842X.1990.tb01010.x)

© 2016 Fukunaga et al. This is an open access article distributed under the terms of the Creative Commons Attribution License: <http://creativecommons.org/licenses/by/4.0/> (the “License”), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.