Power-Law Distribution in Japanese Book Sales Market

Takashi Iba
Faculty of Policy Management, Keio University

Mariko Yoshida
Faculty of Policy Management, Keio University

Yoshiaki Fukami
Keio Research Institute at SFC

Masaru Saitoh
Nippon Shuppan Hanbai, Inc.

In this paper, we analyze the real sales data of the book sales market in Japan. The data which we analyze is the data taken from POS (Point Of Sales) system of over 2,000 bookstores in all areas of Japan. The target term is from April 2005 to March 2006. As a result of analysis, it became clear that the relation between sales volume and sales rank follows power law in both of annual and monthly sales. This is not only an important finding but also the big chance to understand the market mechanism by the analogy from the other phenomena which follow power law. In this paper, we also focus on the books in the top 1.5 percent of sales, and analyze the power index and market share on the time series. We find that the both of index and market share are increasing from April 2005 to March 2006, therefore it shows that the books in top sales are getting to be sold more and more.

1 Introduction

In this paper, we verify that the relation between sales volume and sales rank follows "power law" in the book sales market, which can be considered as a hidden law of the market. Power law in a market means that there are very few top sale products while the rest of them sold few, and surprisingly all of products in the market follow a simple statistical law. Therefore the sales volume can be calculated by the sales rank, and vice versa. Power-law distribution is plotted linearly-graded in double logarithmic graph (Figure 1).

Interestingly, the power-law distributions have been found in various fields such as natural and social sciences[1, 2, 3]. In these studies, the power-law distribution is deemed to be emerged in the system which the elements interact one another and self-organized into "critical state"[4, 1]. Based on the hypothesis, the fact that the power-law distribution is found in the market implies the market is always organized into the critical state.

In this paper, we analyze the real sales data in Japanese book sales market, in order to verify the relation between sales volume and sales rank follows "power law". The result has two significant implications. First, the macroscopic law is emerged from the interactions despite that the customers buy products by their own decision. It means that the law is called as an "emergent order" in market level, which cannot be reduced to individuals level. Second, the fact that the law is the "power law" implies that the fundamental mechanism is similar to other natural and social phenomena. Therefore there is chance to understand the market mechanism analogically by knowing that of other phenomena. In the following sections, we start the brief introduction of related studies, and then show our analyses.

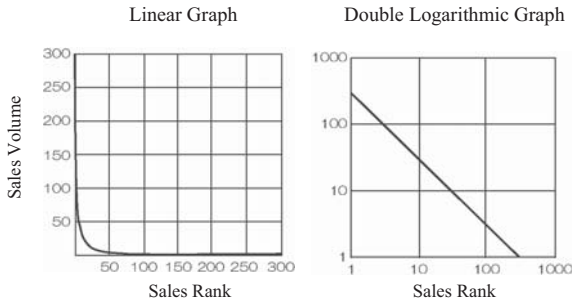

Figure 1: Power-law distribution in linear graph and double logarithmic graph

2 Background

The target in this paper is book sales market. As a product, a book has various value and utility, and there are wide variations in the market. For example, the number of books circulated in Japan is over 1.2 billion and over 77 thousands of new titles are published in a year. In United States, 200 thousands of books are published in every year. Under the situation, the large-scale bookstore always keep over hundreds and thousands of titles in order to meet various demands of customers, however the fact is that sales are polarized. In recent years, "Harry Potter and the Half-Blood Prince" and "The Da Vinci Code" had explosive sales, on the other hand, many other books have not moved their shelves and some books are returned to publishers. Since all books are not equally sold and there is polarization, the book sales market can be considered as "Winner-Take-All" market [5]. "Winner-Take-All" market means that the top-sales products earn overwhelming share. It is well-known that this kind of winner exists in our life, but a whole reality of the book sales market is not clear.

There are some studies trying to analyze book sales market in the viewpoint of power law[1] . Sornette and Deschatres [6] mentioned the relation between sales volume and sales rank, and introduced the presumption by Rosenthal [7]. Rosenthal investigated the sales rank in *Amazon.com* of the books which were published from his publisher, and estimated that the relation between sales volume and sales rank seems to be the power-law distribution. In the analysis, however, it was just a presumption made by some points data[2] , and they are not verified using the real data.

In this paper, the data we used in analysis is the real data which was collected by one of the biggest wholesales of books in Japan. The validity and amount of the data are major features of our analysis[3] .

3 Sales Volume - Rank Distribution in Book Sales Market

We, here, analyze the real sales data in Japanese book sales market, in order to verify the relation between sales volume and sales rank follows "power law". The data we analyze is the real sales data, which is taken from POS (Point Of Sales) system of over 2,000 bookstores in Japan, from April 2005 through March 2006. In this section, we show the results of (1) Analysis of the annual sales, (2) Analysis of monthly sales, (3) Analysis of sales in genres, and (4) Analysis of the tendency of top titles.

3.1 Analysis of the annual sales

The relation between sales volume and sales rank in fiscal year 2005 is shown in Figure 2 (linear graph) and Figure 3 (double logarithmic graph). The vertical axis shows the sales volume, to be exact which is normalized by dividing the sales volume of the title by the sales volume of all books, and the horizontal axis shows the sales rank. These figures show that the top-sales titles merely exist and most of the titles are not very much sold. And we can find that the relation between sales volume and sales rank follows power law, because the plotted values on the straight line of approximation in the double logarithmic graph. The approximate expression of power law distribution is described as follows:

$$v = \alpha r^{-\beta}$$

Here, v is sales volume and r is sales rank. Note that β is a power index, which shows the gradient of approximate line in the double logarithmic graph. In Figure 3, it can be considered that the part which does not fit the approximation is the area of "cut-off" due to the capacity of bookstores.

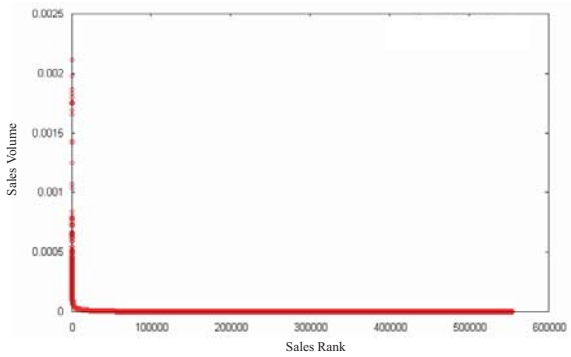

Figure 2: The relation between sales volume and sales rank (from April 2005 to March 2006: linear graph)

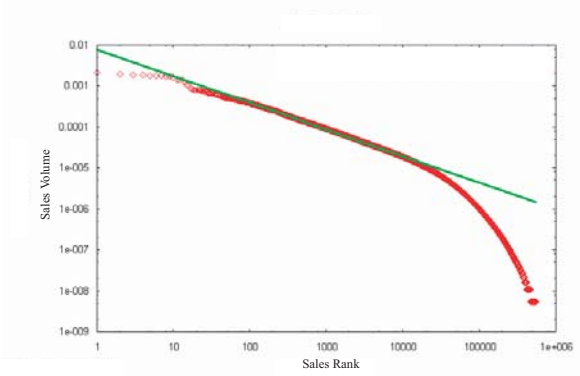

Figure 3: The relation between sales volume and sales rank (from April 2005 to March 2006; double logarithmic graph)

3.2 Analysis of monthly sales

The relation between sales volume and sales rank in the monthly sales from April 2005 to March 2006 is shown in from Figure 4 to Figure 15 respectively. From these figures, it can be said that the power-law distribution in the book sales market is the fractal phenomena, because the monthly distribution shows the almost same distribution of the annual distribution. It also can be said that the power-law distribution is "emergent order" and is kept every time, despite that the customers buy products by their own decision. The customers never imagine that their purchases contribute to such a macroscopic order, because they think that they buying the products by their own decision.

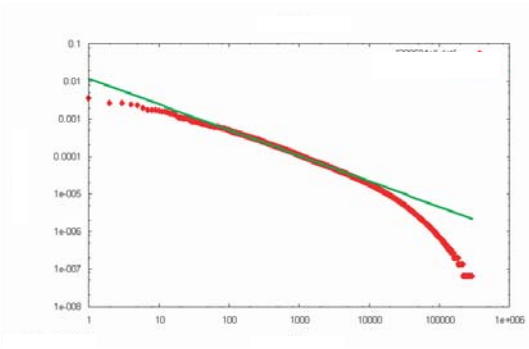

Figure 4: The relation between sales volume and sales rank (April 2005; double logarithmic graph)

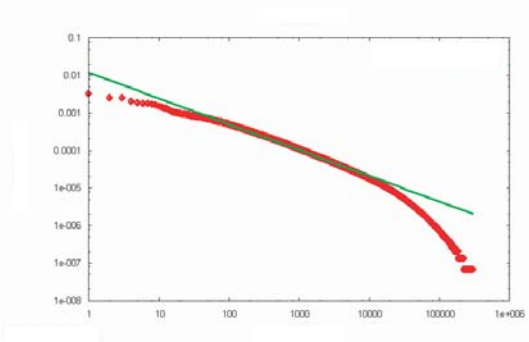

Figure 5: The relation between sales volume and sales rank (May 2005; double logarithmic graph)

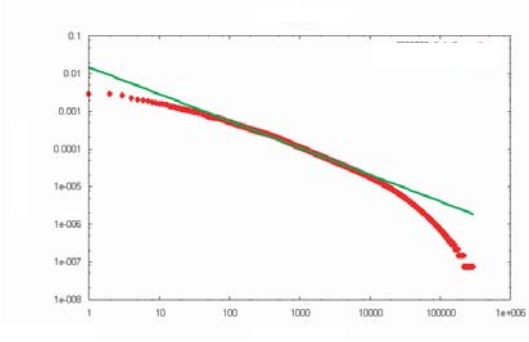

Figure 6: The relation between sales volume and sales rank (June 2005; double logarithmic graph)

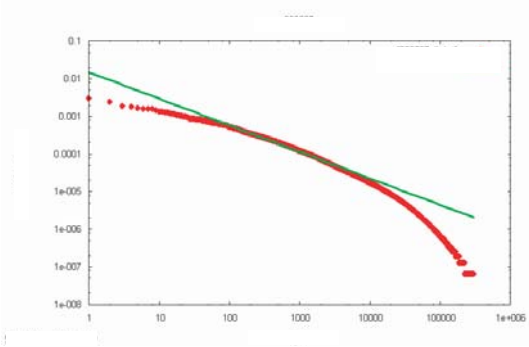

Figure 7: The relation between sales volume and sales rank (July 2005; double logarithmic graph)

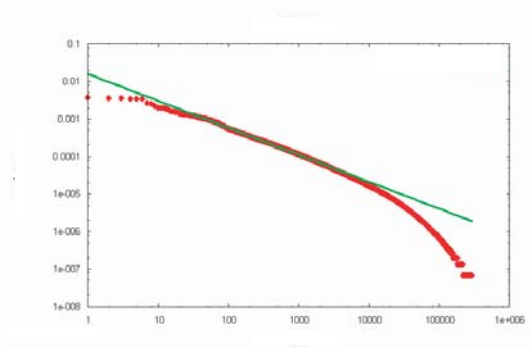

Figure 8: The relation between sales volume and sales rank (August 2005; double logarithmic graph)

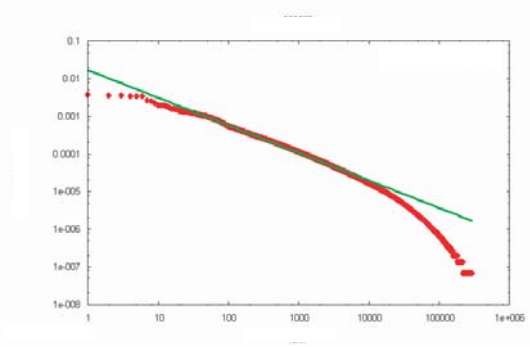

Figure 9: The relation between sales volume and sales rank (September 2005; double logarithmic graph)

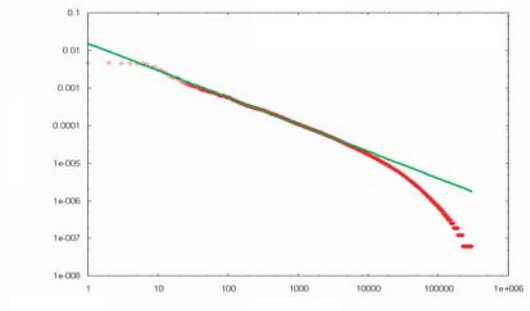

Figure 10: The relation between sales volume and sales rank (October 2005; double logarithmic graph)

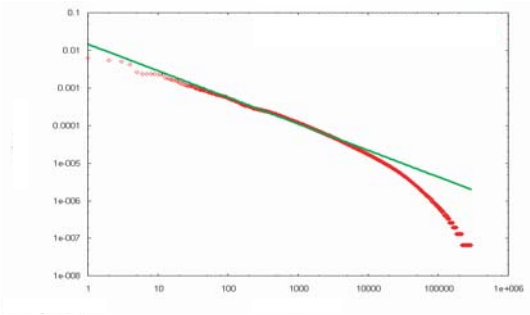

Figure 11: The relation between sales volume and sales rank (November 2005; double logarithmic graph)

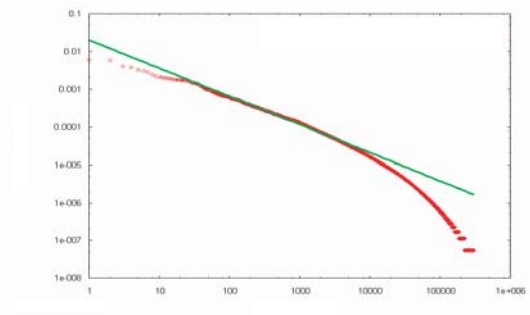

Figure 12: The relation between sales volume and sales rank (December 2005; double logarithmic graph)

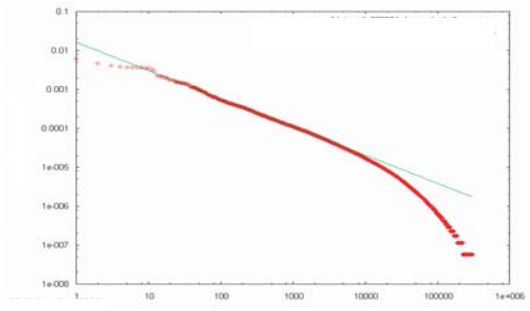

Figure 13: The relation between sales volume and sales rank (January 2006; double logarithmic graph)

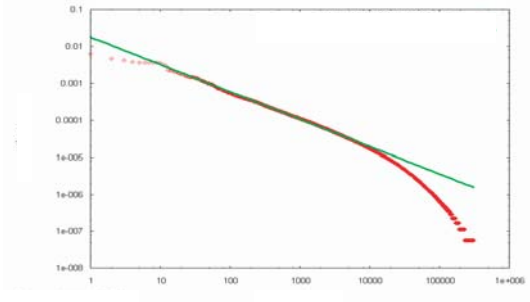

Figure 14: The relation between sales volume and sales rank (February 2006; double logarithmic graph)

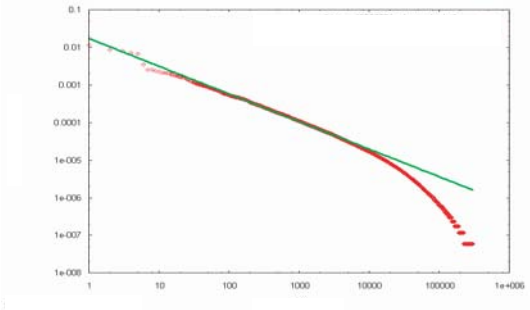

Figure 15: The relation between sales volume and sales rank (March 2006; double logarithmic graph)

Table 1: Power index and market share of upper 1.5% titles

target period	the number of upper 1.5% titles	power index of upper 1.5% titles	market share of upper 1.5% titles
2005s	8324	0.648	48.7465
2005/04	4450	0.684	47.3896
2005/05	4487	0.688	47.2258
2005/06	4411	0.714	48.2682
2005/07	4490	0.706	50.2240
2005/08	4469	0.719	51.5646
2005/09	4397	0.733	51.2183
2005/10	4512	0.718	51.4517
2005/11	4413	0.706	51.0998
2005/12	4461	0.744	53.5941
2006/01	4542	0.725	52.0992
2006/02	4684	0.739	52.5805
2006/03	4536	0.735	52.7173

3.3 Analysis of sales in genres

As a result of analysis in each genres, Figure 16 to 26 show that the relations between sales volume and sales rank in each genres also follow power law. We, here, would like to focus on two genres, Chemistry (Figure 21) and Physics (Figure 22). In the genre of Chemistry, most of the top titles seem to be purchased as textbooks. On the contrary, in the genre of Physics, the top titles are the books about the introduction to Einstein theory or quantum theory. This characteristic may make the difference between the sales volume of the top titles in Chemistry and Physics.

3.4 Analysis of the tendency of top titles

Power law distribution is known to be much different from the normal distribution, therefore statistical values of "average" and "variance" do not make sense under the power-law distribution. For that reason, the other way of capturing the characteristics is required, and the power index β is used as one of the indicator of the distribution.

Strictly speaking, it should be noted that the power index varies depending on the range of approximation, so it is necessary to make a specific baseline for approximation. In order to make the baseline, we investigated the part that along to the vertical axis in the graph. After some explo-

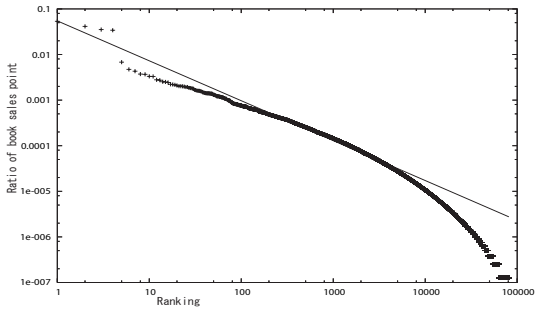

Figure 16: The relation between sales volume and sales rank of "Literature" (May 2006; double logarithmic graph)

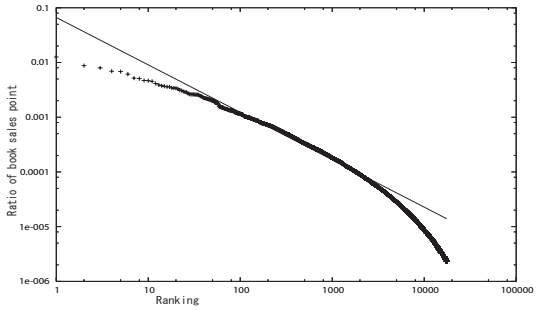

Figure 17: The relation between sales volume and sales rank of "Japanese Literature" (May 2006; double logarithmic graph)

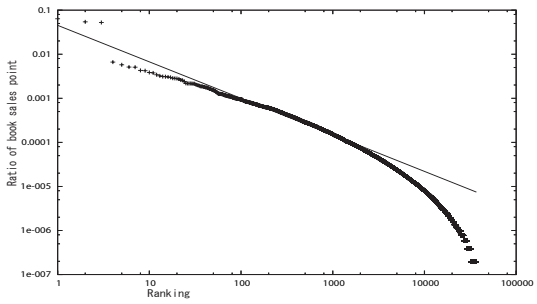

Figure 18: The relation between sales volume and sales rank of "Paperback Novels" (May 2006; double logarithmic graph)

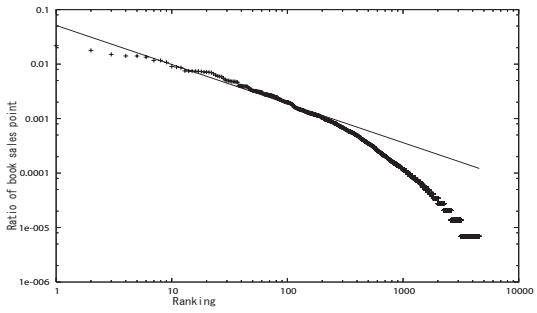

Figure 19: The relation between sales volume and sales rank of "Economics and finance" (May 2006; double logarithmic graph)

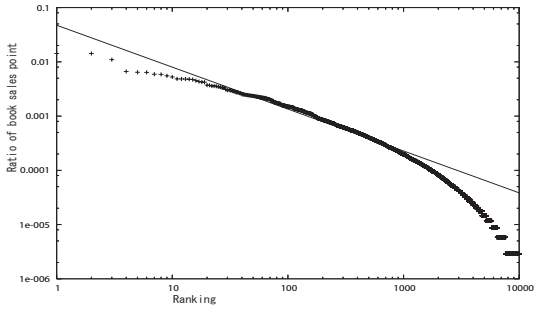

Figure 20: The relation between sales volume and sales rank of “Management” (May 2006; double logarithmic graph)

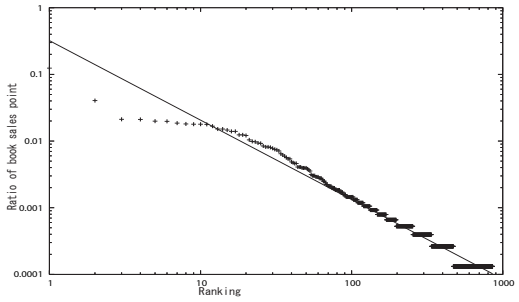

Figure 21: The relation between sales volume and sales rank of “Chemistry” (May 2006; double logarithmic graph)

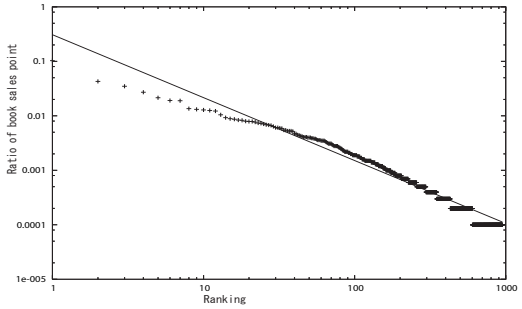

Figure 22: The relation between sales volume and sales rank of “Physics” (May 2006; double logarithmic graph)

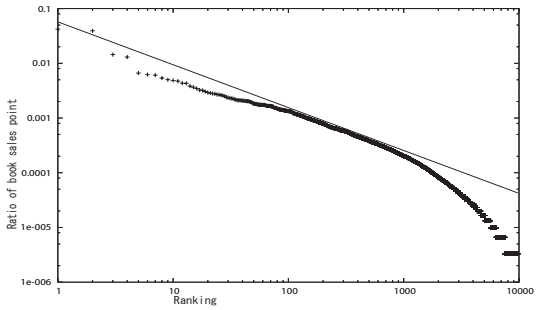

Figure 23: The relation between sales volume and sales rank of “Information and electronics” (May 2006; double logarithmic graph)

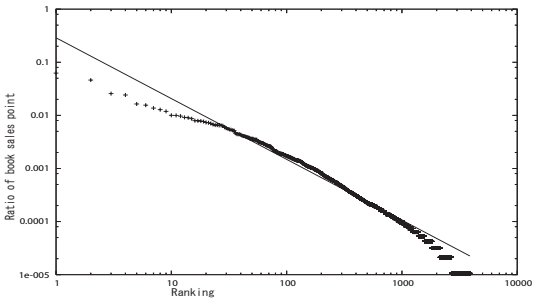

Figure 24: The relation between sales volume and sales rank of “Drawings and scalpture”

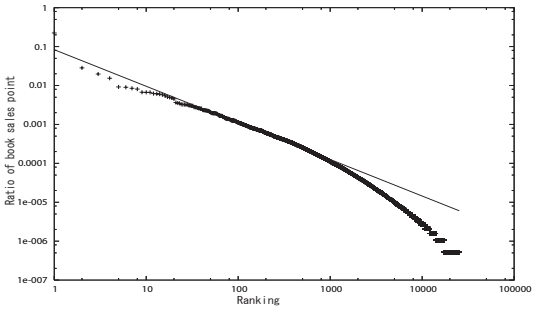

Figure 25: The relation between sales volume and sales rank of Hardcover (May 2006; double logarithmic graph)

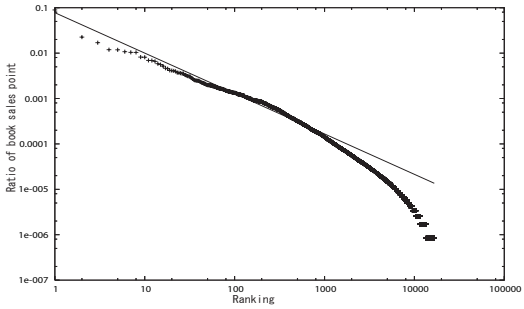

Figure 26: The relation between sales volume and sales rank of Softcover (May 2006; double logarithmic graph)

ration, we found that the range takes over upper 1.5% titles (Figure 27), and these titles make up the share of 50% in book sales market. Thus we decided to set a baseline to 1.5% in this paper). The dotted line in the figures shows the boundary of upper 1.5% titles. Note that upper 1.5% titles indicate about 8300 titles in a year, and about 4400 titles in a month[4] .

The power index of upper 1.5% titles and market share in each months are shown in Table 1. First of all, we find that the number of power index increases gradually (Figure 30). It means that the power index is getting to be bigger, the bigger difference is made among the top titles. Second, we find that the market share of top titles increase more (Figure 31). It means that the gap between the top titles and the other titles are widen.

Combining the results above, it can be drawn as the relation between the change of power index and that of market share (Figure 32). We conclude that the upper 1.5% titles have the tendency to be sold more and more by the analysis of sales from April 2005 to March 2006.

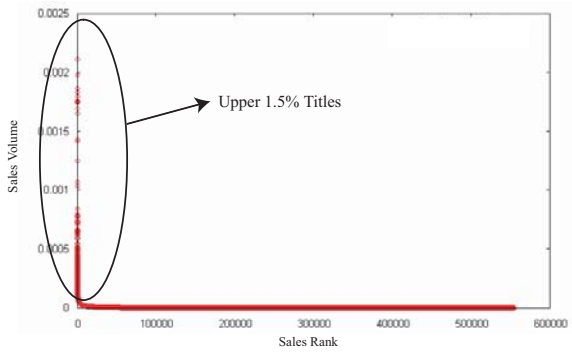

Figure 27: Upper 1.5% titles as a baseline for approximation (from April 2005 to March 2006: linear graph)

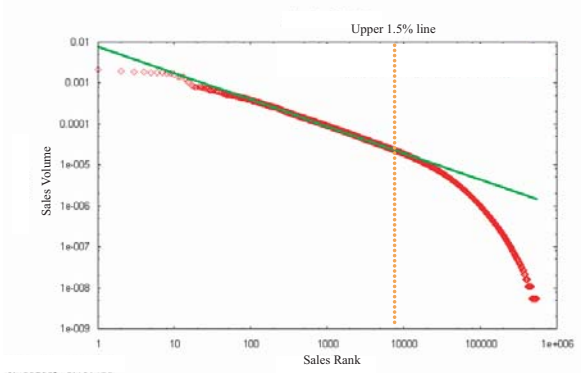

Figure 28: Upper 1.5% titles as a baseline for approximation (from April 2005 to March 2006: double logarithmic graph)

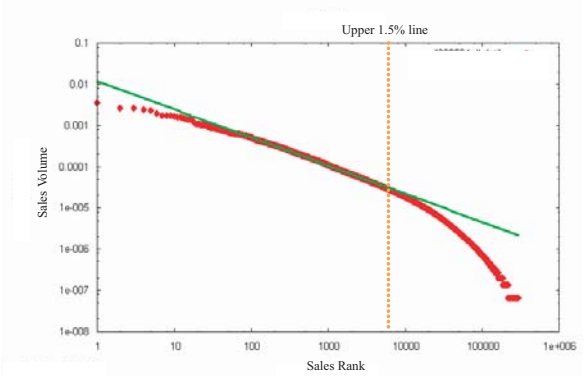

Figure 29: Upper 1.5% titles as a baseline for approximation (April 2005: double logarithmic graph)

4 Discussion: New Approach to Understand Market and Consumers Behavior

We have investigated the book sales market and verified that the relation between sales volume and sales rank follows power law by analyzing real sales data. From the analysis, now we have chance to understand the mechanism of market and consumers behavior based on power-law distribution. In this section, we would like to discuss the future perspective of the study on market and consumers behavior.

When the consumers purchase products, they are mostly influenced by friends' opinions, window advertisements, or mass media (Figure 33). The choice of products by each consumers is "contingent", which means that the individual result of selection is not inevitable and it can be otherwise. However, the order in the market is generated by accumulating of the result of the contingency. It is because the exposure of top sales affects the choice of other consumers. The kind

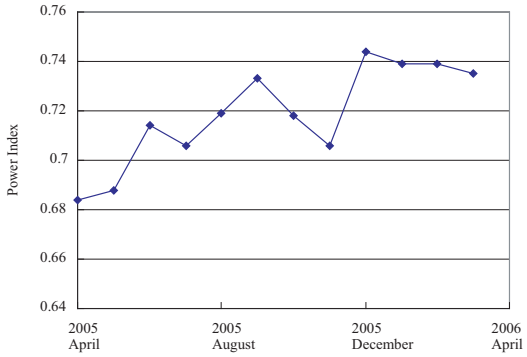

Figure 30: The transition of power index of upper 1.5% titles (from April 2005 to March 2006)

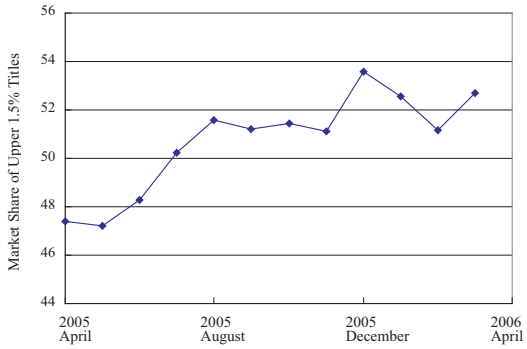

Figure 31: The transition of market share of upper 1.5% titles (from April 2005 to March 2006)

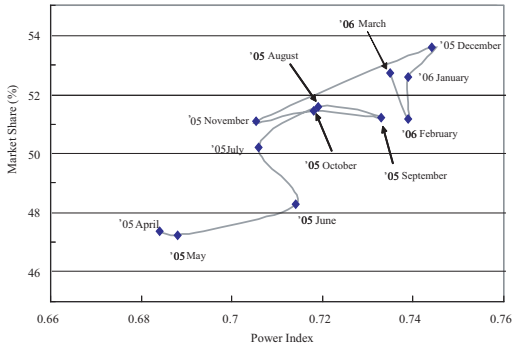

Figure 32: The relation between power index and market share of upper 1.5% titles (from April 2005 to March 2006)

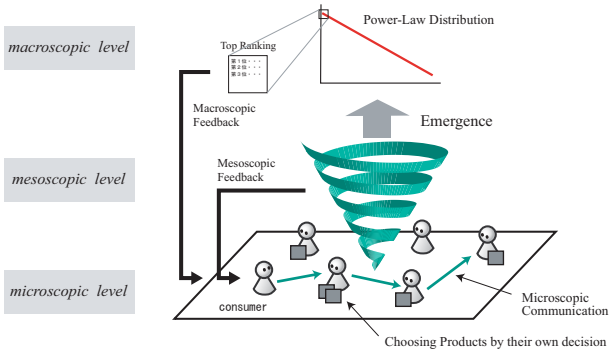

Figure 33: Multilayered interactions in market

of multilayered interactions make the market self-organize to critical state.

The fact that the law we found in the market is "power law" implies that the fundamental principle of self-organization is similar to other natural and social phenomena that follows the power law. From the principle of "universality" in non-equilibrium physics, the phenomena which belong to the same class of universality work in the same mechanism to form the order. It means that there is chance to understand the market mechanism analogically by knowing that of other phenomena, like avalanches on sandpile, earthquakes, or growth of cities. Here is a new way to understand market and consumers behavior (Figure 34).

This approach can lead us to understand the mesoscopic mechanism, which it is difficult to investigate in the study of consumers behavior due to the difficulty of observation. As a conclusion, finding the power-law distribution in the market is important not only because it is the "emergent order", but also because it opens up new chance to study market and consumers behavior with using the analogy from systems belonging to same class.

5 Conclusion

In this paper, we verified the power-law distribution between sales volume and sales rank in book sales market. From the results, we can suggest that the consumers behavior are deeply affected by the market's organizing principle. Therefore, now we have started to develop the multi-agent

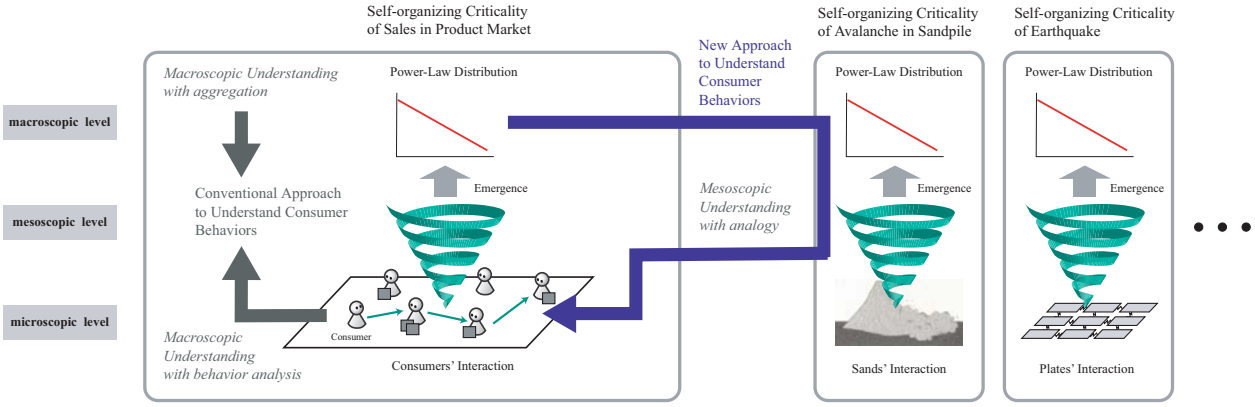

Figure 34: New approach to understand market and consumers behavior

simulation model to understand the mechanism of generating power-law distribution in the market. It is assumed that the consumers' choices are influenced by the information from mass media and their friends. In further studies, we are going to investigate this hypothesis.

Notes

1 The related researches which focus on the other viewpoints are done by Sornette and Deschatres [6, 8], there are some analysis about the tip-tops of books sales, and its forerunner and aftershocks, using the data of ranking system in Amazon. Lambiotte and Ausloos [9] also analyzed the peak of book sales and endogenous peak using the data from Amazon's ranking system. This analysis is addressing the distinction of the peak in the long time scale, so we take this for related research. Furthermore, Brynjolfsson *et. al.*[10] estimated that the online bookstore has wider variety than the real bookstore . But it is suggested that this estimation does not show the real data, because the supposition of algorithm differ from the actual condition.

2 Rosenthal pointed out that the graph which shows the relation between sales volume and sales rank is his "personal guesstimate," and any official recognition or authorization is not given by Amazon [7].

3 The analyses in this paper were originally reported in the papers[11, 12, 13] in Japanese.

4 Upper 10% titles make up the share of 85% (Figure 28), and upper 20% titles make up that of 95% in the market (Figure 29). It means that the book sales market is really "Winners-Take-All" market.

References

[1] Mark Buchanan. *UBIQUITY*. Wiedenfeld & Nicolson, 2000.

[2] M. E. J. Newman. Power laws, pareto distributions and zipf's law. *arXiv:cond-mat/0412004v3*, 2004.

[3] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *arXiv:0706.1062v1*, 2007.

[4] Per Bak. *how nature works*. Springer-Verlag, 1996.

[5] Robert H. Frank and Philip J. Cook. *The Winner-Take-All Society*. The Free Press, 1995.

[6] D. Sornette, F. Deschatres, T. Gilbert, and Y. Ageon. Endogenous versus exogenous shocks in complex networks: an empirical test using book sale ranking. *Physical Review Letters 93 (2004) 228701*, 2004.

[7] Morris Rosenthal. *What Amazon Sales Ranks Mean*. 2006. http://www.fonerbooks.com/surfing.htm.

[8] F. Deschatres and D. Sornette. Dynamics of book sales: Endogenous versus exogenous shocks in complex networks. *Physical Review E 72 (2005) 016112*, 2005.

[9] R. Lambiotte and M. Ausloos. Endo- vs. exogenous shocks and relaxation rates in book and music "sales". *Physica A 362*, pp. 485–494, 2006.

[10] Erik Brynjolfsson, Yu Hu, and Michael D. Smith. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, Vol. 49, No. 11, 2003.

[11] Takashi Iba, Yoshiaki Fukami, and Masaru Saitoh. Hidden law in the book sale market in japanese. In *IPSJ SIG Technical Reports, 2006-MPS-61*. The Information Processing Society of Japan, 2006. in Japanese.

[12] Takashi Iba, Yoshiaki Fukami, Mariko Yoshida, Kohei Yamashita, and Masaru Saitoh. Analyzing the sales distribution in book sale market. 2007. in Japanese.

[13] Takashi Iba, Yoshiaki Fukami, and Masaru Saitoh. Hidden law in the book sale market in japanese. *Transactions on Mathematical Modeling and its Applications*, Vol. 48, No. SIG 6 (TOM 17), 2007. in Japanese.

* *This paper is for Fourth Joint Japan-North America Mathematical Sociology Conference, May, 2008.*
(Contact Us: E-mail to iba@sfc.keio.ac.jp)