

慶應義塾大学経済学部研究プロジェクト

最終成果論文（2011年度）

ロングテール構造の時間変化の分析

経済学部 3年

山下紘史

服部哲弥

要旨

インターネットが社会に浸透し人々の生活との関係が深まることによって、多くの人々が関わり、かつ、順位を付けることに意味のある、巨大なデータを扱ったビジネスが現れるようになり、大きな順位の分析をすることが意味のある時代となった。順位対大きさの関係をパレート分布に当てはめて分析することが、ロングテールビジネスなどの観点から、経済学や経営学上の興味の対象になりつつあると考える。ロングテールとは、膨大な商品を横軸に、その売れ行きを縦軸に大きい順に並べたときにあまり売れない商品で構成される、右側になだらかに長く伸びる部分のことである。ロングテールビジネスとは、売れない商品ひとつひとつの売上は無視できるほどであるが、それらが束になることで売上に貢献するために、ロングテール部分の商品を扱うビジネスのことである。

いわゆるロングテールビジネスが成熟する中で、巨大な順位、特に、下位のランキングが公表されていることはごく稀であり、ロングテールの成立を調査するのに適した時代になったとはいっても研究対象を見つけるのはまだ困難である。以上の背景認識の下で、実際にアクセス可能で興味深いデータを見つけること、および、そのロングテールという視点での、先行研究よりも詳しい研究を行うことを目指した。

本稿では、時間の経過によって分布がロングテール化してゆくのではないかと予想し、膨大な数の動画が投稿されているニコニコ動画における再生数ランキング、大勢のメンバーによって構成されているAKB48の選抜総選挙について、集積期間を複数設けて観察を行なった。そして、得られたデータを実際にパレート分布 ($y = ax^{-\frac{1}{b}}$) に当てはめることで、特に指数 b を得ることで、各データでロングテールが実現しているかどうかを検討した。

ニコニコ動画の再生数ランキングを分析した結果、分布のロングテール化はひと月でもほとんどその変化が分からないほど極めてゆっくり進むことが分かった。そして、AKB48の選抜総選挙を分析したところ、順位による明確な差別化や中間発表の存在が人々の心に作用し、分布に影響を及ぼしていると思われる結果を得ることができた。そのため、ロングテール化は長い年月をかけて進むが、人々の心理的要因が分布に無視できない影響を及ぼすということが分かった。また両研究対象のすべての期間の結果において、1つのべきの曲線 ($y = ax^{-\frac{1}{b}}$) で当てはめた場合、下位にいくほど近似曲線との乖離が見られたため、複数のべきの曲線をつなぎ合わせることで分布をより正確に表すことにした。すると、下位にいくほど不平等つまりロングテールが成立しなくなるという結果を得ることができた。

目次

1.	はじめに	1
2.	第1章 研究の背景と先行研究.....	2
2.1.	テーマの導入.....	2
2.2.	問題提示	3
2.3.	方法の提示	4
2.4.	展開の予告	6
3.	第2章 分析対象の紹介	7
3.1.	なぜニコニコ動画、AKB 選抜総選挙なのか.....	7
3.2.	ニコニコ動画.....	7
3.3.	AKB 選抜総選挙.....	8
4.	第3章 ニコニコ動画再生数ランキングの分析	9
4.1.	分析方法.....	9
4.2.	分析結果.....	10
4.3.	考察	19
4.3.1.	観測期間の違い.....	19
4.3.2.	活動の昼夜差	19
4.3.3.	指数 b の妥当性	20
4.3.4.	指数 b の見方	20
4.3.5.	「フロー」と「ストック」	21
4.3.6.	データ比較の難しさ	21
4.3.7.	それでもロングテール化は進む	22

4.3.8.	注意点	22
5.	第4章 AKB 選抜総選挙の分析	23
5.1.	分析方法	23
5.2.	分析結果	24
5.3.	考察	29
5.3.1.	ファン心理を揺さぶる選挙方式	29
5.3.2.	ロングテールへと向かう力と投票者へ与える影響	29
5.3.3.	非公表の得票分布を暴く	30
5.3.4.	データの揺らぎ	30
5.3.5.	各チームが占める人数	30
5.3.6.	ビッグヒットが AKB 人気を引っ張る	31
6.	第5章 二つの分析対象からの知見	32
6.1.	これまでの分析のまとめ	32
6.2.	論議	32
7.	付録	34
7.1.	大数の弱法則 (WEAK LAW OF LARGE NUMBERS)	34
7.2.	AKB 選抜総選挙の公式発表	35
8.	おわりに	39
8.1.	iPod 内の音楽再生数ランキング	39
8.2.	研究を終えて	40
8.3.	謝辞	41

9. 参考文献・データ出典42

1. はじめに¹

インターネットの成熟により、膨大な数のコンテンツを扱うビジネスが人々の間に浸透することで、いわゆるロングテールビジネスが成立しているかを判断することが意味のある時代となった。だが、膨大な商品を扱い、ロングテールが成立していると言われている多くのウェブサイトでは上位のランキングは発表されているものの、テール部分の情報が開示されているものはほとんどない。これは、テール部分の順位に興味を持つ人などほとんどいないと思われることからある意味当然といえば当然である。

膨大なコンテンツの中で、**350717**位と**350716**位の違いが何なのかについて興味がある人など想定できるはずもないだろう。そのため、研究する意味のある時代になったとはいっても、実際にテール部分の情報を探すことは予想以上に難しかった。しかし、友人からアドバイスをもらうなどして懸命に探した結果、ニコニコ動画という動画提供サービスのウェブサイトに行き着き、他より広範囲でランキングが公表されていたり、また総動画数や総再生数も公表されていたため動画数と再生数についての分布でテール部分の推測を行なうこともできた。

また、ファンからの人気投票でランキングをつけているアイドルグループの **AKB48** についてもメンバーの順位と得票数の関係についてロングテールが成立しているのか否かの調査を行なうことができた。このアイドルグループは所属するメンバーの数が **100** 人を超えており、このような大規模なアイドルグループで、しかも不特定多数のファンによる順位付けが発表されることはこれまでになかった事例として興味がある。

なお、本稿において「インターネット」「ウェブページ」などの情報処理に関する技術的な用語を曖昧に使う。多数の人の参加と大量の情報的高速かつ継続的な更新だけが本稿では重要であり、それ以外の情報処理技術的なことは本稿の議論には関係ない。

¹本稿の作成にあたっては、担当教員である服部哲弥教授（慶應義塾大学）をはじめ、多くの方々から有益かつ熱心なコメントを頂戴した。ここに記して感謝の意を表したい。しかしながら、本稿にあり得る限り、主張の一切の責任は言うまでもなく筆者個人に帰するものである。

2. 第1章 研究の背景と先行研究

2.1. テーマの導入

今日のインターネットの浸透により世の中がますます便利になっている。その上 Wi-Fi の普及やスマートフォン の普及により、いつでもどこでもインターネットに繋がる社会が実現するようになった。家にいながらなんでもできるという時代から、どこにいてもなんでもできるといった時代に変化し、場所を選ぶことなく知りたいことを調べたり、世界中の人とコミュニケーションをとることができるようになっただけでなく、本や日用品を買ったり、動画や音楽などデジタルのコンテンツを視聴したり即購入することすらできる。このようなネット上のコンテンツの最大の特徴は、これまで想像もつかなかったほどの膨大なデータである。インターネットが普及する前の時代では店頭販売では保管コストの面から扱われていなかったほとんど売れないマイナーな商品も今日のネットショップ上にはある。

皆さんは Amazon²というサイトを知っているだろうか。近所の書店ではみつけることができなかった書籍を調べてみて欲しい。新品、中古に関わらず見つけることができるはずだ。実際に私もこのような経験がある。とりあえず Amazon にいけばあるだろう、こういった期待を持っている人々は決して私だけではないはずだ。その上、この新たにネットに出現した膨大な在庫を扱う店舗の出現は、一消費者の生活を変えるだけでなくビジネスの面にも劇的な変化を与えている。マイナーな商品も大量に扱うこの新しいビジネスは、米雑誌大手 Wired の編集長クリス・アンダーソンによって著書「ロングテール―「売れない商品」を宝の山に変える新戦略」（訳:篠森ゆりこ、早川書房、2006年）において“ロングテールビジネス”と名付けられ脚光を浴びている。確かに、人々がニッチな商品を需要していることに対してそのニーズを満たしていることから売上に貢献しているだろうといった想像はつく。しかし、このビジネスは果たして成立するのだろうか。ニッチな商品の売上は全体の売上にどれほど寄与しているのだろうか。この問いに対して、私の担当教授である服部哲弥氏は数学的手法から Amazon の和書について研究していることを知った。著書「Amazon ランキングの謎を解く 確率的な順位付けが教える売上の構造」（化学同人、2011年）によれば、Amazon においてロングテールビジネスは成立していないとの結論がある。Amazon では膨大な商品の売上のランキングがリアルタイムで公表されているのであるが、

² インターネット上で膨大な商品を扱った商取引を行なうロングテールビジネスの代表例とされている。

服部氏は、ニッチな部分にある本一点のランキングの時間推移を観察することだけで書籍の売上全体の分布を鮮やかに照らし出した。ただ、私自身に数学の知識が不足しているためにパレートの分布³やジップの分布⁴その他この本に書かれている内容をあまり理解することができなかつたので、今回まず数学の知識を身につけることが必要であった。

2.2. 問題提示

先ほど紹介したロングテールの名付け親、クリス・アンダーソンは著書において、ロングテールビジネスの代表例として Amazon を挙げている。しかし、これに対して服部哲弥氏はその著書において数学的手法を用いて Amazon の和書のカテゴリーではロングテールビジネスは成立していないと述べている。服部氏は、Amazon の和書を主たる研究対象とし、数学的手法を用いてめったに売れないニッチな商品一点の順位を観察するというだけで全体の売上の様子を浮かび上がらせている。そして明らかになった売上の様子をべきの曲線 ($y = ax^{-\frac{1}{b}}$) で近似することによって、上位と下位の売上への貢献度合いをパレートの平等の指数といったもので数値化し対象がロングテールか否か判断している。この分析によれば Amazon の和書のカテゴリーにおいてロングテールは成立していないという驚くべき結果が得られたというのだ。著書において、

「b は社会的平等性の大きさを表す指数と見ることにする。…とくに平等性の指数 b が 1 より小さい点に注目する。これはアマゾン書店の現実の売上構造においてビッグヒットとロングテールのあいだに大きな売上の格差があることを意味する。総点数 N が大きいから 1 位と最下位の差が大きいのは当然であるが、ここでいう格差は、単に一番売れる本と一番売れない本の格差という意味ではない。ロングテールビジネスモデルの主張は、お荷物となる下位（ロングテール）の本も多数集めれば売上への寄与が無視できない、ということであった。… $b < 1$ ということは書籍点数の大部分を占める普通の本が点数 N の大きさを頼りに束になってかかっても、ひと握りのビッグヒットの売上にはかなわない。この結果を見る限り、書籍に関してロングテールビジネスモデルは成立しないことが、データと理論をつき合わせて見出されたことになる。」

³ 連続型の確率分布。経済学者ヴィルフред・パレートが個人所得高の分布において発見した。この分布は経済学、社会学様々な分野で見出されている。

⁴ 離散型の確率分布。

と記されている。

このように、インターネットの普及により現れるようになった膨大な商品を扱うビジネスが実際に成立しているのかということについて、順位対大きさの関係をパレート分布に当てはめて分析することが、ロングテールビジネスが成立するの否かといった観点から、経済学や経営学上の興味の対象になりつつあるのではないか。

しかし、本稿を執筆しようと決めた当時、なぜべきの曲線で当てはめることができるのかについては理解できていなかった。様々な人が関わってできた分布が1つの曲線で近似できてしまうのかという疑問も抱いていた。

クリス・アンダーソンをはじめ、今ではロングテールビジネスが到来したといった内容の書籍が様々な出版されている。そんな中、服部氏はロングテールは成立していないと主張している。他にも、Amazonのテール部分の商品が注文してみると在庫切れとなっている状況や売り切れの商品があたかも在庫があるかのようにそのままランキングに残っているとといったことが報告されていたりとロングテールの成立には様々な意見が存在している。

それに対し、私は売上頻度の分布に関する法則があると考え、ロングテールと言われている他のものについては一体どのような結果が出るのか、人々の好みに定量的な傾向はあるのかといった疑問が浮かび、データからそれを推測することにした。また、ロングテールの成立とは人々の興味の対象がニッチ部分へ向かうということであると思ったため、商品の売上の形がロングテールになるには時間がかかるのではないかといった仮説をたてた。つまり観測時間が長くなるほどロングテールに近づいてゆくのではないかとということである。

ロングテールの成立に対して観測時間の長さに注目して分析を行なっていることについてはこれまでになかった視点であり、ここに本稿の新規性があるのではないかと考えている。また、のちに登場する研究対象であるニコニコ動画、AKB48についての分析をここまで厳密に行なった論文はなくこの点にも新規性があるのではないかと考えている。

2.3. 方法の提示

分析の方法が正当化される根拠となる数学の知識を得るために、服部氏の研究を拝見し、用いられている数学の知識（大数の法則）を服部氏の著書「Amazon ランキングの謎を解く 確率的な順位付けが教える売上の構造」や熊谷隆氏の「〈シリーズ新しい解析学の流れ〉『確率論』」（共立出版、2003年）を

用いて学んだ。本で理解しにくいところは担当教員である服部氏に直接伺って補うという形をとった。

では、先ほど述べたように膨大な数の商品の売上がなぜべきの曲線で近似できるのか、そのことを理解するために必要な数学の知識などについてこれから説明していく。

まず、数学の確率論という分野において、大数の法則といったものがある。よく用いられる例としては、さいころを投げる回数を増やしていくと各目が出る確率は6分の1に限りなく近づいてゆくといったものである。これは、実数値での大数の法則である。付録にて、数式を使った証明を載せておく。しかし、今回の方法では、ここから一段階上がって、先ほどの実数を‘分布’に置き換えた大数の法則を用いてべきの分布で近似できるということを用いて正当化していることになる。先程の例だと、さいころの出る目のところが数字ではなく分布になっているといえる。このことにより、膨大な数のデータの分布がべきの曲線になるとした。

服部氏の分析の中核部分である、ランキングの時間変化から売上の分布を引き出すのは数学的には高度だが、ロングテールが社会的に重要かどうか（たとえば、ロングテールビジネスが成り立つかどうか）に直接関わるのは売上と順位の関係、典型的にはべき分布に当てはめたときの指数の大小、である。

ビッグヒットは時間が経てば変化してゆくから、特定の本がどれくらい売れるかを予想する際の数学的単純化は難しい。しかし、商品の売れ行きの分布だけが問題ならば、思い切って単純化したところから研究を始めることができる。

服部氏の分析は売上の分布が公開されていないときに、ランキングという公開情報からそれを推測する上で有効だが、今回の研究ではそれは用いず、より直接的に分布が公開されている対象を探すことにした。

すると、この観点から興味深いデータで実際にアクセス可能なデータをいくつか見つけることができた。つまり、服部氏の分析と異なる対象を検討することで、べき分布の指数の値に、対象による異同がどのようにあるのかについて、相補的な研究を行ったことになる。そして、実際にパレート分布に当てはめて、特に指数を得ることでロングテールが実現しているかどうかを検討した。また、より直接的に分布を扱うことで、単純なべき分布よりも細かい分布の構造に立ち入ることもできた。

あるものを順位付けしてグラフにした時、それがロングテールかそうでないかの判断は、ロングテールのそもそもの定義である、べきの曲線 ($y = ax^{-\frac{1}{b}}$) の傾き具合を示す指数の値（先ほどの式の b に該当）を用いる。ロングテールの提唱者である米雑誌 **Wired** の編集長クリス・アンダーソンがロングテールは **Zipf** の法則に従っていると著書「ロングテール―「売れない商品」を宝の山に変える新戦略」にお

いて述べていることから、べきの曲線で近似するということは妥当であると言えよう。

実際、べきの曲線は、社会活動のあらゆるところにおいて大勢の人間が関わって成り立つ現象で観察できることが確認されている。社会学や経済学で扱うような、いろいろな研究で用いられる分布の形なので、本質的に同じ法則に見かけ上違った名前がつくことも多く、例えば人口と所得の関係でパレート分布が用いられていたり、地震の規模や都市の人口、単語の使用頻度などでジップの分布が用いられている。

パレートの分布とジップの分布はどちらもべきの曲線であるが、違いはパレートの分布は連続分布であり、ジップの分布は離散型の分布である。変数変換によって両者は同じ形になることが示されている。

2.4. 展開の予告

これまで、本研究を行なうに至った経緯を述べてきた。以下では、研究内容について述べてゆく。今回、分かりやすさを考慮して研究内容の2つを分けてそれぞれで分析方法、結果、考察を記述した。そして、その後に2つの研究結果からみえてくることについて記述を行なうという形をとった。

第2章において研究対象を選んだ経緯やその詳細について述べる。第3章において1つ目の対象であるニコニコ動画における再生数ランキング、第4章において2つ目の対象であるAKB選抜総選挙についてそれぞれ考察まで行なう。第5章で両分析対象の結果の相違点や、本研究についての議論を批判的に行なっている。

3. 第2章 分析対象の紹介

3.1. なぜニコニコ動画、AKB 選抜総選挙なのか

インターネットの成熟やそれによるコストの削減、またビジネスの多様化で現在では膨大な数を扱うビジネスが行なわれている。しかし、実際に下位の部分の情報は公開されておらず、そのようなデータの詳細を観察するのは困難を極めた。その中で、ニコニコ動画では総動画数が 700 万点、総再生数が 290 億回を超えており（2012 年 1 月現在）、またランキングは上位 1000 位までが公表されていた。そのため、公表されている部分の分布の観察、また公表されていない部分についても分布の推測を行なうことができた。

また、インターネット上のビジネスではないものの、AKB48 といったこれまで芸能界で見られなかった（自分が知らない時代に同じく秋元康氏がプロデュースするおニャン子クラブというグループがいたらしい）数十人という大人数のグループでの活動は非常に特徴的なビジネスだと映り、総選挙という人気投票を行なってランキングをつけていたことから、メンバーと得票数の分布を観ることが可能なため、今回の調査対象とすることにした。

では、今一度、ニコニコ動画、AKB48 の選抜総選挙について説明をしておく。

3.2. ニコニコ動画

ニコニコ動画は、ドワンゴが提供している動画共有サービスである。同種のサービスとしては YouTube などがある。

ニコニコ動画では、一時間ごとに総動画数、総再生数、総コメント数が更新されており、またランキングは毎時、デイリー、週間、月間、合計の期間で分けられており、それぞれの期間での再生数のランキングを見ることができる。毎時の場合は TOP100、デイリーの場合は TOP300、週間、月間、合計の場合は TOP1000 までの動画が載せられている。

合計の場合は、6つのカテゴリーに分けてそれぞれについて上位 300 位までが公表されているので、それら 6つを 1本にまとめることによって上位 1000 位以降のデータも入手することができた。デイリーランキングは毎日午前 6 時更新、週間ランキングは毎週月曜日午前 6 時更新、月間ランキングは毎月

1日午前6時更新となっている。

観測した期間は、毎時は、12月1日の0時～1時、6時～7時。デイリーは12月1日。週間は12月5日～12日。月間は、2011年12月。合計は、2011年10月23日に記録している。

データは、ニコニコ動画のランキングのWebサイト上 (<http://www.nicovideo.jp/ranking>) に全て公表されている。

3.3. AKB 選抜総選挙

AKB48 は、現在巷を賑わせている国民的アイドルグループである。秋元康氏のプロデュースのもと、これまで考えられなかった新しい試みが多くなされている。グループの人数が非常に多いことからメディアに出演できるメンバーやニューシングルに参加するメンバーを絞るため、シングル CD の購入者やファンクラブの会員などに投票権が与えられている。この企画は 2009 年から始まっており、投票対象は、AKB48、SKE48、NMB48 のメンバーおよびその研究生で総勢 150 名余りとなっている。

この選挙がよく知られているような現在の国勢選挙と大きく異なる点は、一人一票とは限らないことである。CD を複数枚買ったり、ファンクラブの会員でありながら CD を購入する（多くのファンクラブ会員は CD を買うだろう）ことによって自分の応援するメンバー（ファンの間では“推しメン”と呼ばれている）に多く投票することができる。

その他の特徴としては最終発表だけでなく速報、中間発表によっても上位 40 位のデータが公表されていることがある。そして、1 位から 12 位までがテレビに出演することができる“メディア選抜”、21 位までがシングル曲の A 面を歌うことができる“シングル選抜”、40 位までが B 面を歌うことができる“アンダーガールズ”だとあらかじめ公表されていることや、自分の“推しメン”の順位を上げたいなどの様々な要因から、速報や中間発表の結果を見て票を入れることができる。実際、中間発表と最終発表の間には得票数に大きな差があり、順位の変動も見られる。

上位 40 位までの得票の分布はマスメディアによって公表されているものの 41 位以下の得票の分布は公表されていない。しかし、本稿では総投票対象数と総投票数が公表されていることから 41 位以下の順位分布はジップの分布を仮定して、分布を推測することができた。

4. 第3章 ニコニコ動画再生数ランキングの分析

4.1. 分析方法

ニコニコ動画のサイトでは動画のランキングを見ることができる。ここでは1時間ごとに総動画数、総再生数、総コメント数が更新されている。公表されていないことであるが、更新のタイミングは30～40分後であることが観測の結果から分かっている。ニコニコ動画のランキングのページ (<http://www.nicovideo.jp/ranking>) にいくと、ランキングの期間や対象を自由に選択できるようになっている。対象には、総合、再生数、コメント数、マイリスト数の計4つがあるのだが今回は人々の関心をより直接的に観察できそうな再生数に絞って観測を行なった。また、期間は、**a.毎時**、**b.デイリー**（2012年から“24時間”に変わっている）、**c.週間**、**d.月間**、**e.合計**の5つがあり、今回それぞれで観測を行なった。観測期間は、**a.毎時**は2011年12月1日0時から1時までと6時から7時までの1時間の2通り、**b.デイリー**は12月1日、**c.週間**は12月5日から12日の1週間、**d.月間**は2011年12月の1ヶ月、**e.合計**は10月23日までのニコニコ動画の累計となっている。ここで注意が必要なこととして、デイリー以上の大きさの期間では、更新のタイミングが午前6時となっていることである。例えば、12月12日のデータとは、12日の午前6時から13日の午前6時までのことである。週間では月曜日の午前6時が区切りとなっており、月間では1日の午前6時が区切りとなっている。

今回、更新が午前6時というのは例えばデイリーだと1. 午前0時から次の午前0時の再生数を6時に更新しているのかそれとも2. 午前6時から次の午前6時の再生数のことなのかという疑問が浮かんだが、午前6時の更新でも30～40分のラグがあったことから2. であると判断した。

また、同じ再生数の動画はいくつか存在していたが、その際投稿日が新しい方が上の順位になっていたため同率0位といったものは存在しなかったと判断している。

今回の分析では、まず、縦軸に再生数、横軸にそれぞれの動画をとり、データを左から再生数が多い順に並べた。よくあることだが、ランキングのうち上位の部分は順位と再生数が公表されているのに対して下位は公表が省略されている。まず上位の公表されている部分については最小2乗法を用いてべき分布に当てはめを行なった。しかし、実際には一つのべき分布で当てはめようとすると、グラフで見て、当てはまりが良くなかった。そして、注意深く分析結果を眺めてみると近似曲線と実際の再生数

との乖離の仕方に共通の傾向があるように見えたので、公表されている上位部分のデータの中をさらに上位下位二つの部分に細分して、各々別のべきの分布に当てはめた。その分点は、まず分点を1つとると分点に対応した2つのべきの分布ができるので、それらの分布が実際の観測結果に最も当てはまる、つまりそのときの残差を最小にする点とした。これは、最小2乗法による統計的当てはめの意味では最善の分点である。

一方、公表されていないデータについては、総再生数と総動画数が公表されているので、総再生数から公表されている動画の再生数の和を引いた残りを、総動画数から公表されている上位を除いた残り動画数でべきの曲線で当てはめを行った。つまり、公表されている中で最も再生数が少ない動画を先頭とし、残りの動画数とそれらの再生数の和が明らかになっているために曲線を当てはめることができるということだ。この方法によって公表されていない部分についてのパレートの不平等性の指数 b を得ることができた。

以上のように、今回一つの対象に対して、三つのべきの曲線の当てはめを行っている。

4.2. 分析結果

a.~e.の各観測期間について、ひとつのべきの曲線で当てはめを行なった結果を以下に示す。

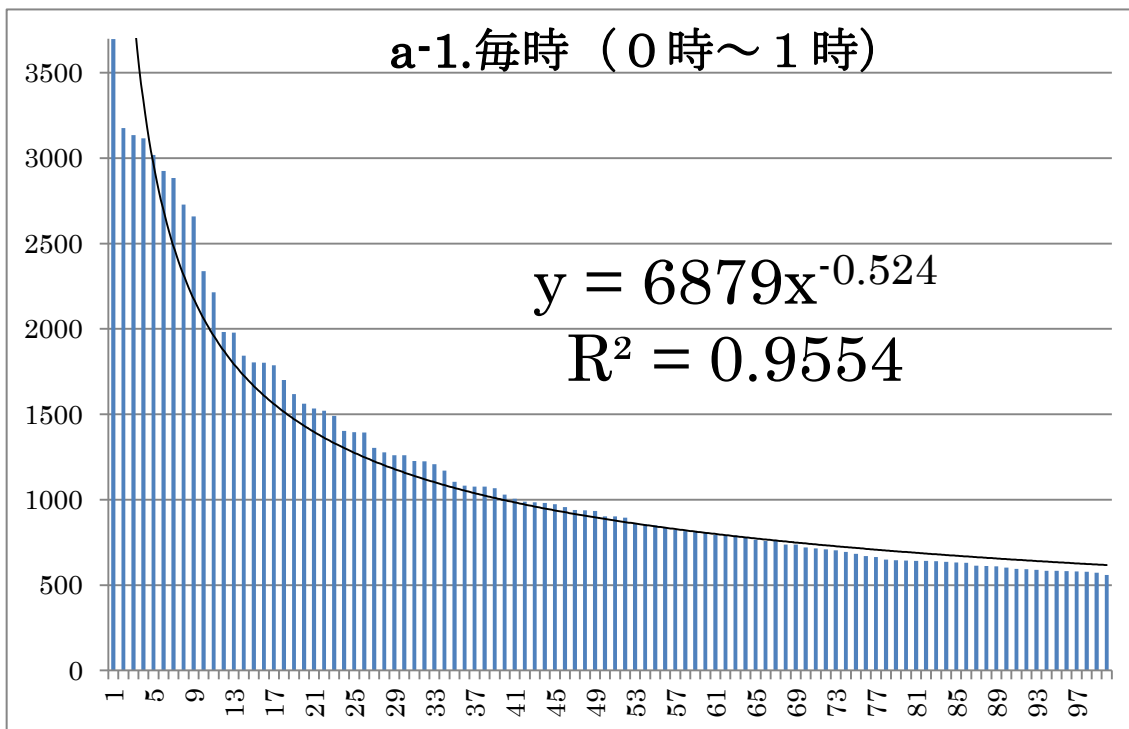


図 1 a-1. 毎時 (0時~1時) の分布

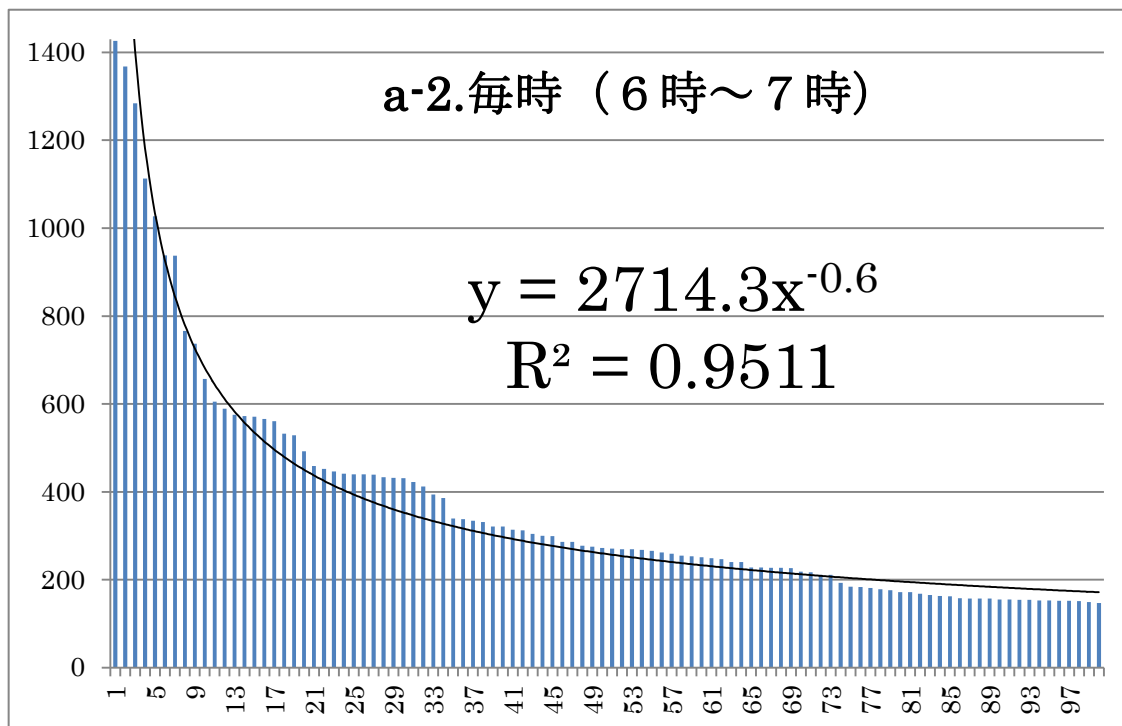


図 2 a-2.毎時 (6時~7時) の分布

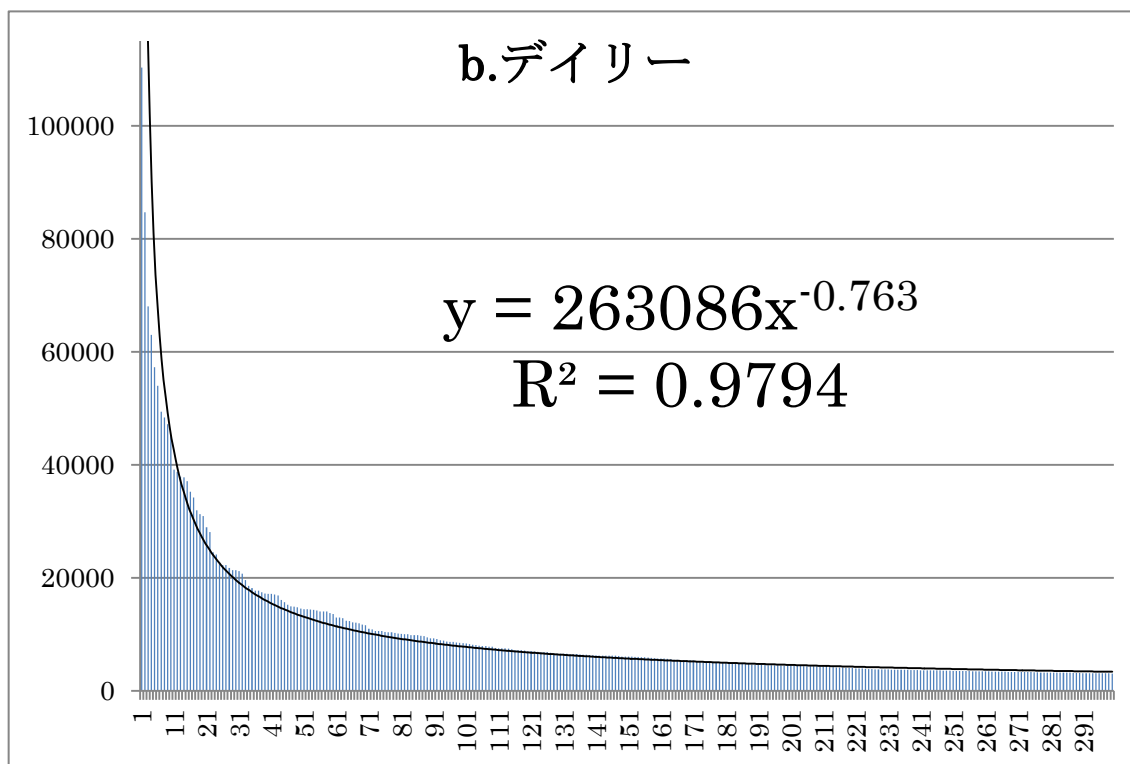


図 3 b.デイリーの分布

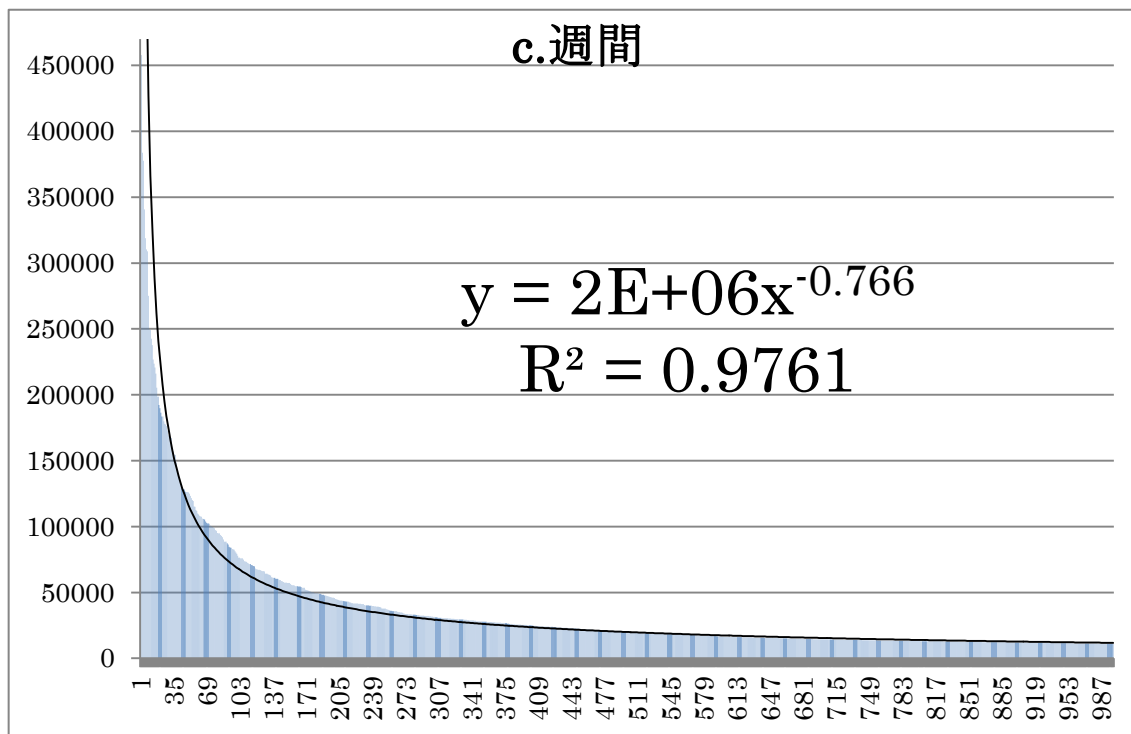


図 4 c.週間の分布

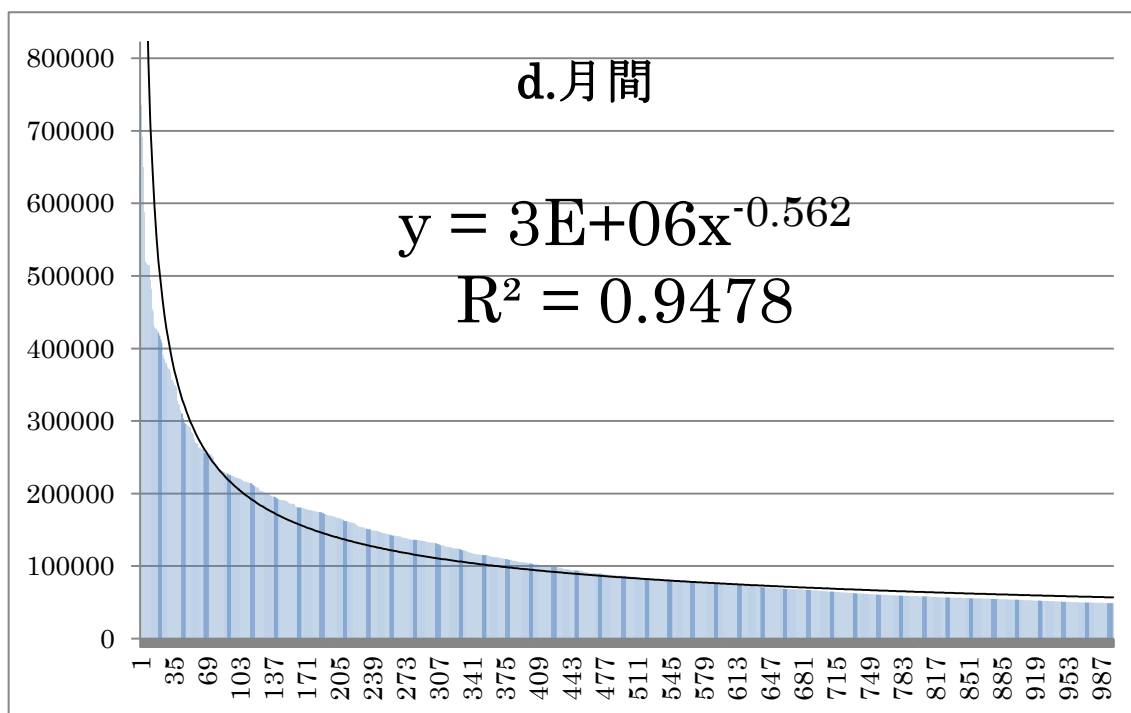


図 5 d.月間の分布

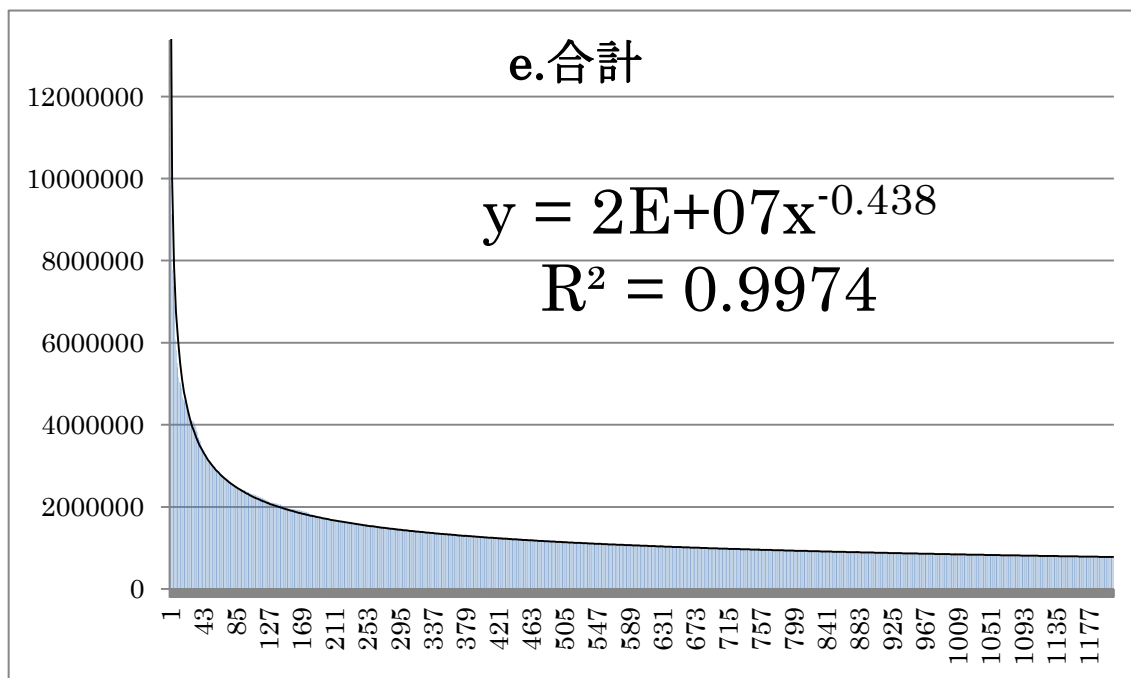


図 6 e.合計の分布

次に、実際に得られたデータとひとつの曲線で近似した場合と、3つの曲線で近似した場合の3つを同時にみることができるグラフをb.~e.について作成したものを載せる。

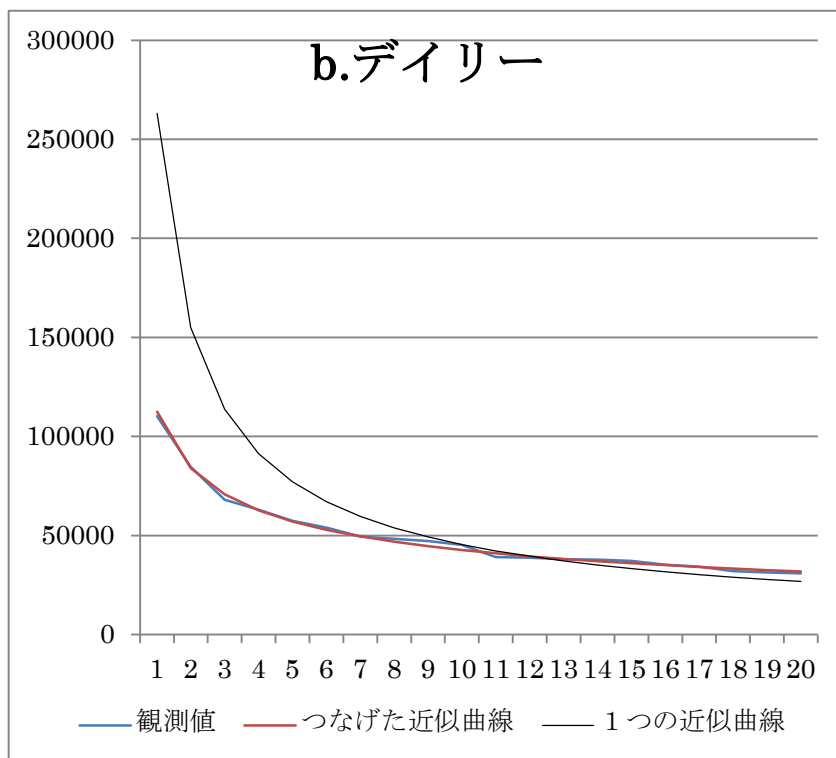


図 7 デイリーでの分布の上位部分

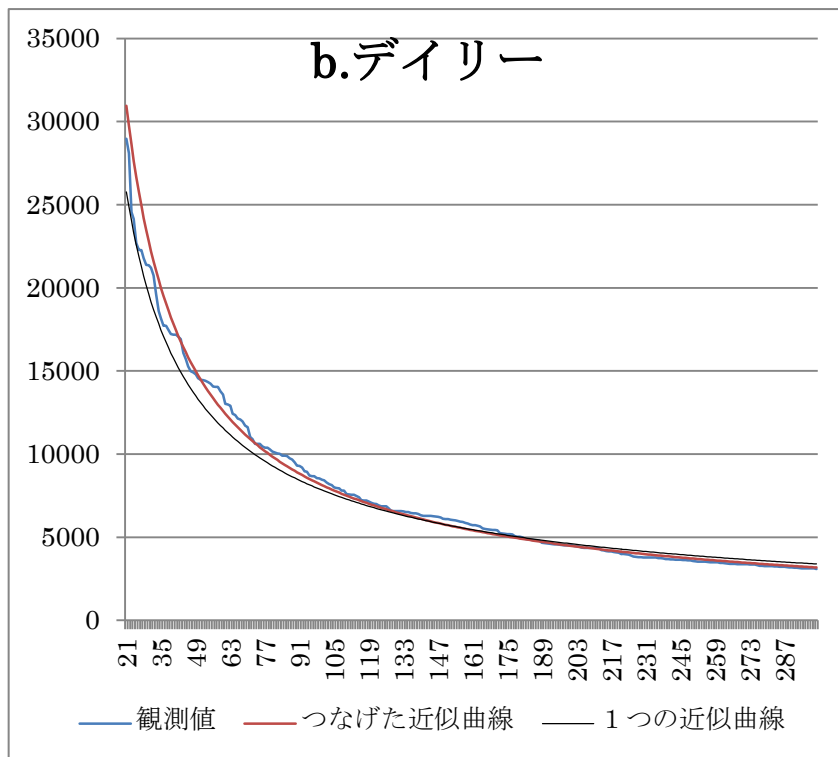


図 8 デイリーの分布の全体像

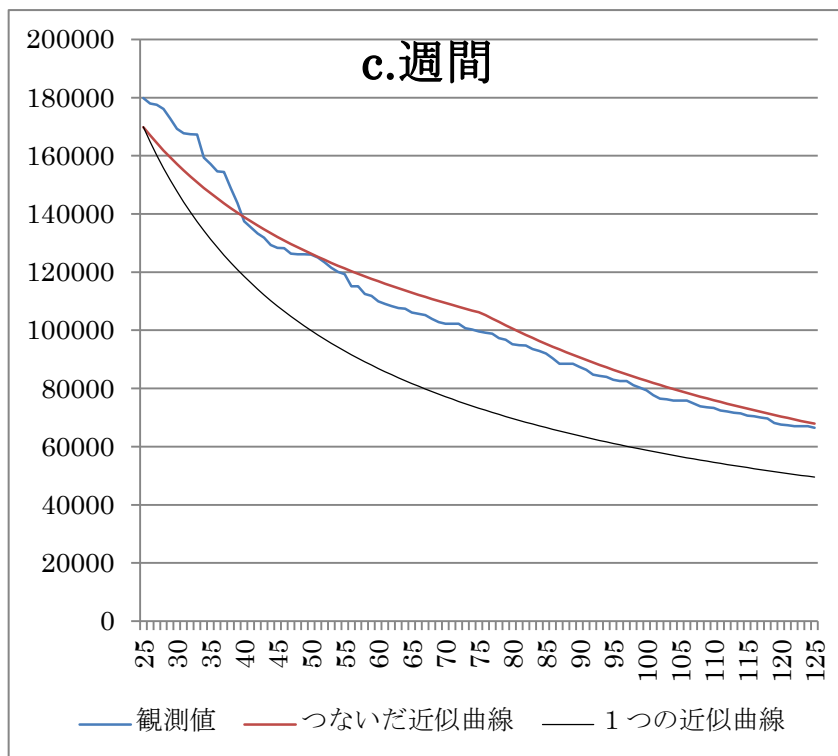


図 9 週間での一つ目のつなぎ目部分の詳細

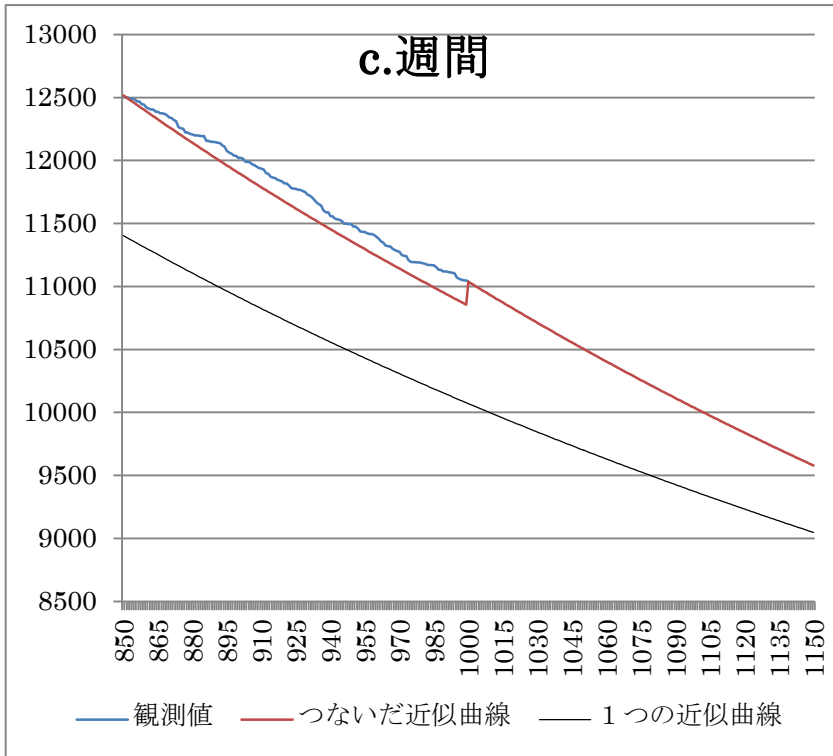


図 10 週間での二つ目のつなぎ目部分の詳細

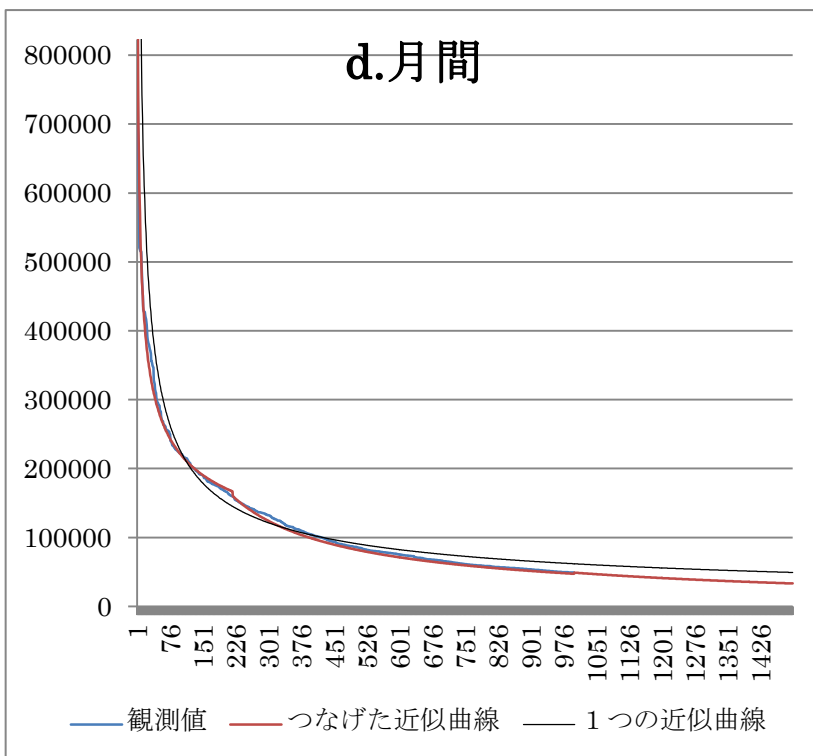


図 11 月間の分布の全体像

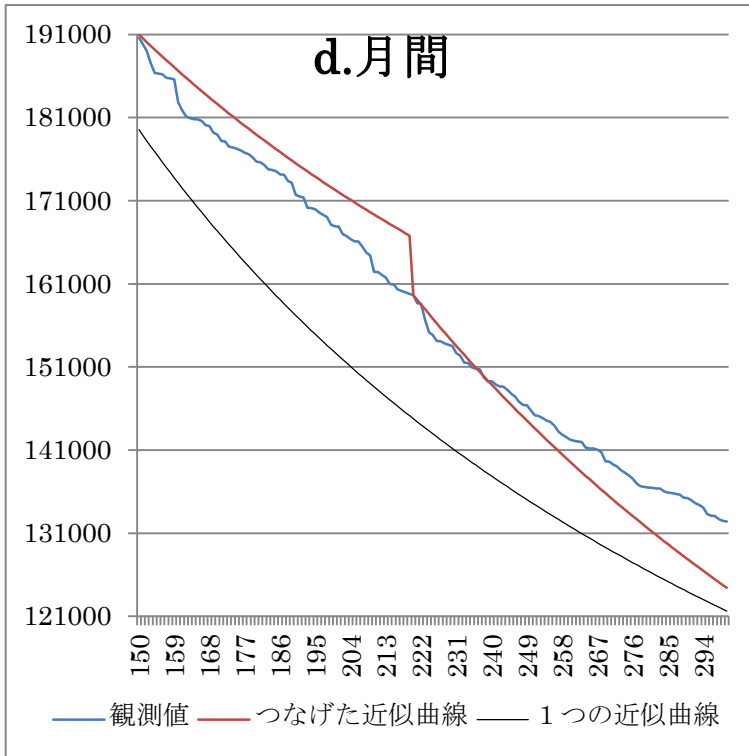


図 12 月間での一つ目のつなぎ目部分の詳細

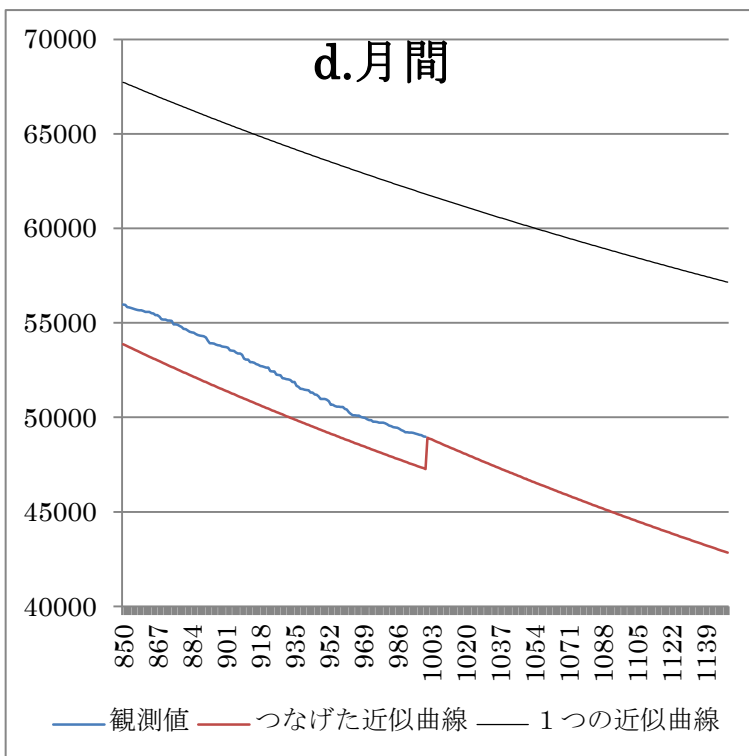


図 13 月間での二つ目のつなぎ目部分の詳細

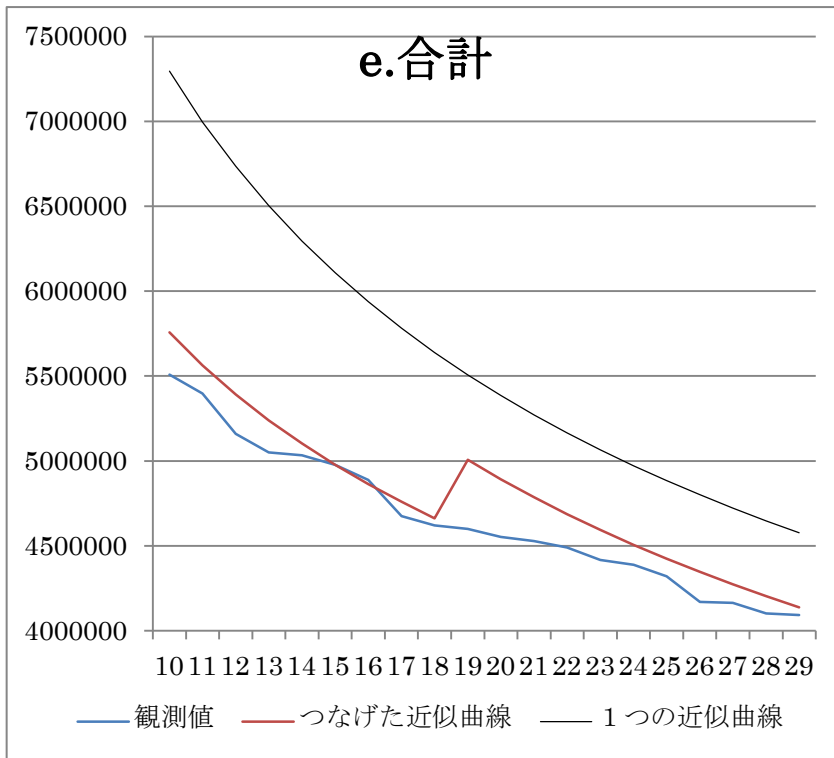


図 14 合計での一つ目のつなぎ目部分の詳細

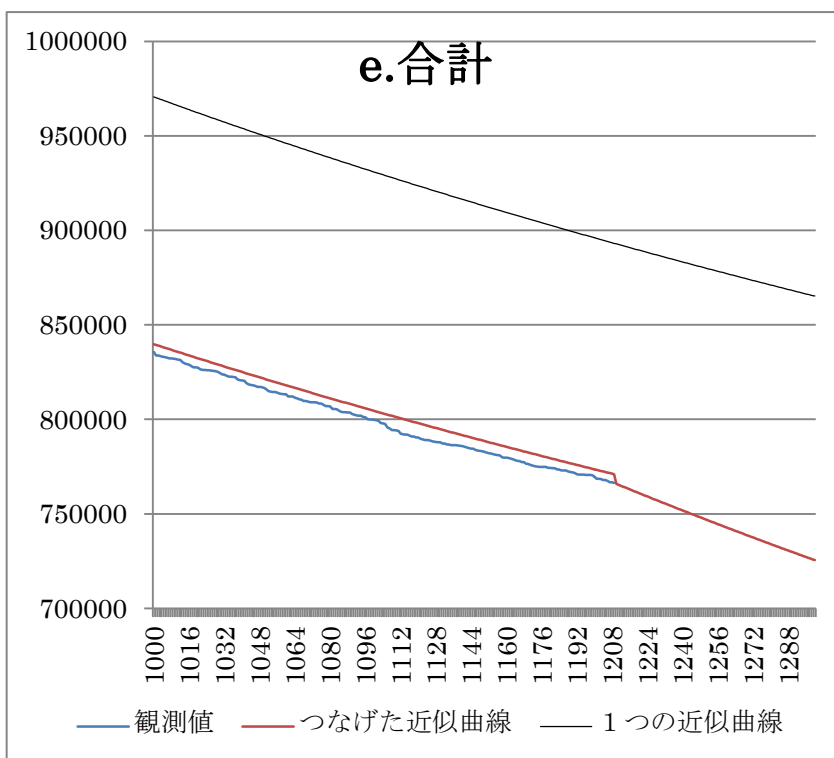


図 15 合計での二つ目のつなぎ目部分の詳細

以上のグラフから分かることとして、e.合計の分布（図 6）ではあまりみられないが、その他 a~d の 4つの期間の分布（図 1~図 5）では、上位に行くほどまた下位に行くほど近似曲線に対して観測値が下に出ている。つまり、順位とともに指数が異なり、また観測期間の長さが違うにも関わらず似たような傾向がみられた。このことから1つの曲線で近似を行なうよりも繋ぎ合わせの分布で当てはめる、つまり近似曲線のある順位を境に複数設けた方がより観測した分布に対して当てはまりの良い結果を得ることができるのではないかと考えられる。

本稿は、ロングテールに関する研究ということで下位の分布に興味があるので、上位のデータの乖離には目を瞑ることにして下位のデータの乖離に注目した。そのため、下位のデータに対してもうひとつのべきの曲線を当てはめることにした。すると、公表されていないさらに下位の部分の予測と合わせて1つの観測期間につき3つの当てはめを行なうことで以下のような指数 **b** を得ることができた。（図 7~図 15）

a.毎時

1-29 位 : $b=2.286$, 30-100 位 : $b=1.129$, 101 位-約 696 万位 : $b=0.442$ (12/01 06 時台)

1- 7 位 : $b=8.680$, 8-100 位 : $b=1.600$, 101 位-約 696 万位 : $b=0.984$ (12/01 00 時台)

b.デイリー

1-20 位 : $b=2.373$, 21-300 位 : $b=1.168$, 301 位-約 696 万位 : $b=1.028$

c.週間

1- 75 位 : $b=2.339$, 76-1000 位 : $b=1.134$, 1001-約 698 万位 : $b=0.986$

d.月間

1-219 位 : $b=2.797$, 220-1000 位 : $b=1.244$, 1001-約 7 百万位 : $b=1.045$

e.合計

1-23 位 : $b=2.789$, 24-1209 位 : $b=2.220$, 1209-6769992 位 : $b=1.326$

上記の結果をみると、テール、つまり分布の裾野へ行くにつれて、パレートの平等性の指数 **b** の値は小さくなっていることが分かる。**b** が $b>1$ だとロングテールが成立しており、 $b<1$ だとロングテールは成立していないと先ほど述べたように、**b** の値が小さければ小さいほどビッグヒット依存ということで不平等性が大きくなり、1 を境にロングテールではないと判断する。

数値をみると、a.~e.のどの観測期間も例外なく分布の裾野へいくほど数値が小さくなって 1 に近づ

いているということがわかる。なかには1より小さくなっているものもある。

このことから、再生数の分布は単純なパレートよりは打ち切り効果、というよりビッグヒット型に傾く、つまり下位に行くほど不平等の度合いが強くなっているということが言える。具体的に指数 b をみると、**b.デイリー**、**d.月間**、**e.合計**では公表されていない部分の b の予測値が1を超えている。しかし、下位にいくほど b の値が小さくなることは今回の結果のどの観測期間でも表れていることから、もしさらに下位の部分のデータを用いた予測を行なうことができるとすれば b の値はいずれは境目である1を下回るのではないかと予想される。

また、次に観測期間の幅すなわち窓の違いによる数値の比較を行うと、いくらか例外はあるものの、全般的に指数 b が大きくなっていることがわかる。このことから、観測期間（観測窓）が長くなるほど分布はロングテール化してゆくのではないかとと思われる。

4.3. 考察

4.3.1. 観測期間の違い

結果の指数 b 特にそれぞれの観測期間の三つ目の近似曲線の指数 b を見ると、**b.デイリー**と**c.週間**の間での誤差はあるものの、観測期間（観測窓）が長くなるほど動画再生数の平等化が進むといえるのではないだろうか。**b.デイリー**と**c.週間**の間での誤差については、考えられる原因としては、近似曲線の左端の順位が異なること（**b.デイリー**では301位、**c.週間**では1001位）、また、べきの曲線で近似する際の順位（グラフでは横軸）の幅が異なることなどがある。これは、全ての観測期間で共通してみられる、下位へ行くほど指数 b が小さくなっている（不平等性が強くなっている）ことから、より高い順位からの近似はその分指数 b の値を高く押し上げるのではないかとということから考えられることである。また、下位へ行くほど指数 b が小さくなっている理由から逆に上位側（左端が1位）では順位の幅が小さい（右端の順位が高い）ほど指数 b の値は高く出してしまう傾向がある。

4.3.2. 活動の昼夜差

観測期間が午前6時から午前7時までの1時間である**a-2**のグラフを見ると、総再生数が小さいことが分かる。これは、午前6時台は寝ていて活動していない人が多いために、総再生数が小さくなっており、統計的には揺らぎの大きい結果だと思われる。毎時のランキングは、**a-1**と**a-2**を比べてみると分かるように、人々の社会活動に昼夜差があることを表している。そのため、観測期間が一日よりも長い

b.~e.のデータとは比較しにくい面を持っている。

4.3.3. 指数 b の妥当性

観測期間が午前0時から午前1時までの一時間である a-1.の最上位の指数も $b=8.680$ と極端に大きい値が出ているが、これは 1-7 位という、ベストヒットの中のベストであるから、これも特殊なことだと考えられる。

残りの観測期間が1日よりも長い b.~e.のデータは指数 b がおおむね 1 弱から 3 以内に収まっていて、妥当であると考えられる。観測期間すなわち「観測窓」との関係では、全累積の結果である e.合計において指数 b が大きい（ロングテールの）のは妥当だと思われる。しかし a.毎時、b.デイリー、c.週間の間では指数 b を見る限りそれほど差はないように思われる。

4.3.4. 指数 b の見方

特に予測値であるそれぞれの指数 b について、三つ目の近似のトップ（例えば、c.週間だと 1001 位）が同じである場合、総動画数が多いほど近似曲線は上方修正する。つまり、この場合において総動画数が増えるほど実際より平等にすなわち指数 b が大きくなっていることになる。このことを考慮に入れると、c.週間と d.月間では d.月間のほうが総動画数が 2 万ほど多いが予測のトップは両方とも 1001 位であり、指数 b はそれぞれ $b=0.986$ 、 $b=1.045$ となっているというのは、実際ほとんど差はないのではないかと考えられ、観測期間が週間から月間になると分布が着実にロングテールに近づくと断定することはできないのではないだろうか。

a.毎時、b.デイリー、c.週間のそれぞれについては総動画数がほとんど変わらない中で予測値の部分のトップが異なるという、c.週間と d.月間の場合と比べもう一方の定数が変化するということが起きているため、これもまた同様に指数 b にそれほど差はないといえる。

これらをまとめると、ビッグヒット達の人気がどのくらい継続するかという期間、つまり流行という観点では、観測期間（観測窓）の幅が a.毎時から d.月間に広がったとしても大した差は見られないように思われる。

つまり、a.毎時から d.月間の指数 b 特に三つ目の近似である予測値の指数 b を見る限り、流行が終わると急速に累積の超長期の観測期間である e.合計の指数 b に近づくのではなく、指数 b から観測される動画のロングテール化（平等化）はひと月という観測期間（観測窓）ですら観察できないほどきわめて

ゆっくり起きているように見える。

流行は週かせいぜい月で終わるのではと予想していたが、**d.月間**が **c.週間**に比べて急に累積に近くなった、もしくは **d.月間**が **c.週間**と **e.合計**の中間的な振る舞いをしているのか、言い換えると、流行の長さや観測期間の長さの大小でデータが急変するのかそれともただらと変化するのか、という点についてははっきりした結果は言いにくい。

4.3.5. 「フロー」と「ストック」

経済ないしは経営上の意味において、累積のロングテール性の意味は短期間の非ロングテール性と違う視点が必要なのではないだろうか。つまり、累積の「人気度」というのはフローとストックで考えるのとストックであり、「文化のストック」ということができる。流行が終わって「倉庫に入った動画たち」がすぐ平等に近くなる、というよりは、長い時間をかけてゆっくりたくさん文化的創造物（今回だと動画）が倉庫（ニコニコ動画のサーバー）に入ってその結果としてパレート指数 **b**（傾き）がゆるやかになっていく、というふうに考えることができるのではないだろうか。一方、短期間の非ロングテール性というのは流行を表していてフローに相当すると思われる。服部氏の先行研究では、Amazon において、一冊の商品のランキングの移り変わりが流行を表しており、その注文頻度というフローを観察することで商品全体の売上の分布を観察しているといえる。

4.3.6. データ比較の難しさ

近似した曲線の境目はそれぞれの観測期間で残差が最小になるように選んでいるため、観測期間による差は、明確なことを言うことはできないが、たいして変わらないのではないかと思われる。公表されていない部分の下位のところは **e.合計**は総動画数と総再生数が相当量あるため $b > 1$ と言っているのではないかと思われるが、これも総動画数が約 7 千万と、他の観測期間と比べて桁多いため、直接比較することは難しい。**e.合計**以外の **a.~d.**の観測期間の公表されていない部分の下位の指数 **b** はほぼ 1 というべきで、ロングテール型とビッグヒット型の境界に位置しているといえるだろう。このことからロングテールが成立しているのかいないのかについて直接的な結論を出すことはできないのではないかと思われる。

4.3.7. それでもロングテール化は進む

しかし、全ての観測期間において、下位にいくほど不平等性が強くなる（指数 b の減少）ことが観察できるので、現時点では公表されていない、より下位の部分に焦点を当てることができた場合、まだ、ロングテールは成立していない可能性があることが推察できる。それでも、e.合計については、下位の部分についての予測の指数 b が $b=1.326$ という十分なデータが出ており、全累積でみた場合ロングテールが成立しているのではないかといえる。

4.3.8. 注意点

服部氏の先行研究では Amazon はロングテールではないという結果が得られた。しかし、ニコニコ動画の再生数ランキングにおいてはこれまでの累積を表す合計のランキングでは、ロングテールが成立しているという結果が得られた。

ここで今一度、ロングテールという言葉を生み出したクリス・アンダーソンの著書に注目してみると、物理的な商品を扱うよりもデジタル商品を扱う会社のほうがテールを伸ばすことができるとある。つまり、服部氏の先行研究である Amazon の和書と今回のニコニコ動画の再生数ランキングでは、コンテンツの保管コストがそもそも異なる。この保管コストが先行研究と今回の分析結果の違いを生み出した可能性もある。

もう1つ注意としてあげられるのは、動画の長さは再生回数に対して影響を及ぼす可能性があるということだ。つまり、同じ時間が与えられているとすれば、短い動画ほど何度も見るのではないかといった疑問が浮かぶかもしれない。しかし、今回の調査では、動画の長さを考慮した研究は行なうことはできないため、それについての考察はできていないことをご了承願いたい。

5. 第4章 AKB 選抜総選挙の分析

5.1. 分析方法

選抜総選挙には東京の秋葉原を拠点とする AKB48 のメンバーだけでなく、AKB ファミリーと呼ばれる AKB48 のチームに入ることを目指す研究生や名古屋の栄を拠点とする SKE48 のメンバー、大阪の難波を拠点とする NMB48 のメンバー（NMB48 は第三回選抜総選挙から）も立候補している。

分析対象である選抜総選挙の最終発表や速報、中間発表の結果は AKB48 のオフィシャルブログである「[～AKB48 TOKYO DOME までの軌跡～](http://ameblo.jp/akihabara48/) (<http://ameblo.jp/akihabara48/>)」において公表されているものを用いた（付録を参照）。公表されていた内容は、1 位から 40 位までのメンバーの名前と所属するチーム、各メンバーの得票数である。

また、今回は、第一回選抜総選挙は投票数が少なかったため、第二回、第三回の選抜総選挙の中間発表、速報と最終発表についてそれぞれ分析を行なった。そして、第二回選抜総選挙の中間発表は 2010 年 5 月 26 日と 6 月 2 日の二度中間発表がされているが、5 月 26 日分は投票数が少ないといった理由から今回分析対象からは外すことにした。

分析方法としては、公表されている 1 位から 40 位までの投票数を用いて、ニコニコ動画ランキングと同様最小 2 乗法によってべきの近似曲線を求めた。そして、パレートの平等性の指数 b が 1 よりも大きい小さいかでロングテールか否かの判断を行なった。繰り返し述べるが、 $b > 1$ のときロングテールが成立しており、 $b < 1$ のときは成立していない。

今回、第三回選抜総選挙では、スポーツ紙であるデイリースポーツオンライン上のページ (<http://www.daily.co.jp/gossip/article/2011/06/11/0004159535.shtml>) にて合計の最終投票数のデータが公表されていたので、これを用いてニコニコ動画ランキングの分析において行なったように公表されていないデータの分布についても予想を立てることができた。つまり、それぞれの得票数が分かっている 1 位から 40 位までだけでなく、41 位から最下位 150 位までの得票の分布の予測を行なうことができた。以下、ページより引用。

『9 日に行われた AKB 48 「第 3 回選抜総選挙」の候補者 150 人全員を対象とした総投票

数が10日発表され、前回の約3倍の116万6145票だった。

投票数の内訳も公表され、投票権付きシングル「Everyday、カチューシャ」（5月25日発売）購入者の投票は77万9090票、その他公式ファンクラブサイト、公式モバイル会員からの投票は38万7055票だったことが分かった。』

ただし、中間発表については、中間発表時点の総得票数のデータを見つけることができなかったため、1位から40位までの得票を元に近似曲線を求めるといった分析のみを行なった。

5.2. 分析結果

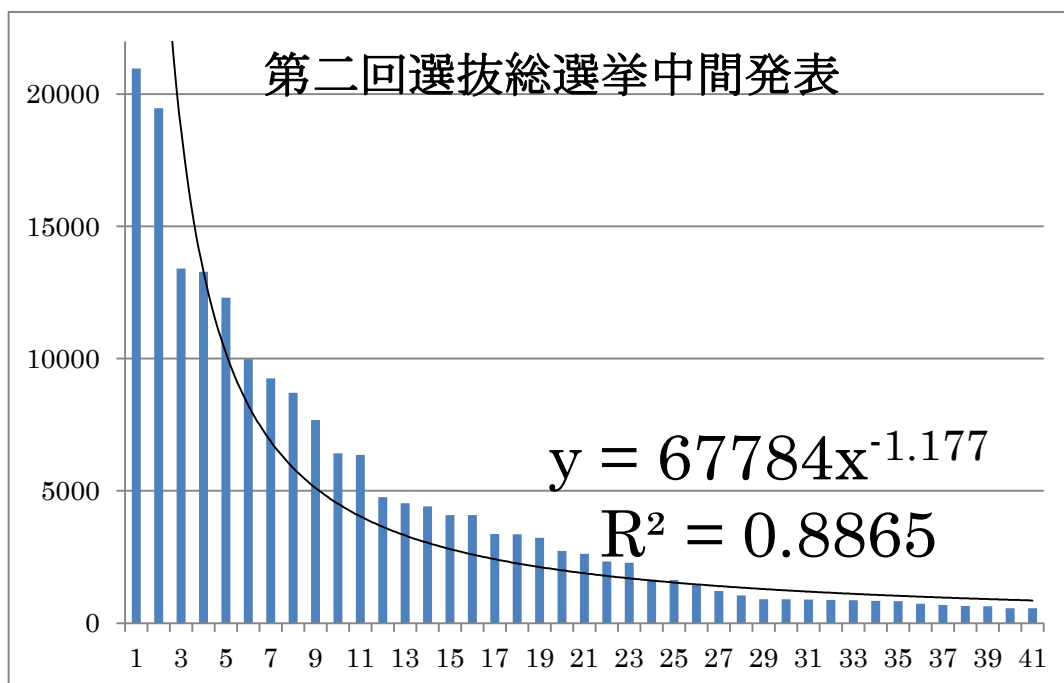


図 16 第二回選抜総選挙中間発表の分布

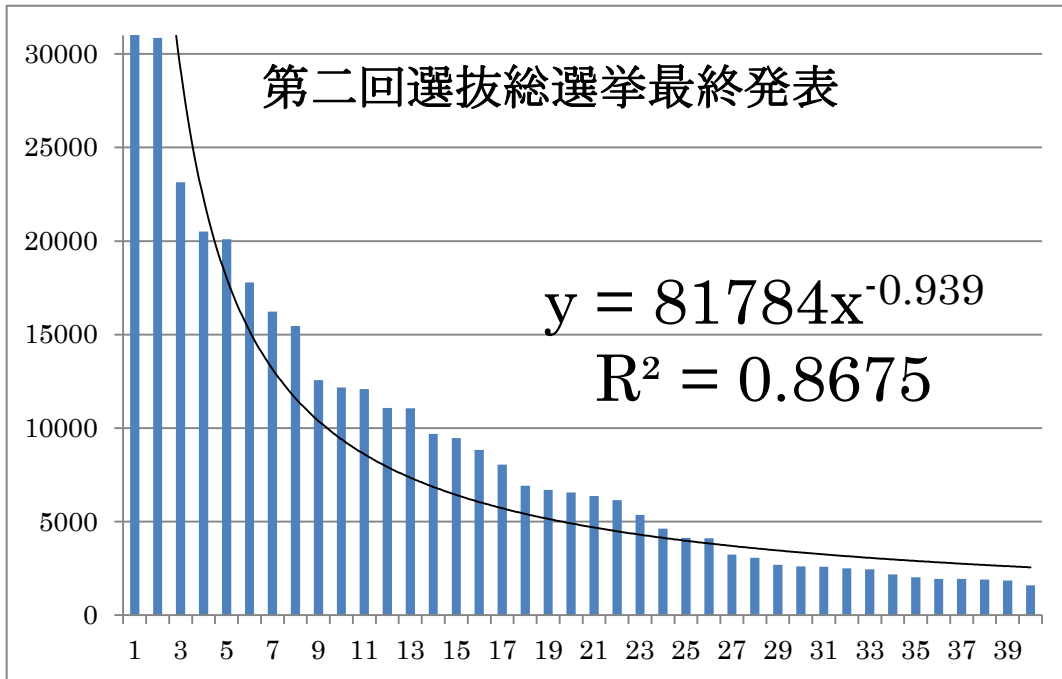


図 17 第二回選抜総選挙最終発表の分布

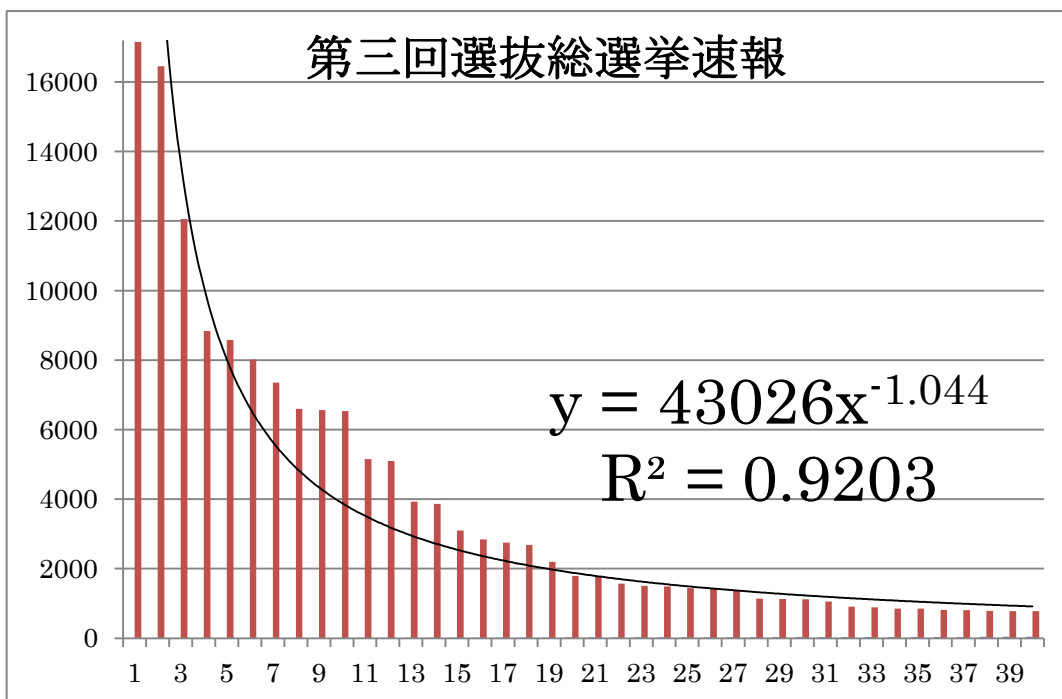


図 18 第三回選抜総選挙速報の分布

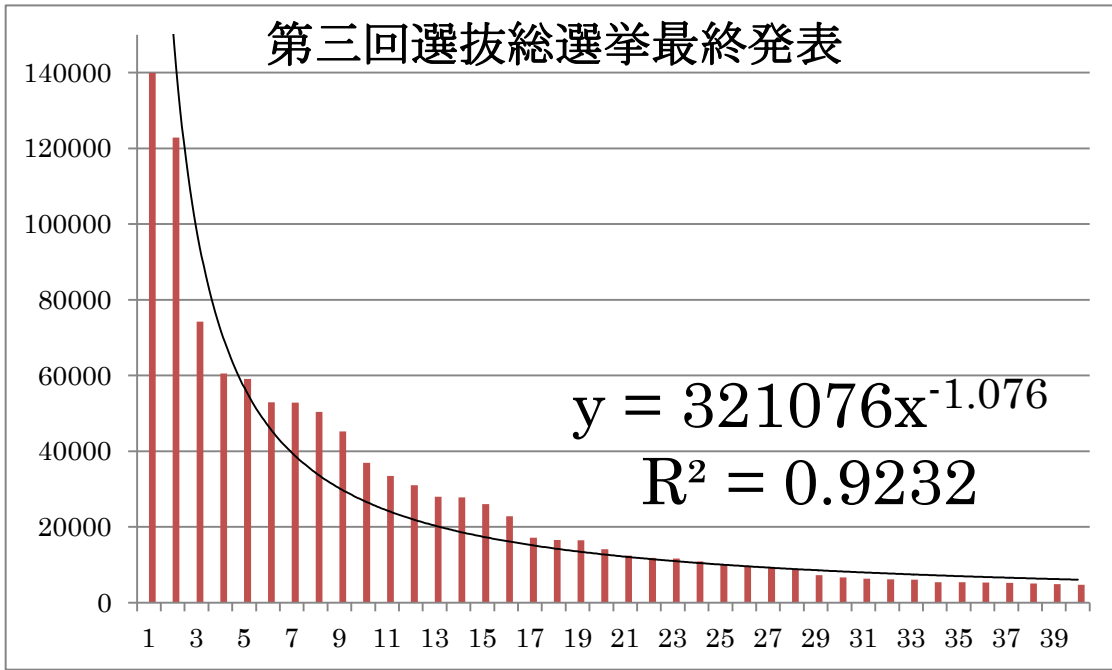


図 19 第三回選抜総選挙最終発表の分布

次に第三回選抜総選挙の最終発表について、3つの近似曲線をつなぎ合わせたもの、1つの曲線で近似したもの、実際のデータの3つをまとめたグラフを以下に示す。

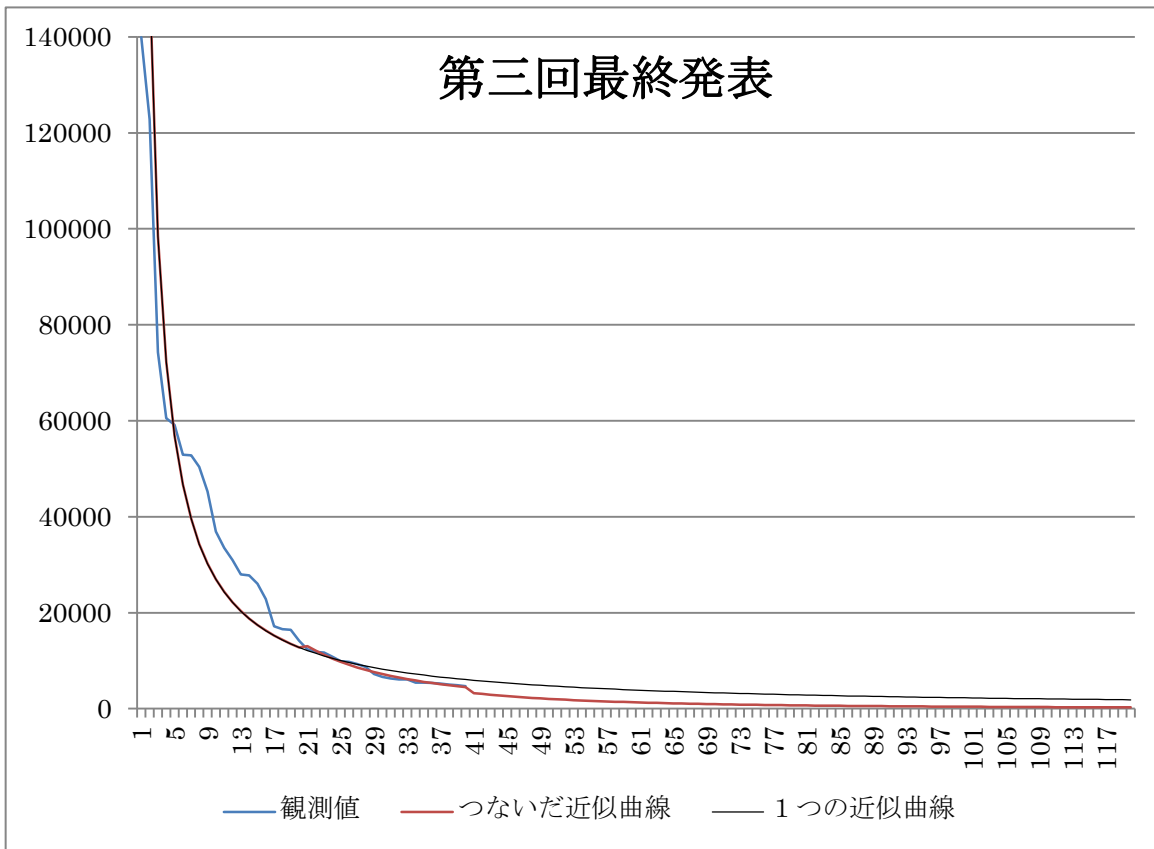


図 20 第三回最終発表の近似曲線

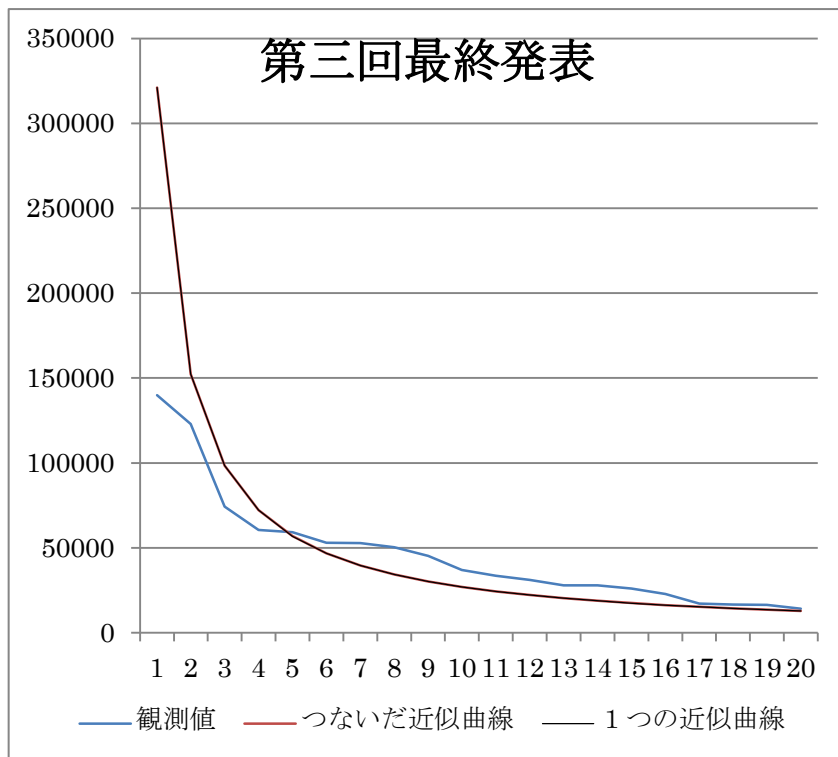


図 21 第三回最終発表の上位 20 位

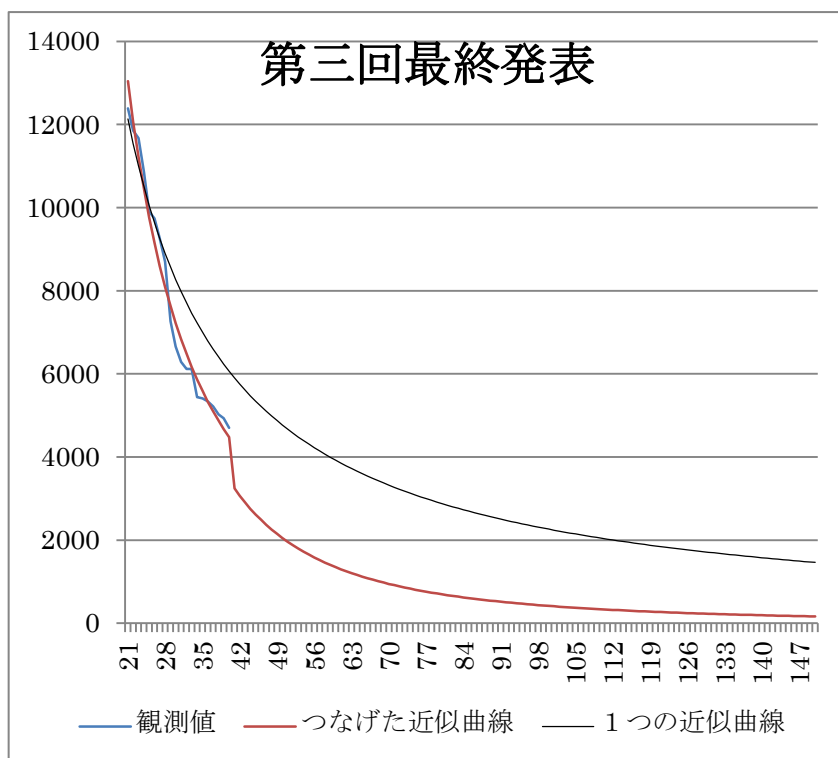


図 22 第三回最終発表の詳細な分析結果

■第二回選抜総選挙

中間発表：1-40位 $b=0.85$

最終発表：1-40位 $b=1.06$

■第三回選抜総選挙

速報：1-40位 $b=0.96$

最終発表：1-20位 $b=0.93$, 21-40位 $b=0.60$, 41-150位 $b=0.43$

公表されている部分の分析結果は以上のグラフのようになり、そのときの指数 b の値は、第三回選抜総選挙における下位 21 位から 40 位までで当てはめたときの指数 b と公表されていない 41 位から 150 位までの分布を予測した場合の指数 b と合わせて以上のような結果になった。

グラフから分かることとして、ニコニコ動画の場合と同様最上位の部分とまた 40 位にいくほど近似したべきの曲線から下方へずれていることが分かる。

パレートの不平等性の指数 b は、1 位から 40 位までで第二回では中間発表から最終発表で 0.85 から 1.06 とわずかながら上昇しており、一方で第三回では速報から最終発表で 0.96 から 0.93 と微妙に減少していることが分かる。これは、総投票者数が十分でないことによるデータの揺らぎだけでなく、先ほど予想した通り一定の順位を基準とした様々な特典であったり中間発表や速報を公表するといったことが投票者に与える心理的な効果があり、その効果の度合いを測ることができないからではないかと予想できる。そのため、今回の場合では、観測期間の長さの違いによるロングテールの成立についての考察は行なうことができないだろう。すなわち、第二回の中間発表では $b=0.85$ であるからロングテールではなく、最終発表では $b=1.06$ であるからロングテールだと決めることはできないということだ。

では、次に、もう一つの興味であるランキングの下位にいくほどロングテールから遠ざかるのではないかといた予想に対する結果を見てみよう。第三回選抜総選挙の最終発表において、1 位から 20 位までの指数 b は、 $b=0.93$ 、21 位から 40 位までの指数 b は $b=0.60$ 、41 位から 150 位までの予測した指数 b は $b=0.43$ となっている。これは、下位にいくほど指数 b は減少しており、ロングテールから遠ざかっているということを意味している。AKB48 については、投票するメンバーを選ぶ際に用いる情報というのはコアなファンではない一般の大勢の人々にはテレビ等のメディアに限られており、そのようなメディアに出演できないメンバーにとって票を得ることは至難の業であるといえよう。そのため、公表されないほど下位だったメンバーの得票はより減少しているのではないだろうか。実際、第三回の最終

発表では“新入り”を表す研究生は40人のうち2人しかランキングに入ることができていなかった。

5.3. 考察

5.3.1. ファン心理を揺さぶる選挙方式

AKB ファミリーの選抜総選挙は先ほどのニコニコ動画の再生数ランキングと比べてメディア選抜枠などが存在しており、純粹に数字による順位付けだけでなく、ある順位を境としてテレビに出れたり、CD シングルの歌を歌うことができる特典があるといった特徴がある。具体的には1位から12位までが“メディア選抜”、13位から21位までが“選抜”、22位から40位までが“アンダーガールズ”となっている。このことは、AKB ファンの心理を揺さぶって中間発表や速報で得票数が拮抗しているメンバーだけでなく、その境目にいるメンバーの最終結果での順位変動にも影響を及ぼしうるのではないだろうか。そのため、境目付近でのデッドヒートが繰り広げられているのではということが予想される。実際に、第二回では、中間発表から最終発表で2位と1位が入れ替わることが起き、第三回では、1位はその2人のメンバーのうちどちらかではないかと前々から予想されていたが、案の定、速報から最終発表にかけて2人の一騎打ちが行なわれている。また、“メディア選抜”と“選抜”の境目や“選抜”と“アンダーガールズ”の境目でも他の順位と比べて変動が激しくなっていた。

5.3.2. ロングテールへと向かう力と投票者へ与える影響

中間発表や速報が存在することにより、中間発表や速報と最終の間では公表されなかったメンバーよりも公表されているメンバーへの投票が増えるのではないかと。公表されていないメンバーに今更投票してもなにも変わらないのだから中間の時点で40位以内に入っている中で自分の好きなメンバーに投票しようといったものだ。そのため、最終発表の分布はその分上方修正されているのではないと思われる。つまり、ニコニコ動画の再生数ランキングの観測結果から予想された、観測期間が長くなるほどロングテールになるのではということに対して負の影響を及ぼし得るということだ。結果の第二回選抜総選挙での中間発表：1-40位 $b=0.85$ 最終発表：1-40位 $b=1.06$ と第三回選抜総選挙での速報：1-40位 $b=0.96$ と最終発表：1-20位 $b=0.93$ はこのような影響を受けている。そのため、観測期間の長さでロングテールの成立について考察を行なうことは難しい。人々への心理的な作用を考えると、ニコニコ動画の再生数ランキングでの「期間が延びた」というよりも、AKB の選抜総選挙では「ゴールが近づいた」という風に考えられるのかもしれない。

5.3.3. 非公表の得票分布を暴く

第三回選抜総選挙最終発表では、総投票数をデイリースポーツの記事から得ることができたので、それをもとに公表されていない41位から150位までの分布の予測を行ない、指数 b を求めることができた。よって、第三回選抜総選挙では、ニコニコ動画の再生数ランキングにおいて行なったように上位と下位での指数 b の変化を考察できた。

結果を述べておくと、1位から20位までのデータを用いた場合の b は、 $b=0.93$ 、21位から40位までのデータを用いた場合は $b=0.60$ 、41位から最下位までの投票数がべきの曲線になるとした場合の予測値は $b=0.43$ となった。すると、ニコニコ動画の再生数ランキングのときと同様、下位に行くほど、指数 b が減少していくといった特徴を得ることができた。これは、下位に行くほど不平等性が強くなるつまりロングテールとは言えず、よりビッグヒット依存型になっているとすることができる。この特徴は、第二回の中間発表や最終発表、第三回の速報のグラフからも観察することができる。1位から40位という限られた順位ではあるが、下位になるにつれて得票数は近似曲線よりも下方にずれていることが分かる。

よって、ランキングの下位になるほどビッグヒット依存の傾向は強くなりロングテールは成立していないということは今回の分析においては確定的だということができる。

5.3.4. データの揺らぎ

これまでAKB48の選抜総選挙についてロングテールは観察されるかといったことを行なってきたが、ニコニコ動画の総動画数（総点数 N にあたる）に比べて今回の投票対象者は公表されているものとして40人、公表されていない部分を含めても150人ということで少ないと言える。そのため、ロングテールの分析に対してデータの揺らぎがあることは否定できない。

5.3.5. 各チームが占める人数

先ほど、そもそもメディアに出演できないメンバーは多くの票を得ることが難しいのではないかと述べたが、実際に第三回選抜総選挙の最終発表において上位40人を所属しているチームごとにまとめてみると、

TeamA : 11、TeamK : 10、TeamB : 10、TeamS : 4、TeamKII : 2、研究生 : 2、TeamN : 1

という結果を得た。TeamA、K、BはAKB48、TeamS、KIIはSKE48、TeamNはNMB48のメンバーが

所属するチームである。

このことから、メディア露出が少なく **AKB48** の陰で頑張っている研究生と、拠点が東京の秋葉原である **AKB48** と比べ、名古屋の栄である **SKE48** や大阪の難波である **NMB48** は全国ネットのメディアに出演する機会が少ないため票を得る機会が失われているのではないかと予想される。また、名古屋圏や大阪圏の投票者は、気に入ったメンバーに投票して公表されないまま終わるより人気がありそうな主要メンバーに投票して **SKE48** や **NMB48** のメンバーを 40 位以内にしようといった心理が働き、それによって **AKB48** 以外では主要メンバーに特に票が集中するのではないかとということも考えられる。

5.3.6. ビッグヒットが AKB 人気を引っ張る

また、もはや言うまでもないことであるが第二回と比べ第三回では総投票数が大幅に上昇している様子を観察することができ、**AKB48** の人気はここでも明らかだということができる。

これに関連して、第二回から第三回にかけての **AKB48** の人気上昇というのはメディアを通して知名度がアップしたからだと考え、メディア露出が多い上位メンバーへの人気の集中を助長しているといえる。

6. 第5章 二つの分析対象からの知見

6.1. これまでの分析のまとめ

これまでの調査の結果をまとめると、ニコニコ動画の再生数ランキングでは膨大な量のランキングは観測時間（観測窓）が長くなるほど、データが増えるほど、人々の興味は下位にまで及び、分布はロングテールになるのではないかと予想される。しかし、AKB48の選抜総選挙では観測時間が長くなるほどロングテールになると断言することはできなかった。このような結果の違いは、前者ではただ時が刻まれ、後者では締切りが近づくといったような観測時間の捉え方の違いが人々の間にあるからではないだろうか。AKB48の選抜総選挙の中間発表と最終発表の間には、中間発表で上位のビッグヒットに位置するメンバーやその他公表される順位に入っているメンバーへ人々の関心が集中することが考えられる。このような場合、つまり順位による差別化など人々に心理的な影響を及ぼしうる場合においては、時間の経過による分布のロングテール化は見られなくなっていたといえる。ただ、これは予想の範囲を超えることはなく、補強するための明確な根拠は得られていない。ただ、その一方で、両方の全観測期間の結果に共通していることとして、下位にいくほどパレート分布の平等性の指数 b が小さくなる傾向があった。これは、下位にいくほど、不平等化が進んでいる、つまりロングテールが成り立たなくなっているということを意味している。以上のことから総合的に考えられることは、流行などによる人々の関心の影響を受けながらも、時間の経過とともにロングテール化は徐々に進んでおり、下位の部分が伸びてゆくのではないだろうか。

6.2. 論議

本章では、本稿の研究方法や結果に対して批判的な立場に立ち、今後の研究につなげていくための議論を行なう。

本稿では、研究対象のニコニコ動画の再生数ランキングや AKB48 の総選挙を多くの人に関わってできたデータの集まりとし、パレート分布としてべきの曲線で当てはめることにし、より当てはまりをよくするために曲線をつなぎ合わせるということを行なった。対数などの代表的な曲線では試してみたものの、もしかしたら、べきではないより当てはまりの良い曲線がある可能性もありその点については課

題が残っている。また、曲線をつなぎ合わせる際の、つなぎ目の決め方に議論の余地が残されていること、公表されていない下位の分布は直接観察することができず予測せざるを得なかったこともあり正確性については、直接的なデータの分析には劣る。

また、AKB48の選抜総選挙では、流行などの人々の心理的な要因が複雑に絡み合うことで分布に影響を与えている可能性が示唆されたが、いったいどれほどの影響を与えているのかを測ることはできなかった。様々な要因を分割してそれぞれの影響を測り数値化するができれば、より精密な分析が行えるのではないかと考えている。

調査の結果を見ることでわかったことなのだが、分布のロングテール化は極めてゆっくり進むということから、もっと長い期間、例えば数か月単位での観察を行なった方が今回よりも時間の経過による変化が顕著に表れ、よりわかりやすい結果を得られるのではないかと考えられる。

7. 付録

7.1. 大数の弱法則 (weak law of large numbers)

$\{X_i\}$ を独立同分布をもつ確率変数の族で $E[(X_1)^2] < \infty$ とする。このとき任意の $\epsilon > 0$ に対して以下が成り立つ。

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - E[X_1]\right| \geq \epsilon\right) = 0$$

証明：まず、以下のシュワルツ (Schwarz) の不等式より $E[|X_1|] \leq \sqrt{E[(X_1)^2]} < \infty$ となることに注意しておく。さて、 $X = |S_n - nE[X_1]|$ 、 $f(a) = a^2 (a \geq 0)$ 、 $= 0 (a \leq 0)$ についてチェビシェフ (Chebyshev) の不等式を用いると、

$$\begin{aligned} P(|S_n - nE[X_1]| \geq n\epsilon) &\leq \frac{E[|S_n - nE[X_1]|^2]}{(n\epsilon)^2} \\ &= \frac{nE[|X_1 - E[X_1]|^2]}{(n\epsilon)^2} \end{aligned}$$

となり、 $n \rightarrow \infty$ により結論を得る。

■シュワルツ (Schwarz) の不等式

確率変数 X_1 、 X_2 が $E[(X_1)^2] < \infty$ 、 $E[(X_2)^2] < \infty$ を満たすとき、

$$E[X_1 X_2] \leq \sqrt{E[(X_1)^2]} \sqrt{E[(X_2)^2]} < \infty$$

が成り立つ。特に、 $X_2 = 1$ として $E[X_1] \leq \sqrt{E[(X_1)^2]} < \infty$ を得る。

■チェビシェフ (Chebyshev) の不等式

f を非負値単調非減少関数 ($x \leq y$ のとき $0 \leq f(x) \leq f(y)$ となる関数) とし、 X を $E[|X|] < \infty$ 、 $E[|f(X)|] < \infty$ を満たす確率変数とすると、 $f(a) > 0$ を満たす $a \in \mathbb{R}$ について以下が成り立つ。

$$P(X \geq a) \leq E[f(X)]/f(a)$$

7.2. AKB 選抜総選挙の公式発表

「第二回選抜総選挙」の中間発表

AKB48選抜総選挙 投票結果中間報告(2010年06月02日時点)

順位	チーム名	タレント名(漢字)	タレント名(よみがな)	投票数
1	A	前田 敦子	まえだ あつこ	20,966
2	K	大島 優子	おおしま ゆうこ	19,465
3	K	板野 友美	いたの ともみ	13,474
4	A	篠田 麻里子	しのだ まりこ	13,289
5	B	渡辺 麻友	わたなべ まゆ	12,307
6	A	高橋 みなみ	たかはし みなみ	9,983
7	A	小嶋 陽菜	こじま はるな	9,252
8	B	柏木 由紀	かしわぎ ゆき	8,709
9	K	宮澤 佐江	みやざわ さえ	7,683
10	S	松井 玲奈	まつい れな	6,419
11	A	高城 亜樹	たかじょう あき	6,357
12	K	小野 恵令奈	おの えれな	4,766
13	B	河西 智美	かさい ともみ	4,532
14	B	北原 里英	きたはら りえ	4,414
15	K	峯岸 みなみ	みねぎし みなみ	4,091
15	S	松井 珠理奈	まつい じゅりな	4,091
17	B	宮崎 美穂	みやざき みほ	3,369
18	A	指原 莉乃	さしはらの りの	3,357
19	B	佐藤 亜美菜	さとう あみな	3,227
20	A	仲川 遥香	なかがわ はるか	2,726
21	K	秋元 才加	あきもと さやか	2,621
22	A	多田 愛佳	おおた あいか	2,332
23	A	倉持 明日香	くらもち あすか	2,278
24	B	増田 有華	ますだ ゆか	1,644
25	B	平嶋 夏海	ひらじま なつみ	1,633
26	B	小森 美果	こもり みか	1,430
27	S	大矢 真那	おおや まさな	1,217
28	K	仁藤 萌乃	にとう もえの	1,050
29	K	梅田 彩佳	うめだ あやか	903
30	S	矢神 久美	やがみ くみ	897
31	B	石田 晴香	いしだ はるか	892
32	K	藤江 れいな	ふじえ れいな	880
33	B	佐藤 すみれ	さとう すみれ	869
34	研究生	山内 鈴蘭	やまうち すずらん	838
35	研究生	島崎 遥香	しまざき はるか	835
36	A	片山 陽加	かたやま はるか	736
37	研究生	石黒 貴己	いしくろ あつき	682
38	K	米沢 瑠美	よねざわ るみ	646
39	K	菊地 あやか	きくち あやか	641
40	A	前田 亜美	まえだ あみ	568
40	K II	高柳 明音	たかやなぎ あかね	568

株式会社メディアシークは、厳正なシステム運用監視体制をとっており、今回の集計結果に対する第三者による不正な票操作や集計結果の改ざん等の事実は認められませんので、ここにご報告申し上げます。

なお、本紙を画像データとして取り込んだ電子ファイル(以下、「本紙データ」といいます。)に対しても偽造や改ざんが行われていないことを証明するため、株式会社メディアシークは、本紙データを主務大臣(総務大臣、法務大臣及び経済産業大臣)より認定を受けている認証認定事業会社である株式会社日本電子公証機構に登録し、作成者および作成日時の公証を受けております。

日付: 平成22年6月2日
 会社名: 株式会社メディアシーク
 住所: 東京都港区麻布台2-3-5ノアビル
 担当役職: マネージャー
 担当: 近藤 力



(URL: http://stat.ameba.jp/user_images/20100602/21/akihabara48/f0/00/j/o0721102410571747209.jpg)

「第二回選抜総選挙」の最終結果

AKB48選抜総選挙 最終集計結果(2010年06月09日)

順位	チーム名	タレント名(漢字)	タレント名(よみがな)	投票数
1	K	大島 優子	おおしま ゆうこ	31,448
2	A	前田 敦子	まえだ あつこ	30,851
3	A	篠田 麻里子	しのだ まりこ	23,139
4	K	板野 友美	いたの ともみ	20,513
5	B	渡辺 麻友	わたなべ まゆ	20,088
6	A	高橋 みなみ	たかはし みなみ	17,787
7	A	小嶋 陽菜	こじま はるな	16,231
8	B	柏木 由紀	かしわぎ ゆき	15,466
9	K	宮澤 佐江	みやざわ さえ	12,560
10	S	松井 珠理奈	まつい じゅりな	12,168
11	S	松井 玲奈	まつい れな	12,082
12	B	河西 智美	かさい ともみ	11,080
13	A	高城 亜樹	たかじょう あき	11,062
14	K	峯岸 みなみ	みねぎし みなみ	9,692
15	K	小野 恵令奈	おの えれな	9,468
16	B	北原 里英	きたはら りえ	8,836
17	K	秋元 才加	あきもと さやか	8,049
18	B	佐藤 亜美菜	さとう あみな	6,921
19	A	指原 莉乃	さしはらりの	6,704
20	A	仲川 遥香	なかがわ はるか	6,567
21	B	宮崎 美穂	みやざき みほ	6,371
22	A	多田 愛佳	おおた あいか	6,145
23	A	倉持 明日香	くらもち あすか	5,355
24	S	大矢 真那	おおや まさな	4,634
25	B	増田 有華	ますだ ゆか	4,137
26	B	平嶋 夏海	ひらしま なつみ	4,106
27	B	石田 晴香	いした はるか	3,235
28	研究生	島崎 遥香	しまざき はるか	3,076
29	K	仁藤 萌乃	にとう もえの	2,693
30	B	小森 美果	こもり みか	2,613
31	B	佐藤 すみれ	さとう すみれ	2,591
32	K	梅田 彩佳	うめだ あやか	2,499
33	K	藤江 れいな	ふじえ れいな	2,460
34	K	米沢 瑠美	よねざわ るみ	2,171
35	KII	高柳 明音	たかやなぎ あかね	2,030
36	研究生	山内 鈴蘭	やまうち すずらん	1,945
37	A	片山 陽加	かたやま はるか	1,935
38	S	矢神 久美	やがみ くみ	1,909
39	A	松原 夏海	まつばら なつみ	1,854
40	研究生	石黒 貴己	いしぐろ あつき	1,603

株式会社メディアシークは、厳正なシステム運用監視体制をとっており、今回の集計結果に対する第三者による不正な票操作や集計結果の改ざん等の事実は認められませんでしたので、ここにご報告申し上げます。

なお、本紙を画像データとして取り込んだ電子ファイル(以下、「本紙データ」といいます。)に対しても偽造や改ざんが行われていないことを証明するため、株式会社メディアシークは、本紙データを主務大臣(総務大臣、法務大臣及び経済産業大臣)より認定を受けている認証認定事業会社である株式会社日本電子公証機構に登録し、作成者および作成日時の公証を受けております。

日付: 平成22年6月9日
 会社名: 株式会社メディアシーク
 住所: 東京都港区麻布台2-3-5 アビッセビル
 担当役職: マネージャー
 担当:

近藤 力



「第三回選抜総選挙」の速報

キングレコード株式会社 御中

平成23年5月25日

P/PED BITS
株式会社「バイドビッツ」

AKB48 22ndシングル 選抜総選挙
速報結果

順位	チーム名	メンバー名(漢字)	メンバー名(読みがな)	得票数
1	K	大島 優子	おおしま ゆうこ	17,136
2	A	大崎 実花	おおさき みか	16,432
3	B	橋本 紗希	はしもと さき	12,056
4	A	高橋 みなみ	たかはし みなみ	8,833
5	B	藤田 奈々	ふじた なな	8,582
6	A	藤原 菜生子	ふじはら なし	8,016
7	A	指原 莉乃	さしはら りの	7,357
8	K	板野 友美	いたの ともみ	6,595
9	S	松井 玲奈	まつい れな	6,509
10	A	小嶋 優菜	こじま はるな	6,534
11	K	菅原 紗由香	すがはら さゆか	5,157
12	A	宮城 真奈美	みやぎ まなみ	5,086
13	K	峯 岸 みなみ	たかねの みなみ	3,931
14	B	北原 美央	きたはら み	3,860
15	B	河野 真美	かひの まみ	3,102
16	S	松井 珠依	まつい じゆい	2,843
17	K	橋山 由莉	はしやま ゆり	2,753
18	B	佐藤 亜美	さとう あみ	2,684
19	B	田嶋 由香	たじま ゆか	2,201
20	A	倉持 明日香	くらもち あすか	1,793
21	K	高柳 明音	たかやなぎ あかね	1,761
22	A	徳川 愛	とくがわ あい	1,571
23	K	秋元 才加	あきもと さやか	1,502
24	B	平塚 穂香	ひらつか ほのか	1,484
25	N	山本 彩	やまもと さやか	1,444
26	K	藤山 佳	ふじやま けい	1,441
27	B	宮崎 美穂	みやざき みほ	1,378
28	S	菊地 亜美	きくち あみ	1,139
29	A	佐藤 亜衣	さとう あい	1,129
30	K	笠原 和子	かさはら かつこ	1,118
31	S	木崎 紗里	きざき さり	1,053
32	A	大塚 志津香	おおつか しづか	904
33	S	大塚 悠妃	おおつか ゆい	885
34	N	宮崎 穂花	みやざき ほな	854
35	B	佐藤 莉子	さとう りこ	852
36	K	松井 咲子	まつい さき	814
37	初参选	大塚 悠花	おおつか ゆか	811
38	B	小森 美穂	こもり みほ	792
39	S	松村 久美	まつむら くみ	779
40	K	藤江 れい	ふじえ れい	778

株式会社「バイドビッツ」は、本システム及びデータを厳格に管理し、不正なアクセスおよびコンピュータ・プログラムの複製、破壊、改竄、漏洩等の危険に對して、技術および新機軸により高度なセキュリティを確保しており、今回の選挙結果に對する第三者による不正な操作や不正な結果の改ざん等のリスクは認められませんでしたので、ここに公開申し上げます。

投票統計開始 平成23年5月24日 午前10時から速報発表まで
 会社名: 株式会社「バイドビッツ」
 住所: 東京都港区赤坂2-9-11 オリックス赤坂2丁目ビル
 役職: 執行役員
 氏名: 古江 恵治

(URL: http://stat.ameba.jp/user_images/20110525/22/akihabara48/ea/3e/j/o0454064011250617846.jpg)

「第三回選抜総選挙」の最終結果

キングレコード株式会社 御中

平成23年6月9日

P/ PED BITS
株式会社パイブドビッツ

AKB48 22ndシングル 選抜総選挙
最終集計結果

順位	チーム名	タレント名(漢字)	タレント名(よみがな)	投票数
1	A	前田 敦子	まえだ あつこ	139,892
2	K	大島 優子	おおしま ゆうこ	122,843
3	B	柏木 由紀	かしわぎ ゆき	74,252
4	A	篠田 麻里子	しのだ まりこ	60,539
5	B	渡辺 麻友	わたなべ まゆ	59,118
6	A	小嶋 陽菜	こじま はるな	52,920
7	A	高橋 みなみ	たかはし みなみ	52,790
8	K	板野 友美	いたの ともみ	50,403
9	A	指原 莉乃	さしはらりの	45,227
10	S	松井 玲奈	まつい れな	36,929
11	K	宮澤 佐江	みやざわ さえ	33,500
12	A	高城 亜樹	たかじょう あき	31,009
13	B	北原 里英	きたはらりえ	27,957
14	S	松井 珠理奈	まつい じゅりな	27,804
15	K	峯岸 みなみ	みねぎし みなみ	26,070
16	B	河西 智美	かさい ともみ	22,857
17	K	秋元 才加	あきもと さやか	17,154
18	B	佐藤 亜美菜	さとう あみな	16,574
19	K	横山 由依	よこやま ゆい	16,455
20	B	増田 有華	ますだ ゆか	14,137
21	A	倉持 明日香	くらもち あすか	12,387
22	K	梅田 彩佳	うめだ あやか	11,860
23	K II	高柳 明音	たかやなぎ あかね	11,674
24	A	仲川 遥香	なかがわ はるか	10,854
25	A	多田 愛佳	おおた あいか	9,910
26	B	平嶋 夏海	ひらじま なつみ	9,742
27	B	宮崎 美穂	みやざき みほ	9,271
28	N	山本 彩	やまもと さやか	8,697
29	A	大家 志津香	おおやしづか	7,264
30	S	大矢 真那	おおや まさな	6,660
31	K	仁藤 萌乃	にともえの	6,288
32	B	小森 美果	こもり みか	6,120
33	K II	秦 佐和子	はた さわこ	6,117
34	B	佐藤 すみれ	さとう すみれ	5,438
35	研究生	大場 美奈	おおば みな	5,411
36	S	須田 亜香里	すだ あかり	5,343
37	A	前田 亜美	まえだ あみ	5,220
38	K	松井 咲子	まつい さきこ	5,020
39	研究生	市川 美織	いちかわ みおり	4,928
40	K	藤江 れいな	ふじえ れいな	4,698

株式会社パイブドビッツは、本件システム及びデータを厳格に管理し、不正なアクセスおよびコンピュータープログラムの紛失、破壊、改竄、漏洩等の危険に対して、技術面および組織面において合理的な安全対策を講じており、今回の集計結果に対する第三者による不正な票操作や集計結果の改ざん等の事実はありませんでしたので、ここにご報告申し上げます。

投票集計期間：平成23年5月24日午前10時から平成23年6月8日午後3時まで
 会社名：株式会社パイブドビッツ
 住所：東京都港区赤坂2-9-11 オリックス赤坂2丁目ビル2階・3階
 役職：執行役員
 氏名：

古江 恵治

立会人：

弁護士 矢部 陽



(URL:http://stat.ameba.jp/user_images/20110609/23/akihabara48/ed/68/j/o0800113111281026729.jpg)

8. おわりに

8.1. iPod 内の音楽再生数ランキング

今回、十分なデータを得ることが出来ずに実現出来なかったこととして、Apple の iPod や SONY のウォークマンなど個人の音楽プレーヤーの再生回数のランキングを集計するといったことがある。個人のデータを多数集めること、つまり曲数や再生数のデータを増やしていくことによってランキングのグラフは大数の法則によって今の社会の挙動に近づいてゆき、それが果たしてロングテール構造に近づいていくのかについての考察を行うことができるのではないかと考えている。今回は多くのサンプルを得ることができなかったが、一個人のデータについてまとめたグラフ（図 23）を用意することができたので以下に載せる。

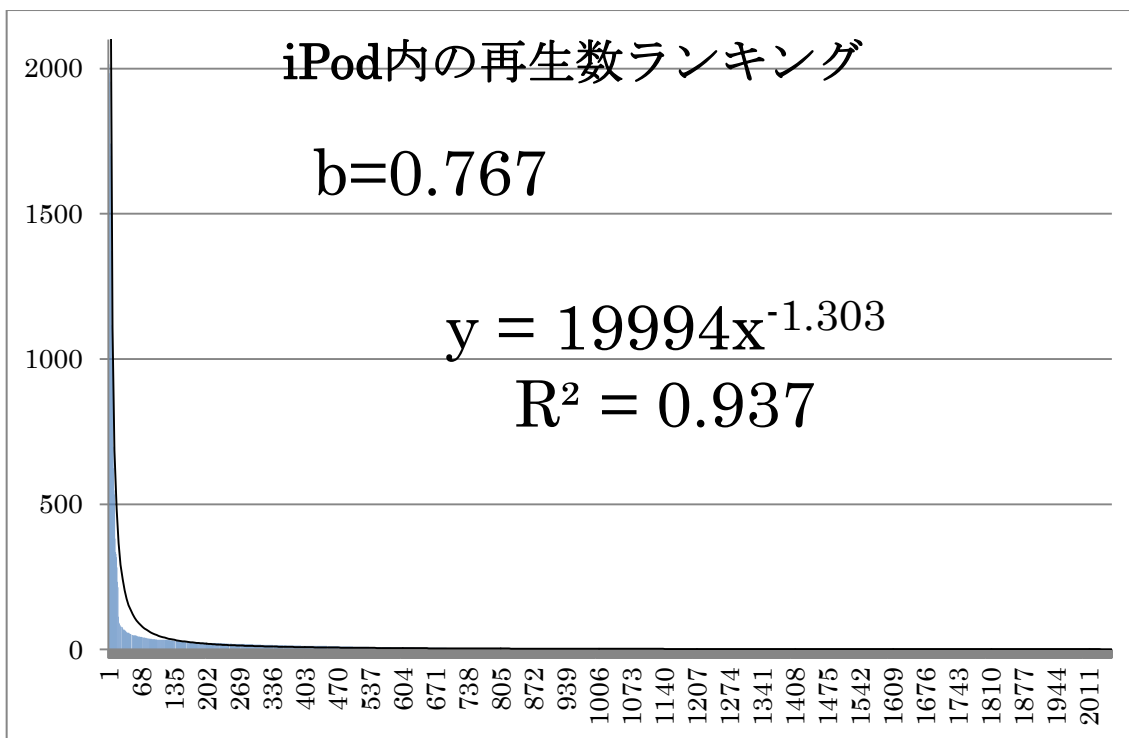


図 23 iPod 内の再生数ランキング

縦軸は再生数であり横軸は曲名である。

期間はおよそ4年であり、ある曲を最初から最後まで聴くことによって再生回数にカウントされるという仕組みである。曲によっては、その曲が聴きたかったわけではないが順番の都合上再生されてしま

うものもあるが、そのようなことは考慮できていない。

グラフでは多少わかりにくいかもしれないが、縦軸の再生数はヘッ드의部分が非常に大きく 2000 を超えるものも存在している一方で、テール部分は再生数が 1 ケタの曲が非常に多く存在しており長くなっている。パレートの分布の平等性の指数をみると、 $b=0.767$ となっており、Zipf の法則に従うとすればロングテールは成立していないと判断できる。ただ、近似曲線の当てはまりの程度を示す R^2 値は 0.937 であるものの、図を見てもらうとわかるようにヘッドとテールの間の部分の当てはまりが決して良いとは言えない。これまでのようにいくつかに分けてそれぞれで近似をするのかも考えたが、AKB48 の選抜総選挙のように〇位以内に特別な権利があったりするわけではなく、また、全てのデータが直接得られたものであるため、今回は行わないこととした。

私がこれまでに行ったニコニコ動画の再生数ランキングや AKB48 の選抜総選挙の調査と比べ、この調査は、テールの部分を実際に測ることができるという特徴がある。そのため、データの予測といった曖昧な部分がなくより現象を正確に表すことができる。しかし、携帯音楽プレーヤーの再生回数といったデータはプライベートなものであること、また入手するためには個人の PC を預らざるをえないということは多くのサンプルを集めるということの非常に大きな障害となっており、今回実現することは出来なかった。これは、本稿で達成できなかったことであり、これから機会があれば研究をしてみたいと感じている。

8.2. 研究を終えて

個人的なこととしてここに挙げさせてもらうが、私は研究を進めていく中で実際に自分のニーズがニッチなところにあったという経験をした。ニコニコ動画では、ランキングには載っていなかったけど自分の好きな音楽を見つけることができたし、AKB48 では、ランキング外に好みのメンバーを見つけたりもした。今回、本稿においてロングテールは成り立っていない、テールはなくても構わないと言ったかもしれないが、ミクロな視点で見た時にニーズが満たされるというのは良いことであると考えており、もう少し時間が経てばロングテールは成立し、ミクロとマクロのどちらの視点からもニッチなところが必要と言えるようになるのではないかと考えており、私はそうなることを願っている。

AKB48 は、論文を書き始めたころから既に人気があったと思うが、論文執筆中の勢いはすさまじいものがありまさに社会現象を巻き起こしているといえる。歌番組はもちろん、バラエティ番組やドラマ、テレビ CM、ニュースや雑誌ついにはコンビニまで見渡せば AKB48 といった感じである。また、論文執

筆中に新しく福岡の博多を拠点とする HKT48 や国内にとどまらず海外に進出してインドネシアのジャカルタで JKT48 が誕生するなどその規模はさらに大きく拡大している。私は、その拡大を大いに歓迎している。それは、お分かりだと思うが次回、選抜総選挙が行われた際の総投票者数が増えるからである。今回の分析と同様のことを行なえば、よりよい結果を得られるのではないかと考えている。次の‘第四回選抜総選挙’が行なわれることを願いたい。ただ、若い女の子が公衆の面前にさらされて順位付けされるというのはどのような気持ちなのだろうかという心配は消えることはない。

8.3. 謝辞

一人で論文を書くということはこれまで経験していなかったこともあり、担当教員である服部哲弥教授をはじめ、友人である阿野喜史さんや他大勢の方々に非常に貴重なアドバイスを頂いたり大変お世話になり、彼らがいなければ本稿を書き上げることは決して出来なかった。ここに今一度御礼を申し上げたい。本当にありがとうございました。今回論文を書き進めるにあたって、どのような内容の論文を書くにしても非常に地味な作業を伴うということは、執筆前の想像を超えており、非常に大変であった。しかし、担当教員の助けがあったものの一人で論文を書き上げたことは私自身非常にいい経験になったと感じており、来年度の卒業論文ではこの経験を活かしていきたい。

9. 参考文献・データ出典

服部哲弥, Amazon ランキングの謎を解く 確率的な順位付けが教える売上の構造(化学同人, 2011.5)

クリス・アンダーソン/篠森ゆりこ=訳, ロングテール 「売れない商品」を宝の山に変える新戦略(早川書房, 2006.9)

熊谷隆, 新しい解析学の流れ 確率論(共立出版, 2003.3)

笠原勇二, 明解 確率論入門(数学書房, 2010.1)

鈴木貴博, アマゾンのロングテールは、二度笑う 「50年勝ち残る会社」をつくる8つの戦略(講談社, 2006.10)

河野哲也, レポート・論文の書き方入門 第3版(慶應義塾出版会, 2002.12)

ニコニコ動画ランキング(<http://www.nicovideo.jp/ranking>)

閲覧日：2011年10月23日, 12月1日, 5日, 12日, 2012年1月1日

～AKB48 TOKYO DOME までの軌跡～(<http://ameblo.jp/akihabara48/>)

閲覧日：2011年6月28日

デイリースポーツオンライン(<http://www.daily.co.jp/gossip/article/2011/06/11/0004159535.shtml>)

閲覧日：2012年1月12日