

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 26 日現在

機関番号：18001

研究種目：若手研究(B)

研究期間：2013～2016

課題番号：25870743

研究課題名(和文) 解ける確率過程を用いたロングテール現象の分析

研究課題名(英文) Analysis of long-tailed phenomena using solvable stochastic processes

研究代表者

山本 健 (Yamamoto, Ken)

琉球大学・理学部・講師

研究者番号：00634693

交付決定額(研究期間全体)：(直接経費) 1,400,000円

研究成果の概要(和文)：本研究では、社会現象および生物現象において裾の長い確率分布が現れる仕組みを理論的に分析した。映画の興行収入の分布にはベキ分布がみられ、ある週に大きな収入を上げた映画は翌週も観客を集めやすいという正のフィードバックの効果を取り入れた確率過程にて記述できることを示した。法律の条文数の分布には対数正規分布がみられ、深さが正規分布で与えられる木構造によって説明できることを示した。バクテリアの長さの分布は対数正規分布と一致することが実験的に示されていたが、本研究では現象論的な確率モデルの解析により、実験結果の理論的な裏付けを与えた。

研究成果の概要(英文)：In this research project, heavy-tailed probability distributions in social and biological phenomena have been analyzed theoretically. First, a simple stochastic model has been constructed and analyzed, which is regarded as a model for the box-office grosse of a movie. This model can explain a power-law decay in the distribution of box-office grosses. Second, we have found that the number of articles within a law follows a lognormal distribution. The tree structure whose depth is normally distributed is an appropriate model for this lognormal behavior. Third, the size distribution of bacterial cells is studied. Our analysis of a phenomenological stochastic model for the bacterial growth has given a theoretical basis for the lognormality of the cell-size distribution.

研究分野：数理物理学

キーワード：確率過程 ベキ分布 対数正規分布

1. 研究開始当初の背景

(1) 社会系や生物系などの、いわゆる複雑系に関するデータには、特徴的な値よりも著しく大きな値が少数ながら存在することが多くある。このようなデータの確率分布は長く裾を引くグラフになる。特に、商品を売上の順に並べたグラフが裾の長い分布になることはロングテール現象とよばれる。

(2) 裾の長い確率分布が現れるためには何らかの仕組みが必要である。商品の売上の過程では、“ある時点までによく売れた商品は、その後もよく売れる”という正のフィードバックによって裾の長い確率分布が生じると予想し、本研究を開始した。正のフィードバックの効果を取り入れたシンプルな確率過程を構築し、その解析をおこなった。さらに、この確率過程を映画の興行収入のデータと比較し、確率過程が実際の売上現象に対する妥当なモデル化となっていることを示した。

2. 研究の目的

本研究の目的は、シンプルな確率モデル・確率過程を解析することで、裾の長い確率分布が出現する仕組みを理論的に分析することである。現象の大まかな特徴をとらえたシンプルなモデル化をおこなうことにより、モデルの厳密な解析が可能となる。そのため、裾の長い確率分布が現れる理論的な仕組みを明確に表現することができると期待できる。裾の長い確率分布の中でも、ベキ分布と対数正規分布の2つは様々な現象で観察され、理論的にも重要な確率分布である。本研究では、特にこの2つの確率分布の性質に注目して分析を進めた。

3. 研究の方法

(1) 社会現象として映画の興行収入額および法律の条文数を、生物現象としてバクテリアのサイズを研究の対象とした。実データの収集と解析、確率過程によるモデル化、およびモデルの解析によって、裾の長い確率分布が現れる仕組みを理論的に研究した。確率過程の数理物理学的な解析自体も、独立した1つの研究成果である。

(2) 映画の興行収入のデータはアメリカの映画に関するデータベース(*Box Office Mojo* および *The Numbers*)を利用した。法律のデータは、法令データ提供システム(日本)、*Juris BMJ* (ドイツ) および *Singapore Statutes Online* (シンガポール)のデータベースから取得した。バクテリアのサイズの研究では、共同研究者から実験データの提供を受けた。

4. 研究成果

(1) 映画の興行収入の時間変化を単純化した確率過程

$$\begin{cases} x(t+1) = \mu(t)x(t) \\ S(t+1) = S(t) + x(t) \end{cases}$$

の解析をおこなった。第1式は、公開後 t 週目の興行収入 $x(t)$ が大きいほど翌週の興行収入 $x(t+1)$ が大きいという正のフィードバックの効果を表す。 $\mu(t)$ は、週ごとの興行収入の増大率を表す確率変数である。 $S(t)$ は、各週の興行収入の合計を表す。この確率過程を模式的に描いたのが図1である。

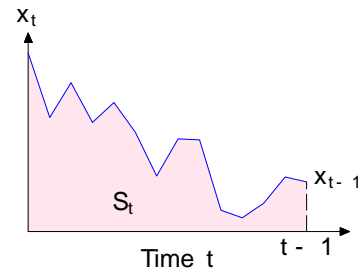


図1 確率過程の模式図。 $x(t)$ のグラフより下の部分の面積が $S(t)$ に相当する。

ケステン過程という確率過程と対応させることにより、 t が十分に大きいときに $S(t)$ の分布は定常になり、

$$P(S(t) > s) \sim s^{-\beta}$$

というべき関数で減衰する裾をもつことを示した。ここに現れるべき指数 β は、増大率 $\mu(t)$ に関する方程式

$$E[\mu(t)^\beta] = 1$$

の正の解として与えられる。解 β は、存在すれば一意であることも示した。

(2) この確率過程の性質として、増大率 $\mu(t)$ が $E[\mu(t)] = 1$ という条件を満たすならば、 $\beta = 1$ であることが分かる。べき指数 $\beta = 1$ に対応するべき分布はジップの法則とよばれ、文章中の単語の出現頻度や都市の人口など様々な量の分布にみられることが知られている。条件 $E[\mu(t)] = 1$ は、「平均的には $x(t)$ が増大も減少もしない」という一種の安定性を意味する。この安定性が、ジップの法則が幅広い現象においてみられる理論的根拠となりうるのではないかと考えた。

(3) 以上の数学的な結果を映画の興行収入のデータに適用した。はじめに、2012年にアメリカで公開された映画の(アメリカ国内での)興行収入額の確率分布を作成した。しかし、分布をみただけでは裾の長い確率分布かどうかを判断することはできなかった。著名な俳優を起用したり多額の広告宣伝費を割いたりすることが可能な、制作規模が大きい映画ほど興行収入をあげやすい。したがって、興行収入額そのものの分布は、映画ごとの制作規模のばらつきを反映して込み入った形になると考えられる。こういった複雑な要因をなるべく排除するため、映画の興行収入を制作費で割った「投資利益率」(return on

investment, ROI) という指標に注目することにした。投資利益率の確率分布は指数 $\beta = 1.12$ 程度のべき分布的な裾をもつことが分かった (図 2)。

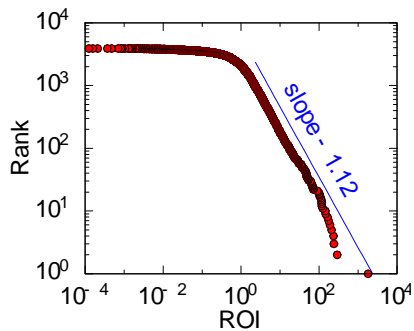


図 2 映画の投資利益率(ROI)の累積分布。裾が $\beta = 1.12$ のべき分布にしたがう。(データベース *The Numbers* に収録された 1915 年から 2014 年 4 月までの 3906 作品。)

一方、各週の興行収入額のデータから増大率 $\mu(t)$ を計算し、(1)の確率過程を仮定してべき指数 β を求めると、 $\beta = 1.10$ という結果が得られた。この値は実際の分布から得られた値 1.12 と非常に近い。以上の定量的な一致から、(1)の確率過程が現実の映画の集客の特徴をうまくとらえていると考えられる。

さらに、(1)の確率過程から、各週の興行収入額 $x(t)$ は平均的には時間とともに指数的に減衰することが示される。実際の映画でも興行収入の指数的な減衰がみられ、その減衰率は理論から導かれる値とよく一致することを確認した。この結果からも(1)の確率過程が妥当なモデルであることが示唆される。

(4) 他の社会現象の例として、法律の条文数の分布の解析をおこなった。日本、ドイツ、シンガポールの法律の条文数の確率分布がみな対数正規分布とよく一致することを示した (日本の法律の分布を図 3 に示す)。

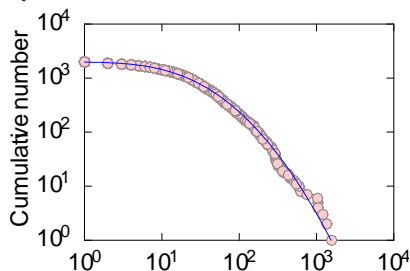


図 3 日本の法律 (2017 年 1 月時点、総数 2000) の条文数の累積分布。実線で示した対数正規分布とよく一致する。

異なる国々の法律の条文数が共通して対数正規分布にしたがうということは、言語や文化的な背景とは無関係な理論的にシンプルな仕組みが存在すると期待される。本研究では、法律の「木構造」に注目した。法律を構成する条文どうしの関係性によって、いくつかの大きなまとまりに分けることができ、そ

れぞれのまとまりはさらにいくつかのまとまりに分けられる。このように、条文の集まりを階層的に分割していくことで、木構造が得られる。条文数は木の末端のノード数に対応する。分岐数が一定であり深さが正規分布で表される木の集団では、末端のノード数が対数正規分布にしたがうことが理論的に示せる。木構造を調べる作業を省略するため、日本の法律の章立ての構造に着目して解析をおこなった。データ解析から得られた量を用いて対数正規分布のパラメータを求めると、実際の条文数分布の値と近くなった。この結果から、木構造によるモデルの妥当性が示唆される。

(5) 生物のサイズ分布として、バクテリアの長さ分布を解析した。過去の研究では、枯草菌 *Bacillus subtilis* の菌体長分布が対数正規分布とよく一致することが過去に調べられており、さらに菌の成長と分裂の効果を取り入れた現象論的な確率モデルが提案された。本研究では、この現象論モデルの数学的性質を調べた。菌の成長速度と分裂周期がともに対数正規分布にしたがうという実験結果を利用し、定常状態における菌体長は対数正規分布にしたがう確率変数の無限和で表されることを示した。対数正規分布の和は対数正規分布にならないことから、現象論モデルを与える菌体長分布は厳密には対数正規分布でない結論される。さらに、現象論モデルによる菌体長分布の対数正規分布からのずれを調べるために、与えられた確率分布と対数正規分布とのずれを計算する手法を導入した。成長速度と分裂周期の確率変数を変化させ、菌体長の定常分布の対数正規分布からのずれを数値的に求めると、枯草菌の実験に対応するパラメータ領域の付近において対数正規分布からのずれが極小となることが示された。この結果は、実際の枯草菌の菌体長分布が対数正規分布に近くなることを意味する。以上の結果より、成長速度と分裂周期が枯草菌と異なる菌のサイズ分布は対数正規分布から大きくずれることが予想される。今後は、実験研究者との共同研究によって、この予想を実験的に検証したい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 6 件)

K. Koyama, K. Yamamoto, and M. Ushio, A lognormal distribution of the length of terminal twigs on self-similar branches of elm trees, Proceedings of the Royal Society of London B 284, 20162395 (2017). [8 ページ、査読有]
DOI: 10.1098/rspb.2016.2395

K. Yamamoto and J. Wakita, Analysis of a stochastic model for bacterial

growth and the lognormality of the cell-size distribution, Journal of the Physical Society of Japan 85, 074004 (2016). [7 ページ、査読有]

DOI: 10.7566/JPSJ.85.074004

K. Yamamoto, Topological analysis of rough surfaces using persistent homology, Journal of the Physical Society of Japan 84, 113001 (2015). [4 ページ、査読有]

DOI: 10.7566/JPSJ.84.113001

K. Yamamoto, Stochastic model of Zipf's law and the universality of the power-law exponent, Physical Review E 89, 042115 (2014). [5 ページ、査読有]

DOI: 10.1103/PhysRevE.89.042115

K. Yamamoto, A simple view of the heavy-tailed sales distributions and application to the box-office grosses of U.S. movies, Europhysics Letters 108, 68004 (2014). [5 ページ、査読有]

DOI: 10.1209/0295-5075/108/68004

K. Yamamoto and J. Wakita, "Average-value paradox" of the lognormal distribution, Journal of the Physical Society of Japan 82, 113001 (2013). [3 ページ、査読有]

DOI: 10.7566/JPSJ.82.113001

[学会発表](計12件)

山本健, 法律の条文数分布とその解析, 統計数理研究所共同研究集会「社会物理学の新展開」, 2017年3月25日, 統計数理研究所(東京都立川市).

山本健, 二分木の分岐解析と分岐比に対する中心極限定理, 2016年度応用数学合同研究集会, 2016年12月15日, 龍谷大学(滋賀県大津市).

山本健, 脇田順一, バクテリア成長の現象論的モデルの解析と菌体長の対数正規性の検証, 日本物理学会2016年秋季大会, 2016年9月14日, 金沢大学(石川県金沢市).

山本健, ベキ分布を導く確率過程と映画の興行収入データの分析, 日本応用数理学会2015年度年会, 2015年9月10日, 金沢大学(石川県金沢市).

山本健, ベキ分布を導く確率モデルと映画の興行収入データの分析, 統計数理研究所共同研究集会「社会物理学の現代的課題」, 2015年3月25日, 統計数理研究所(東京都立川市).

山本健, ベキ乗則を導く確率モデルによる映画の統計データの解析, 日本物理学会第70回年次大会, 2015年3月24日, 早稲田大学(東京都新宿区).

山本健, ベキ乗則を導く確率モデルと映画の統計データへの適用, 第11回数学総合若手研究集会, 2015年3月4日, 北海道大学(北海道札幌市).

山本健, ズブラ過程の合計として導かれるジップの法則, 日本物理学会2014年秋季大会, 2014年9月7日, 中部大学(愛知県春日井市).

山本健, ズップの法則における規模と順位との反比例関係に対する一考察, 第10回数学総合若手研究集会, 2014年3月4日, 北海道大学(北海道札幌市).

山本健, ズブラ過程の蓄積によるベキ分布, 日本物理学会2013年秋季大会, 2013年9月25日, 徳島大学(徳島県徳島市).

山本健, ズップの法則のベキ指数にはなぜ-1が多いのか?, 日本応用数理学会2013年度年会, 2013年9月11日, アクロス福岡(福岡県福岡市).

K. Yamamoto, Power-law behavior in accumulation of the Gibrat process, Mathematical Statistical Physics, 2013年7月30日, 京都大学(京都府京都市).

6. 研究組織

(1) 研究代表者

山本 健 (YAMAMOTO, Ken)

琉球大学・理学部・講師

研究者番号: 00634693