# SNS 言語空間のベキ法則と安定分布[1]

# Power Law in SNS Language Probability Space and Stable Distribution

山内康英（Yasuhide Yamanouchi）[1]・小松正（Tadashi Komatsu）[2]
[1]多摩大学情報社会学研究所　教授・[2]多摩大学情報社会学研究所　客員准教授

［Abstract］

The power Law or Zipf's law appears in many research areas, such as social disparities in personal incomes, price changes in commercial and stock markets, rankings in online product sales and large text data from SNS postings. General agreements has not yet reached among researchers about the generation mechanisms of the power law. Barabasi (2002) proposed the rule of preferential attachment to explain the power law's origination. The rule of preferential attachment is an algorithm for computer simulations and does not investigate the background of the stochastic processes in the power law. According to the model in this paper, the power law distribution observed in SNS language data is the sufficiently large positive values in non-Gaussian α-stable distribution. We can derive the stable distribution from the additive process consisting of a large number of elemental compound Poisson distributions. In this paper, by relying on stochastic process theories in precedence researchers on the stable distribution we verified the power law that appeared in SNS language probability space. We also considered the implication of this model for social sciences.

［キーワード］

ベキ法則、社会的構成主義、安定分布、ホワイトノイズ解析、SNS

## 1. はじめに

個人の所得と社会格差、商品や証券の取引と価格の変動、オンライン・ビジネスと商品販売数のランキング、SNS の投稿とテキスト・データなど、多くの社会現象にベキ法則（Power Law あるいは Zipf's law、Pareto's law）が現れる。現在のところ、異なった分野のデータに、なぜ共通してベキ法則が現れるのかについて、研究者のあいだに一致した見解はない。

ベキ法則の生成のメカニズムについて、バラバシは優先的選択（preferential attachment）の法則を示した。[2] この優先的選択の法則は、コンピュータ・シミュレーションのアルゴリズムであって、ベキ法則の背後にある確率過程のメカニズムを示すものではない。本論文のモデルによれば、SNS の言語データに現れるベキ法則は、非 - ガウス的な安定分布（non Gaussian α-stable distribution）の正の十分に大きな値を観測したものである。この安定分布を、多数の elemental な複合ポアソン分布からなる加法過程から導くことができる。本論文では、安定分布に関する確率過程論の先行研究のモデルを用いて、SNS の言語データに現れるベキ法則を検証した。本論文では、このモデルの社会科学におけるインプリケーションをあわせて検討した。[3]

### 安定分布と先行研究

Feller は 1966 年の著書に安定分布に関する一章を設けて、『安定分布は正規分布の自然な一般化として絶えず増大する役割を演じている』と述べている。[4] 安定分布（=α 安定分布）の一般論は、Paul Lévi の 1920 年代の研究に出発点を持つ。日本では飛田武幸教授のグループが、ホワイトノイズ解析として、この研究を発展させている。[5] 安定分布はガウス分布（=正規分布）、コーシー分布、レヴィ分布、ホルツマーク分布など、われわれが確率密度関数（PDF）のかたちを良く知っている既知の分布を含む概念である。安定分布の特性指数（α）によって、ここでいう既知の分布、およびその間に連続的にある無数の分布を決めることができる。ここでいう既知の分布以外の安定分布の確率密度関数を解析的なかたちで表現することはできない。解析的に表現できない安定分布の確率密度関数は、安定分布の一般特性関数から計算機的に推測する必要がある。

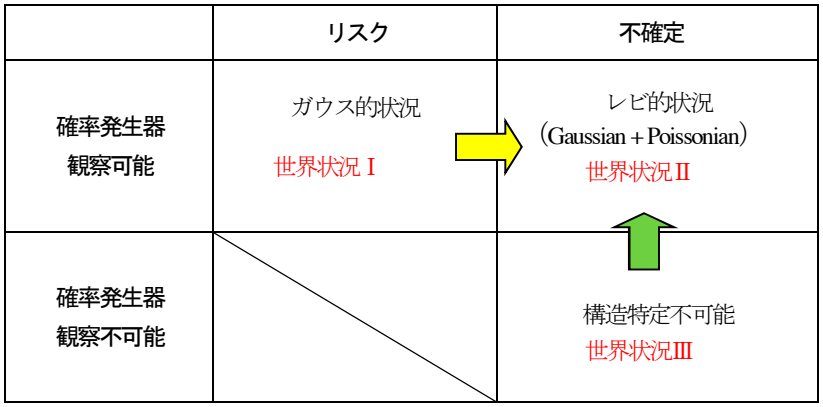
Mandelbrot は 1963 年の論文で、国際綿花市場の価格変化の大きさ（横軸=相対数）と、各価格変化の生起数（縦軸=相対数）の両対数グラフを、価格変化の正と負に分けて作り、価格変化のプロットが α=1.7 の安

定分布の裾野（横軸の十分に大きな値）と自己相似的（self-similar）であることを示した。[6] Taleb の言うファット・テイル現象のように、裾野の厚い非 - ガウス分布を統一的に取り扱う場合には、安定分布の検討が不可欠になる。[7] 服部は、2011 年の研究で、アマゾン書店のランキングについて確率順位付け模型を提示している。この研究は、各書籍が持つ独自のジャンプ率――単位時間当たりランキング 1 位に飛ぶ回数の期待値でポアソン分布の強度に相当する――が、全体として作り出す複合ポアソン過程を前提として、アマゾン書店が取り扱う全書籍のランク図のパレート分布の指数を推計している。[8]

次の理由から社会科学の研究者は、安定分布の重要性を認識する必要がある。社会的な事象とは、毎時間の確率分布の時系列の継起（確率論的には畳み込み（convolution：重畳））である。この継起的累積過程は、権力や制度といった社会システムの主体間の固定的な相互作用に拘束されるために、また社会的な慣習や文化、個人的な毎日の定型的な行動のために、一方では、ほとんど確実に（almost surely：a.s.）同分布的であり、他方では、毎時間の個別の決定に非常におおくの主体の自由意志が働くという点から、ほとんど確実に独立事象的になる。したがって大きな規模の社会事象の多くは、時間とともに増分・蓄積する独立で同分布な事象の継起的累積になる、と予想することができる。ここに社会事象のデータを、確率過程論に基づいて解析し、安定分布やホワイトノイズ解析を適用する根拠を求めることができる。

## 2. 社会的構成主義の安定過程による再定義

この考え方は、社会科学の歴史的構成主義（historical constructivism）の理解にとって重要である。[9] Blyth は、世界状況を「図 1」のように区分した。[10] 世界状況Ⅰは、直接的に観察可能（可視的）な確率発生器を持ち、計算可能な確率の世界＝リスクの世界である。この世界は高次のモーメントを持つ分布からできていて、過去をサンプリングすることは将来に対する良いガイドとなる。これに対して、世界状況Ⅱは、リスクと不確実性が曖昧な fat tail のある『「ガウス的＋ポアソン的」もしくはレヴィ的』[11]であって、2 次のモーメント＝分散が知られておらず、過去のデータを完全にサンプリングすることができたとしても、主体は確率発生器を特定することができない。世界状況Ⅲでは、これらの確率モーメントは単に存在しない。この世界状況Ⅰ、Ⅱ、Ⅲは、安定分布の特性指数が 2（すなわちガウス分布）からより小さくなる状況に対応している。



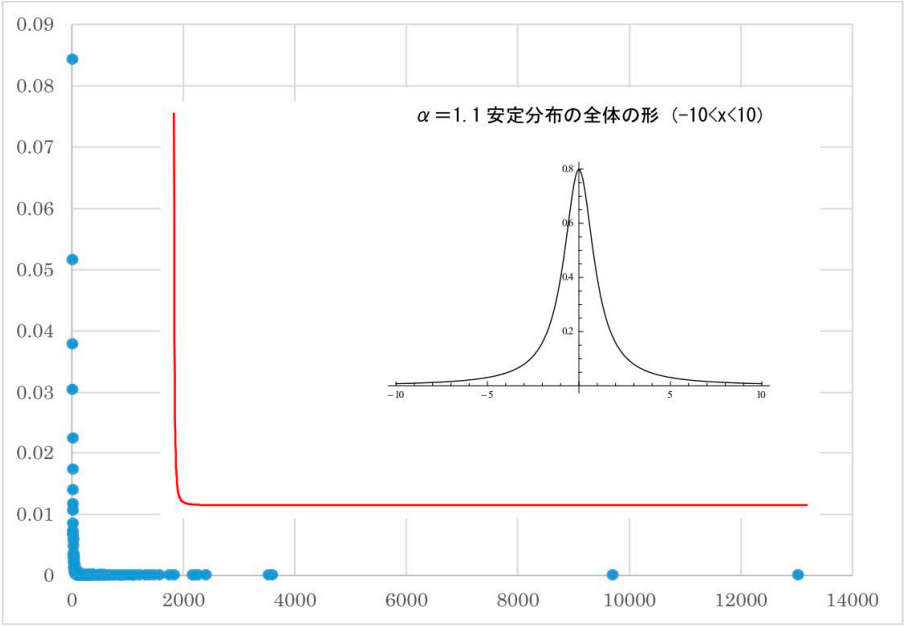

「図 1：世界状況の安定分布による区分」

構成主義とは、アイデアや行為遂行的発言（performative utterance）を共有し、さらにレジームや制度・組織を構築することによって、状況Ⅲを状況Ⅱに変化させようとする人間の一般的な行為を指している。[12] これについて Blyth はつぎのように総括している。[13] 現在の構成主義者は、社会的に重要な安定性や構造性を、リスクの世界状況Ⅰにおける個々の行為主体の特性として、もしくは不確実でリスクのある世界状況Ⅱに対する制度的な経路依存の結果（Ⅰから Ⅱに向かう右向きの矢印）として考えるのではない。そうではなくて構成主義者は、社会システムにおける何らかの安定性を、原理的に構造不特定な世界状況Ⅲから、不確実性のより少ない世界状況Ⅱ（リスク＋不確実性）へと移行せしめる行為的試み（ⅢからⅡに向かう上向の矢印）としての、行為主体に由来する社会的産物である、と考えるのである。構成主義者によれば、人間＝行為主体は、彼らが自明とみなし、しばしば物質に由来すると考える安定性を、実際には performative な言説やアイデアの提示によって、自ら社会的に構成している。

### 安定分布の右側裾野と社会的言語空間および世界状況の変化

本論文の研究グループは 2014 年の研究で、Twitter が蓄積する言語データに着目し、20 日間の合計約 3 億 3 千万個の tweet からなる言語データのベキ分布性を確認した。[14] この Twitter の言語データは、2011 年 3 月 11 日 14 時 46 分に発生した東日本大震災をあいだに挟む期間に採取したものである。この 2014 年の研究では、両対数グラフのランク図を利用して、ベキ分布性を確認したところ、本論文では、このベキ法則が α 安定分布の片側部分であるとの仮説のもとに、同一のデータに確率密度を適用して、地震の前後に生じた安定分布の特性指数 α を決定した。[15]

「図 2」に、3 月 11 日 16 時から 5 分間の、Twitter の名詞の出現頻度（度数）を青点でプロットした。このデータに基づき、安定分布の特性指数の値を、次節で述べる方法を用いて推定し、この安定分布の確率密度関数を特性関数から計算で求め、赤の実線で記載した。両者は自己相似的である。[16] 安定分布の特性指数 α は 1.111 になった。このときのオンラインの発話には「津波」「地震」などの単語が頻出し、投稿の数自体が増えて、出現回数の少ない単語（名詞）の数も増えたため、その密度関数は顕著に高い頭部（tall head）と、厚い裾野（fat tail）を持つことになった。この密度関数のかたちは、ガウス分布（α=2）とは大きく異なっている。しかしながら両者は、特性指数の異なる安定分布という点では共通している。後述のように、地震発災の前後の安定分布の α の値を調べると、1.811→1.111→1.982 と変化している。これは発災直後の社会的混乱、すなわち世界状況Ⅲ的事態を、SNS 言語空間が反映して、安定分布の特性指数 α の値が下がったことを示すと解釈できる。そこではさまざまな言説の交換や共有によって、社会的な事態を確定しようとする主体的行為が活発になるのである。




単語の出現頻度のプロット　3 月 11 日 16 時から 5 分間データ
横軸＝出現回数、縦軸＝相対出現頻度　推定 α＝1.1 1

α＝1.1 の安定分布の確率密度曲線
特性関数　$\psi = exp(-c\mathrm{Abs}[x]^{\alpha}), \{1 < x < 1000\}$　から計算

「図 2：単語の出現頻度と安定分布の確率密度」

### 3. 加法過程、安定過程と Lévi 分解

つぎに、名詞の相対出現回数を確率変数として、SNS の言語確率空間を解析し、安定分布の発生のメカニズムを考察する。きわめて多数の投稿者が、SNS 言語空間に毎時間サンプル（標本）となるテキストを追加する。

統計的な検証によれば、十分に大きなSNS言語空間の個々の単語（名詞：地震、自分、花粉…）は、すべて個別の強度（λ）をもつポアソン分布である。また一定時間（たとえば5分）ごとの単語の出現回数を数えて、隣り合った時間の相関を取ると、ほぼ確実に無相関になる。その理由は、十分に大きな SNS のデータベースが、個別に発話行為を行う多くの参加者から、同時にテキストを受け取るためである。このように、SNS の言語データは独立同分布になっている。

### 加法過程としてのSNS言語データ

確率過程としての加法過程および安定過程の特性を、標本分布の個々の数値の生起の在り方から直接導くことができる。本稿で分析する SNS は、確率法則が時間について定常的となるサービスを提供している。このデータは、きわめて多数の独立な発話（投稿）を、時間毎に重なり合う単一の時系列として蓄積したものである。したがって特定の単語に注目した場合、データ分析におけるこの単語の最初の出現と2回目の出現、2回目の出現と3回目の出現…は、事象としての関係性を失っている。これを分析時間あたりの特定の単語の出現の確率過程としてみれば、生起期間の確率としての無記憶性（memoryless）を持つことになる。このように SNS の言語データは、(1) 独立増分性、(2) 時間的一様性、(3) $X_0 = 0, a.s.$、(4) 確率連続を持ち、確率過程としてみれば加法過程になる。この場合の個々の単語の生起期間は指数分布を、単位時間あたりの出現回数はポアソン分布を成しており、この分布は既述のようにデータから確認することができる。

### 安定過程およびLévy分解

確率過程 $\{X_t : t \in [0,\infty)\}$ が加法過程で、かつ任意の $a > 0$ に対して $b > 0, c \in R$ が存在して、$\{X_{at} : t \in [0,\infty)\}$ と $\{bX_t + tc : t \in [0,\infty)\}$ が法則同値であるとき安定過程になる。これは時間のスケールを変えた場合の分布の同一性を意味しており、定常的なサービスから抽出した SNS の言語データはこの条件を満たしている。安定過程の定める無限分解可能分布は、確率分布 $\{X_t\}$ の $n$ 個の畳み込み $X^{n*} = X_1 * X_2 * \cdots * X_n$ として、指数 $0 < \alpha \le 2$ の安定分布をつくり出す。[17] このとき驚くべき事実として、安定過程を逆に互いに独立な次の3項の和に分解することができる。この分解は一意である。[18]

（i）$mt$、$m$ は定数

（ii）$\sigma B(t)$、$B(t)$ はブラウン運動で $\sigma$ は定数

（iii）複合ポアソン過程 $P_I(t)$ のシステム：$P_I(t)$ は区間 $I$ の独立ポアソン過程

（i）（ii）（iii）による Lévy 分解の式による表現は次のとおりである。[19]

$$Z(t) = mt + \sigma B(t) + \int_{|u|>0} \left( uP_{du}(t) - \frac{tu}{1+u^2} \right) dn(u)$$

$dn(u)$ は Lévy 測度、ただし Lévy 測度の重複度は1とする。その理由は、それぞれの単語が投稿に現れるたびに、この単語の分布関数が大きさ1の正の跳躍的増加をするからである。

### 特定時間でみたときの複合ポアソン分布の時系列としてのTwitterの言語空間

ポアソン過程とは、単位時間当たりの出現率 $\omega$ である単語が、時刻 $s$ 以降時刻 $t$ までに $k$ 回出現する確率

$P[\{k\}]$ が、この場合の平均 $\lambda = (t-s)\omega$ のポアソン分布

$$P[\{k\}] = e^{-\lambda}\frac{e^{\lambda k}}{k!}$$

になり、かつ時間区間毎の出現回数が独立であるような確率過程を言う。既述のように単語の出現率は、独立同分布ではあっても単語によって、また単語毎に時間や社会状況によって変化する。その確率過程は、多くの異なる強度を持つポアソン分布からなる複合ポアソン過程になる。特定の単語が、SNS のすべての投稿からなるデータに出現する回数の平均値が、ここで言う特定の単語のポアソン分布の強度 $\lambda$ になる。

特定時間量（たとえば特定時間から 5 分間）のデータを使って、tweet の名詞の相対出現回数を計算する。このサンプルに対し適切な幅をとって相対出現回数の級数を $x$ 軸に、各級数の出現頻度を $y$ 軸にプロットする。ネットワークからサーバに流入する言語空間は広大であり、単語の数は膨大で、個々の単語の生起（＝発話・投稿）はランダム、かつほとんどの単語の相対出現頻度はごく僅かである。統計的な検証が示すように、単語の出現頻度に対する小確率の法則から、名詞の相対出現回数は、それぞれが異なるきわめて多くのポアソン分布になる。このポアソン分布は、強度（λ）が小さければ大きな歪度と尖度を持ち、$x$ 軸の左側（$y$ 軸に近い側）に偏った分布を持つ。強度が大きくなるにつれて、ポアソン分布は $x$ 軸の右側に移動して分布のかたちがより対称になる。単語の出現頻度は、投稿者の発話行為の話題に依存するため、特定時間で切った言語空間の単語の相対出現回数は、毎時間に異なる複合ポアソン分布の組み合わせを作り出している。

**指数 $\alpha$ の決定**

本稿で用いた $\alpha$ の決定の手法について説明する。[20] 分析の対象とする Twitter の言語データについて、形態素解析ソフト KH Coder を用いて形態素（単語）に分割したうえで名詞を抽出し、各名詞について出現回数を求めた。[21]

(1) 名詞の出現回数を $X$ とする。$X$ について階級幅 $a = 8$ とした度数分布表を作成した。階級幅をこのようにしたのは、予備分析の結果、算出された $\alpha$ の値が $a = 8$ のときに最も安定していることが確認されたためである。ここで言う安定とは、隣接する階級の下記のペア毎に算出した $\alpha$ について、それらのバラツキが小さいことを意味する。

(2) 隣接する階級の度数の比の値が $\alpha$ の下記の関係式になるという性質を利用して、隣接する階級のペア毎に $\alpha$ を算出した。ただし、使用した $X$ の階級は以下の二つの条件を満たすものである。

【条件 1】階級番号の最も小さい階級（$0 < X < 8$）は分析対象から除外した。安定分布において、$X$ の値の小さいところは分布関数（確率密度関数）のかたちがよく分かっていないためである。

【条件 2】階級番号が大きくなるにつれて、階級の度数は減る傾向にある。十分に大きい $X$ についてはサンプルの持つ誤差が大きくなるため利用すべきでない。そこで、隣接する階級の度数の比について、分子となる階級（階級番号の小さいほうの階級）の度数が 30 未満である場合は、分析対象から除外した。

$\alpha$ の関係式：

$$\frac{P((k-1)a < X < ka)}{P(ka < X < (k+1)a)} \approx 1 + \frac{2}{k}\alpha$$

a：階級幅（a>0）、ここでは $a = 8$ を代入。

k：階級番号（$k = 1,2,3,\cdots$）

※ 左辺は隣接する階級の度数の比の値となる。

(3) 隣接する階級のペア毎に $\alpha$ を算出したので、ペア数とおなじ数だけ $\alpha$ が存在する。これらを平均して、最終的な $\alpha$ の値とした。

### 言語空間の$\alpha$ 安定分布

このように、① SNS の時間的・空間的に均質な言語空間の確率過程はLévy 分解が可能で、② ガウス分布の過程（$\sigma B(t)$）が欠落し、③ 独立ポアソン分布のLévy 測度による一次結合（$\int uP_{du}(t)dn(u)$）に分解して記述することができる。ガウス分布の安定過程はガウス分布を、また複合ポアソン分布の安定過程は、Lévy 測度によって安定分布を作り出す。したがって標本空間の例として、SNS 言語空間が独立同分布で、SNS 言語空間の名詞の出現が複合ポアソン分布であれば、安定過程としての SNS 言語空間の名詞の相対出現数は、原理的に（つまりこの空間の構造として）安定分布になる。[22] これは Wiener が、水の分子の衝突から生ずるガウス過程によってブラウン運動を説明したのと同じ原理によって、SNS 言語空間の複合ポアソン過程から$\alpha < 2$ の安定分布を示したことになる。この説明によればSNS の安定分布の特性指標に変化が生じるとは、elemental な複合ポアソン過程の強度（＝平均）λの分布が変化することを意味している。社会的事件が、人々の認識に共通に与える影響によって、特定単語の出現頻度が変化すれば、この変化が複合ポアソン過程の要素となる独立ポアソン分布の強度λの分布に影響し、確率過程の畳み込みの結果として生ずる全体の特性指数αに変化が生ずるのである。

## 4. 分析結果：まとめにかえて

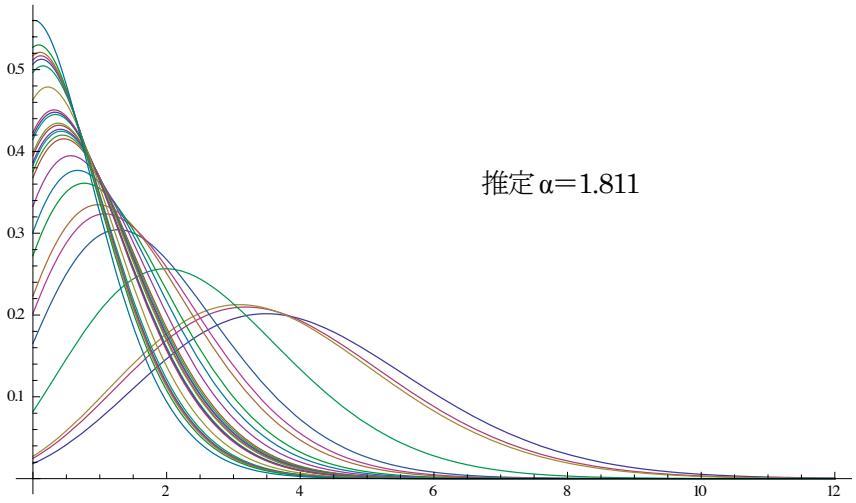
3・11 前後の言語空間の単語のポアソン分布の強度の分布と、主な名詞の相対的出現回数全体のかたちを、「図 3」地震前（3 月 11 日 10 時）、「図 4」地震直後（同 16 時）、「図 5」地震後約 1 週間（3 月 18 日 10 時）の 3 時点の 5 分間のデータで示した。

複合ポアソン分布の確率変数が、安定分布を作り出すとすれば、両者の関係に次の変化を予想することができる。（ア）地震発災の後に、強度λの大きなポアソン分布が、言語空間のポアソン分布を並べたグラフで x 軸の右側に生起し、これがα安定分布の tall head の形成に貢献して、特性関数αの値は小さくなる。（イ）これにより安定分布は、よりベキ関数で近似できるようになり、両対数グラフの ⬚ で囲んだ部分が直線に近づく。（ウ）これに対して、強度λが均等に分布していれば、αの値はより大きく、両対数グラフのかたちは直線から離れて、原点に凹のかたちを取る。（エ）この結果、0 から十分に隔たった部分で、両対数グラフの傾きが大きくなる。この予測結果を、3・11 前後の言語空間に当てはめると、(1) 地震発生にともなって、これまでにない大きな強度をもつポアソン分布が生起する、(2) 地震発災の前後の安定分布のαの値を調べると、大→小→大（1.811→1.111→1.982）となる、(3) データの両対数グラフのプロットに線形近似をすると、その傾きは、急→緩→急となる。実データに基づく検証は、この予想を裏付けている。

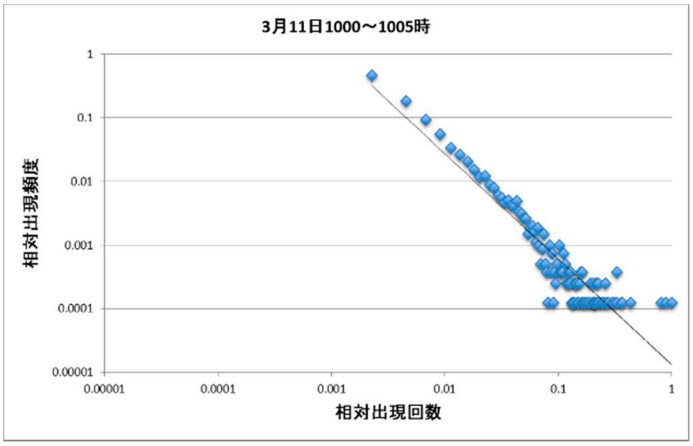
安定過程による安定分布の創発に関する理論的な検討と、実際の言語データの解析は、Twitter の名詞の相対出現回数に現れるα安定分布と、個々に異なる強度を持つ名詞のポアソン分布との繋がりを示唆している。この検証が正しいとすれば、Lévy 分解の主要成分となる複合ポアソン過程の elemental な強度の分布を検出して、情報空間の偶然現象を確率変数系に還元するとの確率的独立還元論（stochastic reductionism using independent stochastic variables' system）の研究を一歩進めたことになるであろう。これは安定分布の研究から、さらに一歩すすんで、その背後にある機序（underlying mechanism）を明らかにしようとするものである。[23]

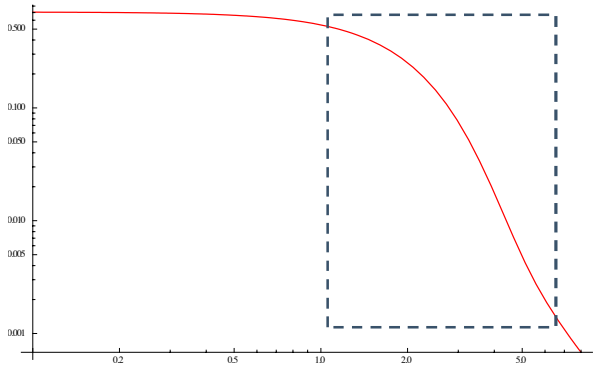
地震発災の前後で SNS 言語空間のα安定分布の特性指数が変化している。これは 3・11 という社会的事象によって単語の出現回数と複合ポアソン分布の強度の分布に変化が生じ、それが日を追って旧に復したためだ、と考えることができる。指標の数値が経時的に旧に復することから、SNS 言語空間の名詞の生起回数に存在する、安定分布の固有のメカニズムと法則性に基づいて、指標に一定期間の変化が生じたものと考えることができる。これは社会事象が人間の共同主観性を構成する情報空間を作りだし、また人間の社会認識が、この情報空間に拘束されるという構成主義的な社会科学の理論を裏付けている。社会事象が作り出す構成主義的な情報空間の分析として、ホワイトノイズ解析と安定分布の連続的な指標の変化を用いた研究が有効である。

「図 3：3 月 11 日 1000～1005 時のλ推定値に基づいたポアソン分布　横軸　0<x<12」



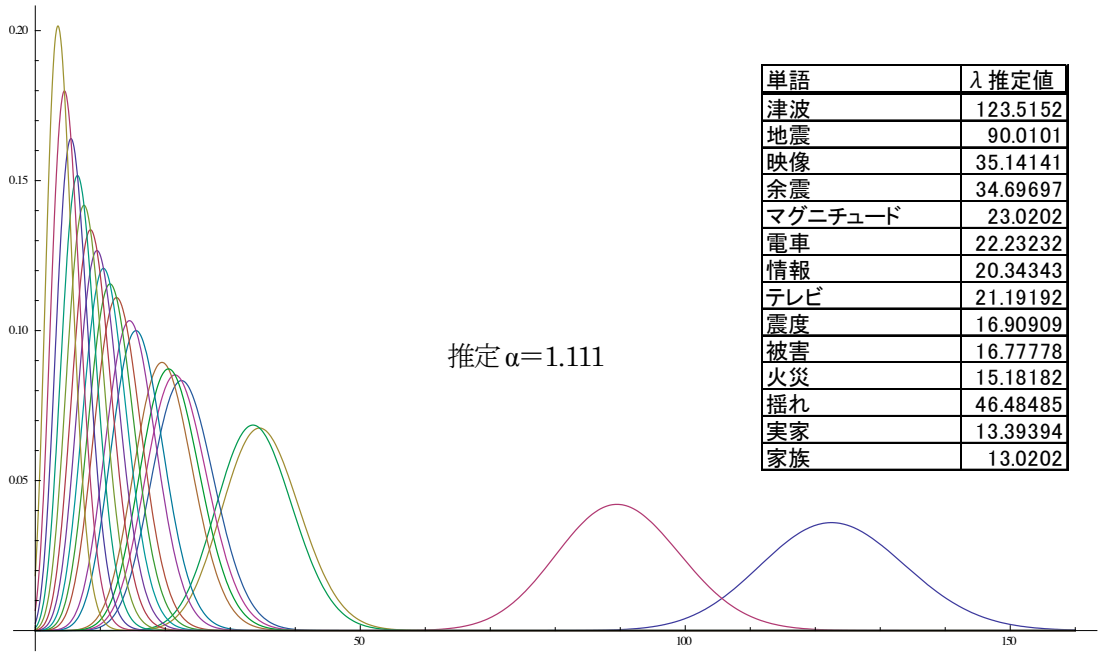

| 単語 | λ推定値 |
|---|---|
| 自分 | 4.070707 |
| 花粉 | 3.676768 |
| 定期 | 3.555556 |
| 感じ | 2.494949 |
| 天気 | 1.585859 |
| 学校 | 1.525253 |
| 楽しみ | 1.878788 |
| 会社 | 1.313131 |
| 休み | 1.616162 |
| 先生 | 1.363636 |
| 電車 | 1.333333 |
| 気持ち | 1.272727 |
| お腹 | 1.212121 |
| 気分 | 1.161616 |



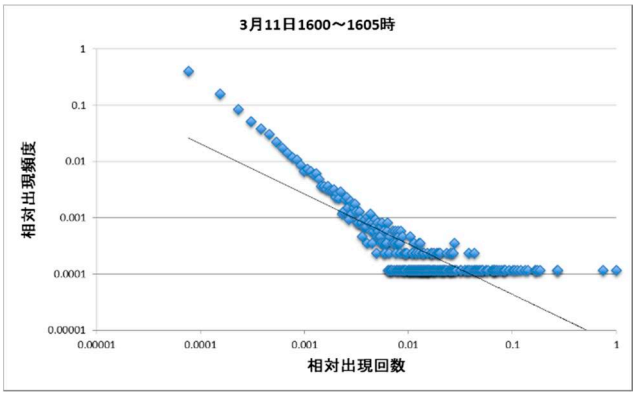



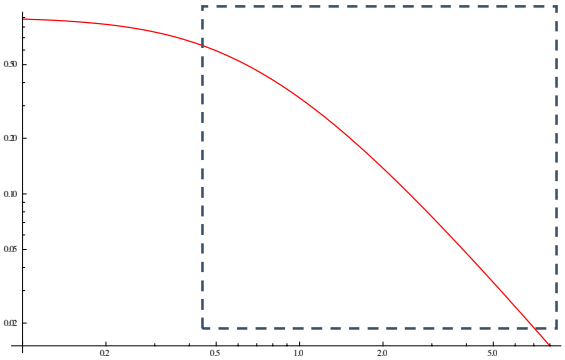
α=1.8 の安定分布（両対数）

「図 4：3 月 11 日 1600～1605 時のλ推定値に基づいたポアソン分布　横軸　0<x<180」



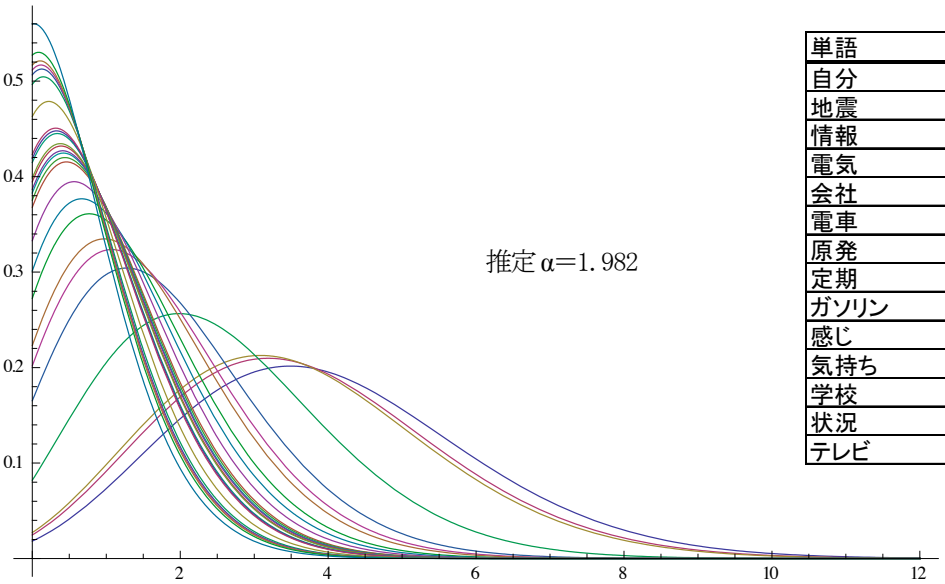

| 単語 | λ推定値 |
|---|---|
| 津波 | 123.5152 |
| 地震 | 90.0101 |
| 映像 | 35.14141 |
| 余震 | 34.69697 |
| マグニチュード | 23.0202 |
| 電車 | 22.23232 |
| 情報 | 20.34343 |
| テレビ | 21.19192 |
| 震度 | 16.90909 |
| 被害 | 16.77778 |
| 火災 | 15.18182 |
| 揺れ | 46.48485 |
| 実家 | 13.39394 |
| 家族 | 13.0202 |



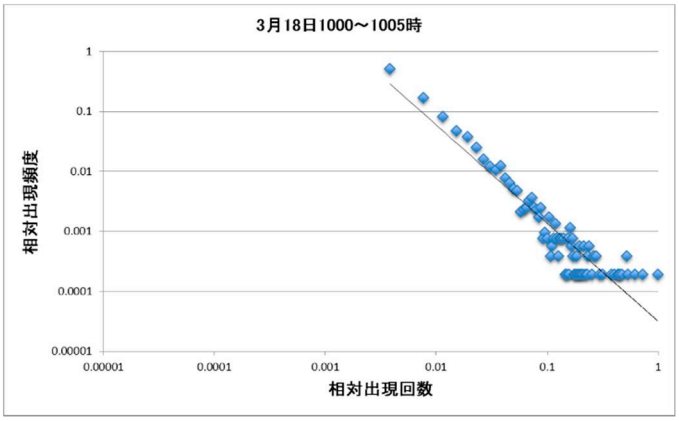



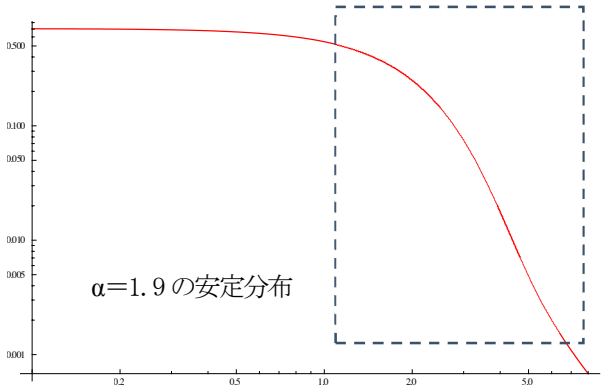
α=1.2の安定分布（両対数）

「図5：3月11日1000～1005時のλ推定値に基づいたポアソン分布　横軸　0<x<12」




| 単語 | λ推定値 |
|---|---|
| 自分 | 5.767677 |
| 地震 | 4.717172 |
| 情報 | 3.282828 |
| 電気 | 2.989899 |
| 会社 | 3.070707 |
| 電車 | 3.535354 |
| 原発 | 2.474747 |
| 定期 | 1.69697 |
| ガソリン | 2.89899 |
| 感じ | 3.353535 |
| 気持ち | 2.484848 |
| 学校 | 2.121212 |
| 状況 | 2.242424 |
| テレビ | 1.666667 |



[1] 本研究に当たって飛田武幸名誉教授（名古屋大学）、公文俊平教授（多摩大学情報社会学研究所）、服部正太取締役社長（株式会社構造計画研究所）、SiSi 教授（愛知県立大学）からご指導を戴いた。本稿のデータは鳥海不二夫准教授（東京大学大学院工学系研究科システム創成学専攻）の共同研究として利用した。形態素解析プログラムの開発は tn-works（永井拓史代表）が担当した。

[2] アルバート=ラズロ・バラバシ『新ネットワーク思考：世界のしくみを読み解く』青木薫訳、NHK 出版、2002 年。127 頁。preferential attachment の発見は 1925 年の G. U. Yule に遡る。

[3] ベキ分布性と社会秩序に関する本質的な検討については以下を参照。公文俊平「情報社会学の諸側面」情報社会学会編『情報社会学会誌』Vol.1 No.1 2006。http://infosocio.org/journal/vol1no1/vol1no1-2.pdf

[4] W.フェラー『確率論とその応用 II 上』国沢清典監訳、紀伊國屋書店、1969 年、第VI章。引用は 229 頁。

[5] 飛田教授は名古屋大学名誉教授、元確率過程論国際学会会長、専門は確率論、関数解析学。

[6] Benoit Mandelbrot, "The Variation of Certain Speculative Prices," *The Journal of Business*, Vol. 36, No. 4, 1963, p. 405.

[7] Nassim Nicholas Taleb, *Silent Risk: Lectures on Fat Tails, (Anti)Fragility, Precaution, and Asymmetric Exposures, Preliminary incomplete draft for error detection*, June, 2014, pp.84-5. この論文はタレブ教授の HP からダウンロードすることができる。http://www.fooledbyrandomness.com/

[8] 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 年。

[9] 情報社会学における社会システムの概念はそもそも構成主義的である。公文俊平『情報文明論』NTT 出版、1994 年。

[10] マーク・ブライス「構成主義理論と政治経済学について」小野耕二編著『構成主義的政治理論と比較政治』ミネルヴァ書房、2009 年。

[11] ブライス、前掲論文、44 頁。

[12] その文を口に出して言うことが、当の行為を実際に行うことに他ならないような発話＝行為（たとえば "I name this ship the *Queen Elizabeth*."）を行為遂行的発話という。J.L.オースティン『言語と行為』坂本百大訳、大修館書店、1978 年、10～11 頁。

[13] 以下の引用は、ブライス、前掲論文の後半の議論を要約したものである。構成主義派政治学者のブライスとファット・テール派経済学者のタレブには、『フォーリン・アフェアーズ』に次の共執筆論文がある。Nassim Nicholas Taleb and Mark Blyth, "The Black Swan of Cairo: How Suppressing Volatility Makes the World Less Predictable and More Dangerous," *Foreign Affairs*, Vol.90, No.3, May/June 2011.

[14] 山内康英、小松正、舟橋正浩「言語空間のベキ分布性に関する確率過程論研究」公文俊平、大橋正和編著『情報社会のソーシャルデザイン：情報社会概論 II』NTT 出版、2014 年。

[15] 密度関数の一部としてのベキ曲線は、一意にランク図のベキ曲線に変換できる。これを Adamic の法則と呼ぶ。SNS 言語空間の名詞の出現頻度のランク図に現れるベキ法則性は、このサンプル空間自体の性質的なものだと考えることができる。Lada A. Adamic, "Zipf, Power-laws, and Pareto - a ranking tutorial." http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html

[16] 両者の概収束の検討が今後の課題になる。

[17] 佐藤健一『加法過程』紀伊國屋書店、1990 年、4 頁および 88 頁。加法過程については、伊藤清『確率過程』岩波書店、2007 年、第 2 章。無限分解可能な分布については、B. V. グネジェンコ『確率論教程 II』鳥居一雄訳、森北出版株式会社、1972 年、第 9 章。安定過程については、飛田武幸『確率論の基礎と発展』共立出版。2011 年、第 10 章。加法過程と Lévi 過程の用語の使い分けについては、佐藤、前掲書、31 頁。

[18] 飛田武幸「時空と偶然、第 1 回～第 4 回」『数学セミナー』日本評論社、2012 年 5～8 月。この式展開については第 2 回（6 月号）2 頁。飛田、前掲書、224～226 頁。

[19] 飛田、シィ・シィ「事物の分布についての新しい見方：特にベキ分布について」公文俊平編著『情報社会学概論』NTT 出版、2011 年、77 頁。本稿は、この論文が提起した解析の手法を SNS 言語空間に適用したもの

である。

[20] この$\alpha$の決定手法については、飛田、シィ・シィ前掲論文、71～80 頁を参照。グラフに最小二乗法を適用して$\alpha$を求めるのは不正確である。

[21] 本研究で扱った「名詞」は、KH Coder における名詞の定義、すなわち「普通名詞であり、かつ、平仮名のみの語と漢字 1 文字の語を除外したもの」という条件を満たすものである。

[22] 確率空間の毎時間の確率分布が、複合ポアソン分布である場合の時系列的な累積過程は、現在ホットな研究テーマになっている。Si Si 前掲書参照。ホワイトノイズ解析については、Takeyuki Hida, Si Si, *An Innovation Approach to Random Fields: Application of White Noise Theory*, World Scientific Pub.Co.Inc., 2004. 飛田武幸『ホワイトノイズ：シュプリンガー量子数理シリーズ』丸善出版、2014 年。

[23] 確率変数系を測度として、偶然現象をこれに還元するアプローチが、飛田教授の研究グループの提唱するホワイトノイズ解析の重要なメルクマールを成している。

(2016 年 7 月 8 日受理)