

# 大規模データに対するべき分布性の確認方法

井 上 寛 康

## Verifying Power-Law Distribution in Empirical Data

INOUE Hiroyasu

### 目 次

1. はじめに
2. 分析方法
3. 大規模データにおける分析例
4. 結論

### Abstract

In social science, it is mostly assumed that there is a typical value for data. This is because normal distribution is supposed. However, it is hardly possible that all data in the natural world as well as human society obey normal distribution. Hence, if people take data for some phenomenon and suppose the data obeys normal distribution, that can lead to a wrong conclusion. In fact, people can have accessed many large scale empirical data in recent years and found those distributions do not obey normal distribution. Some of them obey power-law distribution whose nature has been recently well studied.

This paper studied whether the distribution of joint patent applications for an organization obeys power-law distribution or not. The distribution is far from normal distribution and looks like power-law distribution. Hence, it was judged whether the distribution obeys power-law distribution using Clauset's method. The results showed the distribution in 7 and over obeys power-law distribution, and the exponent is 3.03.

**キーワード** : べき分布, 離散, コルモゴロフ-スミルノフ検定, 最尤法

**Keywords** : Power-Law Distributio, Discrete, Kolmogorov-Smirnov Test, Maximum Likelihood

## 1 はじめに

社会科学で扱われるほとんどのデータ，たとえば経済学における個人消費の金額や，経営学における企業の売上などに対し，その背後にある現象を理解しようとモデルを構築するとき，これらのデータには典型的な値が存在すると仮定される。確率分布を考えると，その典型的な値は大きな確率を持ち，そこから正負の両方に離れた値ほど小さな確率を持つような分布になる。これは一般的に正規分布と呼ばれる。このような正規分布のよくある例は人の身長などである。日本人の平均身長は男性で171cm程度であり，大きくかけ離れることは確率的に小さい。身長は身近なデータで測定が容易であり，実際に統計を得れば正規分布でフィットできることが（正確には，フィットすることが妥当であると）確認できる。しかし，自然や社会経済におけるあらゆる分布が正規分布であるはずはない。したがって，ある現象の一部分のデータを得て，そのデータに対して意識せず正規分布を仮定し，誤った結論に至るケースが数多くある。

現代においてさまざまな大規模データが入手され，分布が明らかになるにつれ，正規分布でない現象が数多く確認されるようになった。そのような現象は単純に正規分布が当てはまらないというだけでなく，学術的に見て極めて重要なものが多い。例は数多く存在するがその中には，人々の富，宇宙に存在する物体の大きさ，インターネット内のルーターにつながれている線の数などが含まれる。これらの分布を正しく理解し，その背後にあるメカニズムを知ることは極めて重要である。

なかでも，特にその現象の発見と分析において盛んに研究されているのが，べき分布である。そのような分布には先にあげた3つの例も含まれる上に，都市の人口，地震の強度など，数多く見られる [1, 2, 3]。たとえば都市の人口の分布において，仮に都市の人口の平均が10,000人だったとしても，その値はまったく意味をなさない。なぜなら都市の人口の分布は正規分布ではフィットできないからである。

べき分布とは次のような式で表される分布のことである。

$$p(x) \propto x^{-\alpha}, \quad (1)$$

ここで $\alpha$ はべき指数であり，一般的に $2 < \alpha < 3$ の値をとることが多い [4]。

べき分布は上記のような式で表されるが，分布の全体においてべき分布であるようなケースは少ない。そこで本論文では， $x_{\min}$ をべき分布に従う最小の値であるとし，それ以上の値においてべき分布になる場合もふくめて議論する。

これまでに述べたように，べき分布は非常に分析対象として興味深いのであるが，本当にべき分布に従っているかどうかについて，これまでの研究 [1, 2, 3] では単に両対

数グラフでの直線性を見た目で判断し、その傾きを最小二乗法などの方法で求めていた。しかしながら、見た目で判断することは明らかに科学的とはいえず、また最小二乗法はどのようなデータに対しても一次関数でのフィットを行うため、それがべき分布であるかどうかに関わらずいつでも傾きが求められてしまう。これを解決するため、与えられた確率分布がどの程度にべき分布に近いのかということを定量化し、妥当な閾値によって判別できる方法が、Clauset らによって最近開発された [5]。本論文では、その方法により、社会科学的に重要でまだ判断されていない例として、企業の共同出願相手数の確率分布を取り上げる。この企業の共同出願相手数の確率分布は、先行研究においてべき分布であることを前提に進められていたり [6, 7]、逆にそのべき分布性自体を否定する研究 [8] もあり、この議論を終結させる意味でも重要である。

## 2 分析方法

本論文では Clauset らがまとめた方法 [5] に従って分析を行う。この方法では以下のような手順で分析を進める。

- ・  $x_{\min}$  と  $\alpha$  を見積もる。
- ・ goodness-of-fit テストを行う。

### 2.1 定義

確率分布は定義域が連続であるか離散であるかで扱い方が大きく異なってくる。一般的には連続の方が議論しやすいが、本論文では企業の共同出願相手数が定義域であるため、離散の方を議論する。

べき分布の式 (1) を

$$p(x) = \Pr(X = x) = Cx^{-\alpha}, \quad (2)$$

と直す。 $x$  が 0 に近づくと発散するので、 $x_{\min} > 0$  である。すると正規化定数  $C$  は

$$C = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})}, \quad (3)$$

とすることができる。ただし

$$\zeta(\alpha, x_{\min}) = \sum_{n=0}^{\infty} (n + x_{\min})^{-\alpha}, \quad (4)$$

は Hurwitz のゼータ関数である。

ここで、確率分布と相補的な関係にある累積確率分布 (Cumulative Distribution

Function (CDF)) を導入する。この関数を用いれば範囲を伴った確率の議論が可能となるため便利である。本論文ではこれを  $P(x)$  とする。離散の場合は

$$P(x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{\min})} \quad (5)$$

となる。

## 2.2 データに対する $x_{\min}$ と $\alpha$ の計算

あるデータに対して  $x_{\min}$  と  $\alpha$  を計算するときに、もっともひどい方法はヒストグラムを用いる方法である。その上で最小二乗法により  $\alpha$  を求めるのがよく取られる方法である。この原因は、19世紀末に行われたべき分布研究の古典であるパレートの研究 [9] がこの方法をとっているためでもある。

この方法はごく一般的なデータでも明確な間違いに導いてしまう。ここではそれとは異なる方法を探る。

$\alpha$  の計算には  $x_{\min}$  が必要である。順番が前後するが  $x_{\min}$  がすでにあるとして、 $\alpha$  を計算する方法を述べる。この  $\alpha$  を求めるには複雑な計算が必要である。その計算過程はここで省くが、データおよび  $x_{\min}$  から推測されるべき指数  $\hat{\alpha}$  は、以下のように表される。

$$\hat{\alpha} \simeq 1 + n \left[ \sum_{i=1}^n \ln \frac{x_i}{x_{\min} - \frac{1}{2}} \right]^{-1}。 \quad (6)$$

$\alpha$  の計算は上記の式 (6) によりかなり正確に求まるが、当然のことながら  $x_{\min}$  も正確に求める必要がある。一般的にこれを正確に求める方法は2つある。1つは Handcock と Jones の方法 [10] である。これはベイズ情報基準を用いる方法であるが、 $x_{\min}$  を小さく見積もりすぎる傾向にあることが指摘されている [5]。

もう1つの方法は Clauset らの方法 [11] である。これは非常に単純な方法で、データが最もよくべき分布に一致するように  $\hat{x}_{\min}$  を決めるという方法である。これは前述の  $\hat{\alpha}$  の議論とあわせると混乱するが、 $\hat{x}_{\min}$  を1として  $\hat{\alpha}$  を求める。次に  $\hat{x}_{\min}$  を2として  $\hat{\alpha}$  を求める。という方法をとるに過ぎない。このとき大事なことは、(1) 真の  $x_{\min}$  に対して  $\hat{x}_{\min}$  を大きくしすぎたなら、確率分布のテールの部分は変動が激しい影響で、一致性が著しく下がる。(2) 逆に真の  $x_{\min}$  に対して  $\hat{x}_{\min}$  を小さくしすぎたなら、仮定しているべき分布とデータとの差が大きくなり、一致性が下がる。これらから、ちょうどいい  $\hat{x}_{\min}$  が見つかるといえる。

2つの確率分布が与えられたとき、それらの一致性の算出にもいろいろな方法がある。

しかし正規分布でないときに使われる最も一般的な方法は、Kolmogorov-Smirnov 検定である [12]。その一致性は距離として表し、

$$D = \max_{x \geq x_{\min}} |S(x) - P(x)| \quad (7)$$

となる。ここで  $S(x)$  は  $\hat{x}_{\min}$  以上における実際のデータの累積確率分布、 $P(x)$  は  $\hat{x}_{\min}$  以上の実際のデータに対して最も一致する ( $\hat{\alpha}$ により求まる) べき分布である。求める  $\hat{x}_{\min}$  は  $D$  を最小化する  $\hat{x}_{\min}$  である。

### 2.3 goodness-of-fit テスト

ここまで述べてきた方法により、与えられた確率分布に最も一致するべき分布、すなわち  $x_{\min}$  と  $\alpha$  を見積もることができる。しかしながら、本論文の問いに答えるにはまだ不十分である。なぜなら、どのような確率分布に対しても  $x_{\min}$  と  $\alpha$  は見積もることができるためである。したがって、べき分布が十分に実際のデータに一致するかを判定する方法が必要となる。それが、goodness-of-fit テストである。

加えて、べき分布の正確な  $x_{\min}$  と  $\alpha$  を従来研究で求めてこなかったこと以上に問題なのが、この節で述べる「べき分布なのかどうか」である。そもそも、対数正規分布も指数分布も両対数のグラフにプロットすれば直線でフィットできるように見える。したがって、単に直線に見えるからといってべき分布であるということはあまりにもいい加減である。

実際のデータの確率分布は、あくまで起きたことに対する確率であるから、多少のずれは生じうるし、そもそも定義域の値が大きい部分の事柄はめったに起きないのでずれやすい。そういったずれがべき分布でないほどずれているのか、そうでないのかを判断する必要がある。

goodness-of-fit テストは具体的には、予測されたべき分布から逆にデータを合成し、そのべき分布からのずれ具合と、実際のデータとべき分布のずれ具合を比較するものである。単純に実際のデータの方が大きくずれていれば、それはべき分布ではない。ここでのずれの計算には再び Kolmogorov-Smirnov 検定を用いる。

合成データのセットは複数作成し、それぞれべき分布とのずれを求める。このとき  $p$  値を次のように定める。

$$p = \frac{\text{ずれが実際のデータより大きかった合成データのセット数}}{\text{合成データのセット数}} \quad (8)$$

この  $p$  値が大きいほど実際のデータがべき分布であるといえる。

上述のように goodness-of-fit テストは合成データが必要である。Clauset らの方法 [5]

では、実際のデータが  $n$  個とすると、 $x_{\min}$  以上の実際のデータを  $n_{\text{tail}}$  個とし、この  $n_{\text{tail}}$  個については推測されたべき分布からデータを合成し、 $n - n_{\text{tail}}$  個については、均一の確率で実際のデータからとるという方法にしている。しかしながら、この処理は複雑であるのと、 $x_{\min}$  以上でのみ比較するのと実質同じである。したがって、本論文では後者の方法、すなわち、次の手続きを必要なデータの個数回繰り返す方法をとる。

1. 0 以上 1 以下の一様な乱数を得る。
2. 元になる分布（この場合、推測されたべき分布の累積確率分布）において、上記 1 の手続きで得た値を初めて超える値をデータとする。

これで得られたデータに対して、確率分布を求めて比較することになる。すなわち、想定される自然なずれを持ったデータになる。当然であるがこのずれはデータの個数が多いほど小さくなっていく。

次にどの程度の合成データセット数が必要か決めねばならない。一般的に  $\epsilon$  の精度がほしい場合、 $\frac{1}{4}\epsilon^{-2}$  の合成データセット数が必要といわれている [5]。

さらに、 $p$  値はいくら以上であれば妥当であるかということだが、多くの文献では  $p > 0.05$  ならば妥当としている。より慎重な判断を行う場合は  $p > 0.1$  が妥当である [13]。

この goodness-of-fit テストにおける信頼性についてであるが、Clauset ら [5] によれば、実際のデータの数が少ない場合、具体的にはだいたい  $n$  が 100 より小さいときは結果はあまり信頼できない。

### 3 大規模データにおける分析

本節では前節までに述べた分析方法を用いて具体的なデータがべき分布に従うのかどうかを検討する。本論文で用いたデータは、特許を共同出願した組織のネットワークに基づくものであり、ある組織（ノード）がデータの期間中に共同出願を他の組織と行った場合、それらの間にリンクがあるとするネットワークである。

現代においては組織の研究開発はその組織に閉じた自前のものではなく、他の組織と連動した、より開かれたものとなっている。これは国際競争力に必要な不可欠である。このような文脈で、組織の共同出願ネットワークはこれまで研究が行われてきた [7]。

この先行研究では、この共同出願ネットワークがどのようなメカニズムで成長するのかを、物理学的なモデルにより研究している。そこで前提とされているのが、各ノードが持つリンクの数（次数）の確率分布がべき分布であることである。このとき、べき分布であるかどうかは、確率分布を両対数でプロットしたときに直線でフィットできるかどうかと

いう、主観的な方法に頼っている。したがって、前節までに述べたように、本当にべき分布であるのかどうかについては不確かである。そこで本論文ではこのデータが実際にべき分布であるかどうかを検討する。

### 3.1 特許データ

ネットワーク構築の元となるデータは、日本の公開特許公報および特許公報において1993年1月から2002年12月の10年間に記載された4,998,464件の特許である。本論文ではTamadaDatabase [14] を利用する。

ネットワークは次の手順で作成した。

#### 1. ノードの抽出

ノードとして用いられるのは組織であるが、この組織の抽出は、各特許に含まれる（複数の）出願人の名称で行った。出願人には個人もありえるが、組織でないためフィルタを用いて除いた<sup>1</sup>。

一般的には特許に含まれる出願人の名称により、組織を区別するだけである。しかしながら、出願人は複数の研究開発拠点をもっており、これら別の組織として区別する必要性が指摘されている [7]。そこで、出願人レベルではなく研究開発拠点ごとにノードを作成する。

#### 2. リンクの抽出

ある1つの特許において、複数のノード（組織、研究開発拠点）間で共同で出願されているとき、これら間に完全グラフになるようにリンクを作成する。すなわち2ノードであればリンクは1本。3ノードであればリンクは3本である。すべての特許についてこれを行うが、重複するリンクは無視される。

### 3.2 形成されたネットワーク

形成されたネットワークはノード数が54,197、リンク数が154,205である。その次数分布は図1のようになる。この図において、横軸は次数であるが、縦軸は順位であり、両対数である。

---

<sup>1</sup> 組織に該当するかの判断は、以下の法人格を表す文字列が含まれるかを条件とした。（これらは特許に出現したものであり、法人格を表す文字列のすべてではない。）それらの文字列は株式会社、有限会社、特定非営利活動法人、財団法人、学校法人、独立行政法人、国立大学法人、医療法人、社団法人、医療法人社団である。

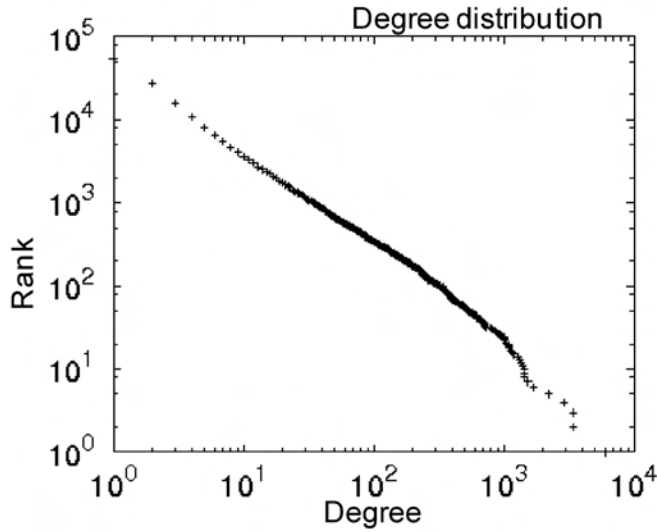


図1 次数分布

順位は累積確率と同様の意味合いがあるため、順位について説明する前に、先に累積確率について説明する。累積確率分布とは、 $k$ 以上の次数をもつノードが出現する確率である。多くの文献<sup>2</sup>では、累積確率分布ではなく確率密度関数 ( $p(k)$ ) を用いて次数分布を議論しているが、確率密度関数と累積確率分布の関係は、次数の連続性を仮定すると

$$P(k) = \int_k^{\infty} dk p(k)$$

である。したがって、もし  $p(k)$  がべき分布、すなわち

$$p(k) \propto k^{-\gamma}$$

ならば、

$$P(k) \propto k^{-(\gamma-1)}$$

である。順位は、同じ値のデータに同じ順位を与えるが、そのときの順位は一般的に（1つ前の順位 + 1）である。しかしここでは（1つ前の順位 + その値のデータの数）とする。この定義ならば各順位を全体のデータの数で割ったときに、累積確率と一致する。さて、図1を一見すれば、累積確率が直線でフィットできることを示している。この分布が実際にべき分布であるかどうかを検討する。

<sup>2</sup> たとえば、論文 [15, 16, 17, 18, 5, 11, 8, 19] など。スケールフリーネットワークに関する議論では必ず累積確率分布が登場する。



### 3.3 べき分布であるかの分析

分析を進めるにあたってはいくつかのパラメータを決定しなければならない。まず  $p$  値の境界であるがこれを0.1以上とする。これを検証するためには精度が0.01必要であるので、 $\epsilon = 0.01$ となり、これに必要なデータセット数は、2,500セットとなる。

表1は分析結果の詳細である。各  $x_{\min}$  に対して、 $\alpha$ 、(1) 実際のデータとべき分布の最大距離、(2) 合成データと推測されたべき分布の間の最大距離の合成データのセット間の平均、(3) 合成データと推測されたべき分布の間に距離が最大になった  $x$  の合成データのセット間の平均が示されている。ここで重要なのは、 $x_{\min}$  が異なれば  $\alpha$  の値も変わ

表1  $x_{\min}$  ごとの分析結果

$x_{\min}$	$\alpha$	(1) Distance from empirical data	(2) Average max distance from synthetic data	(3) Average point in synthetic data
1	2.87 (9)	0.074 (5)	0.0026 (0)	6.6 (4)
2	2.05 (1)	0.029 (4)	0.0039 (8)	7.7 (5)
3	3.09 (6)	0.015 (3)	0.0054 (2)	9.8 (0)
4	3.09 (0)	0.019 (7)	0.0068 (0)	12.6 (5)
5	3.05 (8)	0.014 (2)	0.0081 (5)	15.7 (1)
6	3.05 (6)	0.016 (4)	0.0093 (1)	18.6 (1)
7	3.03 (3)	0.008 (9)	0.0101 (8)	22.1 (1)
8	3.03 (6)	0.011 (4)	0.0111 (0)	24.6 (5)
9	3.03 (3)	0.011 (3)	0.0121 (7)	28.1 (7)
10	3.02 (0)	0.009 (2)	0.0130 (1)	30.5 (2)

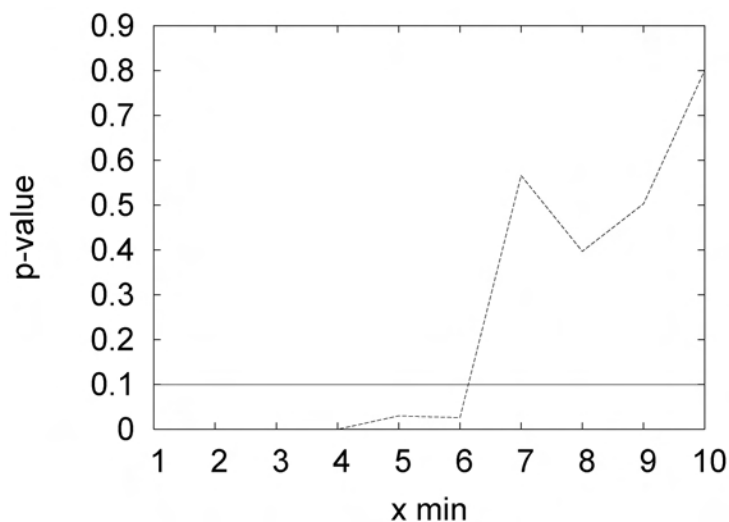


図2 P 値の変化

ることである。したがって推測されるべき分布が $x_{\min}$ ごとに異なるため、Kolmogorov-Smirnov 検定における距離もそれぞれの $x_{\min}$ について調べる必要がある。 $(x_{\min} = 1$ が一番小さいので、この計測をするだけでいいということにはならない。)ここで、これらの値を平均で表していることは少し問題である。平均された元のデータが正規分布である保証はないからである。これらはあくまでも目安として捉えるべきである。

表1を見ると $\alpha$ はほとんど変化しない。(1)は小さくなっていくことから、 $x$ の小さいところがべき分布から外れている傾向が強いことがわかる。ところで、式(6)からわかるように、すべてのデータを元に $\hat{\alpha}$ を決めているが、べき分布においてデータの圧倒的多数は $x$ の小さいところにある。したがって、(2)において距離が大きくなっていくことは、データの個数が少なくなっていることに由来する。また、(3)において最大の点が大きくなっていくことは当然としても、その増え方は $x_{\min}$ が1増える量より大きい。これらの結果はデータの個数が小さくなっていくことによる。前節の分析で述べたように $x_{\min}$ を大きくしていけば、いつかはデータの個数が少なくなり $p$ 値も小さくなる。

図2は分析における $p$ 値の結果である。縦軸が $p$ 値、横軸が $x_{\min}$ である。横軸に並行に引かれているのは $p=0.1$ の境界線である。これを見ると $x_{\min}$ が7以上になったところで、 $p$ 値が0.1を上回るようになっていく。したがって分析の結論としては、共同出願ネットワークは次数が7以上においてべき分布であるといえる。そしてこのときの $\alpha$ は3.03である。前述したように一般的には $2 < \alpha < 3$ であり、わずかにその値を超えている。べき指数は小さいほど寡占(格差がある)状態を示すので、本論文で紹介した他の多くのべき分布の中では比較的格差は少ないと理解できる。

この分析結果はこれまでの研究[7]を肯定する面と否定する面がある。肯定する面からは、確かにべき分布の部分があるということであり、べき分布を仮定した上でのモデルの構築は、そのネットワークの生成過程を理解する上で役に立つといえる。一方で、 $x_{\min}$ が7より小さいところはデータの個数としてほとんど全体であり、9/10程度を占めている。次数の高いノードがリンクの大半を保有してはいるが、ノードの数で見ると次数の低いノードが圧倒的に多い。リンクが生成される数理的なモデルは、高次のノードのリンクの生成の振る舞いについて把握しているとはいえるが、圧倒的多数の次数の低いノードの振る舞いについての理解に用いると誤った結論に陥る可能性が高い。

## 4 結 論

社会科学ではデータに対して、典型的な値が存在するとほとんどの場合に仮定される。これは正規分布を仮定しているからである。しかし、自然はもとより社会経済におけるあらゆる分布が正規分布であるはずはない。したがって、ある現象のデータを得て、そのデータの母分布を意識せず正規分布を仮定した場合、誤った結論に至るケースが数多くある。実際に現代においてさまざまな大規模データが入手され、正規分布でない現象が数多く確認されるようになった。なかでも、近年盛んに研究されているのがべき分布である。本論文ではある分布がべき分布であるかを判定する Clauset らの方法により、ある企業の共同出願相手数の確率分布の判定を行った。

企業の共同出願相手数の確率分布は、まったく正規分布ではフィットできず、べき分布に近い形をしている。実際にその判定において、 $p$  値を0.1に設定したところ、 $x_{\min}$ は7、 $\alpha$ は3.03となった。したがって相手数が7以上のところにおいてはべき分布であるといえる。しかしながら、確率分布全体にわたってべき分布であるということは正しくない。

本論文の例からわかるように、現代においては大量のデータが獲得できるようになり、その確率分布が把握できるようになっている。それらのデータにおける多くの分布が正規分布から外れており、その元となる現象の理解のためには、まず正しく分布を把握しなければならない。本論文で示したべき分布の判定方法は極めて一般的な方法であり、指数分布や対数正規分布などさまざまな分布の判定にも使える。本論文が判定の有用性を示したことは、今後得られる新たな分布に対する判定の有効性と必要性を示したといえる。

## 謝 辞

本研究は科研費 (20730268) の助成を受けたものである。

## 参考文献

- [1] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1 (2): 226–251, 2004.
- [2] M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46 (323), 2005.
- [3] D. Sornette. *Critical Phenomena in Natural Sciences*, chapter14. Springer, Berlin, 2nd edition, 2006.

- [4] A. L. Barabási and Z. N. Oltvai. Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5: 101–113, 2004.
- [5] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *arXiv: 0706. 1062*, 2007.
- [6] H. Inoue, W. Souma, and S. Tamada. Spatial characteristics of joint application networks in japanese patents. *Physica A*, 383: 152–157, 2007.
- [7] 井上 寛康, 相馬 亘, and 玉田 俊平太. 共同研究開発の過程理解に向けた特許のネットワーク分析とモデル構築. 情報処理学会論文誌, 49 (4), 2008.
- [8] V. V. Kryssanov, F. J. Rinaldo, E. L. Kuleshov, and H. Ogawa. A hidden variable approach to analyze “hidden” dynamics of social networks. In, *Why Context Matters*, pages 15–35. VS Verlag für Sozialwissenschaften, 2008.
- [9] B. C. Arnold. *Pareto Distributions*. International Cooperative Publishing House, 1983.
- [10] M. S. Handcock and J. H. Jones. Interval estimates for epidemic thresholds in two-sex network models. *Theoretical Population Biology*, 2006.
- [11] A. Clauset, M. Young, and K. S. Gleditsch. On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 2007.
- [12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.
- [13] D. G. Mayo and D. R. Cox. Frequentist statistics as a theory of inductive inference. In J. Rojo, editor, *Optimality: The Second Erich L. Lehmann Symposium*, pages 77–97, 2006.
- [14] S. Tamada, Y. Naitou, F. Kodama, K. Gemba, and J. Suzuki. Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics* , 68(2): 289–302, 2006.
- [15] W. Souma. 経済における複雑ネットワーク－日本の経済ネットワークは特殊か？－. 人工知能学会, 20 (3) : 289–295, 2005.
- [16] W. Souma, Y. Fujiwara, and H. Aoyama. *The complex networks of Economics Interactions- Essays in Agent-based Economics and Econophysics*, chapter 2, pages 79–92. Springer, 2006.
- [17] 青山 秀明, 家富 洋, 池田 裕一, 相馬 亘, and 藤原 義久. パレートファームズ－企業の興亡とつながりの科学－. 日本経済評論社, 2007.
- [18] 青山 秀明, 家富 洋, 池田 裕一, 相馬 亘, and 藤原 義久. 経済物理学. 共立出版, 2008.
- [19] S. Yook, H Jeong, and A. L. Barabási. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences*, 99 (21) : 13382–13386, 2002.