

オンライン家計簿ココマネの経済分析 -ユーザーの世帯所得とパレート分布の関係について-

村館 靖之[†] 須藤 修[‡]

東京大学大学院情報学環[†] 東京大学大学院情報学環[‡]

1 はじめに

オンライン家計簿ココマネとはネット上でユーザーが家計簿を付けることを可能にした NTTメディアサプライ株式会社の提供するサービスである^[1]。

ココマネのユーザーは約20万人であり、約1:3の比率の男女比で、平均年齢は36.27歳である。メインターゲットは30代の女性層であると推定される。

本研究プロジェクトでは、ココマネユーザーの消費行動の分析によって、リアルタイム性を持った経済分析（オンライン家計簿からの物価指標の取り出しなど）を行うことを目的としている。

本報告では、プロジェクトの一環としてココマネのユーザープロフィールの所得情報をもとに、ココマネユーザーの所得がパレート分布にしていることを明らかにする。

2 ココマネユーザーの所得とパレート分布

約20万のココマネユーザーのプロフィールに関して2011.11.28現在の45780の家計（ユーザー）が世帯年収を記入している。年収の入力は基本的に、1500万円以上、以降100万円おきで、200万円以下と入力されている。分析にあたって、例えば年収3000000万円と入力しているユーザーに関しては、300万円の誤りか、300億円なのかかわからないが、ノイズとして捨象した。また年収ゼロの家計に関しても、分析から外した。

パレート分布とは、個人の所得分布がべき分布にしていることを発見したパレートの研究をもとに命名されている。

もともと19世紀末の所得分配に関する研究であったが、現代においてもパレート分布の研

究は『経済物理学』として復活している^[2]。所得分布だけではなく、コーパスでどの単語が頻出するかという単語の分布などにも見出される。またロングテールの名のもとにパレート分布のしっぽ（裾野）をWEBマーケティングに利用する例もある。

このパレート分布の特徴は所得の分布がべき分布をしており、反比例のグラフのように分布し、社会的富（所得）が非対称的に分布していることを表している^[3]。分布のかたち形状（パレート分布）はどこの社会においてもほぼ同一で、不平等自体は完全にはなくなる（政策的に不平等を緩和することはできる）ことを示している。

次にココマネユーザーの年収と累積家計数の関係を図1でみることにしよう。

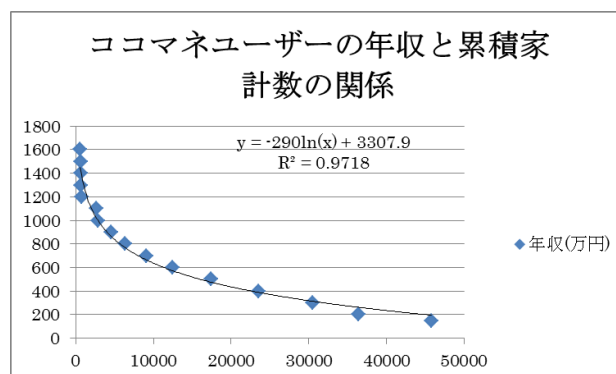


図1 ココマネユーザーの年収と累積家計数の関係（2011.11.28時点でのユーザープロフィール情報をもとに筆者作成）

図は縦軸に年収、横軸に累積家計数を取っている。ラインは近似曲線（対数）で

$$y = -290\ln(x) + 3307.9$$

$$R^2 = 0.9718$$

となっている。

回帰式のp値は $2.99E-12 < 0.05$ となっており、5%水準で有意である。

図1からココマネユーザーの所得分布（累積）が対数の近似曲線で97%の決定係数を越え

Economic Analysis of Cocomane: the relationship between the Pareto distribution and users' income

[†]Yasuyuki Muradate, Graduate School of Interdisciplinary Informatics, The University of Tokyo.

[‡]Osamu Sudoh, Graduate School of Interdisciplinary Informatics, The University of Tokyo.

るあてはまりのよい近似ができることがわかる。

図1は所得 y と所得 y 以上の(累積)家計数 x の数の対数の間に負の相関があるというパレート分布(パレート法則)を忠実に再現している。より正確にはパレート分布は、

$$\ln y = -\beta \ln x + \ln A.$$

と表される。つまりある所得 x 以上の人数 y とある所得 x の間、2つの変数の対数の間に負相関があり、なお y と x の対数の間には線形の関係がある。 β の値を見るために、今度は図2で y も対数を取って、分布をみしてみる。

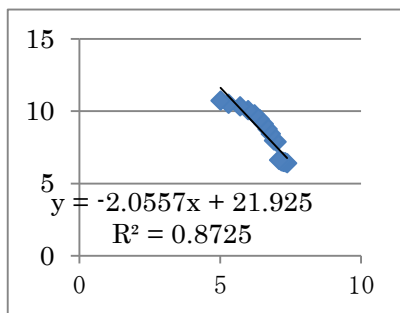


図2 ココマネユーザーの年収と累積家計数の対数の関係(2011.11.28時点でのユーザープロフィール情報をもとに筆者作成)

図2では縦軸に年収の対数、横軸に累積家計数の対数を取っている。ラインは線形の近似式で、

$$y = -2.0557x + 21.925$$

$$R^2 = 0.8725$$

となっている。また回帰式の p 値は $1.21E-07$ (< 0.05) となっており、5%水準で有意である。

3 考察

図1で見たように、ココマネのユーザーの所得はパレート分布にしたがって分布している。

パレート分布は所得分布に関する研究から来ているので、この結果は直感的に受け入れることができる。

オンライン家計簿のユーザー属性(所得)がパレート分布にしたがっていると考えると、オンライン家計簿のデータから得られる他の属性や変数もパレート分布にしたがっている可能性がある。

WEBマーケティングにおけるロングテールの活用で代表されるように、ビッグデータ時代のデータベースに関して、ただデータの容量が大きだけでなく、そのデータが何らかの法則性にしたがっている可能性がある。

パレート分布に代表されるデータの法則性を

つかむことが出来れば、マーケティングに応用が可能である。

またオンライン家計簿の所得累積分布がパレート分布にしたがっているならば、母集団は少ないが高額の所得を持っている富裕層と、人数の多い低所得者層の両者を見据えたマーケティング、商品開発・イノベーションを行うことの必要性が考えられる。

パレート分布が社会の所得の不平等の指数を反映していることを考慮に入れると、静学的なパレート分布の形状だけではなく、動学的にパレート分布の形状がどのように変化していくか見極めることが今後の課題となる。

4 結論と今後の課題

本研究プロジェクトは、オンライン家計簿ココマネを分析することで、リアルタイムの物価水準の予測に代表されるような社会・経済システム分析への成果を生み出すことを最終目標としている。

本報告では、その一環として、ココマネユーザーの所得がパレート分布にしたがっていることを確認した。

所得に代表されるオンライン家計簿のユーザー属性がパレート分布にしたがっているという事実は、オンライン家計簿の他のユーザー属性や変数も、パレート分布やその他の統計分布、法則性にしたがっていることを示唆している。オンライン家計簿上の法則性・統計分布を探求することが今後の課題となる。

またビックデータの中にパレート分布に代表される法則性があるという概念は、データベース検索アプリケーションを開発する際のヒントになるのではないだろうか。

参考文献 URL および註

[1]ココマネ <http://money.cocokulu.jp/money/>
最終訪問時 2012.11.20

[2]青山秀明・家富洋・池田裕一・相馬亘・藤原義久(2008)『経済物理学』共立出版

[3]パレート分布の概念はいわゆるパレート法則(2:8の法則)とロングテールの両者を含んでいるべき分布に関する概念である。

*謝辞

本研究は文部科学省最先端研究開発支援プログラム FIRST 喜連川プロジェクトの一環として行われました。関係者の諸先生方にお礼申し上げます。