

# 確率的な順位付け

服部哲弥

## 1 ランキング

インターネット上には、多数の動画を集めた動画投稿サイトや、大勢の書き手やたくさんの発言を集めるブログサイトや掲示板、そして音楽や書籍などの多様なタイトルが商品として並ぶ通販サイトなど、ウェブサイトや電子的な陳列棚に見立てて、多数の創作物を電子的に並べたサイトがある。

大量の商品や情報を公開し注文を自動で受付けるウェブサイトが比較的容易に作れることや、そのサイトに不特定多数の人が即座に自由にアクセスできる、という特性のおかげだろう。即座にアクセスできるので、こうしたウェブサイトの内容は全体的には流行に敏感になるようだ。

ブログや掲示板や投稿サイトや通販サイトには、流行を、言い換えると人気を、反映した、ランキングと呼ばれる数値が表示されることがある。ここでは、比較的長く研究をしてきた amazon.co.jp という通販サイトが扱う和書のランキングについて、確率論との関わりで発見したことを紹介したい。正式名称でなくて申し訳ないが、簡単のため以下で取り上げる同サイトを「アマゾン書店」と呼ぶ。

アマゾン書店で本を検索すると、その本についての紹介と注文用のボタン（リンク）を含むページが表示される。創作の情熱と努力に対する敬意を脇に置くように申し訳ないが、以下では本の内容には一切興味がないので、「ページ」とは実際の本ではなく、ウェブブラウザで表示されるウェブページを指す。一つのウェブページは町なかの本屋さん並べてある本 1 点に相

当する。アマゾン書店の本のページの中程やや下にアマゾン書店が「ランキング」と呼ぶ、順位を表す数値がある。アマゾン書店は数百万ページ分の和書の表示データを用意しているため、90% 以上の本は十万位前後から数百万位までの巨大な順位の数字が付く。トップ 10 やベスト 100 はアマゾン書店の扱う本の中では例外中の例外、1 万位ですらめったに無い本である。アマゾン書店における「普通の本」とは、普段は 10 万位より悪い順位が付く本だ、と最初に強調しておこう。本稿では「普通の」という形容詞をこの意味に使う。

アマゾン書店のランキングの数値は、ヘルプのページや実際の観測によると、毎時 1 回更新される。ほぼ即座に時間変化が見える巨大な順位という、インターネット時代以前は日常で見ることの無かった、大きな特徴がある。日本全国の多数の読者、いや、日本語が読める世界中の読者、のランダムな注文によって、各々の本のランキングが時々刻々変化する。

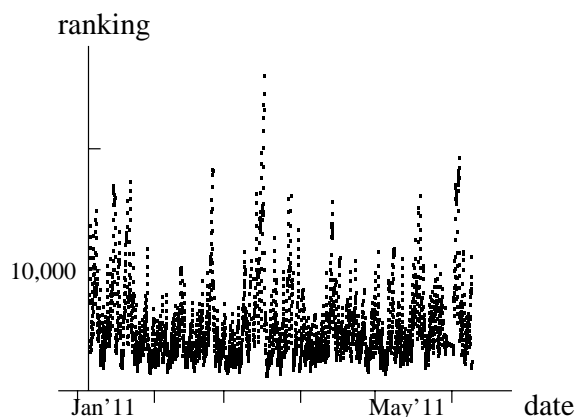


図 1:

アマゾン書店はランキングの具体的な計算方法を公表していない。ランキングに関心のある人の多くは漠然と、最近の売れ行き、言い換えると、流行度を順位づけている、と考えているようだ。しかし、片手間にランキングを眺めてもそれ以上詳しいことはわからない。ほぼ毎時変化すること、数学書のように大して売れない「普通の本」はたいてい見るたびに順位が悪くなること、そして、たまにふと気付くといつのまにか順位が良くなっていることもあること、程度しかわからない。

片手間で見ていてはちがいが分からないというならば、ランキングの数値を毎時自動採取してグラフにしたら規則性が分かるか？図1は、10年近く前にミリオンセラーになったある1点の本の最近半年間の順位変化である。横軸が半年にわたる時刻軸、縦軸は順位である。順位というと通常は図の上のほうを高順位にとるが、ここでは数学のグラフの常に合わせて、順位の数値の小さいほうを図の下のほう、横軸近くにとる。平均すれば5千位を今でも下回る、よく売れる本の座を維持していることがわかるが、恐らくそれ以外の点ではトゲだらけの不規則なグラフに見えるのではないか。

アマゾン書店がランキングをどのような規則で決めているかに関心を持つ人は少なからずいる。そのことを、本稿の背景にある研究を始めた頃に知ったが、明快で定量的な結論は知られていなかったようだ。アマゾン書店という、一民間会社の自由な判断でいくらでも複雑に定義できる数値に単純な数学の入る余地は少ない、という思いこみもあったのかもしれない。しかし、数学的な思考方法の助けを借りて、別の角度から考えることで、ランキングの時間変化についていくつかのことを発見した。その内容の一端を紹介する。興味を持っていただけたら、詳しいことについては拙著[1]を強くお勧めする。原啓介さんが書いて下さった1ページの書評[2]も参考にいただければさいわいである。

## 2 大数の法則

早速だが、ここは数学セミナーなので、現実の現象をいったん忘れて、やや数学的な視点に目を転じよう。結論をひと言で書くと、ランキングへの確率論の関わりは大数の法則である。ランダムな多数の量の平均をとると(しかるべき条件の下では)期待値という、ランダムでない決定論的な値に近づく。この形の定理を大数の法則と呼ぶ。

2008年の数学セミナー[3]で次のクイズを紹介した。『図2の16個の画像は、濃い色で書かれた図形にランダムな濃さのノイズを加えた絵です。ノイズは16枚の絵のあいだで違いますが、元の図形は同一です。16枚の絵に共通する元の図形を当ててください(答えは[4]にあります。)』実際に濃さを数値化して16の図について加えると元の図形が鮮明に浮かび上がる。ランダムなノイズは正負打ち消し合って平均すると0に近く

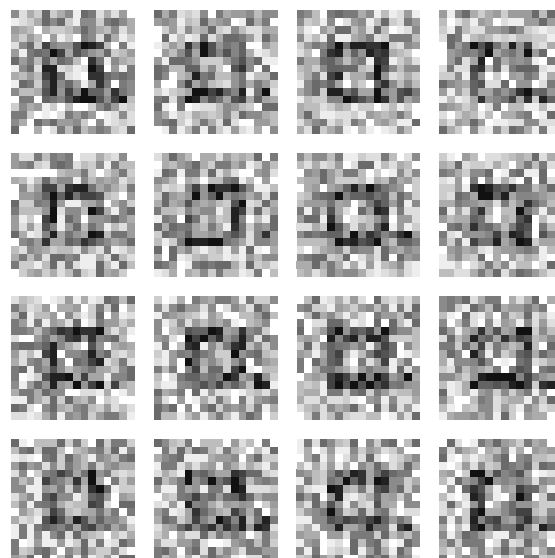


図 2:

なるが、本来の図形は一定値なので平均しても消えないからである。これが大数の法則の一つの説明である。

図1のような、売れている本のランキングを漠然と眺めても大数の法則は見えないが、普通の本に目を転ずると、見える世界が一変する(しつこいが、普通の本とは、アマゾン書店がページを用意する数百万部の本の大多数を占める、数学書を一つの典型例とする、世間一般の表現では『売れない本』のことである。)

まず、ランキングが上がるのは、その本がアマゾン書店で注文されたときである。このことは、実際に普通の本を注文してみれば2時間以内に納得できる。売れている本を注文すると、順位が上がっても自分の注文の効果か、他の人たちの注文の効果かわからないので、薦めない。普通の本に注目することで初めてこの段落の考察が実感できることを宣伝しておきたい。

次に、全ての本について、注文が入ったとき順位が上がるという事実をていねいに考えてみる。注目する本より下位にいた他の本が注文されると、順位を上げて注目する本を追い越すので、注目する本の順位はその分下がらなければならない。

以上の観測と考察を、本質を損なわない範囲で、数学的にもっとも単純化すると、『どの本も注文されたときにランキング1位となり、順位を追い越された他の

本は順位を1ずつ下げることで順位の重複や空きを解消する』というモデルに行き当たる。この規則は先頭に跳ぶ規則として古くから知られている。「積ん読」や「超整理法」の原理でもある。

簡単のため、個々の本はランダムに注文されるが、それぞれ平均的な売れ行きは決まっているとす。  $N$  点の本があって、それらの表題はそっけなく1から  $N$  までの番号とし、  $k = 1, 2, \dots, N$  について、本  $k$  の平均的な売れ行きを  $w_k$  とす。モデルを数学的にもっとも簡単にすべく、1位への跳び(注文)はポワソン過程に従うとすると、  $t$  時間の中に本  $k$  が一度も売れない確率は  $e^{-w_k t}$  となる(これは、半減期  $\frac{1}{w_k} \log 2$  の放射性物質が  $t$  秒間崩壊しない確率に等しい。)

さて、めったに売れない本  $i$  が奇跡的に売れた時刻を  $t = 0$  とす。その後、時刻  $t$  までその本は売れなかったとして、順位が  $X_i(t)$  に後退するとすると、  $X_i(t)$  は時刻0以降  $t$  までに一度でも注文された本の点数に1を加えたものに等しい。注文はランダムなので、この順位  $X_i(t)$  は確率変数だが、  $X_i(t)$  と本の総点数  $N$  の比は、本の総点数  $N$  が大きくなる極限では、大数の法則によって期待値に収束する。本  $k$  についてはその期待値は一度でも注文される確率に等しいので、上に書いた一度も売れない確率  $e^{-w_k t}$  を1から引けばよい。期待値の線形性から  $k$  について足し合わせることで

$$X_i(t) \sim \sum_{k=1}^N (1 - e^{-w_k t}) \quad (1)$$

を得る。  $\sim$  の意味は両辺を  $N$  で割って  $N$  を大きくする極限を考えると等しくなる、という主張である。

右辺が  $i$  によらないことに注意 ( $N$  で割った平均の極限を考慮するので、1点の本の効果は無視できる。) このことから、本  $i$  が売れない間の本  $i$  の順位低下は、  $i$  がどの本であっても共通である。

ここまで、流行度を反映する順位付けの数学的にもっとも簡単なモデルの数学的な結論である。この数理モデルを確率順位付けモデルと名付ける。

### 3 確率的順位付け

実際のアマゾン書店のランキングの時間変化が、確率順位付けモデルという、単純な数理モデルの結論に近いかは数学の中だけではわからない。そこで、データによる統計的当てはめを行うことで適否を確かめる。

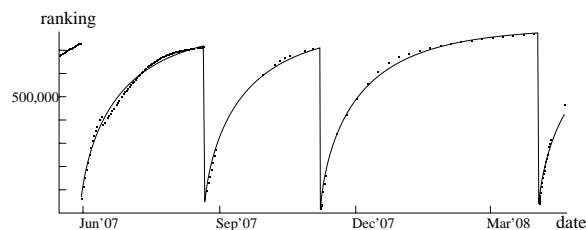


図 3:

図3の点が実測値、ヨットの帆をつないだような曲線が(1)に基づく理論曲線である。横軸は約1年にわたる時間、縦軸は図1と同様にランキングである。ただし、図1に比べて、縦軸は表示を約50分の1倍に圧縮し、横軸は半分に圧縮した。

統計的当てはめの具体的内容は、数学と現実の橋渡しという点で重要である。特に、商品の売上の分布のような、通常は社外秘に属する情報を、ランキングという公開されたデータだけから分析できるという、この研究のもう一つの成果の鍵となる。しかし、詳細に立ち入る必要があるので、この方面の紹介は割愛して[1]に譲ることにする。

図1と図3のデータの見かけの違いが気になるかもしれない。どちらもそれぞれ特定の本について、アマゾン書店のランキングの時間経過を追ったものである。違いは図1がよく売れる珍しいビッグヒットの本、図3は約1年間に4冊しか売れなかった普通の本、である(図1では点が均一なのに対して図3の点に粗密があるのは、前者がデータ収集を自動化した最近のグラフなのに対して、後者が初期の、手で収集していた時期のデータのためである。これは数学とも現実とも無関係なので、目をつぶってほしい。) 図の見かけの違いは、ビッグヒットは順位が下がり始めてもすぐに注文が入って上位に跳ぶので、商品が売れない時間の順位低下(1)を観測することが難しいのに対して、大多数の普通の本は(1)という決定論的な関数形を見だしやすい、ということである。言葉を換えると、図1のグラフの「トゲ」1つを大きく拡大して、グラフの上昇(順位の低下)を子細に観察すれば、図3のヨットの帆と同じ形をしている、ということである。こうして、アマゾン書店のランキングの時間変化の規則性が明らかになった。

前節で実際の現象からいったん数学に退避して、数

学的にもっとも単純な確率順位付けモデルを選んだことを考えると、理論は実際のアマゾン書店のランキングの時間変化を驚くほど正しく捉えていると考える。図3の実測データ（黒点）と理論（曲線）の類似をそのように理解する。それはしかし、アマゾン書店が確率順位付けモデルをそのままそっくりランキングのアルゴリズムとして採用している、という意味ではない。

まず、確率順位モデルは数学的に素直な形で一般化することで、現実をより精密に説明できる。図3に書き込んだ曲線よりも、もう少し精密で実際のデータに合う理論曲線を描くための数学的な基礎準備は整っている、ということは関心ある向きにお伝えしておきたい。

その上でなお、たとえば、アマゾン書店は最新の注文だけではなく、1日程度の過去の注文の効果を残していることが、データと理論の食い違いを子細に観察するとわかる。これについては晋遊舎 MONOQLO 編集部の澤原さんが独自の実験結果を [5] に発表されている。さらに脱線すると、アマゾン書店の注文の仕組みの穴を利用した投機的行為も可能で、ランキングの動きを別の意味でいっそう複雑にする。

現実の複雑さにもかかわらず、図3のように確率順位付けモデルと実際のアマゾン書店のランキングは、少なくとも図に示される程度の概要については、よく合う。これは、『流行度を反映する順位ならば、アルゴリズムの違いが大きく影響するのはよく売れる上位の限られた本のみで、普通の本のランキングの曲線はアルゴリズムの詳細によらず、確率順位付けモデルの理論曲線に近い』という、普遍性があることを意味する。

普通の本はめったに売れない。1冊売れると、その本はその瞬間、大半の普通の本よりも「はやって」いる。大半の普通の本にとって、流行に応じた順位付けは先頭に跳ぶ規則と同じことになる。数学的に徹底的に単純化したモデルを考えるという研究上の作戦が成功したのは流行度という概念の普遍性による。こうして、流行を反映する大規模な順位の時間変化を考えるモデルとして、確率順位付けモデルが有効であることがわかる。

数学としての確率順位付けモデルのより深いおもしろみは、位置とジャンプ率（1位に跳ぶ頻度）の結合経験分布の大数の法則である。本稿で紹介した (1) は極限が数値なのに対して、極限が分布になる問題を扱う。それが非線形偏微分方程式によって特徴付けられる、という意味で、ランダムなミクロの描像と決定論的かつ滑らかなマクロの描像を数学的に結びつけるという

流体力学的極限の側面についての「もっともやさしい例題」になる [1, 2]。

確率論は数学の中では若い分野に属する。書評 [2] を書いて下さった原啓介さんの訳書 [7] の扱う時代から4世紀弱を経ただけである。確率論研究の短い歴史の中で大数の法則は中心的な位置づけであり続けた。流体力学極限の確率論的研究はその最先端の一つに位置づけられることを2011年の数学セミナーで紹介した [6]。若い数学としての確率論と現象の研究上の頻繁な行き来はまだまだ続くと考え。

## 参考文献

- [1] 服部哲弥「Amazon ランキングの謎を解く」, 化学同人出版, 2011年.  
<http://web.econ.keio.ac.jp/staff/hattori/amazonj.htm>
- [2] 原啓介, 書評「Amazon ランキングの謎を解く」, 数理科学 581, 2011年11月号, p.61.
- [3] 服部哲弥「数理統計学」, 数学セミナー(特集 大学数学はこう学べ!) 559 2008年4月号, p.38-41.
- [4] 服部哲弥「統計と確率の基礎」, 学術図書出版, 2006年.  
<http://web.econ.keio.ac.jp/staff/hattori/gakjutu.htm>
- [5] 澤原昇「続アマゾンランキング仕組み検証!」, MONOQLO 2011年11月号, p.26-27.
- [6] 服部哲弥「確率論」, 数学セミナー(特集 大学数学が一望できる数学ランドへようこそ) 597 2011年6月号, p.11-15.
- [7] K. デブリン, 原啓介訳「世界を変えた手紙」, 岩波書店, 2010年.

服部哲弥(慶應義塾大学・経済)

URL: <http://web.econ.keio.ac.jp/staff/hattori/hattori.htm>