広島大学　大学教育研究センター
大学論集　第11集（1982）：1－21

科学の生産性とその階層化過程

―「マタイ効果」の基礎理論―

山崎博敏

目　次

はじめに
1. 科学の生産性の分布に関する理論
2. 日本の化学者の生産性の分布
3. 研究活動の過程における強化
4. 科学者間の異質性
5. 科学社会学理論との関連性

科学の生産性とその階層化過程

――「マタイ効果」の基礎理論 ――

山　崎　博　敏*

は じ め に

科学の生産性，すなわち科学者が公表する論文の数量には個人による大きな差が存在する。そのような差異は科学者個人の能力や，置かれている社会的条件の相違が複雑に影響しあって生じるものと想定される。アメリカの統計学者A.J.ロトカは1926年に科学の生産性に一定の規則性があることを明らかにし，数学的に定式化した。それ以来今日まで，多くの学者によってより正確な分布法則を確立しようとする試みがなされてきた。科学社会学の発展とともに，これまで主として統計学者などによってなされてきた研究はさらに精緻化され，今日では理論的にも整備されたものになっている。そして，科学の生産性の分布の数理モデルは科学社会学理論の中核概念である，「マタイ効果」論と関連をもつことが明らかにされている。

小稿はまず，第一章でロトカの逆自剰法則からプライスの修正法則，サイモンの分布関数，負の二項分布などに至る主要な科学の生産性の分布モデルを整理し，次いで第二章ではこれらの諸モデルを日本の大学の化学者 124 人の生産性のデータに適用して，各モデルの適合性を実証的に検討し，その妥当性を明らかにする。その分析の結果，負の二項分布の適合度が最も高いことが明らかになった。つづく第三章と四章においては，この事実から導き出される 2 つの仮説，すなわち「研究活動の過程における強化」と「科学者間の異質性」を検討する。前者は，論文を公表すればするだけ益々その後再び論文を公表する確率は高まる，という仮説であり，マートンの「マタイ効果」の理論と関連がある。他方，後者は科学者の間に能力や研究条件等において大きな差異が存在する，という仮説である。この 2 つの仮説のいずれかが妥当すれば，生産性の分布は負の二項分布に従う。第三章と四章では，データに即してこの 2 つの仮説の妥当性を検討し，最後に，第五章では少くとも「強化モデル」の解釈の正しさが証明され，それがマートンの科学社会学理論に対してもつ意味と，日本の学界の社会構造の性格について論じる。

* 広島大学・大学教育研究センター助手

1 科学の生産性の分布に関する理論

ロトカの逆自乗法則

科学の生産性の分布に関するパイオニア的研究はアメリカの人口統計学者A.J.ロトカによって1926年になされた。これは「n篇の論文を公表する学者の数は$1/n^2$に比例する」というもので，ロトカの生産性の逆自乗法則と呼ばれるものである。これを数式で表現すれば次のようになる。

$$f(n) = c/n^2 \quad (n \geqq 1)$$

ここで$f(n)$はn篇の論文を公表した科学者の数，cは定数である。これから明らかなように，公表論文数の少ない学者の数は大きい。しかし，nが大きくなるにつれて$f(n)$は急速に減少してゆき，公表論文数の大きい学者の数は少なくなる。従って，$f(n)$は極めて非対称な逆J型の形をしている。

上の式から，1篇の論文を公表する科学者数が100人いるとすれば，2篇の論文を公表する者の数は25人，3篇の論文を公表する者は11人いることになる。また，1篇しか論文を公表していない者の割合は全体の60.8％と計算される。なぜならば公表論文総数を100とすれば

$$\sum_{n=1}^{\infty} f(n) = C\left(\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots\cdots\right)$$

$$= C\sum_{1}^{\infty}\frac{1}{x^2} = C\frac{\pi^2}{6} = 100$$

ゆえに，$C = 100 \times \frac{6}{\pi^2} = 60.8$，$f(1) = c/1^2 = 60.8$ となるからである。[1)]

このように，ロトカの逆自乗法則はきわめてシンプルな逆自乗法則である。これを起点として今日まで50余年のあいだ，科学の生産性の分布の形状を確定しようとする研究がなされてきた。その際，この法則は必ず議論の出発点となっている。これほど長い間，多くの学者にとりあげられてきたのは，この法則が科学の生産性の分布の概略的な形を単刀直入にとらえているからであるが，他方，細かく検討してみれば，実際の観測データと相容れないところが見い出されるからでもある。そこで，本節では以後ロトカ以降の諸条件をレビューし，ロトカの逆自乗法則がどのように修正され，展開していったかを跡づけてゆくことにしたい。

プライスの修正法則

プライスは1963年に17世紀～18世紀初期にかけての英国の王立協会の『哲学年報（Philosophical Transactions)』（第1巻～70巻）と，『ケミカル・アブストラクト』（1907～1916）からとったデータから，ロトカの法則の正確さを確認した。しかし，詳細な検討の結果，高い生産者のところでは逆自乗はフィットせず，むしろ逆3乗に近づく，と主張した。[2)] つまり，ロトカの逆自乗法則は多生産者のところでは人数$f(n)$を過大に見積っている，というわけである。そして彼は次のような修正法則を提出する。ある時間間隔（ここでは生涯をとる）内で，少なくともP篇の論文を公表する人の累積数をNとすれば，Nは

$$N = k\left(\frac{1}{p} - \frac{1}{a+p}\right) = \frac{ak}{p(a+p)}$$

であらわされる。ただし，k，aは定数で，ふつうaは約15である。

サイモンの分布関数

H．サイモンは，1955年に，ある確率モデルから出発して，科学生産物の分布のほかにも，文章中における単語の使用頻度の分布，都市人口の分布，所得分布，生物学上の種の分布をも含む広範囲な分布を記述・説明のできる一つの非対称な分布関数を導いた。[4]それによると，n篇の論文を公表する科学者の数$f(n)$は

$$f(n)=(1+\rho)\,\mathrm{B}(n,\ \rho+1)\,f(1)$$

で与えられる。ただし，$\mathrm{B}(n,\ \rho+1)$はベータ関数，

$$n_k=\sum_{n=1}^{k}f(n),\quad \alpha=\frac{n_k}{\sum_{n=1}^{k}nf(n)},\quad f(1)=\frac{n_k}{2-\alpha},\quad \rho=\frac{1}{1-\alpha}$$

である。このように，$f(n)$の形状はただ1つのパラメータρの値によってきまる。

サイモンは数学，化学，物理学，および計量経済学の4つのデータについて観測値と推定値を比較した。その結果，概して生産性の低い者については過大推定，生産性の高いものについては過小推定の傾向がみられた。[5]

ところで，プライスは1976年に，新たに蓄積的優位分布 (cumulative advantage distribution) という，ベータ関数によって記述される科学の生産性の分布を提案した。これは「成功が成功を生む」状況を数学的にモデル化しているもので，成功しないこと，つまり事象が生起しないことは失敗のチャンスが増大することになる罰を受けない，という点において負の二項分布とは異なるものである。[6]伝幡（強化）は成功の場合にのみおこり，失敗の場合にはおこらないという，正のフィードバックのみを仮定する。論文の生産過程に即して言えば，論文公表をすれば次に論文を公表する確率は増大するが，論文公表をしない場合には次に論文公表をする確率は不変であるという，単方向の正のフィードバックのみを想定している。[7]一人の科学者がn篇の論文を公表する確率密度関数は次式で与えられる。

$$f^{*}(n)=(m+1)\,\mathrm{B}(n,\ m+2)$$

ここでパラメータmは

$$m=1/(\mu-1)$$

で与えられる。ただし，μはサンプル科学者の公表論文数平均値である。[8]

しかし，この蓄積的優位分布は実質的にはサイモンの分布関数と等価である。なぜならば，プライスの蓄積的優位分布の予測式は，

$$f_p(n)=\mathrm{P}(x)\cdot\mathrm{S}=(m+1)\,\mathrm{B}(n,\ m+2)\,\mathrm{S}$$

ここで，$m=1/(\mu-1)$ （Sはサンプル科学者総数）

から，$m+1=\mu/(\mu-1)$，$f_p(n)=(\mu/(\mu-1))\mathrm{B}(n,\ m+2)\,\mathrm{S}$

一方，サイモンの分布関数は

$$f_s(n)=(\rho+1)\,\mathrm{B}(n,\ \rho+1)\,f(1)=(\rho+1)\,\mathrm{B}(n,\ \rho+1)\frac{\mathrm{S}}{2-\alpha}$$

ここで，$\rho=1/(1-\alpha)=1/(1-1/\mu)=\mu/(\mu-1)=m+1$
ゆえに，$f_s(n)=((2\mu-1)S\mu/(\mu-1)(2\mu-1))B(n, m+2)$
$=S\mu/(\mu-1)B(n, m+2)=f_p(n)$
となるからである。

対数級数分布と幾何分布，幾何級数分布

C.B.ウィリアムスは1944年に生物学者の論文公表数の分布を幾何分布，幾何級数(geometric series)分布，および対数級数 (logarithmic series) 分布の3つそれぞれに適用して比較研究した。[9]彼のデータでは対数級数分布が最も適合していた。対数級数分布は生物学において，種の分布の理論モデルとしてフィッシャーらによって1943年に提出されたものである。
幾何分布の確率分布は

$$P(n)=\left(\frac{1}{1+\lambda}\right)\left(\frac{\lambda}{1+\lambda}\right)^n \quad (n=0, 1, 2, \cdots\cdots)$$

ここで，$\lambda=1-N/S=(S-N)/S$ である。なお，幾何分布は負の二項分布の特殊な場合である。[10]
幾何級数分布による公表数推定値は

$$f(n)=n_1 x^{n-1}$$

ここで，$n_1=S^2/N$，$x=(N-S)/N$
対数級数分布の公表数推定値は

$$f(n)=n_1 x^n/n$$

パラメータ n_1 と x は $S=n_1(log(1-x))/x$
$N=n_1(1-x)$ で求められる。[11]
ここにSは科学者数，Nはその科学者たちによって公表された総論文数である。

逆自乗法則，プライスの修正法則，サイモンの分布関数，対数級数分布の比較検討

（理論）モデルを比較する一つの方法は現実のデータとの適合度を調べることである。ここではロトカが用いたのと同じアウエルバッハ (Auerbach) の物理学史数表に掲載されている19世紀以前の物理学者の生産性のデータ[12] を用いて計算してみることにする。表1は観測値と4つのモデルそれぞれにおける推定値である。公表論文数 $n=1$ ではプライスの分布関数と対数級数の推定値は観測値よりかなり低いが，ロトカの逆自乗法則はほぼ観測値に近似している。$n=2$ でもロトカの逆自乗法則は正確であるが，残る3つのモデルは過大推定の傾向がある。しかし，ロトカ逆自乗法則は，プライスが指摘した通り，n が大になると過大推定になる。しかし，プライスの修正法則と対数級数分布は n が大のときの推定値は観測値に近いものの，$4<n<10$ の範囲において過大推定になっている。サイモンの分布関数は $n=1, 2$ で過大推定，n が6を越えるころから過少推定の傾向がある。この4つの分布モデルの推定値の観測値に対する適合度をカイ2乗検定したところ，いずれも5％水準で適合していなかった。

この4つの分布モデルの特徴をより明確に把握するために，全体で165人の科学者がおり，ロトカ

の逆自乗法則にしたがって科学の生産性が分布している場合を想定して，これを観測値とした際に残る3つの分布モデルがとる推定値を計算することによって4つのモデル相互を比較してみよう。図1は横軸に公表論文数，縦軸に科学者数をプロットしたものである。さきのアウエルバッハのデータの推定値の比較と同じような関係があることがよりいっそう明確にわかる。さらに，図2はロトカの逆自乗法則を基準にして，これに対する残る3つの分布モデルの推定値の偏差(%)をプロットしたものである。対数級数分布は5＜n＜13の範囲ではロトカの逆自乗法則から100%以上もの偏差がある。

3つの分布はnがある一定以上の数より大きくなると，偏差（マイナス）は大きくなり，ロトカの逆自乗法則よりも推定値は小さくなってゆく。

公表論文数	観測値	推定値 $f(n)$			
n	$f(n)$	ロトカ	プライス	サイモン	対数級数
1	784	806	701	823	645
2	204	201	231	226	261
3	127	90	114	98	141
4	50	50	67	52	85
5	33	32	44	31	55
6	28	22	31	20	38
7	19	16	22	14	26
8	19	13	17	10	18
9	6	10	13	8	13
10	7	8	11	6	10
11−12	13	12	16	9	12
13−15	13	12	15	8	9
16−20	7	13	15	7	
21−30	11	13	13	6	7
31−40	2	7	6	7	
41−	2	20	9		

表1　逆自乗法則，プライスの修正法則，サイモンの分布関数，対数級数分布の比較（1）

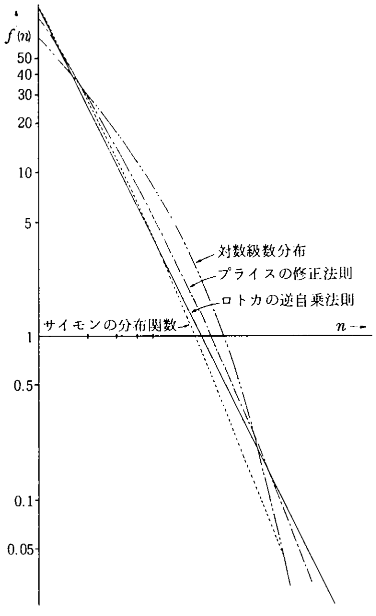


図1　逆自乗法則，プライスの修正法則，サイモンの分布関数および対数級数の比較（2）

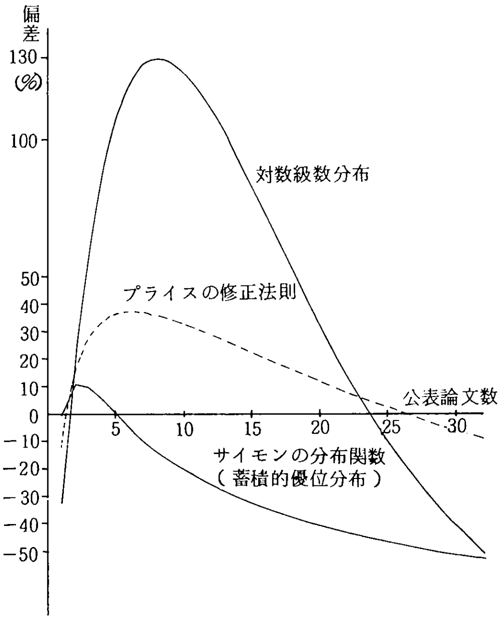


図2　プライスの修正法則，サイモンの分布関数，対数級数分布の逆自乗法則からの偏差(%)

対数正規分布

所得の分布が対数正規分布によく適合することは知られている。W. ショックレーは科学の生産性の分布も対数正規分布に従うことを主張した。[13] アメリカの自然科学系の3つの研究所と有名大学物理学科のスタッフ数年間の生産性の分布を対数正規確率紙にプロットし，そこからこのような結論を出したものである。ショックレーのデータをみてみると，中程度の生産者には対数正規分布がうまくフィットしているが，両極端すなわち生産性の高い者と低い者についてはそれほど適合していないようである。

負の二項分布

1970年代の半ばころから，科学の生産性の分布は負の二項分布に従うという見解が現われた。負の二項分布はいろいろなデータにあてはめてみても，ほとんど統計的に有意な適合を示していることから，これまでの諸モデルよりもかなりの程度の説得力をもっている。P.アリソン[14]はアメリカの239人の科学者について，I.K.R.ラオ[15] は生物学をはじめとする5種類のデータについて負の二項分布が最もよく適合することを報告している。

負の二項分布の確率分布は次式であらわされる。

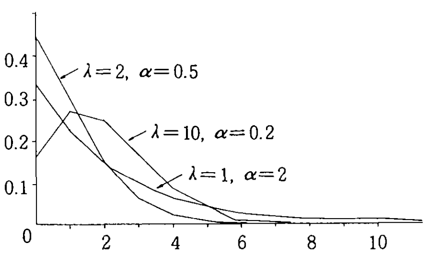
$$\mathrm{P}(k)=\binom{\lambda+k-1}{k}\left(\frac{1}{1-\alpha}\right)^{\lambda}\left(\frac{\alpha}{1+\alpha}\right)^{k}\quad(k=0,\ 1,\ 2,\ \cdots\cdots\cdots)$$

または

$$\mathrm{P}(k)=\binom{\lambda+k-1}{k}p^{\lambda}(1-p)^{k}\quad(k=0,\ 1,\ 2,\ \cdots\cdots\cdots)$$

その形は下図のようになる。

図3, 負の二項分布 NB(λ, α)



負の二項分布は Polya の壷の問題と呼ばれる確率モデルから導かれる。b 個の黒玉と r 個の紅玉を含む壷があり，いま，その中から無作為に1個取り出し，取り出された色と同色の玉を c 個付け加えて壷に戻す，というモデルを考える。この場合，最初に取り出された玉が黒であれば，次に黒玉が出る確率は，第1回目に黒の出る確率 $b/(b+r)$ よりも高い $(b+c)/(b+c+r)$である。この例は正の強化の事例であるが，これを一般化すれば，負の二項分布型の確率分布が導かれる。[17]

また，個体の異質性 (heterogeneity) を前提とした確率モデルからも負の二項分布は導かれる。これは稀にしか生起しない事象において，個体間である事象の生起する確率が違うという場合を想定したモデルである。いまポアソン分布 $\frac{\lambda^x}{x!}e^{-\lambda}$ において，λ が一定でなくて確率変数であるという場合に，λ の密度関数を $g(\lambda)$とおくと

$$\int_0^{\infty}\frac{\lambda^x}{x!}e^{-\lambda}g(\lambda)\,d\lambda$$

はまた　$x=\{0, 1, 2, \cdots\cdots\}$ 上の確率分布を与える。$g(\lambda)$が特にガンマ分布

$$g(\lambda)=\frac{1}{\Gamma(\alpha)}e^{-\frac{\lambda}{\beta}}\left(\frac{\lambda}{\beta}\right)^{\alpha-1}\frac{1}{\beta}, \quad \left(\begin{matrix}0<\lambda\leq\infty\\ \alpha,\beta>0\end{matrix}\right)$$

に従うとき，

$$\int_0^\infty \frac{\lambda^x}{x!}e^{-\lambda}g(\lambda)d\lambda=\frac{\Gamma(\alpha+x)}{x!\Gamma(\alpha)}(1-p)^x p^\alpha$$

（但し　$p=\dfrac{1}{1+\beta}$）

となり，αが特に正の実数であるとき，上式は，

$$\binom{\alpha+x-1}{x}p^\alpha(1-p)^x$$ となる。[18]

若干の問題

これまでとりあげた諸研究は，モデルそれ自体およびそれを検証するためて採用したデータの双方において，それぞれ異なった点が見られる。まず，対象とするデータについて述べることにしよう。データには大きく分けて，何年分かのアブストラクトや雑誌に登場するすべての科学者について論文公表数をカウントしたものと，ある一定の科学者をはじめから決めておいて，彼らが公表した論文数をカウントしたものとの2つの種類がある。したがって，前者のタイプのデータは通常，多くの科学者を対象とすることが多く，どのような経歴をもち，どのような地位にある者でも，アブストラクト等に掲載されている限りすべてサンプルとしている。そのなかには一回だけ登場するが，（例えば共同研究論文で名を連ねたような）最終的には研究者にはならなかったものも含まれているし，他の分野の研究者が1回だけその学問分野の雑誌に寄稿したような場合もあるであろう。このタイプのデータは公表論文数が1以上の者について対象としているのが特徴である。他方，後者のタイプのデータは，予めサンプルをフィックスしているため，サンプルとなる科学者の経歴や地位を知ることができるという有利な点がある。この場合は，公表論文数がゼロの者も出てくる可能性があり，公表論文数がゼロ以上の者を対象とすることになる。ロトカ，プライス，サイモン，ウィリアムス，ラオが使用したデータはすべて前者のタイプのデータであり，アリソンとショックレーの使用したデータは後者のタイプのデータである。

分布モデルにも公表論文数1以上を対象とするものと，ゼロ以上を対象とすることのできるものとの2つのタイプがある。前者は，逆自乗法則，プライスの修正法則，サイモンの分布関数，対数級数分布，幾何級数分布，対数正規分布であり，後者は，幾何分布と負の二項分布のような離散量の確率分布である。

さいごに，測定期間の問題がある。何年間の論文公表数を測定しているかは人によってまちまちである。ロトカのデータには1907年から1916年までのケミカルアブストラクトから得たものとアウエルバッハの作成した19世紀の約100年間にわたる物理学者のデータの2種類がある。これに対してウィリアムスは1年間だけ測定したデータを使用している。このように，測定期間については各人の研究によって千差万別である。測定期間によって観測値の分布の形が変わることも十分考えられる。これ

までのモデルでは，測定期間 t を考慮せずとも支し支えがないかどうか，あるいはそれをモデルに組みことが可能かどうかなどについては多くの場合不問のままにされている。

2. 日本の化学者の生産性の分布

これまで述べてきた科学の生産性に関する諸説はそれぞれの程度，実際の観測値に適合するのか。以下では，この問題を改めて日本の大学で活動している科学者の生産性のデータにあてはめ，検討する。サンプルは1979年現在日本の国公私立大学の理学系学部の化学科および類似の学科と化学系の大学付置研究所に勤務する講師以上の者を地位，設置者，大学のランクで 2×3×4 の 24（実際は16）のクラスターに層化した上で，それぞれを均等に 1/7 の割合で無作為抽出した合計 124 のサンプルである。[19] そして，各サンプルごとにケミカル・アブストラクトで1972年から1979年までの 8 年間の特許をを除く論文公表数（会議議事録も含む）をカウントした。共著論文も一つとしてカウントしている。したがって，本サンプルは fix されたサンプルであり，ロトカななどのような，ある雑誌に論文を公表した者全員を対象にしたものとは異なったタイプのデータである。

表 2 は 124 人の科学者の1972年から1979年までの生産性の分布である。明らかに，公表論文数の小さい方に偏よった極めて非対称な分布をしており，公表論文数 $n=0$ で人数が最大で，n が増大するにつれて人数は急激に減少し，長いすそをもっている。各年の公表論文数の平均値は1972年から1978年の 7 年間においては 2.0 から 2.7 の間にあり，[20] 1 年間の平均では 2.3 である。つまり，日本の大学の理学部化学科（研究所も含む）の講師以上の教員は一人あたり平均して一年間に 2 篇強の論文を公表していることになる。また，その分散はかなり大きい。分布を時系列的にみると，時間が経るにつれて，つまり，学者の年令が増すにつれて平均値と分散は大きくなっている。標本集団全体をみれば，年令が増すとともに，地位も講師から助教授へ，助教授から教授へと昇進する者もあり，一人あたりの論文生産能力が高まる結果，平均公表論文数は大きくなっていくのであろう。

公表論文数	1972	1973	1974	1975	1976	1977	1978	1979
0	40	39	36	35	33	39	39	43
1	24	23	33	29	27	29	25	33
2	23	26	24	16	21	21	17	16
3	16	12	6	16	10	13	12	13
4	9	6	6	11	11	4	7	4
5	4	4	5	2	3	4	7	8
6	1	2	2	6	3	1	7	2
7	4	3	4	2	3	4	3	1
8	0	2	2	1	7	3	1	1
9	0	3	0	1	2	0	1	1
10	0	2	3	1	1	2	0	0
11	0	0	1	1	0	3	1	0
12	0	1	0	1	0	0	2	0
13	1	0	0	1	0	0	0	1
14	0	0	2	0	1	0	1	0
15	0	1	0	0	1	0	0	1
16 − 20	2	0	0	1	0	0	1	0
21 − 30	0	0	0	0	1	1	0	0
計	124	124	124	124	124	124	124	124
平　均	2.0	2.2	2.2	2.3	2.7	2.3	2.5	1.8
標準偏差	7.6	7.9	8.3	8.4	14.2	10.1	9.6	6.0
対称度*	3.1	2.0	2.1	2.2	3.5	2.9	2.1	2.6

*　三次モーメント (Skewness)

表 2　日本の化学者 124 人の生産性の分布

さて，このような生産性の分布はこれまでに述べた分布のモデルのどれに最も適合するであろうか。その結果によって，我々は学問生産の背後にある社会的メカニズムを推測することができる。表2は1975年の公表論文数のデータを各モデルにあてはめたものである。なお，サイモンの分布関数，ロトカの逆自乗法則，対数正規分布，対数級数分布は公表論文数ゼロにおいては定義されていないので，その部分の観測データは除外して計算している。観測値に対して最も良い適合度を示しているのは負の二項分布で，5％水準で有意である。(カイ二乗検定による)。対数級数分布は1％水準で有意であるが，他はすべて適合性は棄却される。これを1972年から1979年まで8つのサンプルについてそれぞれ調べてみると，表3に示すとおり，負の二項分布は1976年を除くのこる7つで5％水準で適合性が認められ，1％水準では8つすべてで適合性が認められる。対数級数分布では1％水準では8つすべてで，5％水準では3つで適合が認められた。しかし，そのほかのモデルは全く適合していなかった。図4は1977年1年間のデータを例にして負の二項分布の観測値と推定値を図示したものである。さらに，8年間の公表論文数の総数について同様の分析を行なってみると，負の二項分布だけが5％水準で適合していた（表4参照）。なお，ここで一方向のみの強化を前提としてつくられているプライスの蓄積的優位分布の適合性が棄却されたのは注目される。研究活動の過程では論文を公表しない場合においても強化がはたらくのである。

公表論文数	観測値(1975年)	推定値					
		ロトカ	サイモン	対数正規	対数級数	負の二項	ポアソン(参考)
0	35	−	−	−	−	39	12
1	29	53	52	11	38	26	29
2	16	13	15	25	16	18	33
3	16	6	7	18	9	12	25
4	11	3	4	12	6	9	15
5	2	2	2	7	4	6	7
6	6	1	2	4	3	4	3
7	2	1	1	4	2	3	1
8	1	1	1	2	2	2	0
9	1	1	1	1	1	2	0
10	1	1	0	1	1	1	0
11	1	0	0	1	1	1	0
12	1	0	0	1	1	1	0
13	1	0	0	0	1	0	0
14	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0
16	1	0	0	0	0	0	0
N (0を除く)	124 (88)	82	85	87	85	124	125
検定 (※)	5％水準	×	×	×	×	○	×
	1％水準	×	×	×	○	○	×

※ ○は仮説採択，×は仮説棄却をあらわす。

表2 科学の生産性の分布の適合性
（日本の化学者124人，1975年）

以上の結果から，日本の化学者の生産性の分布はラオやアリソンの研究結果と同様に負の二項分布に最も適合することが明らかになった。それでは，なぜ科学の生産性は負の二項分布に従うのであろうか。

負の二項分布は，前節で述べたことからも予想されるように，生起頻度が小さい事象において，(1)事象間に伝幡 (contagion) ないし学習や強化 (reinforcement) が働いている場合と，(2)母集団が非均一

な成分から構成されている場合，つまり個体間に異質性 (heterogeneity) が存在する場合の２通りの場合において生ずると考えられる。[21] 経験的な分布が負の二項型になっているときには，個体間の異質性という解釈と，事象間の伝幡の解釈という２つの解釈が可能なわけである。以後，両者の解釈に従うモデルをそれぞれ「強化モデル」，「異質性モデル」と呼ぶことにする。

		1972		1973		1974		1975		1976		1977		1978		1979	
公表論文数		観	推	観	推	観	推	観	推	観	推	観	推	観	推	観	推
0		40	46	39	40	36	43	35	38	33	42	39	47	39	39	43	48
1		24	26	23	26	33	25	29	26	27	22	29	24	25	25	33	27
2		23	16	26	18	24	17	16	18	21	15	21	15	17	17	16	17
3		16	11	12	12	6	12	16	12	10	11	13	10	12	12	13	11
4		9	8	6	8	6	8	11	9	11	8	4	7	7	9	4	7
5		4	5	4	6	5	6	2	6	3	6	4	5	7	6	8	5
6		1	6	2	7	2	7	6	7	3	8	1	7	7	8	2	5
7		4		3		4		2		3		4		3		1	
8		0	6	2	7	2	6	1	7	7	5	3	9	1	9	1	4
9		0		3		0		1		2		0		1		1	
10		0		2		3		1		1	7	2		0		0	
11		0		0		1		1		0		3		1		0	
12		0		1		0		1		0		0		2		0	
13		1		0		0		1		0		0		0		1	
14		0		0		2		0		1		0		1		0	
15		0		1		0		0		1		0		0		1	
16 − 20		2		0		0		1		0		0		1		0	
21 − 30		0		0		0		0		1		1		0		0	
平　均		2.0		2.2		2.2		2.3		2.7		2.3		2.5		1.8	
標準偏差		2.7		2.8		2.9		2.9		3.8		3.2		3.1		2.4	
パラ	p	0.27		0.29		0.27		0.29		0.19		0.23		0.26		0.31	
メータ	α	0.77		0.90		0.81		0.94		0.66		0.66		0.85		0.79	

表３　負の二項分布への適合性（各年１年間の論文公表数）

公表論文数		観測値	推定値
0		6	7
1		1	6
2		3	6
3		4	6
4		8	5
5		4	5
6		5	5
7 − 8		13	9
9 − 10		12	8
11 − 12		9	7
13 − 14		10	6
15 − 16		4	6
17 − 18		7	5
19 − 21		3	6
22 − 24		8	6
25 − 29		8	8
30 − 35		3	6
36 − 45		4	7
46 −		12	11
平　均		18.1	
標準偏差		18.8	
パラ	p	0.051	
メータ	α	0.98	

表４　負の二項分布への適合性（1972～1979年の論文公表数の合計）

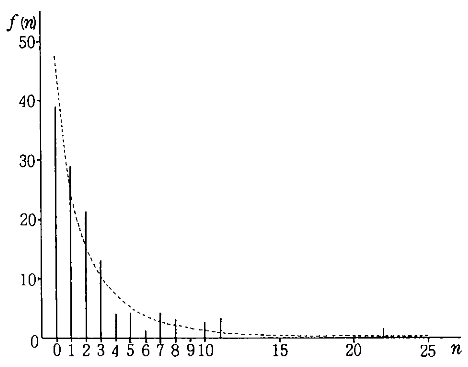


図４　負の二項分布の観測値と推定値（点線）（1977年）

さて，われわれの科学の生産性の分布はどのように解釈されるべきであろうか。まず強化モデルの方から考えてゆこう。このモデルは，ある人がある行為をしたあと次にその行為をする確率は変化する，ということを意味している。一度行為をしたあと，再びこの行為をする確率が高くなる場合と，低くなる場合とが考えられる。前者は学習とくにオペラント条件づけにおこる「正」に強化された状況，後者は「負」に強化された状況である。多くの社会的ないし個人的行動は学習の伝幡の形をとる。[22] これを科学者の行動に適用すると，科学者が論文を公表すればするだけそのあとに論文を公表する確率は高くなり，論文を公表しなければそのあとに論文を公表する確率は低くなる，ということができる。この命題は科学社会学にとって極めて重要な意味をもっている。それは科学社会学の創設者R．K．マートンによって概念化された科学社会学の中核的な理論である「科学におけるマタイ効果」[23] を数学的にいい換えたものにほかならないからである。

他方，異質性モデルはどうであろうか。すでに明らかになったように，ふつうひとりの科学者が一年間に公表する論文はせいぜい2篇強で，毎日の学問研究の努力が結実し，論文として公表に至る確率は高くはない。それは丁度，毎日自動車を運転していて，一歩誤れば交通事故をひきおこすことにさらされているが，安全運転に努める結果，事故の確率が低くなるのと同じことである。ポアソンは昔，プロシアの兵士が馬に蹴られて死亡する確率の分布を計算して，ポアソン分布を導いた。この例や，自動車事故のようにまれにしかおきない生起の確率の低いランダムな事象はポアソン分布に従う。したがって，もし，論文公表が全くランダムに生起するなら，公表論文の分布はポアソン分布に従うであろう。ところが，科学者はその能力や置かれている研究上の環境において多様である。大学院の博士講座の教授と大学院をもたない大学の教授や助教授では，研究費，研究設備，助手や大学院学生やその他の学生の数では大きなちがいがある。これらの条件は組織化された共同研究が主流の自然科学では学問生産に大きなちがいとなって現われる。このように，各科学者やその研究上の環境に異質性があり，個人によって論文公表頻度をあらわすパラメータが異なっているという状況は，科学者社会には極めて一般的な状況であり，科学の生産性の分布の説明を個体間の異質性に求めることは不合理ではない。これを証明するかのように，データはたしかに，個体間の生起頻度の均質性を前提とするポアソン分布ではなく，負の二項分布に従っていた。

3. 研究活動の過程における強化

前節において，科学の生産性の分布が負の二項分布に従い，科学者の学問研究のプロセスにマタイ効果と深い関連をもつ，伝幡ないし強化の現象がみられることを仮定した。

ここでは，その学問研究プロセスにおけるパブリケーション行動のさいの強化の存在を観察すべく，別の側面からアプローチしてみたい。もし強化がはたらいているとすれば，年とともに論文公表数のちらばりの程度は大きくなるであろうということが予想される。生産力のある者は益々多く生産するが，生産しない者は徐々に研究を行わなくなるであろうと判断されるからである。そこで，実際に年齢が高くなるにつれて生産性の不平等が拡大していくかどうか，を種々の不平等指標によって検討してゆくことにしよう。分析データは，ある年における各年齢層の学者の生産性のばらつきの程度をみ

るというようなクロスセクショナルなデータによる分析ではなく，1931年生れから1948年生れの者の集団からなる「準」コーホート集団の生産性の不平等度の推移をみるという方法である。

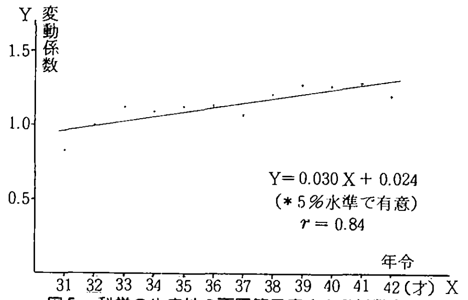


図5　科学の生産性の不平等尺度（変動係数）の年令による変化

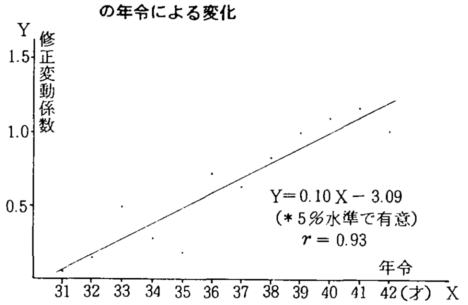
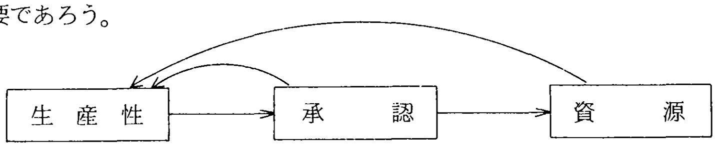


図6　科学の生産性の不平等尺度（修正変動係数）の年令による変化

図5と6は，そのサンプル集団に属する科学者の31才から42才までの論文公表数の不平等指数の変化で，図5は変動係数,[24] 図6は修正変動係数の数値をプロットしたものである。両者とも，年齢が増大するにつれて大きくなっている。すなわち，年齢が増すとともに論文公表数の不平等は拡大している。統計的にみても，カイ2乗検定の結果，直線の勾配の大きさをあらわす回帰式の偏回帰係数は5％水準で，0であるとはいえない。つまり，この場合は直線の傾きは正である。このことは，科学論文の生産プロセスに，正の強化が働いていることを示している。

時間とともに増大する公表論文数の不平等は，科学社会学の報賞体系論のなかでは，どのように解釈されるであろうか。アリソンらは，強化の過程は承認 (recognition) と資源という媒介変数を含むフィードバックループからなるものと考えている。第一に，重要な学問の発展に寄与したとして承認をうけた科学者は，さらに論文を公表することによってその承認を維持し増大するように動機づけられる。第二に，承認によって，研究を促進する資源への接近の可能性が増大するとともに，生産性に対しては正の効果をもつ。換言すれば，「公表は承認をもたらす。承認は動機づけを拡大し，より大きな研究資源をもたらす。両者は次の公表の可能性を増大させる」となる。[25] これを図示すると，下図のようになる。なお，本論文では承認や資源に関する変数は考慮されていないが，マートン学派の報賞体系論を厳密に検証するためには，それらの変数をくみこんで分析することが必要であろう。



4.　科学者間の異質性

科学の生産性の分布が負の二項分布に従うときに想定される2つの可能な解釈のうち，のこる個体

の異質性の仮説をやや詳しく検討しよう。全国の国公私立大学理学部化学科と研究所の教員から抽出された124のサンプルは，それぞれその勤務する大学の特性や各科学者個人の属性において異質である。これらの異質性はどの程度，科学の生産性の違いとなってあらわれてくるのであろうか。ここではそれを明らかにするために，異質性を構成すると想定される多くの変数のうち，性，年令，出身大学のタイプ，博士学位所有の有無，地位，下位専攻分野の6種類の個人的属性，勤務する大学のタイプ（大学院博士課程や修士課程の有無），所在地域，設置者，所属学科の教員の学位所有率の4つの研究環境上の特性の合計10個の変数によって，科学の生産性の大きさとちらばりがどの程度説明され，さらに科学の生産性を規定する大きな要因としてはどのようなものがあるかを分析することにする。各説明変数はほとんどが属性変数であるので，林の数量化理論Ⅰ類を用いる。外的規準変数として用いるものは1972年から1979年までの8年間の論文公表数である。

表5はその分析結果である。重相関係数は0.674となっており，10個の変数で1972年から1979年までの8年間の124人の論文公表数の変動の約67.4％を説明している。残る32.6％はこの10変数以外の要因に求められることになる。科学の生産性に影響を及ぼす最大の要因は，科学者の大学における地位であり（レンジ27.23），教授は助教授より，助教授は講師よりも，論文公表の頻度が高い。この地位による論文公表頻度の格差は主として研究組織におけるヒエラルキーを反映していると考えられる。教授は研究室の主任として共同研究を指揮する立場にあり，そこにおいていくつもに分業された研究の成果を公表する際には，各論文の著者の一人として名を連ねるのが普通であるからである。ショックレーはこのような現象から「組織仮説」(organizational hypothesis)' を提案した。[26] 彼は結局，他のデータを考察するなかでこの仮説を放棄したが，我々のデータではこの仮説は妥当するように思われる。1972～1978年の教授の一人あたり一年間の平均論文公表数は3.0，助教授では1.8，講師では1.0と地位による格差は大きい。ショ

変数	カテゴリー	スコア	レンジ	順位	サンプル
性	男	－0.03	1.33	10	121
	女	1.30			3
年令	60～	－7.35	18.41	4	11
	50～59	－5.72			41
	40～50	1.81			52
	～40	11.06			20
出身大学	東大・京大	2.91	5.70	7	54
	旧帝大・官大	－2.79			51
	その他	－0.80			19
学位	博士	0.44	7.80	6	117
	修士・学士	－7.36			7
地位	教授	8.79	27.23	1	63
	助教授	－7.24			51
	講師	－18.44			10
専攻分野	物化・構造	－0.93	12.31	5	40
	有機	2.34			15
	天然物有機	7.03			13
	分析・地球	5.37			14
	無機・錯塩・放射	－5.28			22
	その他	－2.50			20
設置者	国立	6.17	18.60	3	80
	公立	－1.80			5
	私立	－12.43			39
地域	東京	3.14	4.23	8	32
	東京以外	－1.09			92
所学有位率	～80％	1.55	2.00	9	28
	～100％	－0.45			96
勤務大学	博士	6.01	20.95	2	74
	修士	－13.80			28
	学士	－14.09			10
	研究所	6.86			12
重相関係数		0.674			124

表5 科学の生産性の規定要因

ックレーが組織仮説を棄てた理由の一つは，調査した科学者の勤務する研究所においては，高生産者は単独論文の割合が高く，多量の公表論文数のかなりの部分は単独論文によるものである，という事実であった。ところが我々のデータでは，公表論文総数に占める単独論文数の割合は教授で 7.5 %，助教授・講師で 5.1 %と共に低く，両者とも論文のほとんどが共著論文である。さらに，高生産者を調べてみたが，彼らがとくに単独論文の割合が高いともいえなかった。このような事実をみれば，やはり地位による効果は否定できないように思われる。組織化され，分業化されている現代の科学研究においては，研究組織の長の影響力は大きく，組織上の地位はおのずから公表論文数にも反映されるであろう。この意味で「組織仮説」は地位による科学者の異質性を表現する概念であるといえる。

次に大きな要因は，勤務する大学の大学院の種類である。大学院博士課程の付置されている化学科と研究所に勤務する化学者のスコアはそれぞれ6.01と6.86であるのに対し，修士課程および学士課程のみの化学科のそれは－13.80および－14.09と大きな格差があり，2つのグループに分れている。すなわち，博士講座と研究所はそれぞれ同等の論文生産能力をもっており，修士講座と学部のみの化学科もほぼ同等であるという二極構造になっている。これは博士講座や研究所には高い論文生産能力をもつ科学者が配置されていることのほかに，研究費や研究設備，それに大学院学生などの人的資源に恵まれているからでもあろう。現代の化学の研究は分析機器に依存するところが大であるが，ある日本の化学者は次のように述べている。（従来の有機化学的方法よりも）「機器分析を使う方がより速く，より正確にできるとなると，また機器以外では知りえないとなると，構造決定のやり方はこれらの機器を持っている人，持っていないまでも，使える人しかやれなくなってしまう。」[27]「それに，千万円の桁の値段のする高級な分析機器を何台も何種類もそろえられる所はごく限られた所となり，小企業の研究者，地方大学で研究者が1～2名しかいない所では，機器分析とは縁のない仕事でも探さねばならない。それは飛行機・戦車なしで戦争するようなもので，苦しい困難なたたかいである。………むしろ年毎に分析機器の性能が高度化し，大型になり価格が高価になるにつれ，大きな研究所，一部の大学に大型分析機器がますます集中，較差は一層ひろがっている。」[28] 従来なら何年もかかってやっていたような分析を数週間でやってしまうような分析機器からは短期間にぼう大なデータが生み出される。このような機器をもつ研究室からは多数の論文が生まれることになる。先の研究所や博士講座のグループと修士講座や学士課程のみの学科のグループの間に存在する論文生産力の二極構造は研究資源の集中と独占によるところが大きいと思われる。

論文公表数に影響を与えている第3に大きな要因は勤務大学の設置者によるちがいである。国立大学勤務者は公立大学勤務者よりも，そして公立大学勤務者は私立大学勤務者よりも論文を多く生産するという結果である。とくに国立大学と私立大学では教員一人あたりの学生数に大きな違いがあり，教育負担の差が研究に影響を与えているということもその原因のひとつと考えられる。しかし，それにしてもスコアの差は大きい。

第4に大きな要因は年令である。ここでは純粋に年齢のもつ影響の大きさをあらわしていることを考慮して考えるべきなのであるが，年齢が低いほど論文生産能力が高いという結果になっている。第5位は下位専攻分野のちがいである。天然物有機化学と有機化学では比較的多くの論文が出やすく，無機・錯塩・放射化学ではそうでもないようである。以下，博士学位の有無，出身大学のタイプ，大

学の所在地域，当該学科の教員の博士号所有率，性の順に影響力が大きい。上位の3つの要因だけで同様の分析を行ったら，重相関係数は 0.59となった。地位，勤務大学のタイプ，設置者の3つだけで，科学の生産性の変動の60%近くを説明することができるのである。この3つが日本の大学の化学者の生産性を規定する3大要因であるといえる。

これまでに述べてきたように，日本の化学者124人の生産性の分布が負の二項分布に従うという事実は，科学論文の生産プロセスにおいて強化がはたらいているということと，一人ひとりの科学者のあいだに置かれた研究条件のちがいを反映した異質性が存在することの2つに原因を求めることができた。この2つの解釈は，科学社会学理論からみても受け入れ可能な解釈であるように思われる。しかし，これまでの分析だけでは，この2つの要因のうちどちらがより正しい解釈であるのか，あるいは，どちらも同等に評価されるべきかについては必ずしも明らかではない。しかもこの2つの要因は以下に示すように，相互に関連しているからなお面倒である。つまり，科学者の異質性は個人間の科学の生産性における不平等（不均一性）を生むが，この差異はすぐれた業績を達成した者を研究条件のすぐれた大学への再配分（＝移動）をもたらすと共に，次の論文生産への動機づけを強化し，そうでない者に対しては逆の強化をすることによって，科学者間の異質性をさらに増大するからである。

そこで，前節で明らかにしたように，科学者の生産性は彼の置かれている研究条件，なかんずく大学院のタイプと地位によって大きな違いが生まれることから，このような研究条件の等しい科学者のグループについて生産性の分布を調べてみることにする。つまり，異質性の要因をできる限り排除した場合に，なお生産性の分布が負の二項分布に従うかどうかを分析するのである。サンプルの大きさの関係から，博士課程の設置されている学科に勤務している教授（N＝39）についてのみ調べてみたが，その結果，表6と7に示すように1年間単位のデータでは1972年と1979年のデータを除く，1973, 74, 75, 76, 77, 78 年の6つについてと，1972－79年の総計データについては生産性の分布は負の二項分布に従っていた（5%レベル）。

公表論文数	1972 観	1972 推	1973 観	1973 推	1974 観	1974 推	1975 観	1975 推	1976 観	1976 推	1977 観	1977 推	1978 観	1978 推	1979 観	1979 推
0	6	8	8	9	7	6	5	6	6	6	6	7	6	6	5	9
1	5	7	9	7	7	6	7	7	8	6	9	7	6	6	14	8
2	8	6	7	6	7	6	6	6	8	6	7	6	5	6	6	6
3	7	8	2	8	1	9	8	5	1	9	4	8	7	8	4	8
4	7		3		2		3	7	3		2		4		0	
5	2		2		5	5	1		2	5	4		2	5	5	
6	1		2		2		4		1		0		3		1	
7	1		3		2		1		3		0		2		1	
8	0	10	0	9	2		0		3		1	11	0		1	9
9	0		1		0	7	1	8	2	7	0		1	7	0	
10	0		1		3		1		1		2		0		0	
11 − 15	0		1		1		2		1		3		2		2	
16 − 20	2		0		0		0		0		0		1		0	
平均	3.2		3.0		3.7		3.4		3.7		3.3		3.7		2.8	
標準偏差	3.5		3.3		3.5		3.1		3.5		3.4		3.6		3.1	
パラメータ p	0.26		0.28		0.30		0.35		0.30		0.29		0.28		0.29	
パラメータ α	1.13		1.18		1.60		1.87		1.53		1.33		1.44		1.16	

表6 同等の研究条件件にある化学者（博士課程・教授）の主産性の分布（各年 N＝39）

公表論文数	観測値	推定値
0	1	
1	0	
2	0	
3	0	5
4	1	
5	1	
6	1	
7	1	
8	3	
9	1	9
10	1	
11 − 15	4	
16 − 20	4	8
21 − 25	7	
26 − 30	2	6
31 − 35	2	
36 − 50	4	5
51 −	6	5
平均	27.5	
標準偏差	22.5	
パラメータ p	0.054	
パラメータ α	1.57	

表7 同等の研究条件にある化学者の生産性の分布（1972 — 1979）

5. 科学社会学理論との関連性

このことから明らかなことは，われわれは少くとも科学論文の生産プロセスにおいて強化がはたらいているということを否定できない，ということである。したがって，マートンの「マタイ効果」理論は日本の科学者の研究活動のメカニズムのある部分を説明できると結論される。

科学者が論文を公表すればするだけ，そのあとに論文を公表する確率は高くなり，論文を公表しなければ，そのあとに論文を公表する確率は低くなる，負の二項分布から導出される「強化モデル」の解釈は科学社会学にとって極めて重要な意味をもっている。なぜならば，それはマートンによって建設された科学社会学理論の中核を構成する「科学におけるマタイ効果」を確率モデルという純粋に数学レベルにおいて記述するものであるからである。また，日本の科学者の実際のデータがその確率モデルの正当性を裏づけたことは，ひるがえってマートン理論の現実的妥当性とそれが日本の学界の社会構造をも説明しうるという普遍性を意味している。

ただし，若干の留保が必要である。それはわれわれがここで使用したデータが科学の「生産性」つまり科学者の論文公表数であるからである。マートンのマタイ効果の本質はつづめていえば，「顕著な評判を獲得している科学者に対しての特定の科学上の貢献に対する承認の増大と，まだ評価を獲得していない科学者からのそのような承認の留保」[29] であるといえる。これはマートンがノーベル賞受賞者をインタビューするなかで，科学者の社会に，新約聖書のマタイによる福音書に出てくる「持つ者には与えられ，豊かになるが，持たざる者からは彼が持っているものさえも奪い取られる」という聖マタイのことばがよく符合するところからつけられたものである。[30] そもそも，マートンはマタイ効果を科学の報賞体系 (reward system of science)，コミュニケーション体系および科学上の資源の配分の側面についてのみ言及していた。しかしその弟子のJ．コール＆S．コールは，これを科学の生産性の領域にまで拡張して，「累積的優位」(accumulative advantage) の概念を提出したという事情がある。[31] したがって，厳密にいえば，マートン科学社会学理論の中核部分まで完全に検証したとはいえない。そこに達するには，論文公表数のほかに，承認 (recognition) という変数が同様に分析され，検証されねばならない。これは今後に残された課題である。

いづれにせよ，小論において負の二項分布がマートン理論と現実のデータの間のリンクの役割をはたすことを確認し，マートン学派の科学社会学理論が，日本の化学者の社会構造の一端を説明することが可能であるということを示したことは確かである。それはまた，アメリカの学界の社会構造と日本の学界の社会構造の類似性を示唆しているとともに，マートンの科学社会学理論の普遍性を証明している，ともいえるであろう。

〔注〕

1) A. J. Lotka, "The frequency distribution of scientific productivity" *Journal of the Washington Academy of Sciences,* 16 (June 19), 1926.
2) D. J. de Solla Price, *Little Science, Big Science,* Columbia Univ. Press, 1963.
（D．J．プライス，島尾永康訳 「リトルサイエンス・ビッグサイエンス」，創元社，1969， 52～75ページ）

3) D. J. プライス 上掲書 57ページ

4) H. A. Simon, "On a Class of Skew Distribution Functions", in; *Models of Man,* John Wiley & Sons, 1957. (H. A.サイモン，宮沢光一監訳 「人間行動のモデル」同文館，1970年，241～262ページ)

5) H. A. サイモン 上掲書 262ページ

6) D. J. Price "A General Theory of Bibliometric and Other Cumulative Advantage Processes", *Journal of the American Society for Information Science,* Vol. 27, Number 5, 1976, P.292.
プライスはこの論文で，科学の生産性の分布は負の二項分布にはならない，といっている。

7) 富む者はますます富み，次の論文公表が容易になる，という意味で「蓄積的優位」という名が付けられているが，これはマートンのマタイ効果の概念を拡張してコール兄弟が1970年代のはじめに作りだしたAccumulative advantage とほぼ類似の概念である。
(J. Cole & S. Cole, *Social Stratification in Science,* Univ. of Chicago Press, 1973, p.72–75)

8) I. K. R. Rao "The Distribution of Scientific Productivity and Social Change" *Journal of the American Society for Information Science,* Vol. 31, 1980 (March), p.121.

9) C. B. Williams "The Number of Publications Written by Biologists" *Annals of Eugenics,* Vol. 12, 1944, p.143–146.

10) 国沢清典「確率論とその応用」，岩波全書，52ページ参照。本論中の図3において $\lambda=1$ の場合がこれである。

11) x と n_1 を求めるための計算は面倒である。s と n から容易に x と n_1 を求める数表が木元新作氏によって作成されている。(木元新作 「動物群集研究法Ⅰ — 多様性と種類組成 —」(生態学研究法講座14)，共立出版，1976年，168ページ参照。)

12) A. J. Lotka, op. cit. p.318.
なお，以下の計算では，ベータ関数の値は $B(x,\ y)=\Gamma(x)\Gamma(y)/\Gamma(x+y)$ およびスターリングの公式 $\Gamma(x)=\sqrt{2\pi}\,x^{x-\frac{1}{2}}\,exp\left(-x+\frac{1}{12x}-\frac{1}{288x^3}\right)$ によって計算している。ここで，$\Gamma(x)$はガンマ関数である。

13) W. Shockley "On the statistics of individual variations of productivity in research laboratories", *Proceedings of the Institute of Radio Engineers,* Vol. 45, 1957 (March), p. 283

14) P. Allison, *Processes of Stratification in Science,* Arno Press, 1980.

15) I. K. R. Rao, op. cit. p. 111

16) この図は河田・鍋谷「大学演習 数理統計」，裳華房，1962年，P 118からの転載である。なお，2つのパラメータPと α の推定式は，$\hat{p}=\frac{1}{s_k^2}$，$\hat{\alpha}=\frac{(\bar{k})^2}{s_k^2-\bar{k}}$ である。(P. Allison, op. cit., p.91)

17) 国沢清典，上掲書，50～53ページ

18) 石井吾郎「数理統計入門」，培風館，1968年，17～18ページ および国沢清典，上掲書，52ページ

19) 抽出台帳としては，広潤社論「全国大学職員録」(昭和55年版) を使用した。各クラスターから抽出したサンプルは次の通りである。

		国立		公立		私立		計
		教授	助教授・講師	教授	助教授・講師	教授	助教授・講師	
大学のタイプ	博士	18	20	2	3	19	12	74
	修士	12	12	−	−	2	2	28
	学士	3	3	−	−	1	3	10
	研究所	6	6	−	−	−	−	12
計		39	41	2	3	22	17	124
		80		5		39		

20）1979年の平均値は小さいのは，1979年に公表された論文の一部が筆者のサーチした期間（1972 － 1979）以後のケミカルアブストラクトに掲載されているためである。

21）J. S. Coleman, *Introduction to Mathematical Sociology,* Free Press, 1964, p.301およびP. Allison op. cit. p.134.

22）J. S. Coleman, ditto, p.299.

23）本論文18ページを参照。

24）変動係数は s/μ，修正変動係数は $s^2/(\mu^2-s)$ である。後者は scale invariant な尺度である点ですぐれており，アリソンが積極的に提案している（P. Allison "Inequality and Scientific Productivity", *Social Studies of Science,* Vol. 10, 1980). ここで s は標準偏差，μ は平均である。

25）P. Allison and J. Stewart, op. cit., p.603–604.

26）W. Shockley, op. cit., p.284.

27）中川直哉「機器分析」，井本・大沼・道家・中川編『化学のすすめ』（第二版），筑摩書房，1980年，170ページ

28）中川直哉，上掲書，193ページ

29）R. K. Merton "The Matthew Effect in Science", (1968), in; R. K. Merton, *Sociology of Science,* The University of Chicago Press, 1973, p.446.

30）R. K. Merton, ditto, p.445.

31）J. Cole and S. Sole, *Social Stratification in Science,* The University of Chicago Press, 1973, p.72. P. Allison, *Processes of Stratification in Science,* Arno Press, 1980, p.135.

Distribution of Scientific Productivity and the Processes of Stratification

HIROTOSHI YAMASAKI*

This paper reports that the distribution of productivity of 124 university chemists in Japan shows the best fittness to a negative binomial distribution, and then considers the reasons for and the sociological implications of the results.

Since the inverse square law of A.J. Lotka (1926), several mathematical models on the distribution of productivity have been proposed by Williams (1944), Simon (1955), Shockley (1957), Price (1963, 1976), Allison (1976) and Rao (1980) et al. The characteristics of these models were examined in comparative perspective. The negative binominal distribution showed the best fittness to our data among these models. This result proposes the hypotheses – reinforcement in the process of research activity and heterogeneity among each scientist. It is difficult, however, to judge which hypothesis is more appropriate, mainly because both models correlate with each other. Heterogeneity of the ability and the socialization process of each scientist causes an inequality of productivity among them. And this inequality reallocates the productive scientists to research oriented-universities and strengthens their motivation to the further research on the one hand, and weakens the motivation of less productive scientists. This process increases the differences of productivity among scientists.

However these difficulties were solved by the two findings in that

(i) the distribution of the productivity in the subsample of full professors in graduate schools with doctoral programs (N = 39) shows a good fittness to the negative binomial distribution, and

(ii) the coefficient of variation of productivity increases as the age of scientists increases.

From these two facts, we can accept the reinforcement hypothesis at least. This implies that the more the scientist publishes, the more the probability to publish later increases, while the less he publishes, the more the probability decreases. This hypothesis has a significant meaning for the theory of sociology of science, because "reinforcement model" describe the Merton's Matthew Effect on the mathematical level. Our result also confirms the empirical validity and international universality of the Mertonian theory of sociology of science.

* Research Assistant, R.I.H.E.