

Title	科研費採択の統計解析による政策意義に関する予備研究
Author(s)	細坪, 護拳
Citation	年次学術大会講演要旨集, 26: 659-664
Issue Date	2011-10-15
Type	Conference Paper
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/10205">http://hdl.handle.net/10119/10205</a>
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

# 科研費採択の統計解析による政策意義に関する予備研究

○細坪 護孝（文部科学省 科学技術政策研究所）

## 1. はじめに

本研究では、国公立大学教員の不完全パネルデータより構成されるデータベース[1][2][3][4][5](Higher-education teachers' Mobility Data Base:HM-DB)に対して科研費採択情報を接続し、科研費採択数と国公立大学教員数の統計解析を行い、その構造解明を目指す。

科学研究生産性に関する統計学的検討は戦前から歴史があり、A.J.Lotka[6]は論文数と学者数の関係を示した。これにWilliams[7], Simon[8], Shockley[9], Price[10][11], Allison[12]らは統計的改良モデルを提案し、Rao[13], 山崎[14]は負の二項分布の適合を示した。

本研究は、論文の研究資金である科研費採択数と国公立大学教員数の関係究明のため予備的統計解析を行う。

## 2. HM-DB と科研費採択データ

HM-DBと科研費採択データ間の接続には、相互のデータ同定が必要である。両者の共有データは教員氏名、所属大学、掲載時点で留まる。科研費採択側から同定能力の高い科研費番号は、HM-DBには必ずしも掲載がなく、一方HM-DBが有する学歴や年齢は、科研費採択データにはほとんど関係ない。年齢に限り、「若手研究領域での課題採択が不自然か否か」等の消極的同定能力を持つ。

HM-DBの観測年88,91,94,97,00,03,05,06年の8時点に接続処理した。この際、工程数とデータ数節約のため、接続科研費を基本的に当概年に採択・継続中の案件とした(但し、同一課題が2期間に跨ることはしない)ため、採択課題が確率的に抜ける可能性はある。国立情報学研究所(NII)のKAKEN(科学研究費補助金データベース)が現在の形に洗練される以前に、データを入手して、手作業等でデータクレンジングしながら上記作業を行ったため、データの精度に甘さは残る。全体的傾向を把握できる水準にはあると考えられるが、本点の検証も必要である。国公立大学や分野別の採択件数であれば、毎年出版される「文部科学省科学研究費補助金採択課題・公募審査要覧」(ぎょうせい)や文部科学省、日本学術振興会(JSPS)などの公開資料で明らかである。KAKENを用いれば、大学群別等でも仮説検定を行うことができよう。今後補完したい。

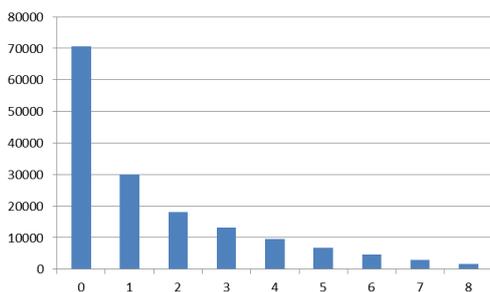


図1 科研費採択数に対する国公立大学教員数の観測値

上記8時点の科研費採択数に対する国公立大学教員数の観測値は図1となる。事前の想定を整理すると、

- 1) 観測時点19年間、競争的資金重点化により科研費の性格が変化した。
- 2) HM-DBは不完全パネルデータだが、接続時点数と採択数間の関係は不自然かもしれない。
- 3) 大学教員全員が科研費に申請するインセンティブを持つと限らない。分野等では生涯申請しないかもしれない。

以上の点は論文数とは異なる側面もあるものの、先行研究[6]-[14]等では十分には議論されていない。しかし、「論文を執筆しないと決めていた」教員の存在は否定できない。

(例:分野依存性、昇格インセンティブに乏しい教授職等)

例えば、3)に関連しては、図1から採択数0件が非常に多い。つまり、採否だけでなく、「そもそも科研費に申請したか否か」を調べる必要がある。HM-DBやKAKENでも申請問題は解明されない。以下の2点から不可能である。

- ・個人情報保護: 科研費申請情報は基本的に当該申請目的以外には使用されない。但し、採択課題は別。
- ・個人情報保護の課題をクリアしても、多くの大学で科研費申請を半ば義務付ける現状では、毎年同じ数行の申請書類を出す等のlatent(潜在的な)不申請行為等は観測不可能である。なお、これは1)も関係する。

具体的に、科研費採択は統計学的モデルで決まるとする。すると、「決定的に科研費申請する必要がない・できない・する気がない教員」と「確率的に科研費申請する教員」に分けられる。前者を構造的0と置き、教員に応じて申請か否か確率が異なるとする。採択数と接続時点しか分からない現状で本推定は不可能だが、先行研究成果を元に分布を仮定できたら、推定できるかもしれない。図1の分布は、申請者数分布と科研費採択数分布の混合分布と想定する。本分布を推定するため、教員の接続時点別の科研費採択数を調べた(図2)。

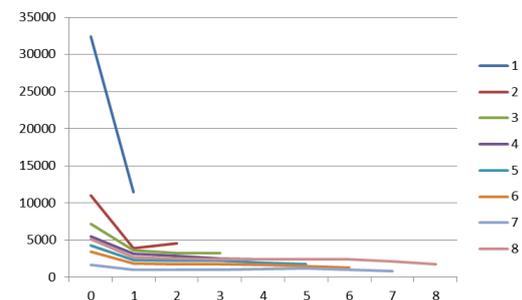


図2 大学教員の接続時点数に対する科研費採択数

科研費採否が接続時点数に依存しないとし、不完全パネルデータ時点数は観測時点に応じて確率的とすると、近似的に、科研費採否は二項分布に従う。確率 $p$ は接続時点に不変のはずである。二項分布の期待値 $E(X)=np$ ( $n$ : 試行回数)から、各時点の科研費採択確率 $p=E(X)/n$ と推定される(図3)。 $p$ の推定値は

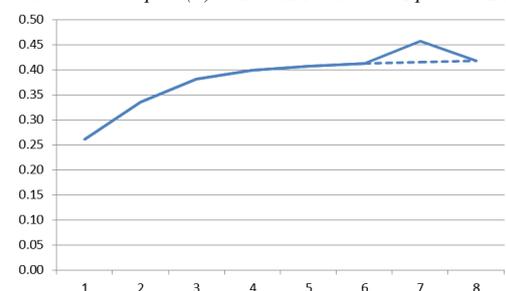


図3 接続時点数に対する科研費採否二項確率 $p$ の推定値(点線は線形補間を示す)

接続時点数に依存する。時点数7で外れ値を示すが、観測時点の時間間隔が等しくない(05年)ためと考えられる。時点数6と8とで線形補間したデータが点線である。点線と他の実線は滑らかな曲線を示す。大学教員の在任期間が長いほど、科研費採択され

やすく( $p=0.42$ :8 時点)、短期間に任教員の確率は低い( $p=0.26$ :1 時点)。観測は 06 年以前なので、任期付教員や非常勤講師は今ほどおらず、データ掲載元の全国大学職員録でも非常勤教員数は極めて少ない。即ち、これは主に切断データの附随要因、具体的には、採用直後の教員が業務に不慣れな一方、停年退職前の教員が申請を控えた可能性が考えられる。また、専門職の性格の強い学部等では、在任期間がパネルの切断とは異なる意味で不連続になる(例:医学部教員の地域病院等への一定期間の意向)。その場合、将来の研究時間の確保が不透明なため、研究能力があっても科研費に申請しない可能性もある。

図 3 の曲線(点線)は、採択数に対する 1・2 次階差は単調減少を示し、一定確率に漸近するように見える。一方、図 2 から、採択数は二項分布に従っていない。

### 3. 確率分布適合の試行

データに適合する他の確率分布を調べる。適合方法として、カイ二乗最小化法、最尤推定法、古典的モーメント法を使用する。

適合分布は、二項確率の 1 回試行確率  $p \rightarrow 0$ 、試行回数  $n \rightarrow \infty$  とした極限であるポアソン分布である。本分布では科研費採択を宝くじ当選のようなものと仮定する。適合度検定の結果でも、全ての接続時点数、カイ二乗最小化法・最尤推定法いずれの場合でも、ポアソン分布に適合しない。そこで、先行研究[13][14]にある負の二項分布を考える。負の二項分布の定義では、統計的に独立なベルヌーイ試行を行ったとき、 $r$  回の「成功」を得るのに必要な試行回数の分布の方を使う。結局、P 値はいずれの分布のどの時点でも 0.000 となり、適合しない。

先行研究の論文数分布に関する研究では、負の二項分布適合が確認された[13][14]。本研究の仮説は、

1. 研究者数と論文数のいずれも確率的要因が支配してきた可能性: 人意の介在はあっても、それは全体的には特定の研究者や論文に偏らないかもしれない。
2. 標本数が少ないため内生的因子の影響が顕在化せず、適合した可能性[13][14]: 実証的に標本数が増えるほど、内生因子の影響により適合が難しくなる一方、適合実現は論拠を更に強固にする。無論、先行研究の時期に KAKEN のような良質 DB は存在しない。

	1	2	3	4	5	6	7	8	全部
ポアソン分布									
モーメント法	13724	6798	1496	970	1729	755	796	1781	50494
カイ二乗法	0	2081	2490	3723	5745	7726	6336	29557	162933
最尤推定法	3284	4534	4593	5250	6337	7068	5204	19440	99529
自由度	0	1	2	3	4	5	6	7	7
負の二項分布									
モーメント法	-	6833	1051	1323	885	708	464	921	80228
自由度	1	2	3	4	5	6	7	8	8
カイ二乗法	-	-	-	-	-	1456	-	3900	4027
最尤推定法	-	-	-	-	-	-	-	-	8100
自由度	-	-	-	-	-	4	-	6	6

表 1 接続時点数別の分布のあてはめのカイ二乗統計量  
(- は計算不可能)

分析を進めるには以下の方法が考えられる。

1. ポアソン分布・負の二項分布を拡張した確率分布に対して、あてはめを行い、意味を解釈する。
2. 科研費採択者-不採択を応答変数、当該者の属性を説明変数とし、応答変数はポアソン分布・負の二項分布に従うと仮定。最尤推定法により係数を推定し、どのモデルが的確にデータを説明し、かつどの説明変数が効いているのか調べる。と考えられる。様々な方法で 1. を試した(詳細は付録参照)。例えば、接続時点を統合したデータに対して、モーメント法により負の二項分布をあてはめた(図 4)。適合度検定では P 値=0.000 で、適合の帰無仮説は棄却された。本パラメータ  $r$  の推定値は 0.708、

成功確率パラメータ  $p$  の推定値は 0.304 であり、推定には R[15] の goodfit 関数を用いた。これらは拡張分布に対する数値計算の初期値の役割を果たす。

解釈上の一つの問題は、負の二項分布で表現できない図 4 中のズレの意味の解明にある。比較的大きな接続時点には一貫性が見られることから、このズレは少数接続数で大きな効果を持つのだろう。これらが統計学的モデルとして表現できれば、本研究の前進となる。

しかし、より自由な拡張分布への適合は、一般的アプローチとは言い難い。本研究が考える適合手法(G 最小化法)を導入しても、適合度は十分とはいえない。理由は、

1. どのように分布が拡張されればより良い適合が得られるのか事前に知らないと網羅的試行の繰り返しになる。
2. 1. を事前を知るには、核心となる統計的メカニズムについてある程度目星が必要である。仮に、科研費採択が負の二項分布としても、適切な分布を推定し(直観的なアタリとしてジブ分布などが考えられる)、かつ分布パラメータも推定する必要がある。

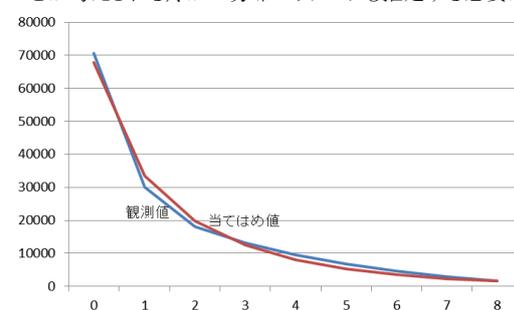


図 4 全科研費採択数に対する負の二項分布のあてはめ

ポアソン分布・負の二項分布の拡張分布への適合は今後の課題である。

### 4. ポアソンモデル・負の二項モデルへの回帰分析

以降、本稿では、前述の 2. に沿って、従属変数を科研費採択件数として、負の二項分布・ポアソン分布の回帰モデルを検討する。HM-DB は不完全パネルデータで、観測年数を offset 項とする[16]。教員に関する観察可能な因子を整理した。

- 1) **観察年間**: 観察開始年・終了年の組み合わせ
- 2) **最近大学カテゴリ**: 国立大学 4 群[1][2][3][4][5]、公立大学、私立大学、大学校等の 7 種
- 3) **世代カテゴリ**: 一般的とされる、大正・昭和・一桁・焼跡・団塊・しらせ・バブル・氷河期の 7 分類
- 4) **最近大学の立地地域カテゴリ**: 北海道・東北・関東/中部/近畿/中国・四国/九州の 6 地域
- 5) **最近職位**: 教授・准(助)教授・講師・助教(助手)・教務職員/医局員等の 5 職位
- 6) **最終学歴**: 博士・修士・学士等の 3 種
- 7) **出身大学カテゴリ**: 自校出身、国立大学 4 群(2)と同じ、公立大学、私立大学、大学校等、外国大学の 9 種
- 8) **出身大学地域カテゴリ**: 北海道・東北・関東/中部/近畿/中国・四国/九州と外国の 7 地域
- 9) **分野カテゴリ**: 総合領域、複合新領域、人文学、社会科学、数学、物理系科学、化学、工学、生物学、農学、医歯薬学の 11 分野
- 10) **異動回数**: 大学間の異動回数。スカラー
- 11) **昇格回数**: スカラー
- 12) **昇格パス**: 観察最初・最後の職位の組み合わせ

推定式は上記を主効果とする一次線形モデルとし、AIC を用いた変数増減法(ステップワイズ法)で最適モデルを選んだ。デー

タ量や説明変数のカテゴリーや水準数が多く、交互作用項は入れられない。

ポアソン分布・負の二項分布でもAICは当初の線形モデルから説明変数を全く減少させなかった。これは、交互作用項や未導入の説明変数の存在可能性を示唆する。

採択モデルの観測データ説明状況を検討するために逸脱度分析[17]を行う。逸脱度分析には主に4種類の計算方法があり、ここではType IとIIの仮説検定を実施した。Type Iの逸脱度分析は計算で除去する変数の順序に結果が依存してしまうことがある。ポアソン分布、負の二項分布でもType Iの逸脱度分析では昇格回数の有意差が比較的顕著でなかった(ポアソン分布:P=0.018, 負の二項分布:P=0.033)。しかし、除去順序に依存しないType IIの逸脱度分析では全ての変数に有意差が認められた(表2)。逸脱度分析はデータ数に対して説明変数が過剰傾向の場合により有効と思われる。

次に、over-dispersion(過分散)を調べる。ポアソン分布では理論的に期待値と分散値が同じで、再現性の判断に特に重要となる。過分散の明確な基準はない[17][18]が、残差逸脱度と残差自由度の比、より正確には、ピアソンのカイ二乗統計量と残差自由度の比( $\phi$ : dispersion parameter)が1.0をある程度超えれば過分散となる。

$$\phi_{poisson} = \frac{162707.1}{157654} = 1.032051$$

となり、ポアソン分布モデルでも再現されるように考えられる。

	ポアソン分布モデル			負の二項分布モデル		
	逸脱度の増加量	自由度	カイ二乗検定のP値	逸脱度の増加量	自由度	カイ二乗検定のP値
1) 観察年間	6769.6	35	0.000	6140.4	35	0.000
2) 最近大学カテゴリー	7684.6	6	0.000	6555.5	6	0.000
3) 世代カテゴリー	2838.3	7	0.000	2574.5	7	0.000
4) 最近大学地域カテゴリー	117.9	5	0.000	109.5	5	0.000
5) 最近職位	134.3	6	0.000	113.9	6	0.000
6) 最終学歴	6633.6	3	0.000	5887	3	0.000
7) 出身大学カテゴリー	1115.9	9	0.000	968.1	9	0.000
8) 出身大学地域カテゴリー	114.9	7	0.000	97.1	7	0.000
9) 分野カテゴリー	9858.3	11	0.000	9056.8	11	0.000
10) 異動回数	352.1	1	0.000	324.2	1	0.000
11) 昇格回数	76.2	1	0.000	60.8	1	0.000
12) 昇格パス	823.4	21	0.000	693.8	21	0.000

表2 TYPE II 逸脱度分析結果

こうして、応答変数をポアソン分布でも負の二項分布としても、推定できそうである。それぞれの係数推定値が表3である。

	切片	ポアソン分布モデル			負の二項分布モデル		
		推定値	標準誤差	z値のP値	推定値	標準誤差	z値のP値
1) 観察年間	1988-1991	-0.865	0.045	0.000	-0.831	0.049	0.000
	1988-1994	-0.941	0.027	0.000	-0.929	0.028	0.000
	1988-1997	-1.016	0.025	0.000	-1.001	0.026	0.000
	1988-2000	-1.031	0.024	0.000	-1.016	0.025	0.000
	1988-2003	-1.025	0.024	0.000	-1.011	0.025	0.000
	1988-2005	-0.978	0.026	0.000	-0.959	0.028	0.000
	1988-2006	-0.874	0.022	0.000	-0.858	0.023	0.000
	1991-1991	-0.022	0.038	0.571	-0.024	0.039	0.533
	1991-1994	-0.697	0.043	0.000	-0.691	0.045	0.000
	1991-1997	-0.887	0.041	0.000	-0.879	0.043	0.000
	1991-2000	-0.932	0.038	0.000	-0.921	0.040	0.000
	1991-2003	-1.006	0.036	0.000	-0.993	0.038	0.000
	1991-2005	-0.940	0.045	0.000	-0.923	0.049	0.000
	1991-2006	-0.789	0.023	0.000	-0.772	0.024	0.000
	1994-1994	0.099	0.035	0.005	0.097	0.036	0.007
	1994-1997	-0.627	0.040	0.000	-0.622	0.041	0.000
1994-2000	-0.770	0.036	0.000	-0.761	0.038	0.000	

	1994-2003	-0.964	0.036	0.000	-0.951	0.038	0.000
	1994-2005	-0.929	0.044	0.000	-0.915	0.047	0.000
	1994-2006	-0.727	0.023	0.000	-0.710	0.023	0.000
	1997-1997	-0.084	0.036	0.018	-0.087	0.037	0.017
	1997-2000	-0.696	0.038	0.000	-0.690	0.039	0.000
	1997-2003	-0.864	0.035	0.000	-0.856	0.037	0.000
	1997-2005	-0.814	0.043	0.000	-0.802	0.046	0.000
	1997-2006	-0.664	0.023	0.000	-0.649	0.023	0.000
	2000-2000	-0.200	0.029	0.000	-0.221	0.030	0.000
	2000-2003	-0.722	0.036	0.000	-0.721	0.038	0.000
	2000-2005	-0.816	0.044	0.000	-0.812	0.046	0.000
	2000-2006	-0.599	0.023	0.000	-0.589	0.023	0.000
	2003-2003	-0.237	0.038	0.000	-0.246	0.039	0.000
	2003-2005	-0.535	0.046	0.000	-0.539	0.047	0.000
	2003-2006	-0.369	0.023	0.000	-0.367	0.024	0.000
	2005-2005	-0.832	0.065	0.000	-0.864	0.067	0.000
	2005-2006	-0.501	0.025	0.000	-0.518	0.026	0.000
	2006-2006	-1.480	0.035	0.000	-1.504	0.036	0.000
2) 最近大学カテゴリー	G1	-0.200	0.031	0.000	-0.209	0.036	0.000
	G2	-0.353	0.031	0.000	-0.362	0.036	0.000
	G3	-0.583	0.031	0.000	-0.596	0.036	0.000
	G4	-0.677	0.031	0.000	-0.689	0.036	0.000
	公立大学 私立大学 大学校等	-0.742 -0.717 -0.910	0.032 0.031 0.034	0.000 0.000 0.000	-0.760 -0.735 -0.935	0.036 0.036 0.039	0.000 0.000 0.000
3) 世代カテゴリー	大正(12-26年)	-1.336	0.053	0.000	-1.375	0.055	0.000
	昭一(27-34年)	-0.893	0.025	0.000	-0.942	0.027	0.000
	焼跡(35-46年)	-0.648	0.021	0.000	-0.706	0.023	0.000
	団塊(47-49年)	-0.581	0.021	0.000	-0.639	0.024	0.000
	しらけ(50-64年)	-0.410	0.020	0.000	-0.459	0.022	0.000
	バブル(65-69年)	-0.242	0.021	0.000	-0.281	0.023	0.000
氷河期(70-86年)	-0.083	0.022	0.000	-0.118	0.024	0.000	
4) 最近大学地域カテゴリー	北海道・東北	NA	NA	NA	NA	NA	NA
	関東	0.058	0.008	0.000	0.062	0.009	0.000
	中部	0.047	0.009	0.000	0.048	0.010	0.000
	近畿	0.018	0.009	0.035	0.024	0.010	0.012
	中国・四国 九州	-0.015 -0.003	0.010 0.010	0.134 0.757	-0.017 -0.003	0.011 0.011	0.125 0.761
5) 最近職位	教務職員等	0.132	0.059	0.026	0.134	0.065	0.040
	助手・助教	-0.187	0.025	0.000	-0.190	0.028	0.000
	講師	-0.244	0.046	0.000	-0.250	0.051	0.000
	准(助)教授	-0.181	0.041	0.000	-0.186	0.046	0.000
	教授 学長等	-0.108 0.096	0.027 0.045	0.000 0.030	-0.124 0.087	0.031 0.051	0.000 0.086
6) 最終学歴	学士等	-0.276	0.022	0.000	-0.267	0.023	0.000
	修士	0.045	0.014	0.001	0.057	0.015	0.000
	博士	0.523	0.012	0.000	0.535	0.013	0.000
7) 出身大学カテゴリー	自校	0.015	0.067	0.824	0.011	0.074	0.885
	G1	0.090	0.068	0.183	0.090	0.074	0.226
	G2	-0.047	0.068	0.492	-0.050	0.074	0.498
	G3	-0.075	0.069	0.271	-0.080	0.075	0.289
	G4	-0.161	0.068	0.018	-0.166	0.075	0.027
公立大学 私立大学 大学校等 外国大学	-0.100 -0.170 -0.792 1.165	0.069 0.068 0.148 0.706	0.146 0.013 0.000 0.099	-0.103 -0.174 -0.798 1.121	0.075 0.075 0.156 0.711	0.172 0.020 0.000 0.115	
8) 出身大学地域カテゴリー	北海道・東北	-0.361	0.071	0.000	-0.349	0.077	0.000
	関東	-0.299	0.070	0.000	-0.285	0.077	0.000
	中部	-0.317	0.071	0.000	-0.303	0.078	0.000
	近畿	-0.328	0.071	0.000	-0.316	0.077	0.000
	中国・四国 九州 外国	-0.369 -0.345 -1.668	0.071 0.071 0.707	0.000 0.000 0.018	-0.355 -0.334 -1.625	0.078 0.078 0.711	0.000 0.000 0.022
9) 分野カテゴリー	総合領域	1.400	0.026	0.000	1.410	0.026	0.000
	複合新領域	1.451	0.026	0.000	1.462	0.027	0.000
	人文学	1.065	0.026	0.000	1.070	0.026	0.000
	社会科学	1.167	0.026	0.000	1.176	0.026	0.000
	数学	1.563	0.027	0.000	1.572	0.028	0.000
	物理系科学	1.449	0.026	0.000	1.458	0.026	0.000
	化学	1.410	0.026	0.000	1.415	0.027	0.000
	工学	1.478	0.025	0.000	1.489	0.026	0.000
	生物学	1.505	0.026	0.000	1.514	0.027	0.000
	農学 医歯薬学	1.568 1.589	0.026 0.025	0.000 0.000	1.584 1.606	0.027 0.025	0.000 0.000
10) 異動回数	0.069	0.004	0.000	0.075	0.004	0.000	

11) 昇格回数		0.057	0.007	0.000	0.058	0.007	0.000
12) 昇格パス	教務職員等	-0.672	0.062	0.000	-0.706	0.068	0.000
	教務職員等→助手・助教	-0.562	0.042	0.000	-0.580	0.047	0.000
	教務職員等→講師	-0.663	0.075	0.000	-0.685	0.083	0.000
	教務職員等→准(助)教授	-0.538	0.068	0.000	-0.567	0.076	0.000
	教務職員等→教授	-0.575	0.069	0.000	-0.602	0.079	0.000
	教務職員等→学長等	-0.865	0.502	0.085	-0.872	0.574	0.128
	助手・助教	-0.618	0.030	0.000	-0.636	0.033	0.000
	助手・助教→講師	-0.617	0.051	0.000	-0.636	0.056	0.000
	助手・助教→准(助)教授	-0.552	0.047	0.000	-0.573	0.052	0.000
	助手・助教→教授	-0.500	0.038	0.000	-0.508	0.043	0.000
	助手・助教→学長等	-0.634	0.067	0.000	-0.643	0.078	0.000
	講師	-0.668	0.050	0.000	-0.682	0.055	0.000
	講師→准(助)教授	-0.587	0.047	0.000	-0.602	0.052	0.000
	講師→教授	-0.535	0.038	0.000	-0.545	0.042	0.000
	講師→学長等	-0.686	0.088	0.000	-0.704	0.102	0.000
	准(助)教授	-0.606	0.045	0.000	-0.621	0.050	0.000
	准(助)教授→教授	-0.425	0.035	0.000	-0.431	0.039	0.000
	准(助)教授→学長等	-0.488	0.055	0.000	-0.496	0.063	0.000
	教授	-0.344	0.033	0.000	-0.354	0.037	0.000
	教授→学長等	-0.340	0.051	0.000	-0.344	0.057	0.000
学長等	-0.409	0.080	0.000	-0.397	0.085	0.000	

表3 科研費採択数を応答変数としたポアソン分布、負の二項分布の一般化線形モデルの係数推定値

説明変数 12 のうち、10 変数が質的、2 変数が量的である。量的変数の数が少なく有意だが、異動回数や昇格回数の効果は小さい。

以下、各変数の係数から科研費採択への意味を考察する。断りない限り、考察対象はポアソン分布モデルとするが、負の二項分布の係数も大差ないことが分かる(表3)。

- 観察年間:** 教員の観測時期と時点数効果。本時間間隔は offset 項に入れたため年数による単純な採択数の違いはモデルから除去。しかし、推定値を見ると、単年度の場合(94年:0.099, 91年:-0.022, 97年:-0.084, 00年:-0.200, 03年:-0.237)で高く、長期間 88-06年 (-0.874)、88-05年 (-0.978)等で低くなる。これは科研費採択データ接続の際、分担研究者等として含まれる教員情報を補完するためである。時期と時点数を統合しているため、構造的要因の影響が存在する可能性がある。
- 最近大学カテゴリー:** G1・2・3・4、私立・公立大学、大学校等の順に効果が高い。国立大学群はそもそも論文数シェアに基づく分類で[1][2][3][4][5]、本結果に特に大きな違和感はないだろう。
- 世代カテゴリー:** 本推定値を見る限り若い世代ほど科研費は採択されやすく見えるが、申請件数の情報が必要である(若い研究者ほど積極的に科研費に申請しているかもしれない)。確かに前述の要覧や文科省・JSPSの公表資料から申請件数自体は判明し、件数・代表者ベースでは議論可能である。しかし、分担研究者数は分からないため、本研究のスコープを縮小する必要がある。また、科研費自体の経年的変化の影響も考えられる。
- 最近大学地域カテゴリー:** 明確な傾向は見られないが、相対的に中国・四国、九州地域の効果が小さい模様。
- 最近職位:** 教務職員等や学長等は標本数が十分ではなく、P値も大きい。それ以外では、教授、助教(助手)、准(助)教授、講師の順に効果がある。全般的に偉い方が有利であると考えられよう。
- 最終学歴:** 圧倒的に博士が優位であり、修士が続く。近年の大学院重点化傾向等から直観的に理解されよう。
- 出身大学カテゴリー・8) 出身大学地域カテゴリー:** 2変数をまとめた。カテゴリー化で外国を細分化しなかったため、出身大学では外国大学が大きな効果を持つ一方、地域では大きく負となった。これは多重共線性的一种と考えられ、解釈に工夫が必要である。即ち、両者を足すと 1.165-1.668=-0.503 となり、

実は外国大学出身は G1 や自校出身に及ばない。G1 や自校出身者はどの地域でも外国出身者の推定値を上回る。

- 分野カテゴリー:** 科研費では制度設計上、分野内で採択数・金額を調整するため、分野間の差はあまりない。ここで影響するのは採択数であるため、本推定値が高い分野(医歯薬学、農学、数学、生物学など)は、採択数は大きい、配分金額は小さい又は申請件数が大きい若しくは分担研究者が多い可能性が推察される。逆に、推定値が低い分野(人文学、社会科学など)は採択数が少なく、配分金額は大きい又は申請件数は少ない若しくは分担研究者が少ない可能性が推察される。分野の構造的な問題が潜むと推察される。例えば、臨床医歯薬学などでは観察データを得るため多数の分担研究者が不可欠である。数学のような分野内の領域が非常に深く、研究者一人でカバーできる範囲が相対的に小さい分野では、分担研究者との共同研究は重要な役割を果たすだろう。
- 異動回数・11) 昇格回数:** とともに科研費採択効果は有意に正だが、大きくはない。科研費採択に対して教員異動・昇格は大きな影響を及ぼすとは言えない。
- 昇格パス:** 昇格経路の効果を調べた。効果が大きい順から、教授→学長等(-0.340)、教授(-0.344)、学長等(-0.409)となっており、偉さ・偉くなりやすさと採択数には明確な関係がある。5)と併せると、科研費は研究者のキャリア形成に大きく貢献すると考えられる。

## 5. 対数線形モデル・多項ロジットモデル

本章では採択件数を質的変数とし、対数線形モデル[1][2][3][4][5]、多項ロジットモデルを試行した。説明変数はポアソン回帰等の 12 変数のうち、観察期間の自由度が大きく、終年と時点数に分解した 13 変数を使う。この修正点は直前のポアソン回帰等にも反映したい。

### (1) 対数線形モデル

因果関係が明確でない変数間分析を行う対数線形モデルでは、推定モデルの自由度が大きくなる傾向があり、変数・水準数を絞る必要がある。そこで採択数と 13 変数のうち科研費採択数以外の 4 変数の組み合わせを網羅的に AIC の変数増減法で調べたところ、

①[0]科研採択数[4]大学地域[10]異動回数[11]昇格回数[12]昇格パス

②[0]科研採択数[1]時点数[10]異動回数[11]昇格回数[12]昇格パス

③[0]科研採択数[1]時点数[4]大学地域[11]昇格回数[12]昇格パス

④[0]科研採択数[1]時点数[4]大学地域[10]異動回数[12]昇格パス

の解釈可能な 4 つのモデル構成が得られた。これらの最適モデルは、3 次交互作用項を 1 つ含み、簡略表記では次である。

①[0]科研採択数:10)異動回数:12)昇格パス[4]大学地域[11]昇格回数

②[1]時点数:10)異動回数:11)昇格回数[0]科研採択数[12]昇格パス

③[0]科研採択数:4)大学地域:12)昇格パス[1]時点数[11]昇格回数

④[1]時点数:10)異動回数:12)昇格パス[0]科研採択数[4]大学地域

①と③、②と④の最適モデルは最高次の交互作用項が類似する。特に前者は科研費採択数と、異動回数又は大学地域と昇格パスの関係を示し、回帰分析では明確でなかった異動回数や大学地域の科研費採択との関係が明らかである。また、最適モデルの非適合度指標  $D[19]$ は次となる。

$$D = \sum_i n_i - \hat{\mu}_i / (2N) = ①:0.061, ②:0.058, ③:0.065, ④:0.094$$

①-④は 6 種変数の組み合わせである。即ち、6 因子モデル

[0]科研採択数[1]時点数[4]大学地域[10]異動回数[11]昇格回数

[12]昇格パス]から AIC の変数増減法で得た最適モデルは

解釈可能となり、[1]時点数:10)異動回数:11)昇格回数[0]科研

採択数[4]大学地域[12]昇格パス]、 $D = 0.113$ となり、データの

約 89%が最適モデルで説明され、再現性も概ね十分である。

今後、係数推定値からオッズ比等を調べ、最適モデルの意味を解明する。上記分析は人数を度数としているが、今後度数を科研費採択数としたモデルとポアソン回帰の結果とも比較する。

上記では、説明変量は名義尺度だが、変量や水準の一部を順序尺度にして対数線形モデルを行う。再現性悪化も見込まれるが、順序尺度とする因子の適切な選択により、容易な解釈・最適モデルの簡略化・簡略化による仮説検定の検出力強化等が期待される[20][21]。交互作用項も含むため、具体的には不均質/均質線形連関モデル等が考えられる。3 因子モデルの ANOVA coding として[22]、①,②,③: 因子  $i, j, k$ : 水準  $\lambda, \mu, \tau$ : モデル推定量  $u_i, v_j, w_k$ : 順序スコア  $\beta$ : 順序モデル推定量とすると、部分連関モデル:

$$\ln F_{ijk} = \lambda_0 + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \beta^{(12)} \mu_i (v_j - \bar{v}) + \beta^{(13)} \mu_i (w_k - \bar{w}) + \beta^{(23)} (v_j - \bar{v})(w_k - \bar{w})$$

不均質一様連関モデル:

$$\ln F_{ijk} = \lambda_0 + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \beta_k (u_i - \bar{u})(v_j - \bar{v}) + \lambda_{ik}^{(13)} + \lambda_{jk}^{(23)}$$

均質一様連関モデル:

$$\ln F_{ijk} = \lambda_0 + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \beta (u_i - \bar{u})(v_j - \bar{v}) + \lambda_{ik}^{(13)} + \lambda_{jk}^{(23)}$$

一様相互作用モデル:

$$\ln F_{ijk} = \lambda_0 + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \beta^{(12)} (u_i - \bar{u})(v_j - \bar{v}) + \beta^{(13)} (u_i - \bar{u})(w_k - \bar{w}) + \beta^{(23)} (v_j - \bar{v})(w_k - \bar{w}) + \beta^{(123)} (u_i - \bar{u})(v_j - \bar{v})(w_k - \bar{w})$$

これらは最適モデルの尤度比検定等により採否が決まる。主従を明確にしない下での変量関係や構造の分析が対数線形モデルの特徴であり、本研究の検討の主力である。

## (2) 多項ロジットモデル(多項ロジスティック回帰)

対数線形モデルの特殊形である 2 値応答変数のロジスティック回帰を更に拡張した多項ロジットモデルを検討する。本モデルでも説明変量 12 は多いため、表 2 から、P 値の大きな[8]出身大学地域[11]昇格回数]を省き、期間を分解した 11 説明変量、従属変量は名義尺度化した科研費採択数として線形モデルとした。すると 4) 大学地域と 5) 職位が落ち、

$$\ln \frac{P(\text{科研費採択数} = [2, 3, \dots, 8])}{P(\text{科研費採択数} = 0)} = [\text{切片}] + [1] \cdot \text{終年} + [1] \cdot \text{時点数} + [2] \cdot \text{大学}$$

$$+ [3] \cdot \text{世代} + [6] \cdot \text{学歴} + [7] \cdot \text{出身大学} + [9] \cdot \text{分野} + [10] \cdot \text{異動回数} + [12] \cdot \text{昇格パス}$$

が最適モデルとなる。科研費採択数毎に係数推定値があり、巨大になるため推定値の記載は省略するが、標準誤差の小ささと標本数の大きさから、各係数の P 値は小と考えられ推定値は信頼できる。また、採択数と係数の推定値の増減傾向は必ずしも一致しない。これは詳細な変量効果の調査には向くが、全体的傾向把握には情報が多すぎるともいえる。そこで、科研費採択数には順序があるとしたモデルに変える。

$$\ln \frac{P(\text{採択数} = 1) + P(\text{採択数} = 2) + \dots + P(\text{採択数} = 8)}{P(\text{採択数} = 0)} = a_{01} + \beta X$$

$$\ln \frac{P(\text{採択数} = 2) + \dots + P(\text{採択数} = 8)}{P(\text{採択数} = 0) + P(\text{採択数} = 1)} = a_{12} + \beta X$$

:

$$\ln \frac{P(\text{採択数} = 8)}{P(\text{採択数} = 0) + P(\text{採択数} = 1) + \dots + P(\text{採択数} = 7)} = a_{78} + \beta X$$

この比例オッズモデルの係数を推定すると表 4 となる。

	推定値	標準誤差	t 値	
1) 終年	-0.046	0.000	-5.562	
1) 観測数	0.619	0.003	198	
2) 最近大学カテゴリー	G2	-0.415	0.012	-34
	G3	-0.991	0.013	-79
	G4	-1.261	0.011	-110
	公立大学	-1.300	0.014	-92
	私立大学	-1.461	0.014	-105
	大学校等	-1.911	0.002	-927
	不明	0.857	0.000	5,102
3) 世代カテゴリー	大正(12-26 年)	-1.518	0.000	-4,432
	昭一(27-34 年)	-1.039	0.009	-118
	昭和(35-46 年)	-0.821	0.014	-61
	団塊(47-49 年)	-0.707	0.006	-115
	バブル(65-69 年)	0.599	0.014	41
	氷河期(70-86 年)	0.872	0.012	71

	不明	0.817	0.007	119
6) 最終学歴	学士等	-0.487	0.002	-198
	修士	-0.163	0.008	-20
	博士	0.951	0.010	99
7) 出身大学カテゴリー	自校	-0.235	0.012	-20
	G2	-0.416	0.015	-28
	G3	-0.425	0.002	-186
	G4	-0.598	0.003	-190
	公立大学	-0.479	0.001	-373
	私立大学	-0.515	0.009	-57
	大学校等	-1.575	0.000	-49,370
	外国大学	-0.667	0.003	-217
	不明	0.148	0.007	22
9) 分野カテゴリー	総合領域	-0.459	0.018	-26
	複合新領域	-0.344	0.003	-113
	人文学	-1.284	0.016	-82
	社会科学	-1.142	0.015	-76
	数学	-0.028	0.001	-25
	物理系科学	-0.330	0.019	-18
	化学	-0.487	0.003	-144
	工学	-0.221	0.015	-14
	生物学	-0.216	0.002	-101
	農学	-0.018	0.011	-2
	不明	-2.938	0.003	-1,040
10) 異動回数		0.308	0.011	28
12) 昇格パス	教務職員等→助手・助教	0.560	0.000	2,864
	教務職員等→講師	0.502	0.000	11,390
	教務職員等→准(助)教授	0.499	0.000	7,041
	教務職員等→教授	0.705	0.000	16,470
	教務職員等→学長等	-0.203	0.000	-140,600
	助手・助教	-0.371	0.012	-31
	助手・助教→講師	-0.230	0.007	-34
	助手・助教→准(助)教授	0.057	0.015	4
	助手・助教→教授	0.480	0.010	50
	助手・助教→学長等	0.655	0.000	17,410
	講師	-0.505	0.003	-190
	講師→准(助)教授	-0.004	0.007	-1
	講師→教授	0.135	0.003	45
	講師→学長等	0.420	0.000	19,350
	准(助)教授	-0.314	0.018	-17
	准(助)教授→教授	0.309	0.014	22
	准(助)教授→学長等	0.866	0.000	6,260
	教授	0.317	0.012	27
	教授→学長等	0.976	0.001	1,679
	学長等	0.752	0.000	7,675
	不明	0.535	0.008	64

切片	推定値	標準誤差	t 値
a01	-91.6	0.001	-153,057
a12	-90.3	0.007	-13,324
a23	-89.4	0.009	-10,288
a34	-88.5	0.010	-8,488
a45	-87.7	0.012	-7,122
a56	-86.9	0.015	-5,936
a67	-86.0	0.018	-4,697
a78	-84.9	0.027	-3,204

表 4 多項ロジット(比例オッズモデル)における係数推定値

比例オッズモデルは順序を仮定しない多項ロジットより制約を仮定する分、AIC や残差尤度比は大きくなり、再現性は悪くなる。このモデルはポアソン回帰等と異なり、データの頻度を元に構成されており、表 4 の結果はその傾向が顕著に表れるようだ。これは比例オッズの仮定のためと考えられる。こうして、ポアソン回帰・対数線形モデル・多項ロジットモデルはそれぞれの統計学的関係に基づき解釈でき、科研費構造の解明に貢献するだろう。

第 4 章 5)12)の結果は新約聖書の「おおよそ、持っている人は与えられて、いよいよ豊かになるが、持っていない人は、持っているものまでも取り上げられるであろう」(マタイ福音書の第 13 章第 12 節)に因むマタイ効果[4]の傍証となる。能力には分散が存在し、程度に応じ得手不得手がある。何でもできる人もいるが、だか

らと言って研究を含む全業務を担わせたら、直ぐに業務が回らなくなる。一方、低評価とされ研究費を失った研究者は存在意義否定を感じて失望し、他業務にも熱意を持たなくなる。現状は研究の多様性に必ずしも対応できておらず、同じく統計的最適化にも離反する。本研究はそういった状況を説明し、Evidence-Based Policy (EBP)の基礎資料となることを目指す。

## 6. 謝辞

本研究の遂行に当たり、御協力いただいた方々にお礼申し上げます。また、筆者が国立試験研究機関所属のため、本研究で構築・使用するHM-DBは行政機関の保有する個人情報保護に関する法律等の規定の適用を受ける。本研究者は当該法令規定を遵守して研究を実施する。なお、本研究における主張等の責任は専ら筆者が負う。また、本研究は科学研究費補助金(19710139)の助成を受けた。

### 【附録:分布の拡張とパラメータ推定法】

ポアソン分布・負の二項分布を拡張した分布のバリエーションを調べた。ポアソン分布の拡張である一般化ポアソン分布[23]や、負の二項分布の拡張類[24][25]は複雑であるため、ほとんど使用されていない。なお、負の二項分布の一般化[25]は、モデルのパラメータ数が4つあり、不定解に陥る可能性が高いため省略する。

### 【ポアソン分布の拡張】

(1) 一般化ポアソン分布

$$p(x) = \lambda_1 (\lambda_1 + \lambda_2 x)^{x-1} \exp(-(\lambda_1 + \lambda_2 x)) / x! | \lambda_2 < 1$$

$$E(x) = \lambda_1 / (1 - \lambda_2), \sigma^2(x) = \lambda_1 / (1 - \lambda_2)^2$$

$$\lambda_1 = E(x)^2 / \sigma(x), \lambda_2 = 1 - E(x)^2 / \sigma(x)$$

### 【負の二項分布の拡張】

(2) B3型一般超幾何分布: ベータ負の二項分布

$$GHGB3(\alpha, \beta; \gamma) \quad p(x) = \frac{Be(\alpha + \gamma, \beta + x) \Gamma(\alpha + x)}{Be(\beta, \gamma) \Gamma(\alpha x)}$$

(3) 0で修正された負の二項分布

$$p_k = \frac{1 - p_0}{1 - (1 - p)^\alpha} p^k (1 - p)^\alpha \frac{\Gamma(\alpha + k)}{\Gamma(\alpha) k!}, k \geq 1, 0 < p < 1$$

(4) 非心負の二項分布

$$NNBD(p, v, \lambda) \quad p_k = e^{-\lambda p} (1 - p)^p \frac{\Gamma(v + k)}{k! \Gamma(-k)} \sum_{j=0}^{\infty} \frac{\Gamma(-k + j)}{\Gamma(v + j)} \frac{(\lambda(p-1))^j}{j!}$$

以上の負の二項分布の拡張に対してあてはめを試みた。あてはめ法は次のカイ二乗最小化法、G最小化法、最尤推定法、古典的モーメント法が考えられた。簡便性と時間的制約からG最小化法と古典的モーメント法を重点的に使用した。

### 【観測値との当てはめによる離散確率分布のパラメータ推定方法】

(1) カイ二乗最小化法

推定したパラメータを元に期待度数を推定して、観測度数との適合度を検定する際、多くの場合、 $\chi^2$ 分布に従うカイ二乗検定統計量が使われることを利用し、当該統計量を最小化するパラメータを推定する方法。カイ二乗検定統計量は

$$\chi^2 = \sum_{i=1}^n \frac{(\text{expected}_i - \text{observed}_i)^2}{\text{expected}_i}$$

となる。ここで推定すべきモデルのパラメータを  $\alpha_j$  とすると、

$$\frac{\partial}{\partial \alpha_j} \chi^2 = \sum_{i=1}^n \left[ 1 - \frac{(\text{observed}_i)^2}{(\text{expected}_i)^2} \right] \frac{\partial (\text{expected}_i)}{\partial \alpha_j} = 0$$

通常、解の導出には Newton-Raphson 法、共役勾配法などの数値計算法を使用する。簡便法として疑似的な最小二乗法も考えられる。

(2) G最小化法

カイ二乗検定統計量を最小化するアイデアから類推して、G検定に使用されるGを最小化してモデルのパラメータを推定する方法を考えた。

$$G = 2 \sum_{i=1}^n (\text{observed}_i) \ln \left( \frac{\text{observed}_i}{\text{expected}_i} \right)$$

$$\frac{\partial}{\partial \alpha_j} G = 2 \sum_{i=1}^n \left( \frac{\text{observed}_i}{\text{expected}_i} \right) \frac{\partial}{\partial \alpha_j} (\text{expected}_i) = 0$$

と(1)より簡単になる。

(1)(2)の共通点として、期待値としてモデル式をパラメータで偏微分する必要がある。しかし、(2)では分母に同次数の期待値(モデル式)があるため、解析的な計算での大幅な簡略化が期待できる。

(3) 最尤推定法

対象とする確率分布から尤度関数を設定し、その尤度を最大にする分布のパラメータを推定する。

(4) 古典的モーメント法

本稿の古典的モーメントとは、 $E(X^m), m=1, 2, \dots$  を指す。

低次モーメントほど重要な分布情報を持つため、仮にモデルのパラメータ数が3であれば、その数に等しい次数まで、即ち理論値と観測値の1・2・3次モーメントが等しくなるようパラメータを推定する。

最終の結果ではないが、一般化ポアソン分布の計算は容易だが、そもそも適合しないと思われる。一方、一般超幾何分布や、0で修正された負の二項分布は更に改良を重ねれば適合する可能性があると考えられる。一般超幾何分布に比して、0で修正された負の二項分布の制約が強いと考えられるため、後者の適合が認められれば、前者の適合検討は必要なく、後者の解釈に入る。実現値0を特別に取り扱う点に関しては、本文中で触れたとおり、後者の解釈の目的も立ち得る。

0で修正された負の二項分布に対してG最小化法を適用する。

$$p_k = \frac{1 - p_0}{1 - (1 - p)^\alpha} p^k (1 - p)^\alpha \frac{\Gamma(\alpha + k)}{\Gamma(\alpha) k!}, k \geq 1, 0 < p < 1$$

$$\frac{1}{p_k} = \frac{1}{\frac{1 - p_0}{1 - (1 - p)^\alpha} p^k (1 - p)^\alpha \frac{\Gamma(\alpha + k)}{\Gamma(\alpha) k!}} = \frac{[1 - (1 - p)^\alpha] \Gamma(\alpha) k!}{(1 - p_0) p^k (1 - p)^\alpha \Gamma(\alpha + k)}$$

$$\frac{\partial}{\partial p_k} p_k = \frac{1}{p_0 - 1}, k \geq 1, \quad \frac{\partial}{\partial p_k} p_k = \frac{1}{p_0}, k = 0$$

$$\frac{\partial}{\partial p_k} p_k = \frac{k}{p} \frac{\alpha}{[1 - (1 - p)^\alpha] (1 - p)}, k \geq 1, \quad \frac{\partial}{\partial p_k} p_k = 0, k = 0$$

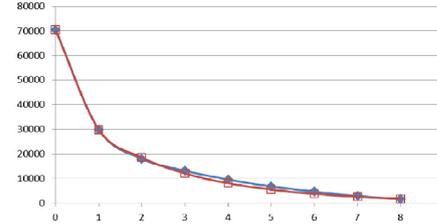
$$\frac{\partial}{\partial p_k} p_k = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha + k)} \frac{\Gamma(\alpha)}{\Gamma(\alpha)} \frac{[1 - 2(1 - p)^\alpha] \ln \alpha}{\{1 - (1 - p)^\alpha\}}, k \geq 1, \quad \frac{\partial}{\partial p_k} p_k = 0, k = 0$$

理論度数を確率密度関数に総度数を掛けかけたものとし、G最小化法を使うと、

$$\sum_{i=1}^n \left( \frac{n_i}{np_i} \right) \frac{\partial}{\partial \alpha_j} (np_i) = \sum_{i=1}^n \left( \frac{n_i}{np_i} \right) \frac{\partial}{\partial \alpha_j} p_i = 0$$

$$\Leftrightarrow p_0 = \frac{n_0}{n_0 + n_1} \sum_{i=1}^n i n_i = \frac{p \alpha n_0}{[1 - (1 - p)^\alpha] (1 - p)} \sum_{i=1}^n \frac{\Gamma(\alpha + i)}{\Gamma(\alpha + i)} n_i = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} \frac{[1 - 2(1 - p)^\alpha] \ln \alpha}{[1 - (1 - p)^\alpha]} p_{n_0}$$

$p_0$ は観測値と同じと推定され、 $p, \alpha$ の推定値を上式後半2式から近似的に求める。ここからは数値計算だが、参考までに現時点で最も再現性の高い解が下図である。これでもカイ二乗適合度検定のP値は0.000となり適合を示さない。



また、非心負の二項分布は、元々負の二項分布がポアソン分布とガンマ分布の混合分布であり、本ガンマ分布を非心ガンマ分布としたものだが、確率分布の式が簡単にならず、以上のような解析的アプローチが難しい。

### 【参考文献】

- [1] 国立大学教授へのキャリアパス—国立大学間異動と昇格の実態に関する分析—, 科学技術政策研究所 Discussion Paper No.60  
www.nistep.go.jp/achiev/ftx/jpn/ds060j/idx060j.html (2010年2月)
- [2] Moritaka Hosotsubo, A statistical study of transferral and promotion mechanisms relating to the appointment of professors at Japanese universities based on cross tabulation and log-linear model analysis, Scientometrics, Springer Volume 86, Number 2  
www.springerlink.com/content/0k170151h6146n66/ (Aug. 2010)
- [3] 細坪謙, 国立大学教授就任に係る異動・昇格に関するカテゴリーカルデータ分析, 研究・技術計画学会第25回年次学術大会講演要旨  
http://hdl.handle.net/10119/9275 (2010年10月)
- [4] 細坪謙, 日本の国立大学教授の異動・昇格に関する統計学的データ分析, 第15回情報・統計科学シンポジウム, http://bic.math.kyushu-u.ac.jp/ (2010年12月)
- [5] 細坪謙, 日本の国立大学教授の昇格・異動に関する統計解析, 応用統計学会2011年度年会  
www.applstat.gr.jp/annual/ (2011年6月)
- [6] A. J. Lotka, "The frequency distribution of scientific productivity" Journal of the Washington Academy of Sciences, 16(june 19), 1926.
- [7] C. B. Williams, "The Number of Publications Written by Biologists" Annals of Eugenics, Vol. 12, 1944, p.143-146.
- [8] H. A. Simon, "On a Class of Skew Distribution Functions", in: Models of Man, John Wiley & Sons, 1957.
- [9] 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.
- [10] D. J. de Solla Price, Little Science, Big Science, Columbia Univ. Press, 1963.
- [11] D. J. Price, "A General Theory of Bibliometric and Other Cumulative Advantage Processes", Journal of American Society for Information Science, Vol. 27, Number 5, 1976, p.292.
- [12] P. Allison, Processes of Stratification in Science, Arno Press, 1980
- [13] 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.
- [14] 山崎博敏, 科学の生産性とその階層化過程-「マイ効果」の基礎理論-, 広島大学 大学教育研究センター 大学論集 第11集(1982), pp.1-21.
- [15] R Development Core Team. (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL (http://www.R-project.org.)
- [16] 下野嘉子, 特集 統計解析(再?)入門 Rを用いた一般化線形モデル(回帰係数編): カウントデータを例に, 総説, 雑草研究 Vol. 55 (4), pp.287-294 (2010), (独)農業環境技術研究所
- [17] 今泉智通, 特集 統計解析(再?)入門 Rを用いた一般化線形モデル(仮説検定編): 割合データを例に, 総説, 雑草研究 Vol. 55 (4), pp.275-286 (2010), (独)農業・食品産業技術総合研究機構
- [18] McCullagh, Peter; Nelder, John. Generalized Linear Models, Second Edition. Boca Raton: Chapman and Hall/CRC. (1989)
- [19] Agresti, A. (1996). An Introduction to Categorical Data Analysis, pp. 201-231. John Wiley & Sons, Inc.
- [20] Agresti, A. (1984). Analysis of Ordinal Categorical Data, John Wiley & Sons, Inc.
- [21] 田中豊, 垂水共之, 藤本和昌. (1990). パソコン統計解析ハンドブックV 多変量分散分析・線形モデル編, pp.174-215. 共立出版株式会社.
- [22] MASAKO ISHII-KUNTZ. (1994). ORDINAL LOG-LINEAR MODELS, SAGE PUBLICATIONS, International Educational and Professional Publisher.
- [23] 北野原春彦, 一般化ポアソン分布におけるテストデータ解析のためのパラメータの変換, 心理学研究 (1996), 66(6), pp.425-430.
- [24] 清水邦夫, 浜谷政昭, 横内大介, 高橋睦, 確率分布公式集, 慶應義塾大学理工学部, 東京歯科大学 (2006)
- [25] 北野昌志<sup>1</sup>, 青山一基<sup>2</sup>, 清水邦夫<sup>3</sup>, リスクモデルにおける離散確率分布の漸化式, 統計数理 (2006), 第54巻第1号 147-175, 住友信託銀行年金信託部<sup>1</sup>, 慶應義塾大学理工学部数理工学研究所<sup>2</sup>, 慶應義塾大学理工学部数理工学科学科