研究分野による特性の違いを考慮した論文数の正規化手法

山本 鉱，石川 眞澄（九州工業大学）

１．　はじめに

大学は近年大きな競争の場に置かれていることから，各大学とも教育・研究の質について，各段の向上を目指している．本稿ではこの中の研究を対象としている．研究の質の向上を目指すうえで，大学は自身の研究力を正確に把握する必要がある．そのためには，大学全体，大学の部局あるいは学科単位，個人レベルの全てにおいて，研究力の把握が不可欠である．

通常は論文数と，被引用数を研究力の指標とすることが考えられる．しかし，表１からも分かるように，研究分野毎の論文数や論文数/著者数などが異なるだけではなく，論文当たり被引用数も異なる[1].大学の研究分野分布はさまざまであるため，このような分野間の特性の違いは，大学レベルでの論文数や被引用数の正確かつ公平な比較を困難にする．部局や学科，あるいは個人レベルでの研究力を評価する際にも，対象群の中で研究分野分布が異なれば同様の問題が生じる．以上のことから，研究分野による特性の違いを考慮した評価方法の確立が強く望まれる．

表1 Article, Reviewで見る分野毎の特性(2010～2014年)

Essential Science Indicators (22分野)	論文数	論文数/著者数(のべ数)	被引用数/論文数
Clinical Medicine	1,262,028	0.35	8.10
Chemistry	759,247	0.47	9.83
Engineering	554,365	0.50	5.33
Physics	554,001	0.46	8.04
Social Sciences, general	421,163	0.53	4.04
Materials Science	344,724	0.41	8.08
Biology & Biochemistry	342,015	0.27	10.55
Plant & Animal Science	339,737	0.38	5.66
Neuroscience & Behavior	241,327	0.31	11.25
Molecular Biology & Genetics	210,943	0.22	15.59
Environment/Ecology	202,152	0.35	8.06
Geosciences	201,970	0.43	7.56
Mathematics	199,798	0.90	2.67
Agricultural Sciences	192,282	0.35	5.32
Psychiatry/Psychology	184,443	0.46	7.26
Pharmacology & Toxicology	181,708	0.27	8.21
Computer Science	161,317	0.46	4.90
Economics & Business	125,865	0.72	4.52
Immunology	120,654	0.22	12.16
Microbiology	97,229	0.26	9.76
Space Science	68,714	0.42	12.81
Multidisciplinary	9,691	0.18	9.41

出典：Thomson Reuters InCites (2016年3月時点)

被引用数は，他の論文から引用された回数であり，論文の注目度を表す．従って，これは論文の質の近似指標と言えるだろう．ただし，先述の通り研究分野によって異なるだけでなく，出版年によっても異なり，古い論文ほど被引用数が大きくなるという傾向がある[2]．そのため，学術文献データベースを提供しているトムソン・ロイター社やエルゼビア社からは，上述の問題点を考慮した指標が，研究分析ツールなどを通して提供されている[2]．例えば，分野，出版年，論文タイプが同じ集団において，一定期間の平均被引用数が1 となるように正規化された指標として，トムソン・ロイター社から NCI（Normalized Citation Impact）が[3]，エルゼビア社から FWCI（Field Weighted Citation Impact）が提供されている[4]．そのため，論文の質に関しては比較的正確な評価が可能となってきている．しかし，被引用数だけでなく論文数も研究力の重要な要素である．一定期間に発表

された論文数は，研究の生産性の指標と言える[5]．90%を超える国立大学で個人業績評価が実施され，さらにその 90%において論文・総説数が評価項目として用いられている[6]．このことからも，論文数が研究力を測る指標として定着していることが分かる．他にも，他大学や諸外国とのベンチマーキングの際にも論文数は使われている．このような重要性があるにも関わらず，NCI や FWCI のように分野を超えて比較できるような指標は考案されていない．

そこで本稿では，研究分野間の論文生産性の違いを考慮した論文数の正規化方法を提案する．研究者単位で高精度に名寄せされた論文情報を，世界の研究者について全て入手することは困難である．また，各研究分野において，研究者の母集団が明確でないため，研究者当たりの論文数についてのデータも公表されていない．ただ，著者が所属する九州工業大学のように母集団が明確な場合には、各研究分野における研究者当たりの論文数を求めることは可能である．そこで本学のデータを利用し，研究分野による特性の違いを考慮した論文数の正規化が可能であることを示す．さらにその結果と今後の課題について検討する．なお，本学では平成 27 年度の教育職員評価において，ここで提案する手法を基に正規化した論文数を，新たな指標として試行的に導入した．

２．　データ

表 2 に本稿で使用するデータの詳細を示す．論文情報は，トムソン・ロイター社が提供する学術文献・引用索引データベース，Web of Science（以下，WoS と称する）の自然科学分野に収録されているもののうち，原著論文と総説に限って使用する．単年の分析では論文数が少なく正しい評価を行えないこと，また，論文が掲載されるまでの時間遅れなどを踏まえ，2010 年から 2014 年の 5 年間に出版された論文を分析対象とする．これらの条件に該当する論文数は 1,430 本，著者総数は 288 人であった．

表2 本稿で使用したデータの詳細

データベース	Science Citation Index Expanded
論文タイプ	原著論文，総説（Article，Review）
出版年	2010年～2014年（5年間）
研究分野	Essential Science Indicators（ESI）の22分野
その他	所属機関名称に九州工業大学が含まれている教員の論文 （精度を高めるために，2015年6月に本学独自で名寄せを実施）

３．　手法

３－１．論文数の正規化における定義

第一に，論文作成に要する労力について定義する．ある教員（以下，教員 i と称する）がある期間（本稿では 5 年間）に発表した全論文の作成に要した労力を 1 とし，単位をマンパワーと称する．

第二に，任意の研究分野に投入した労力について定義する．研究分野は雑誌単位で決められており，複数の雑誌へ論文を発表している教員の研究分野は，一般に複数の研究分野となる．任意の 1 分野（以下，分野 j と称する）への投入マンパワーは分野 j の論文数を全分野の論文数で割ったものと仮定する．なお，WoS 収録論文数が 0 の研究者の研究分野は未知なので除外する．これは誤差となるが，やむを得ないと考えている．

第三に，共著者間の論文数の配分について定義する．学内に共著者が存在する場合は，論文数を共著者数で割った値を各教員に等分する．つまり，任意の論文作成に学内の 2 人の教員が関係していた場合，論文数は 1/2 本ずつとなる．

３－２．正規化手法

先述の定義を基に，論文数の正規化の手順を記す．なお，教員 i の分野 j における論文数を$\alpha_{ij} = \{\alpha_{i1}, \cdots, \alpha_{iM}\}$と表す．ここで$M$は分野の総数である．

【手順１】教員 i が分野 j に投入したマンパワーβ_{ij}を計算する．

$$\beta_{ij} = \frac{\alpha_{ij}}{\alpha_{i.}} \quad (1)$$

ここで$\alpha_{i.}$はα_{ij}の分野に関する総和，つまり教員 i の総論文数である．

【手順２】分野 j における 1 マンパワーあたりの論文数 P_j を計算する．

$$P_j = \frac{\sum_{i=1}^{N} \alpha_{ij}}{\sum_{i=1}^{N} \beta_{ij}} \quad (2)$$

ここでNは本学の研究者の総数である．分子は分野 j の総論文数，分母は分野 j への全投入マンパワーを表す．

【手順３】教員 i の正規化論文数$\gamma_{i.}$を計算する．

教員 i の分野 j における論文数を P_j で正規化し，分野に関する総和をとる．

$$\gamma_{i.} = \sum_{j=1}^{M} \frac{\alpha_{ij}}{P_j} \quad (3)$$

$\gamma_{i.}$の平均値は 1 であり，1 を上回る研究者は，研究分野の特性の違いを考慮した場合，学内の平均よりも論文を多く発表していることになる．

３－３．ブートストラップ法による信頼区間

P_j は確率変数であり，ここではブートストラップ法を用いて P_j の信頼区間を求めた．この手法はモンテカルロ法の一種であり，一つの標本から復元抽出により複数のブートストラップ標本を構成し，母集団の性質を推定するものである．推定量の形を制限したり，パラメトリックモデルを仮定したりする必要がない[7]ため,利用可能なデータの個数が限られ分布が正規分布とは限らない本研究においても，適用可能と考えられる．

計算手順を以下に示す．1)論文情報データ（ここでは論文数とマンパワー）から無作為抽出により大きさ n の標本を構成する．2)標本から大きさ n のブートストラップ標本を復元抽出し，P_j の推定値を求める．3)これを 2000 回繰り返し，P_j のブートストラップ分布を

得る．4)パーセント点を用いて 95%信頼区間を求める．本計算には統計解析ソフト R を使用した．

４．　結果と考察

４－１．P_jの分野間のばらつきと信頼度

図 1 に各分野の論文発表に関わった著者数，平均値に対する相対偏差，及び P_j の推定結果を示す．P_j のエラーバーを求める際に用いた標本は，各分野の全データであり，標本の大きさ n はその分野の著者数とした．

ここで，P_j の統計的なばらつきが非常に大きいことが予想されることから，著者数が 10 人未満の分野は除外した．信頼性の高い結果を得るための必要最低限の標本数については，４－３で検討する．

図 1 に示した分野の P_j の平均値は 4.8 であり，標準偏差は σ =1.74 であった．95%信頼区間を考慮すると，平均よりも有意に高い P_j を持つ研究分野（以下，高 P_j 分野と称する）は，CHEMISTRY，PHYSICS，BIOLOGY & BIOCHEMISTRY であり，低い研究分野（以下，低 P_j 分野と称する）は ENGINEERING，COMPUTER SCIENCE，NEUROSCIENCE & BEHAVIOR であることが分かる．論文数の平均値に対する相対偏差では，高 P_j 分野で実論文数よりも正規化論文数が小さくなり，低 P_j 分野でその逆となっていることから，正規化の効果が適切に現れていることを確認できた．

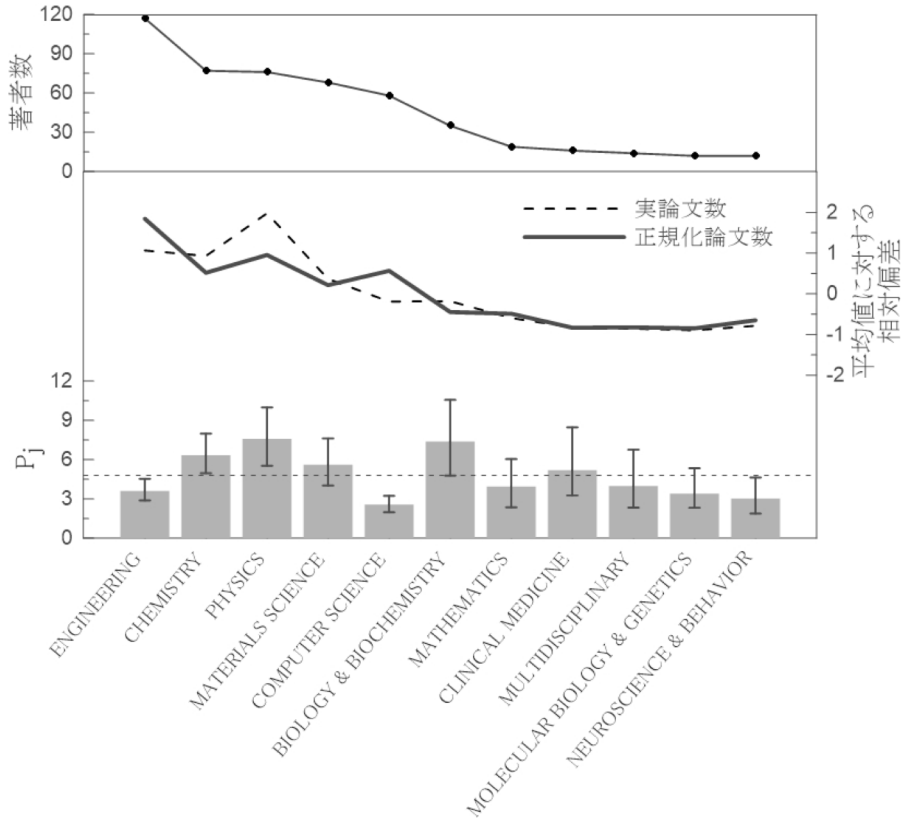


図 1　各分野における著者数，実論分数と正規化論文数のそれぞれの平均値に対する相対偏差，および P_j（点線は全 P_j の平均値，エラーバーは 95%信頼区間）

４－２．論文数を基にした教員の順位の変化

次に，正規化の前後で論文数を基にした順位がどのように変化したかを図２に示す．４－１では少数分野を省いたが，それらが順位決定に及ぼす影響が小さいこと，教員の順位を決定するために全論文を利用する必要があることから，ここでは，全ての分野に対して正規化を行い，順位を決定した．図２から分かるように，正規化前の論文数順位が高い教員，あるいは低い教員ほど順位変化は小さかった．大きく順位が変化したのは上位 21～80%に属する中位の教員で，最大の変化幅は-80～60 であった．このことから，論文の正規化は平均的な論文数を持つ教員に対して順位への影響が大きく，顕著に論文数が多い教員や，少ない教員に対しては順位への影響が小さくなることが分かった．

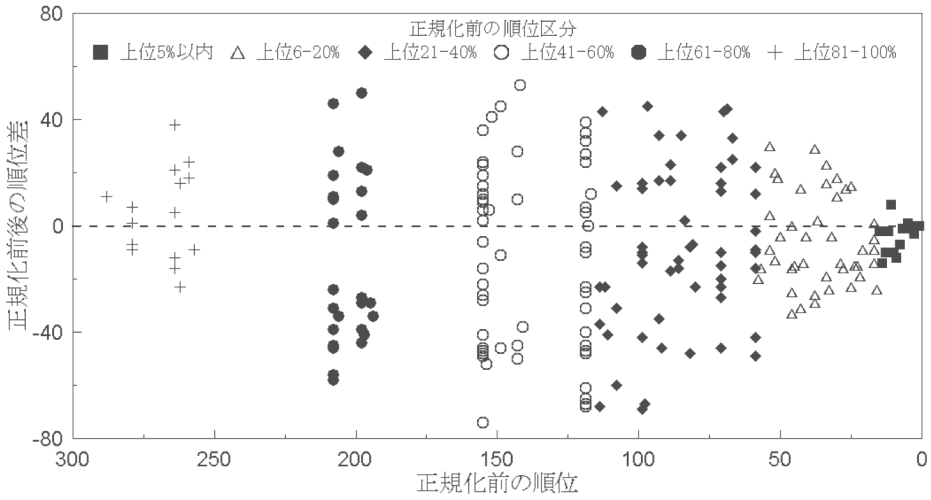


図２　正規化前と正規化後の順位差

４－３．必要最低限の標本数に関する検討

分野間の P_j を比較するには，信頼区間の正しい設定が不可欠である．しかし，それに必要な標本数が明らかではなく，図１に示した信頼区間に基づいて P_j の大小を検討すること

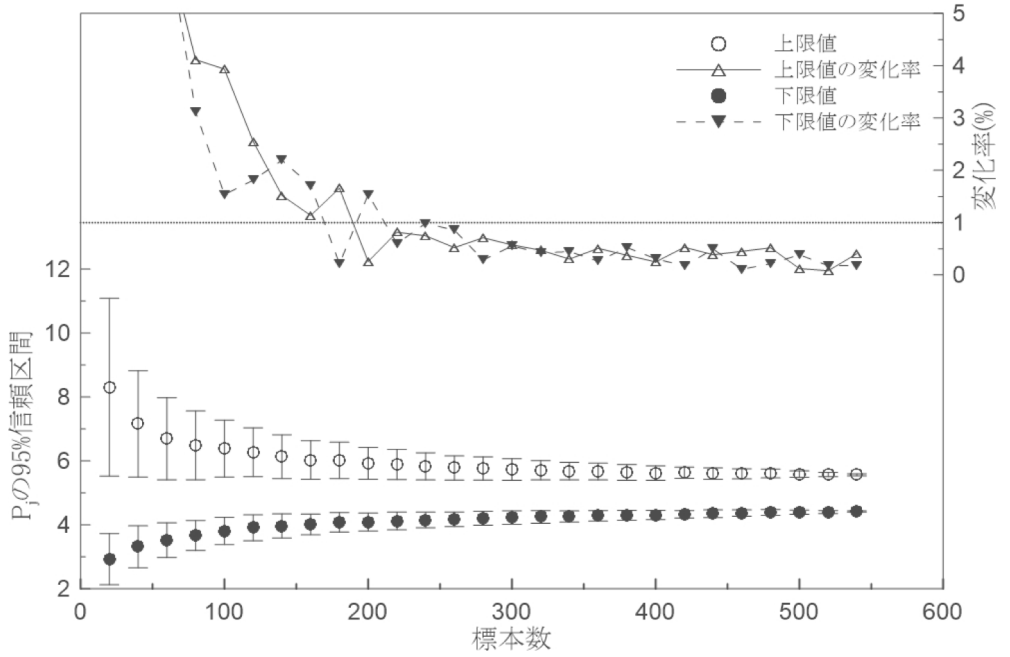


図３　95%信頼区間の標本数による違いとその変化率
（エラーバーは上限，下限の標準偏差）

が難しい．そこで，全分野のデータ（標本数 542）を母集団とみなし，そこから無作為抽出する標本の大きさを 20 から 540 まで 20 刻みで変化させ，それぞれの標本を基にブートストラップ標本を構成した．各標本数に対する 95%信頼区間を求めた結果を図 3 に示す．この際，異なる乱数の種で無作為に抽出した標本を 100 回構成し，上限値，下限値の平均値と標準偏差を求めた．また，収束を確認するため，平均値の変化率を求めた．

その結果，標本数が少ないと信頼区間の幅が過大になるだけでなく，その不確定性も大きくなることが分かった．また，上限値と下限値の変化率は，標本数が 200 を超えると 1%以下に抑えられることも分かった．これにより，各研究分野の P_j を正しく推定するためには，標本数が最低でも 200 程度必要であると考えられる．しかし，本学だけでは全研究分野において，この標本数を確保することは容易ではない．そのため，他大学との連携を通じて標本数を増やし，P_j を正しく推定することが必要になると考えている．

５．　まとめ

研究分野毎で異なる特性を考慮するため，論文数の正規化を行った．これにより，高 P_j 分野の論文数は正規化により相対的に減少し，逆に低 P_j 分野では増加することを確認した．正規化前後で論文数を基にした順位の変化を見ると，正規化前の論文数順位が高い教員，あるいは低い教員ほど順位変化は小さかった．大きく順位が変化したのは上位 21～80%に属する中位の教員で，最大の変化幅は-80～60 であった．このことから，論文数が顕著に多い教員や，少ない教員に対しては正規化の影響が大きくないことが示唆された．

次に論文数を正規化するうえで必要最低限の標本数を検討した．その結果，標本数が少ないと信頼区間の幅が過大になるだけでなく,その不確定性も大きくなることが分かった．これにより，各研究分野の P_j を正しく推定するためには，標本数が最低でも 200 程度は必要であることが示唆された．本稿で使用した九州工業大学のデータは，最も標本数が多い分野でも 120 に満たず，P_j を基にした分野特性についての詳細な分析ができなかった．今後は，教員単位で研究業績を把握している他大学との連携を通じて，全ての研究分野の P_j を正しく推定することが課題となる．その上で正規化論文数を用い，正確かつ公平な研究力の評価に繋げていきたい．

【参考文献】

[1] 福成洋(2014),研究戦略のための計量書誌学の実践的活用と応用,情報管理,57,6
[2] 古林奈保子(2014),引用情報に基づく研究の分析・評価,情報の科学と技術,64,12
[3] Thomson Reuters(2014),InCites Indicators Handbook
[4] Elsevier(2014),SciVal Metrics Guidebook
[5] トムソン･ロイター学術情報事業(2014),研究評価のためのビブリオメトリクス手法早わかり
[6] 三菱総合研究所(2015),研究者等の業績に関する評価に関する調査・分析報告書
[7] 汪金芳,桜井裕二(2011),ブートストラップ入門,共立出版, pp.236