

原著

病院のサイズ分布について —実証研究と確率モデル—

田中昌昭*¹

要 約

地域における医療資源の適切な配置や行政機関による地域医療計画の策定において医療提供体制の現状を正確に把握することは不可欠である。従来から厚生労働省が実施している医療施設調査を利用すれば、都道府県別の病床数や病床規模別の病院数を把握することはできる。しかしながら、これらの調査で分かるのは設備としての病床規模で、しかも10~100床刻みの大雑把なものである。綿密な医療政策の実現に必要なのは、病床数ベースの病院規模ではなく、むしろ実際の患者数ベースの病院規模で、しかも疾患領域別に病院規模を把握できるのが望ましい。そこで、厚生労働省が公開しているDPC (Diagnosis Procedure Combination) データを用いて、疾患領域別・都道府県別に退院患者数で見た病院のサイズ分布を調べた。その結果、多くの場合、病院のサイズ分布が幾何分布になっていることがわかった。その原因を解明するために単純な確率モデルを構築してコンピュータシミュレーションを行った。そのシミュレーション結果と確率モデルの解析によって、患者が病院を選択するときに、病院までの距離を優先した場合は二項分布、病院の規模(症例実績)を優先した場合は対数正規分布、そして両方を考慮した場合は幾何分布になることが示された。これらの傾向は、実際のDPCデータを用いて確認された。従来、病院のサイズ分布にはあまり関心が寄せられてこなかったが、本研究は、疾患領域別や都道府県別に病院のサイズ分布を初めて明らかにしただけでなく、病院のサイズ分布を幾何分布や対数正規分布などのパラメトリックな分布関数で表現できるようにした最初の研究でもある。これにより、医療政策や医療評価に必要な諸量を統計的に計算することができるようになるため、これまで以上に精緻な病院規模別の各種分析や統計処理を行うことが可能になる。

1. 緒言

病院の診療実績を表す臨床指標の一つに退院患者数がある。地域医療を担う病院には大小様々な規模があり、通常、その規模は病床数で分類されるが、退院患者数は病床数とも密接な関係があり、その意味で病院の規模(以降、病院のサイズと記す)を表す指標と考えることもできる。2003年に始まった診断群分類包括評価制度によって、診断群分類(DPC: Diagnosis Procedure Combination)ごとの退院患者数が公開されるようになり、DPC参加病院に限定されているものの、疾患領域ごとの病院のサイズを把握できるようになった。これによって、たとえば、この医療圏は、~疾患領域の年間退院患者数が~人以上の病院は全体の~%であるといった定量

化が可能になった。

さて、では、この病院のサイズは地域内でどのような分布をしているのだろうか。地域の医療政策の決定や医療提供体制の評価のために、たとえば100床未満を小規模、100床以上500床未満を中規模、それ以上を大規模といったように、病院をその病床数で大雑把に分類して地域間で比較するといったことがよく行われる。しかし、上述したようにDPC公開データのおかげで病院ごとの詳細な退院患者数が把握できるようになった今日においてもなお、病院のサイズ分布がどのようなになっているかを調査した研究はあまり例がない。

退院患者数を尺度として測った病院のサイズの分布の様子は医療圏によって異なるのであろうか。ま

*1 川崎医療福祉大学 医療福祉マネジメント学部 医療情報学科
(連絡先) 田中昌昭 〒701-0193 倉敷市松島288 川崎医療福祉大学
E-mail: mtanaka@mw.kawasaki-m.ac.jp

た、疾患領域によっても違いがあるのだろうか。そもそも医療圏を超えた普遍的な「病院サイズの分布関数」なるものは存在するのであるだろうか。もし、そのような分布関数が存在すれば、それを用いて病院サイズの平均や分散、あるいはより高次のモーメントを導出でき、それによって医療政策や医療評価に必要な諸量を統計的に計算できるかもしれない。そうすれば、恣意的な病院規模の分類による曖昧な値を用いることなく、より正確な政策なり評価を行うことができるようになる。

本研究の目的は、実際のDPC公開データを用いてそのような分布関数が存在することを実証的に確かめ、その発生メカニズムを確率モデルによるコンピュータシミュレーションとそれに対する理論的解析により解明することである。なお、本研究では統計処理には統計解析ソフトRバージョン3.0.2を、コンピュータシミュレーションにはExcel VBA (Visual Basic for Applications) を用いた。

2. 先行研究

著者は、病院の機能分化を測る新しい指標として、ハーシュマン・ハーフィンダール指数 (HHI: Hirshman-Herfindahl Index) を用いて「病院機能分化指数 (FDI: Functional Differentiation Index)」を考案した¹⁾。これは、直感的には、「特定の医療圏におけるある疾患領域の患者数を、その医療圏を独占している病院の数で割った値」、すなわち「医療圏を独占している病院 (患者シェアを獲得した病院) を受療するその疾患領域の平均患者数」を表しており、その導出の過程で患者数の分布関数 (厳密には確率質量関数) を用いた。具体的には、医療圏におけるある疾患領域の全患者数を y 、全病院数を n 、HHI の期待値を h すると、

$$h = \frac{n}{y^2} \sum_{k=0}^y k^2 P(k)$$

と表される。ここで、 $P(k)$ が退院患者数 k の分布関数である。著者は、 $P(k)$ として、パラメタが $\lambda = n/y$ の幾何分布を仮定して、 yh (地域内の全患者数に HHI を乗じたもの) という FDI を得た。しかしながら、この幾何分布を得るために「任意の医療圏にいる y 人の患者の列がランダムに n 病院に分割される」というあまり現実的とは言えない仮定を置いた。

Gao and Chan は、American Hospital Directory 2012からのデータを用いて米国の病院のサイズ分布が“sub-lognormal” (サブ対数正規) 分布になることを実証的に示した²⁾。ここで、彼らは sub-lognormal 分布を急尖的で歪度が正規分布と対数正規分布の中間であるような連続分布だと定義してい

る。そして、この sub-lognormal 分布を生み出す背後にあるメカニズムを再現するために、エージェント・ベース・モデルを構築してシミュレーションを行った。エージェント・ベース・モデルとは、システムを構成する自律的なエージェントが相互に作用し合い、その結果、システム全体にどのような影響を及ぼすかをシミュレートするコンピュータモデルである。Gao and Chan は、患者と病院をエージェントとしてシステムに組み入れ、それらの間の相互作用の結果、病院のサイズ分布がどのようになるかを調べたところ、米国の病院のサイズ分布に似た sub-lognormal 分布が得られたと報告している。ただし、Gao and Chan は病院のサイズとして退院患者数ではなく患者一人あたり一日当たりの医療費 (Patient days) を使っている。これは、医療費が、病床数などの施設の規模とその利用状況の両方を測る物差しだからであると述べている。

病院のサイズ分布に関する研究は洋の東西を問わずほとんど見られないが、他の領域では企業サイズ (firm size) の分布や都市サイズ (city size) の分布に関する研究が長年にわたって行われてきた³⁻⁶⁾。その理論的な研究の先駆けとなったのがジブラの法則 (Gibrat's law) である⁷⁾。ジブラの法則は「企業や都市のサイズが成長する割合は、サイズそのものには依存しない」という法則で、 $k_{i,t}$ を時刻 t における i 番目の企業あるいは都市のサイズ、 $\beta_{i,t}$ をランダムな正の成長係数とすると、

$$k_{i,t} = \beta_{i,t} \cdot k_{i,t-1} \cdots (1)$$

と定式化される。両辺の対数を取って、 $\ln k_{i,t}$ について漸化式を解くと

$$\ln k_{i,t} = \ln k_{i,0} + \ln \beta_{i,1} + \ln \beta_{i,2} + \cdots + \ln \beta_{i,t}$$

が得られるが⁸⁾、 $\ln \beta_{i,t}$ が平均 μ 、分散 σ^2 の独立同一分布に従うとみなせば、中心極限定理より $t \rightarrow \infty$ で

$$\ln k_{i,t} \sim N(\mu t, \sigma/\sqrt{t})$$

となる。つまり、 $k_{i,t}$ の極限分布は対数正規分布になる⁷⁻⁹⁾。Gibrat は最初これを所得分布に適用し、ついで製造業における工場のサイズ分布に適用し、きわめて良い一致を得た⁷⁾。また、Eeckhout は米国の人口調査のデータを用いて都市のサイズが対数正規分布になると報告した^{9,10)}。それに対して、Levy は、人口の多い方から0.6%の大都市 (これらの都市が全人口の23%を占める) では対数正規分布ではなくパレート分布 (ベキ乗則) に従うことを示した¹¹⁾。Malevergne 等は、ある条件下ではパレート分布と対数正規分布を区別することが難しい

ことに言及したうえで、UMP 不偏検定 (uniformly most powerful unbiased test) を用いて Levy の結論が正しいことを示した¹²⁾。他にもサイズ分布の上端 (upper tail) で対数正規分布からずれる例が数多く報告されている^{8,13-15)}。

Gabaix は確率モデル (1) の右辺に小さな正のランダム項 $\epsilon_{i,t}$ を加えた次のモデルを提案した¹⁶⁾。

$$k_{i,t} = \beta_{i,t} \cdot k_{i,t-1} + \epsilon_{i,t} \cdot \dots (2)$$

このモデルでは、たとえ $\epsilon_{i,t}$ がどんなに小さくとも、 $\epsilon_{i,t} > 0$ である限り、少なくとも上端 (upper tail) ではサイズ分布がべき乗則に近づくことが Kesten によって証明されている¹⁷⁾。

一方、Simon は、ジブラの法則に加えて、ある仮定を置くことにより、サイズ分布 $P(k)$ が次のようなユール分布 (Yule distribution) になることを示した¹⁸⁾。

$$P(k) = K \cdot B(k, \rho + 1) \cdot \dots (3)$$

ここで、 $B(k, \rho + 1)$ は k と $\rho + 1$ のベータ関数で、 K は規格化定数、そして、 ρ はパラメタである。なお、ある仮定とは、「一定確率 θ で新しい企業や都市が“誕生する”」という仮定で、 θ は ρ と $\rho = 1/(1 - \theta)$ の関係にある。ユール分布は、 $k \rightarrow \infty$ で $P(k) \sim k^{-(\rho+1)}$ と漸近近似できるので、大きな k 、すなわち分布の上端でサイズ分布はべき乗則に従うことが示され、漸化式 (2) から導かれる結論と同じ結果が得られる。

以上、サイズ分布に関する先行研究を見てきたが、現在では、サイズ分布の上端はべき乗則に従うというのが定説になっている。さらに驚くべきことに、べき乗則に従うのは何も企業や都市のサイズ分布だけでなく、単語の出現頻度、論文の引用数、地震の大きさ、富裕層の富、生物の種の数などもべき乗則に従っているとされている¹⁹⁻²²⁾。では、病院のサイズ分布はどうであろう。次節では公開 DPC データを使って都道府県ごとの疾患領域別退院患者数分布を具体的に見ていくことにする。

3. 実証研究

平成25年度の公開 DPC データは、全国にある 1,804 の急性期医療機関 (DPC 対象病院: 1,497, DPC 準備病院: 244, 出来高算定病院: 66) から提出されたデータを中央社会保険医療協議会の診療報酬調査専門組織・DPC 評価分科会が分析・集計したもので、厚生労働省のサイト²³⁾ から入手できる。このサイトでは、90以上の集計ファイルが公開され、医療機関別の退院患者数や救急車で搬送された患者

数、他院からの紹介患者数、平均在院日数などが疾患別に集計されている。このうち本研究では、主要診断群 (MDC: Major Diagnostic Category) 別・医療機関別に1年間 (平成25年4月1日～平成26年3月31日) の退院患者数を手術の有無別に集計した「MDC 別医療機関別件数 (割合)」の手術ありのデータを利用して都道府県ごとに疾患領域別の退院患者数分布を求めた。ここで、手術ありのデータを用いたのは、先行研究¹⁾ で病院の機能分化を測る指標を求める際に手術ありのデータを用いていたので、それと整合させるためである。ただし、予備実験によって手術の有無で結果は変わらないことを確認している。なお、疾患領域は本論文執筆時点で18種類ある MDC によって分類した²⁴⁾。

図1は、北海道における循環器系疾患 (MDC05) の退院患者数分布を様々な方法で描いたものである。図の左上は退院患者数のヒストグラム、中央上は経験累積分布関数 (ECDF: Empirical Cumulative Distribution Function)、そして右上は分位数プロット (Q-Q Plot) である。なお、中央上の図中の鎖線はパラメタが $\lambda = 74/22908 = 0.0032$ の幾何分布の累積分布関数である。ここで、22908は、公開 DPC データから得られた北海道における循環器系疾患の退院患者総数で、74は病院数である (実際の DPC 参加病院数は116であるが、循環器系疾患の退院患者数が0の病院は除いてある)。また、右上の図は、パラメタが同じ $\lambda = 0.0032$ の幾何分布を理論分布とした分位数プロットである。次に、左下は退院患者数の常用対数をとってヒストグラムを描いたものである。中央下はその経験累積分布関数 (鎖線は標本平均と標本分散を用いて得られた正規分布の累積分布関数)、そして右下はその正規 Q-Q プロットである。図1の上段は、得られた退院患者数分布が幾何分布かどうかを視覚的に確認するための図で、下段は同じ分布が対数正規分布に従っているかどうかを確かめるために描いたものである。図から、この分布は対数正規分布というよりも、どちらかといえば幾何分布に近いように思われる。そこで、分布の適合度検定を行った。有意水準として5%を設定し、対数正規分布の適合度検定については退院患者数の対数に対して Shapiro-Wilk 検定 (以下 SW-test と記す) [帰無仮説: 退院患者数の対数は正規分布に従う (以降、これを「退院患者数は対数正規分布に従う」と表現する)] を、幾何分布の適合度検定については Kolmogorov-Smirnov 検定 (以下 KS-test と記す) [帰無仮説: 退院患者数はパラメタが $\lambda = 0.0032$ の幾何分布に従う] を行った。検定の結果、SW-test では $p < 0.001$ となり、有意水準5%のもとで退院患

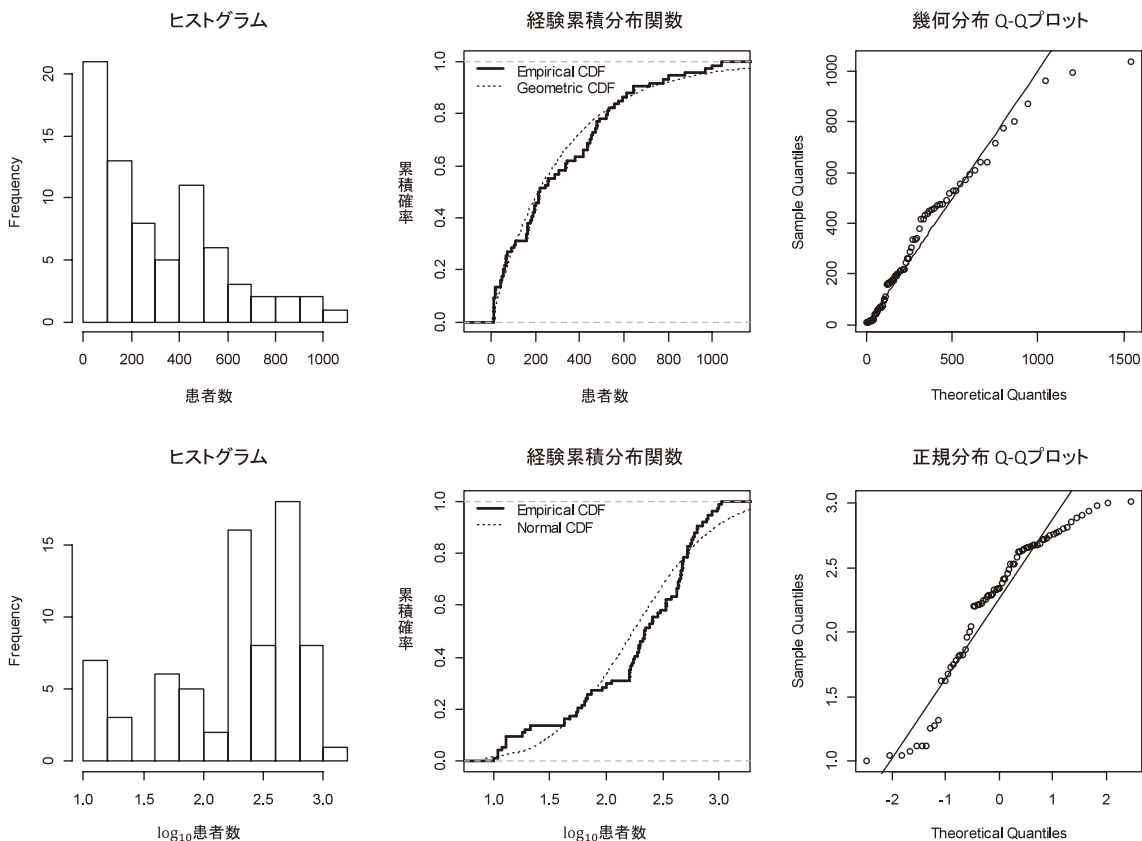


図1 北海道の循環器系疾患 (MDC05) の退院患者数分布 (平成25年度)

者数が対数正規分布に従うという帰無仮説が棄却された。一方、KS-testでは $p=0.3736$ となり、退院患者数が幾何分布に従うという帰無仮説は棄却されなかった。

北海道以外の都道府県に対しても循環器系疾患の退院患者数の分布に対して同様の検定を行った。その結果を表1に示す。なお、表中の y はその都道府県の循環器系疾患の退院患者数で、 n はその退院患者数が0でない病院の数である。そして、 λ はKS-testで用いる幾何分布のパラメタで、 $\lambda = n/y$ によって求めている。表から、すべての都道府県で幾何分布への適合度を検定したKS-testの結果は $p > 0.05$ となっており、5%有意水準のもとで循環器系疾患の退院患者数が幾何分布に従うという帰無仮説が棄却されないという結果となった。一方、対数正規分布への適合度を検定したSW-testでも、いくつかを除くほとんどの都道府県で $p > 0.05$ となっており、5%有意水準のもとで循環器系疾患の退院患者数が対数正規分布に従うという帰無仮説が棄却されないという結果となった。さらに、他の疾患領域に対しても適合度検定を行ったところ、ほぼ同様な結果が得られた。すなわち、ほとんどの都道府県において、有意水準5%のもとでは、退院患者数が幾何分布に従うとい

う帰無仮説も対数正規分布に従うという帰無仮説も棄却されないという結果である。

対数正規分布と幾何分布はともに左に偏って、右に裾が広がる、似たような形状をした分布である。そのため、KS-testが誤って対数正規分布を幾何分布と判定するケース(偽陽性 FP: False Positive)や、逆に、SW-testが誤って幾何分布を対数正規分布と判定するケース(偽陰性 FN: False Negative)が考えられる(ここでは、幾何分布と判定することを陽性としている)。図2は、そのような誤りが、サンプルサイズが少ない場合に生じることを示している。図2の上図はKS-testにおける偽陽性のサンプルサイズ依存性を、下図はSW-testにおける偽陰性のサンプルサイズ依存性を示している。5%有意水準のもとでは、 $p > 0.05$ になるサンプルサイズで誤りが起きていると解釈できるので、KS-testでは、サンプルサイズが200以下で偽陽性が生じており、SW-testでは、30以下で偽陰性が生じている。表1が示すように、大多数の都道府県で病院数 n (これがサンプルサイズに相当する)は30以下なので、有意水準5%のもとでは、ほとんどの都道府県において退院患者数は対数正規分布であるという帰無仮説も幾何分布であるという帰無仮説も棄却されないという表

表1 循環器系疾患 (MDC05) の退院患者数分布の適合度検定

都道府県	y	n	λ	p_{LN}	p_{GE}	都道府県	y	n	λ	p_{LN}	p_{GE}
北海道	22,908	74	0.0032	0.0000	0.3736	滋賀県	6,088	15	0.0025	0.0560	0.6095
青森県	4,028	13	0.0032	0.1115	0.9652	京都府	12,932	33	0.0026	0.1237	0.8583
岩手県	3,673	12	0.0033	0.6868	0.9316	大阪府	34,849	98	0.0028	0.0041	0.5259
宮城県	7,724	21	0.0027	0.9286	0.4788	兵庫県	21,525	62	0.0029	0.0664	0.2254
秋田県	2,317	15	0.0065	0.0520	0.7444	奈良県	6,100	16	0.0026	0.1177	0.3063
山形県	4,137	14	0.0034	0.7321	0.7284	和歌山県	3,713	13	0.0035	0.1791	0.7156
福島県	5,728	19	0.0033	0.3385	0.5390	鳥取県	2,285	9	0.0039	0.4991	0.9342
茨城県	8,149	22	0.0027	0.5224	0.5939	島根県	2,457	11	0.0045	0.1122	0.8146
栃木県	5,512	14	0.0025	0.4274	0.9080	岡山県	8,185	18	0.0022	0.3754	0.0962
群馬県	7,223	19	0.0026	0.1873	0.9497	広島県	9,596	27	0.0028	0.7446	0.8342
埼玉県	19,821	49	0.0025	0.0127	0.9117	山口県	5,433	20	0.0037	0.1075	0.4392
千葉県	21,547	42	0.0020	0.1545	0.9331	徳島県	3,804	10	0.0026	0.6356	0.8436
東京都	47,775	114	0.0024	0.0011	0.5302	香川県	3,511	15	0.0043	0.0541	0.5491
神奈川県	32,271	75	0.0023	0.0168	0.7623	愛媛県	4,119	13	0.0032	0.5921	0.7577
新潟県	5,473	21	0.0038	0.4123	0.9650	高知県	3,144	7	0.0022	0.0883	0.6954
富山県	3,647	15	0.0041	0.6596	0.9113	福岡県	22,805	67	0.0029	0.7169	0.0543
石川県	4,756	22	0.0046	0.3219	0.4691	佐賀県	2,489	10	0.0040	0.2106	0.6248
福井県	3,013	10	0.0033	0.7694	0.9048	長崎県	4,652	23	0.0049	0.1029	0.3551
山梨県	1,780	7	0.0039	0.9026	0.9836	熊本県	6,513	22	0.0034	0.4060	0.1184
長野県	8,108	29	0.0036	0.1046	0.9736	大分県	3,961	16	0.0040	0.0414	0.3462
岐阜県	6,695	27	0.0040	0.5832	0.0887	宮崎県	3,842	11	0.0029	0.5446	0.5972
静岡県	12,408	34	0.0027	0.0259	0.9175	鹿児島県	5,142	26	0.0051	0.8370	0.1363
愛知県	22,674	54	0.0024	0.0089	0.4892	沖縄県	5,265	17	0.0032	0.0863	0.6089
三重県	6,257	20	0.0032	0.3853	0.9589						

y, n, λ については本文中の説明を参照。 p_{LN} および p_{GE} は、それぞれ、SW-test, KS-test で退院患者数の対数正規分布性および幾何分布性を検定した際の p 値。また、太字斜字体にしてあるのは ROC 解析の判定で有意になったもの。

1の結果は信頼できない。

図3の左側は、図2の上側と同じ方法で、 $n=74$ の場合に、 $\lambda=0.0032$ に「対応」する対数正規分布が、同じパラメタ λ の幾何分布と同一の確率分布であるかどうかを KS-test で検定した際に得られた p 値のばらつき具合を示したものである。つまり、KS-test が誤って対数正規分布を幾何分布と判定する度合い (FP) を、p 値を尺度として見たものである (ここでの「対応」の意味については図2のキャプションを参照)。一方、右側は、与えられたパラメタ λ の幾何分布から乱数を n 個生成し、それが同じパラメタ λ の幾何分布と同一の確率分布であるかどうかを KS-test で検定した際に得られた p 値のばらつき具合を示したものである。つまり、KS-test が正しく幾何分布を幾何分布と判定する度合い (TP: True Positive) を、p 値を尺度として見たものである。もともと同じ分布なのだから、当然、帰無仮説「同一確率分布である」は棄却されず、p 値は左側に比べて大きな値をとる。図3はそれを示している。これを利用すれば、p 値の相对比较でその分布が対数正規分布か幾何分布かを判別できる。具体的には、与えられた p 値 (これを α とする) に対して感度

と特異度を求め、 α をパラメタとして変化させながら ROC 曲線を描き、感度 = 特異度となるような α を求め、KS-test で得られた p 値がそれよりも大きければ幾何分布と判定する。なお、図3を描くには、図2のキャプションにも書いたように、1,000回の適合度検定を行っているので、感度と特異度は次のようにして求めることができる。

$$\text{感度} = \frac{\text{Count}_{\text{geo}}(p > \alpha)}{1000}$$

$$\text{特異度} = \frac{\text{Count}_{\text{ln}}(p \leq \alpha)}{1000}$$

ここで、 $\text{Count}_{\text{geo}}(p > \alpha)$ は、図3の右側を描くときに行った KS-test のうち、p 値が α より大きくなった回数で、 $\text{Count}_{\text{ln}}(p \leq \alpha)$ は、左側を描くときに行った KS-test のうち、p 値が α 以下になった回数である。また、 α は 0~1 の値をとるパラメタである。こうして描いた ROC 曲線を図4に示す。

図4の破線は感度と特異度が一致する直線を表し、この直線と ROC 曲線の交点に対応する α が対数正規分布と幾何分布を分離する閾値である。同様にして SW-test に対しても ROC 分析を行い、本当に対

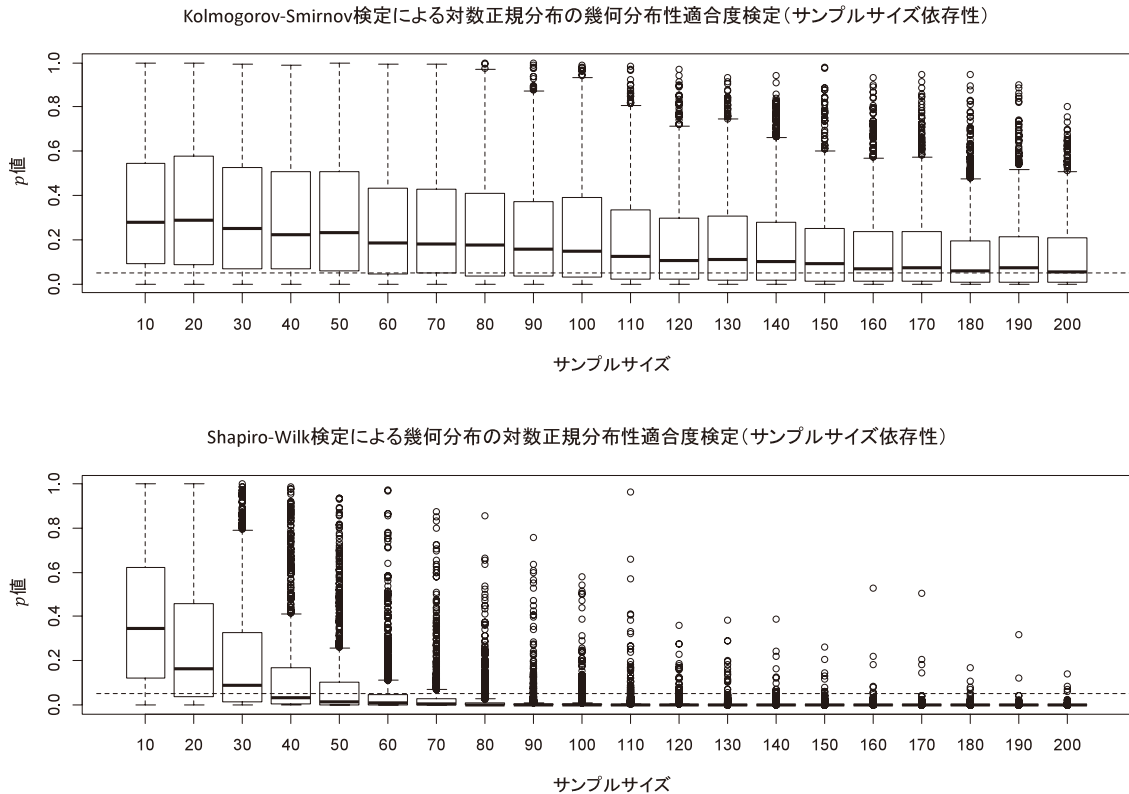


図2 適合度検定のサンプルサイズ依存性

上図は、与えられた対数正規分布の幾何分布性を Kolmogorov-Smirnov 検定で検定したときの p 値が、サンプルサイズにどのように依存するかを調べた図。まず、パラメタが $\lambda = 0.003$ の幾何分布から、指定したサンプルサイズだけ乱数を生成してその対数を求める。次いで、その平均値と標準偏差をパラメタとする対数正規分布（これをパラメタが λ の幾何分布に「対応」する対数正規分布と呼ぶことにする）から同じ数だけ乱数を生成し、それを Kolmogorov-Smirnov 検定を用いて、パラメタが $\lambda = 0.003$ の幾何分布と同一の確率分布かどうかを検定する（帰無仮説：与えられた分布は幾何分布である）。

下図は、与えられた幾何分布の対数正規分布性を Shapiro-Wilk 検定で検定したときの p 値が、サンプルサイズにどのように依存するかを調べた図。パラメタが $\lambda = 0.003$ の幾何分布から、指定したサンプルサイズだけ乱数を生成してその対数を求める。次いで、Shapiro-Wilk 検定でその正規性を検定する（帰無仮説：与えられた分布は対数正規分布である）。いずれも各サンプルサイズに対して1,000回の適合度検定を行って、そのときの p 値のばらつき具合をプロットしている。また、図中の破線は5%有意水準に対応する $p=0.05$ の直線を示している。

数正規分布なのか、それとも誤って幾何分布と判定したのかを分離する閾値を計算することができる。こうして、表1の各都道府県の n と λ に対してこれらの閾値を計算し、それによって各々の都道府県の退院患者の分布が対数正規分布と幾何分布のどちらになるか、あるいはそのどちらでもないかを判定した。表1で p 値を太字斜字体にしてあるのが ROC 解析の判定で有意になったものである。その結果、対数正規分布と判定された都道府県のは5、幾何分布は22、その両方は16、どちらでもないは4であった。両方の分布に従うというのはいり得ないので、解像度不足でこの方法では判定できないと見るべきで、実際にはいずれかの分布に従うものと考えられる。

循環器系疾患以外の疾患領域に対しても同様な解析を行ったところ、表2に示す結果を得た。この表から、ほとんどの疾患領域で、退院患者数が幾何分布に従うと判定された都道府県の方が多いことがわかる。特に耳鼻咽喉科系疾患（MDC03）、循環器系疾患（MDC05）、腎・尿路系疾患及び男性生殖器系疾患（MDC11）、女性生殖器系疾患及び産褥期疾患・異常妊娠分娩（MDC12）、外傷・熱傷・中毒（MDC16）などの疾患領域でこの傾向が際立っている。反対に、対数正規分布に従うと判定された都道府県が多い疾患領域は、多い順に、その他（MDC18）、乳房の疾患（MDC09）、そして血液・造血器・免疫臓器の疾患（MDC13）であった。

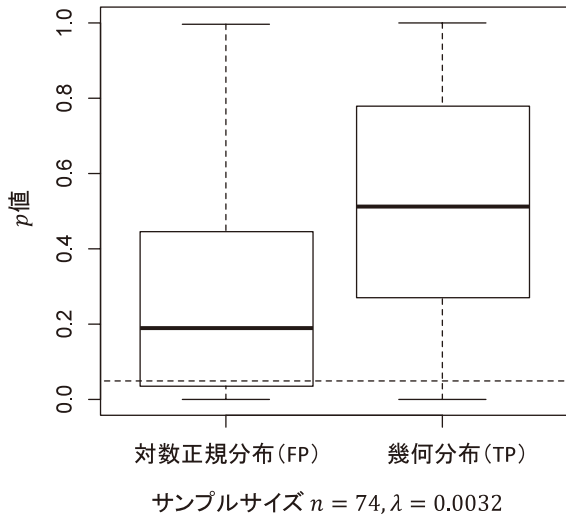


図3 幾何分布性に対する KS-test の検定結果

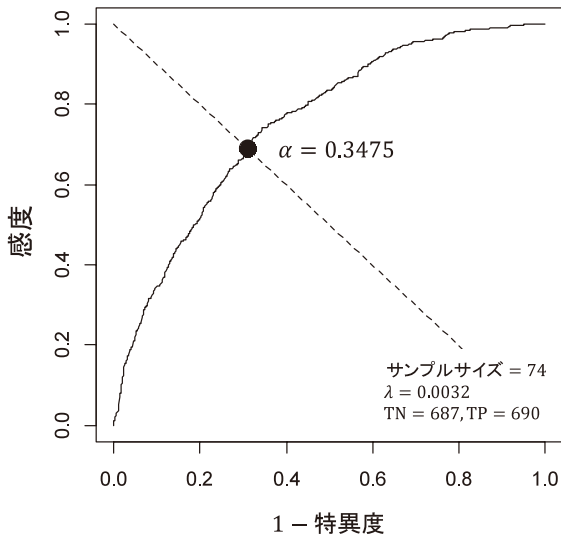


図4 KS-test の精度を評価する ROC 曲線

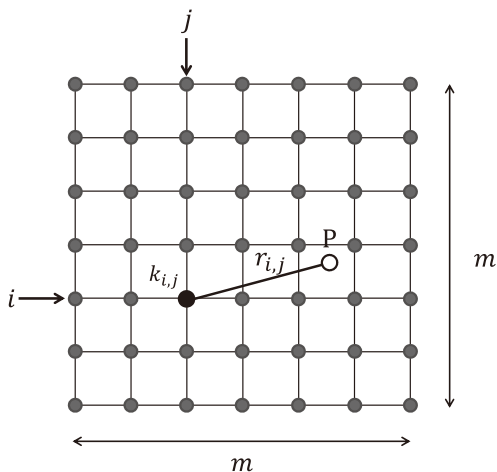


図5 格子病院モデル

表2 MDC 別に行った ROC 解析の結果

MDC	N_{LN}	N_{GE}	N_{BOTH}	N_{NTHR}
MDC01	1	20	17	9
MDC02	5	12	22	8
MDC03	1	23	12	11
MDC04	1	18	9	19
MDC05	5	22	16	4
MDC06	7	18	6	16
MDC07	4	20	16	7
MDC08	8	11	5	23
MDC09	13	9	10	15
MDC10	5	9	5	28
MDC11	0	25	6	16
MDC12	1	21	11	14
MDC13	9	6	3	29
MDC14	0	12	14	21
MDC15	-	-	-	-
MDC16	3	21	4	19
MDC17	-	-	-	-
MDC18	16	4	6	21

N_{LN} , および N_{GE} はそれぞれ ROC 解析によって退院患者数分布が対数正規分布, および幾何分布と判定された都道府県の数で, N_{BOTH} はその両方に, N_{NTHR} はそのいずれでもない判定された都道府県数.

4. シミュレーション

実証研究によって, 退院患者数の分布は, 幾何分布や対数正規分布など, 様々な様相を呈することがわかった. 退院患者数の分布がなぜこのような振る舞いをするのか, その発生メカニズムを明らかにするために確率モデルを考案し, シミュレーションを行った.

図5のような, 等間隔に並んだ格子の交差する点上に $(m+1) \times (m+1)$ の病院があるような格子病院モデル (以降, LHM: Lattice Hospital Model と記す) を考える. 今, 格子平面で表される地域内のある点 P に住んでいる患者が, 病院 (i,j) に引きつけられる (入院しようとする) 力が, 病院との距離に反比例し, 病院サイズ (ここでは入院患者数を病院サイズとする) が大きいほど強いとするモデルを考える. すなわち

$$F_{i,j} = \frac{1 + a \cdot k_{i,j}}{r_{i,j}} \dots (4)$$

という吸引力が病院 (i,j) と患者の間に働いているとする. ここで $k_{i,j}$ は病院 (i,j) の入院患者数, $r_{i,j}$ は病院 (i,j) と患者の居住地 P との距離で, a は吸引力の強さを表す正のパラメタである. そして, 患者はこの吸引力がもっとも強い病院に入院するものとする. この LHM を用いて次のようなシミュレーションを行った.

初期条件として, 時刻 $t=0$ に, すべての病院の

入院患者数を0にする。時刻 t に、この格子平面上のランダムな位置に患者が「発生」し、格子点上にある $n = (m+1)^2$ の病院のうち、吸引力 F_{ij} が最も強い病院 (i,j) に入院する。これによって、病院 (i,j) の入院患者が一人増えて k_{ij} から $k_{ij}+1$ になる。次に、時刻 $t+1$ に、また、格子平面上のランダムな位置に患者が発生し、同じように吸引力の最も強い病院に入院する。これを繰り返して、格子点上の全病院に入院している患者総数が y になったら終了する。この一連の処理を様々なパラメタ a の値に対して実行し、入院患者数の分布が a の値によってどのように変わるかを調べた。

図6は、 $m=14, y=45000$ に対して行ったシミュレーション結果である。上から順に、 $a=0.000, 0.048, 0.080$

の場合の結果で、左から、ヒストグラム、経験累積分布関数、そして分位数プロットを描いてある。中段中央の図中の鎖線は、パラメタ λ が $n/y=0.005$ の幾何分布の累積分布関数である。また、下段のグラフは、入院患者数そのものではなく、入院患者数の常用対数を描いたものである（したがって、これが正規分布の特徴を示せば、もとの入院患者数の分布は対数正規分布である可能性が高い）。下段中央の図中の鎖線は、その標本平均と標本分散を用いて得られた正規分布の累積分布関数である。これらの経験累積分布関数は、参照とする累積分布関数（それぞれ幾何分布と正規分布）と重なってほとんど見分けがつかない上、分位数プロットもほぼ参照直線上にある。これらの結果から、 $a=0.048$ の場合は幾何

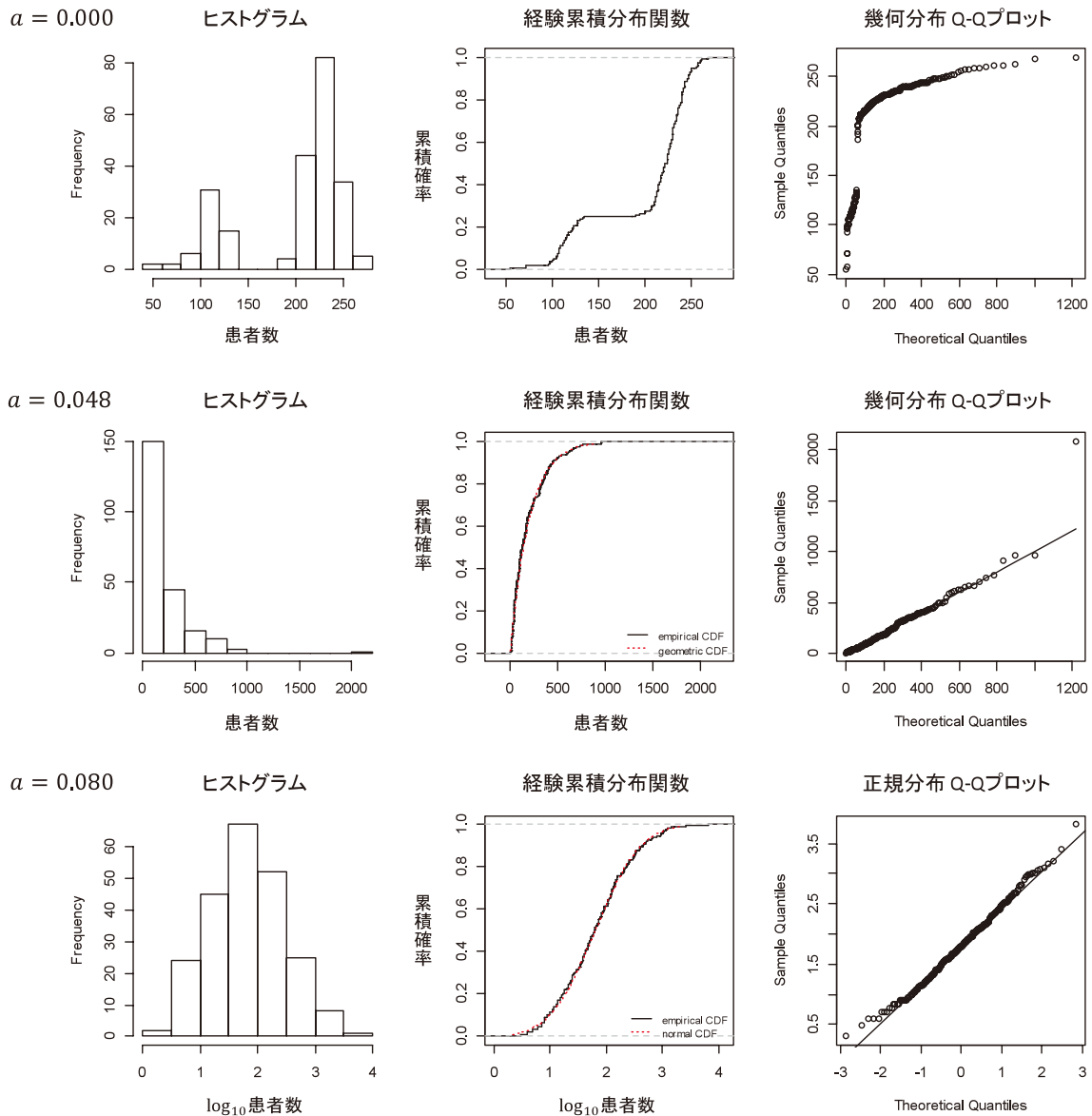


図6 格子病院モデルのシミュレーション結果

分布, $a=0.080$ の場合は対数正規分布と推測される。実際, $a=0.048$ の場合, KS-testによる幾何分布適合度検定を行ったところ, $p=0.577$ となり, この分布が幾何分布であることが棄却されなかった。また, 入院患者数の対数に対してSW-testで正規分布適合度検定を行ったところ, $p=0.009$ となって, 有意水準5%のもとで, 正規分布であることが棄却された(つまり, 入院患者数が対数正規分布であるという帰無仮説が棄却された)。同様に, $a=0.080$ の場合に対してもこれらの適合度検定を行った結果, それぞれ $p<0.001$ および $p=0.453$ となり, 入院患者数の分布が対数正規分布であることが示唆された。一方, $a=0.000$ の場合, KS-test, SW-testいずれも $p<0.001$ で, 有意水準5%のもとで, 幾何分布も対数正規分布も棄却された。ヒストグラムを見ると, 複数の単峰性分布が合成されているように見える。

このように, LHMでパラメタ a の大きさを変えて入院患者数に比例して大きくなる病院と患者の吸引力を強くしていくと入院患者数の分布が変わっていく。この変化をより明確に捉えるためにSW-test, KS-testの結果得られた p 値のパラメタ a 依存

性を描いたのが図7である。いずれも a の値を大きくしていくと, p 値が系統的に変化していることがわかる。なお, SW-testは, 入院患者数そのものではなく, その対数に対して行ったものである。有意水準を5%に設定した場合, 鎖線で描いた $p=0.05$ の直線よりも上では帰無仮説が棄却されないので, 図から $0.032<a<0.052$ で入院患者数の分布が幾何分布に, $0.052<a$ で対数正規分布になっていることが示唆される。そして, $a<0.032$ では, それらのいずれでもない分布になっているものと考えられる。

5. モデルの解析と評価

公開DPCデータを用いた実証研究から, 多くの都道府県において, 疾患領域別の受療患者数を尺度とした病院サイズが幾何分布や対数正規分布になっていることがわかった。そして, LHMを使ってシミュレーションを行うことにより, それを再現することができた。このモデルは, 患者と病院の間に式(4)で示す吸引力を仮定しているが, これは「患者は近くの規模の大きな病院に行きたがる」ということを最も単純に定式化したものである。そして,

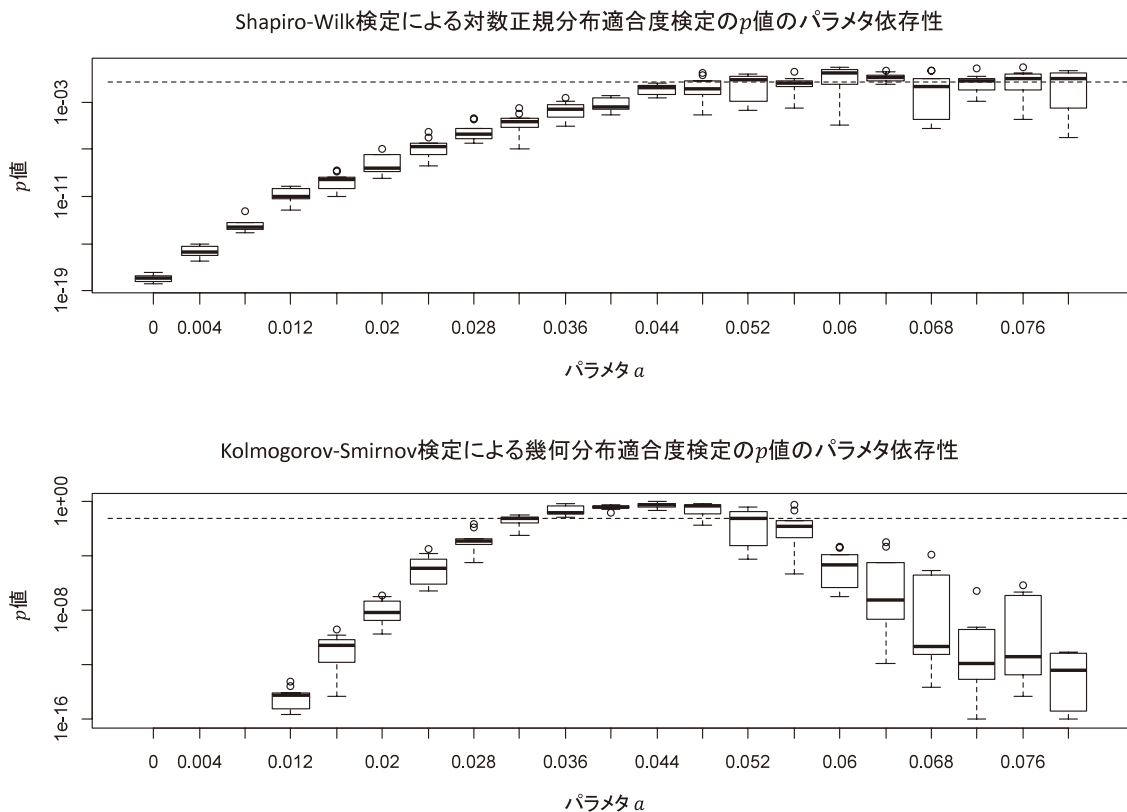


図7 適合度検定のパラメタ依存性

格子病院モデル(LHM)において, パラメタ a を0~0.08に0.004刻みで変化させた時の入院患者数分布に対してSW-test(上), KS-test(下)を行った際に得られた p 値をプロットしたもの。ここで, SW-testは入院患者数そのものではなく, 対数をとったものに対して行っている。図中の鎖線は有意水準を表す $p=0.05$ 。

パラメタ a は、距離か規模かどちらをより優先するかを表しており、 a が大きいほど病院の規模が優先されることを意味している。シミュレーションでは、 a が小さい、すなわち距離優先の場合は、入院患者数の分布は複数の単峰性分布からなっていた。この理由について考えてみよう。

今、 a が十分に小さくて $a \cdot k_{ij} \ll 1$ となる場合を考える。この場合、式 (4) から吸引力は $F_{ij} \sim 1/r_{ij}$ と近似でき、患者がその病院に入院するかどうかは病院からの距離によってのみ決まる。患者は格子平面上にランダムに発生すると仮定しているので、時刻 t における病院 (i, j) の入院患者数を $k_{ij}[t]$ とすると

$$k_{ij}[t] = k_{ij}[t-1] + \varepsilon_{ij}[t] \cdot \dots \quad (5)$$

となる。ただし、 $\varepsilon_{ij}[t]$ は次のような確率変数である。

$$\varepsilon_{ij}[t] = \begin{cases} 1 & p = \frac{l}{4 \cdot m^2} \\ 0 & p = 1 - \frac{l}{4 \cdot m^2} \end{cases}$$

ここで、 p は確率、 l は各々の病院を取り囲む格子の数で、格子平面のコーナーにある4病院では $l=1$ 、それらを除く周辺の $4(m-1)$ 病院では $l=2$ 、そして、格子平面内部にある $(m-1)^2$ 病院では $l=4$ である。初期条件 $k_{ij}[0]=0$ を使って漸化式 (5) を解くと

$$k_{ij}[t] = \varepsilon_{ij}[1] + \varepsilon_{ij}[2] + \dots + \varepsilon_{ij}[t]$$

となり、 $k_{ij}[t]$ はベルヌーイ列の和になる。これは、パラメタ $p = l/(4 \cdot m^2)$ の二項分布になる。したがって、図6に示す $m=14$, $y=45000$ の場合、全体では、パラメタが $p=1/784$, $1/392$, $1/196$ の3つの二項分布がそれぞれ4:52:169の比率で合成された分布関数になる。二項分布の平均値は $y \cdot p$ であるから、各々の平均値はそれぞれ、57, 115, 230となり、図6左上のヒストグラムとよく一致している。このように、 a が小さい、すなわち距離優先の場合は、入院患者数の分布は二項分布になることが示された。

次に、 a が大きい場合を考えてみよう。今、 $1 \ll a \cdot k_{ij}$ とすると、式 (4) から吸引力は $F_{ij} \sim a \cdot k_{ij}/r_{ij}$ と近似できる。したがって、漸化式は次のようになる。

$$k_{ij}[t] = k_{ij}[t-1] + \psi(a \cdot k_{ij}[t-1]/r_{ij}[t]) \cdot \dots \quad (6)$$

ここで、 $\psi(F_{ij})$ は、 F_{ij} が最大になる場合に1、それ以外では0をとる確率変数であるが、このままではこの漸化式を解くことはできないので、 $\psi(F_{ij}) \propto a \cdot k_{ij}[t-1]/r_{ij}[t]$ として、これ自体を確率とみなし、式 (6) を次のように変形する。

$$k_{ij}[t] = (1 + \varepsilon_{ij}^1[t]) k_{ij}[t-1] \cdot \dots \quad (7)$$

ここで、 $\varepsilon_{ij}^1[t] = C \cdot a/r_{ij}[t]$ は確率変数である (C は規格化のためにつけた定数)。式 (7) はジブラの法則を定式化した式(1)と同等な漸化式であるから、十分時間が経過したのちに病院 (i, j) の入院患者数 k_{ij} は対数正規分布に従うことになる。こうして a が大きい、すなわち病院までの距離よりも病院の規模が優先される場合、入院患者数の分布は対数正規分布になることが解析的にも示され、シミュレーション結果を裏付けることができた。では、 a がどの程度の大きさになれば入院患者数の分布が対数正規分布になるか見積もってみよう。先述したように $1/k_{ij} \ll a$ であるから、 k_{ij} をその平均値 $\bar{k}_{ij} = y/n$ で置き換えると $a \gg n/y$ になる。これを満たす a が n/y の10倍程度だとすれば、図6のシミュレーションの場合、すなわち $n=(m+1)^2=255$, $y=45000$ の場合、 $a \approx 0.05$ になる。これは図7の結果にほぼ一致する。つまり、 $a \geq 0.05$ で入院患者数の分布が対数正規分布になる。

それでは、 a が小さ過ぎず、また、大き過ぎない場合、すなわち病院に入院するうえで、病院までの距離と病院の規模が選択条件として拮抗するような場合はどうなるだろうか。それを明らかにするために、Suttonの方法⁷⁾ にならって、次のようなモデルを考える。

時刻 $t=1$ において入院患者が1名しかいない病院が1つだけあるとする。その後、時刻が $t=2, 3, \dots$ と進むにつれて、患者が一人ずつ発生し、確率 θ で「新しい病院」へ入院し、確率 $(1-\theta)$ で既に入院患者のいる病院 (以後、これを「既存の病院」と呼ぶことにする) へ入院すると仮定する。ただし、ここで言う「新しい病院」とは、まだ一人も入院患者がいない病院という意味である。今、時刻 t において、入院患者数が k である病院 (以後、これをクラス $[k]$ の病院と呼ぶ) の数の期待値を $\Lambda(k, t)$ と表すことにする。次に、患者が「既存の病院」に入院する確率は、病院サイズ (入院患者数) に拘わらず、等しく $(1-\theta)/n_t$ であるとする。ただし、 n_t は時刻 t における「既存の病院」の数の期待値で、 $n_t = 1 + \theta \cdot (t-1)$ である。以上から、クラス $[k]$ の病院に関する時間発展式は次のようになる。

$k \geq 2$ の場合 :

$$\Lambda(k, t+1) - \Lambda(k, t) = \frac{1-\theta}{n_t} \cdot \{\Lambda(k-1, t) - \Lambda(k, t)\}$$

$k=1$ の場合 :

$$\Lambda(1, t+1) - \Lambda(1, t) = \theta - \frac{1-\theta}{n_t} \cdot \Lambda(1, t)$$

ここで、時刻 t におけるクラス $[k]$ の病院の存在確率を $p(k,t) = \Lambda(k,t)/n_t$ とすると、 $p(k,t)$ に関する次の時間発展式が得られる。

$k \geq 2$ の場合：

$$p(k,t+1) + \theta t \cdot \{p(k,t+1) - p(k,t)\} = (1 - \theta) \cdot p(k-1,t)$$

$k = 1$ の場合：

$$p(1,t+1) + \theta t \cdot \{p(1,t+1) - p(1,t)\} = \theta$$

系が定常状態、つまり新しい患者が発生しても入院患者数の分布が変わらない状態に達していると仮定すると、 $p(k) \equiv p(k,t+1) = p(k,t)$ が成り立つので、 $p(k) = (1 - \theta) \cdot p(k-1)$ 、 $p(1) = \theta$ が得られる。この漸化式を解くと、 $p(k) = \theta(1 - \theta)^{k-1}$ という、パラメタが θ の幾何分布が得られる。また、時刻が $t = 2, 3, \dots$ と進むにつれて、患者が一人ずつ発生するとしているので、患者総数が y のときの病院数を n とすれば、 $n = 1 + \theta \cdot (y-1)$ が成り立つ。したがって、 y が十分大きければ、 $\theta \sim n/y$ と近似できるので、図6の中段のように、入院患者数の分布がパラメタ n/y の幾何分布に従うことになる。

6. 考察

まず、前節で議論した LHM の解析結果に基づいて、公開 DPC データを用いた実証研究の結果を考察する。LHM では、パラメタ a を 0 から次第に大きくしていくと、最初は二項分布だった入院患者数の分布が次第に幾何分布へと変わり、さらに a を大きくすると対数正規分布が現れる (図7)。そして、それぞれのパラメタの領域は、患者が病院を選択するときの基準 (病院までの距離と病院規模のどちらを優先するかという基準) に対応し、それぞれ、距離優先、両方考慮、規模優先であった。公開 DPC データを用いた実証研究の結果では、多くの疾患領域・都道府県において退院患者数の分布が幾何分布になっていた。これを LHM の観点から解釈すると、患者は病院までの距離と病院の規模 (あるいは症例実績) の両方を考慮して病院を選択していると解釈できる。これはきわめて妥当な選択基準なので、退院患者数分布の多くは幾何分布になっているのは理に適っている。一方、乳房の疾患 (MDC09) や血液・造血器・免疫臓器の疾患 (MDC13) などは、幾何分布よりも対数正規分布になる都道府県が多かった。これらの疾患領域は専門性が高いため、病院選択の基準として、より病院規模 (症例実績) が優先されているものと解釈できる。

次に、先行研究で述べた Gao and Chan の研究²⁾

との関連性について考察を行う。彼らは、米国の病院のサイズ分布は sub-lognormal であると主張した。sub-lognormal は彼らの造語で、正規分布と対数正規分布の中間的な分布のことを、彼らはそう呼んでいる。対数をとったときの分布が正規分布に比べて先が尖っていて (尖度 >3)、右に偏っている (歪度 <0) のがその特徴である。しかし、著者は、これは幾何分布あるいはそれに近い分布ではないかと考えている。というのも、幾何分布に従う乱数を生成してその対数を取り、尖度と歪度を計算すると尖度 >3 、歪度 <0 となり、彼らが言う sub-lognormal 分布と同じ特徴が見られることに加え、ヒストグラムと分位数プロットを描くと、彼らの論文²⁾にある Figure 1(b), (c) と非常によく似た図が得られるからである。

彼らは、また、彼らの実験データを説明するために手の込んだ確率モデルを作った。格子平面上に患者と病院を配置する点は LHM と同じであるが、彼らは自律的に活動する患者と病院の相互作用を取り入れ、より現実に近い形でシミュレーションを行った。たとえば、病院を決定するために患者は病院までの距離 (正確には病院を探す範囲) や病院のサイズだけでなく待ち時間も考慮に入れている。それだけでなく、現実の世界では当然考えないといけない患者の「退院」も考慮している。これに対して LHM は病院へ入るだけで、出ていくことを想定していない。さらに、彼らの病院エージェントは提供するサービスを状況に応じて最適化する機能を持っている。すなわち、待ち患者が多い場合はサービスステーション (たとえば診察室、医師、病床など) を増やし、少ない場合は減らす。このようなより現実に即したモデルを構築してシミュレーションを行った結果、彼らは sub-lognormal 分布を再現することに成功した。さらに、患者が病院を選択する際の基準として病院サイズや病院間の競合がどのように影響するかについての感度分析を行った。その結果、分布が sub-lognormal になるためには患者の大病院指向と適度な病院間の競争 (受診する病院を探す範囲をほどよい距離に設定すること) が重要な要因になっていることを見出した。ところが、この知見はいずれも LHM でも得られている。同じ知見を得るのに彼らの構築したエージェントモデルは少しコストがかかり過ぎているばかりか、複雑すぎて背後にあるメカニズムを捉えるには不向きである。その点、LHM はモデルのパラメタが1つしかなく、したがってより単純であるばかりか、サイズ分布がパラメタによってどのように変わっていくかを容易に、しかも低コストで追跡できる (図7)。さらに、

彼らのモデルでは、サイズ分布に関してこれまで先人たちによって蓄積されてきた知見（ジブラの法則や Simon 過程など）との関係性を議論できない。

その意味でも LHM の方が優れたモデルだと言える。

最後に、その LHM と Simon 過程の関係について考察する。Sutton は、Simon の提案したモデル¹⁸⁾に修正を加えることで企業のサイズ分布が幾何分布になることを示した⁷⁾。その際、Sutton は、Simon モデルにおけるジブラの法則「企業が次の市場機会を獲得する確率は、その企業サイズに比例する」を、「企業が次の市場機会を獲得する確率は、企業サイズの増加に伴って減少することはない」という条件（条件①と記す）に置き換えた。一方で、「一定の確率 θ で新しい企業が市場に参入する」という Simon が設定したもう一つの条件（条件②と記す）は維持したままにした。そして、Sutton は、条件①の特別な場合として「企業が次の市場機会を獲得する確率は、企業サイズに拘わらず同じである」という条件（これを条件①' と記す）を考え、これと条件②から企業サイズが幾何分布になることを示したのである（ジブラの法則をそのまま使うと企業分布はユール分布になる^{18,19)}）。著者は、あるパラメタ領域において、LHM から幾何分布が生まれることを、Sutton が提示したこの考え方に基づいて導いた。LHM から幾何分布が生まれるのは、パラメタ a が小さ過ぎもせず、大き過ぎもしない場合であった。これが Sutton の提示した条件①' と等価だと考えると、そのようなパラメタ a の領域が「病院に次の患者が入院する確率は、病院サイズに拘わらず同じである」状況を生み出していることになる。つまり、式 (4) において、絶妙なパラメタ a の値によって、距離による斥力と入院患者数による引力がうまく調和した状態、すなわち吸引力 F_{ij} がすべての既存の ($k_{ij} > 0$) 病院 (i, j) でほぼ等しくなるような状態が出現しているのかもしれない。

いずれにしても、LHM は、病院が格子状に配置されている、患者が病院を選択する基準が病院までの距離と病院のサイズ（規模）だけしかない、退院を考慮していない、などかなり単純化された現実離れたモデルであるにも拘わらず、実際の DPC データから得られる退院患者数の分布の特徴をうまく捉えている。逆に言えば、分布を形成する要因には様々な要因があるが、些末なものを取り除き、本質的なものだけを残すと LHM のようになるのかもしれない。

7. 結論

本研究では、公開 DPC データから病院サイズの分布を調べ、多くの場合、幾何分布になっていることを示した。そして、患者が病院に入院するプロセスをシミュレートする単純な確率モデルを考案して分布の発生メカニズムを調べた。その結果、患者が病院選択時に考慮する基準として、病院までの距離と病院の規模（症例実績）のいずれかに偏るのではなく、その両方をほどよく考慮した結果として幾何分布が現れることを解明した。逆に、病院の規模（症例実績）を優先した場合は対数正規分布になることが示唆された。これは、乳房の疾患（MDC09）や血液・造血器・免疫臓器の疾患（MDC13）など専門性が高く症例実績が豊富な病院ほど患者が集まりやすい疾患領域において実際のデータで確認された。

ところで、DPC データは我が国の全病院を網羅しているわけではないので、これらの結果には何らかのバイアスがかかっているかもしれない。これについては、全国にあるすべての病院に関するデータが入手できない以上、確認する術はないが、著者はそのようなバイアスはないと考えている。なぜならば、LHM において、たとえば格子点上の病院を一つ置きに間引きした場合を考えると、これは単に距離のスケールを2倍にただけ（式 (4) において、 r を $r/2$ で置き換えただけ）で、モデルの構造自体は変わっていないので、本質的には同じ結果が得られると考えられるからである。

従来、病院のサイズ分布にはあまり関心が寄せられてこなかった。それは、分布を求めるだけの詳細なデータがなかったからである。そのため、病院は病床数で粗く区分され、大雑把な統計処理しか行われてこなかった。しかし、DPC の登場で状況は一変した。DPC データを利用することで、疾患領域別や都道府県別に病院のサイズ分布を求めることができるようになった。本研究は、おそらく本邦で初めてそれを行ったものである。それだけでなく、病院のサイズ分布を幾何分布や対数正規分布などのパラメトリックな分布関数で表現できるようにした最初の研究でもある。特に、世界中で多くのサイズ分布について研究がされている中で多くの病院のサイズ分布が幾何分布になっていることを初めて示した。本研究で得られたこれらの知見によって、これまで以上に精緻な病院規模別の各種分析や統計処理が可能になるものと期待される。

文 献

- 1) 田中昌昭：病院の機能分化を測る新しい指標の開発．川崎医療福祉学会誌，**25**(1)，121–132，2015.
- 2) Gao B and Chan WKV：Sub-lognormal size distribution of hospitals：An agent-based approach and empirical study. *Winter Simulation Conference 2013*, 1443–1454, 2013.
- 3) Simon HA and Bonini CP：The size distribution of business firms. *American Economic Review*, **48**(4)，607–617，1958.
- 4) Hart PE and Paris SJ：The analysis of business concentration: A statistical approach. *Journal of the Royal Statistical Society. Series A (General)*, **119**(2)，150–191，1956.
- 5) Hart PE：The size and growth of firms. *Economica*, **29**(113)，29–39，1962.
- 6) Quandt RE：On the size distribution of firms. *American Economic Review*, **56**(3)，416–432，1966.
- 7) Sutton J：Gibrat's legacy. *Journal of Economic Literature*, **35**(1)，40–59，1997.
- 8) Hart PE and Oulton N：Growth and size of firms. *The Economic Journal*, **106**(438)，1242–1252，1996.
- 9) Eeckhout J：Gibrat's law for (all) cities. *American Economic Review*, **94**(5)，1429–1451，2004.
- 10) Eeckhout J：Gibrat's law for (all) cities: Reply. *American Economic Review*, **99**(4)，1676–1683，2009.
- 11) Levy M：Gibrat's law for (all) cities: Comment. *American Economic Review*, **99**(4)，1672–1675，2009.
- 12) Malevergne Y, Pisarenko V and Sornette D：Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E*, **83**, 036111, 2011.
- 13) Stanley MHR, Buldyrev SV, Havlin S, Mantegna RN, Salinger MA and Stanley HE：Zipf plots and the size distribution of firms. *Economics Letters*, **49**(4)，453–457，1995.
- 14) Axtell RL：Zipf distribution of U.S. firm sizes. *Science*, **293**(5536)，1818–1820，2001.
- 15) Cabral LMB and Mata J：On the evolution of the firm size distribution: Facts and theory. *American Economic Review*, **93**(4)，1075–1090，2003.
- 16) Gabaix X：Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, **114**(3)，739–767，1999.
- 17) Kesten H：Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, **131**(1)，207–248，1973.
- 18) Simon HA：On a class of skew distribution functions. *Biometrika*, **42**(3–4)，425–440，1955.
- 19) Newman MEJ：Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, **46**(5)，323–351，2005.
- 20) 小野寺夏生：Zipf 分布と Poisson 分布を統合する確率モデル．情報科学技術研究集会論文集，**20**，173–183，1984.
- 21) 小野寺夏生：Zipf-Bradford 分布．オペレーションズ・リサーチ，**31**(3)，144–151，1986.
- 22) 松葉育雄：複雑系の数理．初版，朝倉書店，東京，2004.
- 23) 厚生労働省：平成26年度 第5回 診療報酬調査専門組織・DPC 評価分科会．<http://www.mhlw.go.jp/stf/shingi/0000056344.html>，2014．(2015.8.31確認)
- 24) 厚生労働省：診断群分類 (DPC) 電子点数表について．<http://www.mhlw.go.jp/topics/2012/03/tp0305-02.html>，2013．(2015.8.31確認)

(平成27年11月6日受理)

On the Size Distribution of Hospitals: Empirical Study and Stochastic Model

Masaaki TANAKA

(Accepted Nov. 6, 2015)

Key words : size distribution of hospitals, stochastic model, DPC (diagnosis procedure combination), geometric distribution, computer simulation

Abstract

It is necessary to make an accurate understanding of the current state of the healthcare delivery system for the proper placement of regional healthcare resources and for the formulation of a regional healthcare plan. Indeed we can obtain the number of hospitals by the number of beds using the health statistics provided by the Ministry of Health, Labour and Welfare (MHLW), however, these tell us not only hospital size as facilities but rough data on a 10 to 100 beds basis. The true necessity for the realization of in-depth medical policies is not the facility-based hospital size but the patient-based hospital size. So the author investigated the size distribution of hospitals by prefecture and disease domain measured by the number of discharged patients using the Diagnosis Procedure Combination (DPC) data. As a result in most cases the size distribution of hospitals turned out to be geometric distribution. In order to elucidate the mechanism, a simple stochastic model was build and computer simulation was conducted. By the results of the simulation and the analysis of the stochastic model, it was shown that the size distribution of hospitals will be binomial when patients prefer the distance to hospitals to be short, and will be log-normal when they prefer the size of hospitals (the number of cases of disease) to be large, and will be geometric when they take into account the both. These tendencies were confirmed to be consistent with DPC data. While the size distribution of hospitals had not received great attention in the past, the current research not only clarifies it in detail but also represents it in a parametric distribution function such as geometric or log-normal for the first time. Herewith a variety of quantity necessary for a medical policy and a medical evaluation can be calculated statistically which will make various analysis and statistical works more elaborate than ever.

Correspondence to : Masaaki TANAKA

Department of Health Informatics

Faculty of Health and Welfare Services Administration

Kawasaki University of Medical Welfare

Kurashiki, 701-0193, Japan

E-mail : mtanaka@mw.kawasaki-m.ac.jp

(Kawasaki Medical Welfare Journal Vol.25, No.2, 2016 281 – 294)