Library and Information Science No. 27 1989

図書の貸出頻度を記述する負の二項分布モデルの演繹的導出とその一般化

A Deductive Derivation and a Generalization of a Negative Binomial Distribution Model Describing the Frequency of Book Circulation

岸 田 和 明
Kazuaki Kishida

Résumé

Recently, some models describing the frequency distribution of the book circulation are proposed. In a previous paper, authors also propose a new model of book circulation distribution which is based on a negative binomial distribution. This model is able to describe varians of book circulation distribution with time.

This paper has two purposes; 1) as a model of the book circulation, mathematically deducting a negative binomial distribution, 2) discussing the validity of the book circulation model which authors proposed and generalizing this model for the change with time.

Consequently, two components of a compound Poisson distribution, a Poisson distribution and a gamma distribution, are deductively derived by considering the nature of book loans, and a negative binomial distribution are deductively derived perfectly.

And the two parameters of this gamma distribution are assumed as functions of time variation for describing varians of book circulation with time. The validity of this assumption and the function of these parameters are discussed. As this function, two expressions are used; 1) the usual expression of obsolescence; a exponential function, 2) an alternative for the expression of obsolescence which is derived from a extended obsolescence model. As a result, a generalized model of book circulation distribution with a extended expression of obsolescence is constructed.

岸田和明：慶應義塾大学大学院文学研究科図書館・情報学専攻博士課程，東京都港区三田 2-15-45.
Kazuaki Kishida: Graduate School of Library and Information Science, Keio University, 2-15-45, Mita, Minato-ku, Tokyo.
1989年8月3日受付

I. はじめに

図書館における図書の利用状況を把握する1つの方法として，それを利用頻度を独立変数とする度数分布で記述するということがしばしば行われる。この場合の度数とは図書のタイトル数または冊数を指す。つまり，各利用回数ごとに図書を計数してそれを分布として捉えるのである。特に，各図書の利用頻度として館外貸出の頻度をとったものは「貸出頻度分布」と称され，これを確率分布を用いてモデル化する試みが多数なされている（このようなモデルを，本論では便宜的に「貸出頻度分布モデル」と称する)。

筆者は，以前に確率分布を用いて，この「貸出頻度分布」のモデル化を試み，時間パラメターを含んだ負の二項分布をベースとした貸出頻度分布についての新たなモデルを提案した[1]。そしてさらにそれを慶應義塾大学日吉情報センターの貸出データを用いて検証することによって，ある程度そのモデルの妥当性を確認した。

この時間パラメターを含む貸出頻度分布モデルとは，ある図書の集合の貸出頻度分布の経年的な変化を記述することを可能としたモデルである。つまりこのモデルに第1年次および第2年次の実際の貸出頻度分布の平均（すなわち図書一冊あたりの平均貸出数）と分散さえ与えれば，このモデルは第3年次以降の貸出頻度分布の予測を可能にする。これは図書館の計画・運営等に非常に役に立つであろう。これと同様な試みは，Burrell がすでに行っている[2]が，筆者の提案したモデルはそれをさらに改良したものとなっている。

その負の二項分布に基づく具体的なモデルは，以下のようなものである。まず第 t 年次の貸出頻度を確率変数 x_t とおき，その貸出頻度が r 回である確率を $P(x_t=r)$ と表記することにする。このとき，

$$P(x_t=r)=\binom{r+k(t)-1}{r}p(t)^{k(t)}[1-p(t)]^r \quad (1.1)$$

ここで，

$$k(t)=k_1\gamma^{t-1} \quad (1.2)$$

$$p(t)=\frac{k_1\gamma^{t-1}}{k_1\gamma^{t-1}+\mu_1\theta^{t-1}} \quad (1.3)$$

によって貸出頻度分布を表現する。ただし，γ, θ は数式を簡単にするため，$\gamma=e^{-c}$, $\theta=e^{-d}$ (c, d は定数) と置いたものであるから，(1.2), (1.3) 式は，

$$k(t)=k_1e^{-c(t-1)} \quad (1.2)'$$

$$p(t)=\frac{k_1e^{-c(t-1)}}{k_1e^{-c(t-1)}+\mu_1e^{-d(t-1)}} \quad (1.3)'$$

と表すこともできる。また，(1.2), (1.3) および (1.2)′, (1.3)′ 式中の k_1 は，第1年次における負の二項分布のパラメター k の値であり（すなわち，パラメター $k(t)$ の初期値)，μ_1 はやはり第1年次の分布の期待値（簡単にいうならば，蔵書一冊あたりの平均貸出数）である。

ここでモデル中に μ_1 が現れるのは，$p(t)$ が，負の二項分布において $k(t)$ が確定している場合の，最尤推定量として定義されているためである。その結果，$p(t)$ は $k(t)$ と μ_1 とに依存した関数となっている。

モデル中の k_1, μ_1 はデータより直接計算できるので，(1.1)式を貸出頻度分布モデルとして利用するためには，あとは (1.2), (1.3) 式中の γ, θ を推定するだけでよい。この推定法としては，積率推定法を利用した，

$$\hat{\gamma}=\frac{(s_1^2-\bar{x}_1)\bar{x}_2^2}{(s_2^2-\bar{x}_2)\bar{x}_1^2} \quad (1.4)$$

$$\hat{\theta}=\bar{x}_2/\bar{x}_1 \quad (1.5)$$

を提案した[1]。ここで，$\bar{x}_i, s_i^2, i=1,2$ は，それぞれ第 i 年次の貸出頻度分布の標本平均，標本分散である。

このモデルは，現実のデータにかなりよく適合した[1]が，かなり直感的あるいは経験的に導出されたものであり，その理論的背景や，いくつか置いたモデルの前提の妥当性などは十分論じることができなかった。

そこで本論文では，まず (1.1) 式の貸出頻度分布モデルを前論文とは逆に演繹的に導出することにより，このモデルの理論的な根拠を明らかにする。そのうえで，その結果に基づいて，モデルのより広い適用を可能にするために，モデルの一般化を試み，実際に，従来のモデル (1.1) 式をその特殊形として含む，時間変化に対してより一般的な貸出頻度分布モデルを提示する。また，その一般的なモデルを実際のデータに適用する方法を示す。

II. 貸出頻度分布モデルとしての負の二項分布の演繹的な導出

その数式のかたちを見れば明らかなように，(1.1) 式は，基本的には通常の統計学で問題にされている負の二項分布そのものである。つまり (1.1) 式は，負の二項分布をベースとして，図書の貸出頻度分布の時間変化を記述できるようにそのパラメターに工夫がなされたものとみなすことができる。

そのため，まず貸出頻度分布を記述する確率分布としての，この負の二項分布の妥当性を調べる必要がある。

確かに，負の二項分布が実際の貸出頻度分布のデータに適合するという事実はすでにいくつか報告されている (例えば，岸田ら[1]，Burrell[2]，Bagust[3] 等)。この意味で，負の二項分布の貸出頻度分布モデルとしての優位性はある程度は一般的に受け入れられている。

しかし，上に掲げた諸研究が負の二項分布を採用した根拠は，あくまでデータの観察から導かれた経験的なものに過ぎず，理論あるいは原理から演繹的に導出されたものではない。つまり，データから発案し，データによる検証で終わっている。

そこで本章では，この負の二項分布が貸出頻度分布をよく記述するということの理由あるいは原理を明らかにするために，この負の二項分布を貸出頻度分布として演繹的に導出する。これによって，先行研究の帰納的な根拠が補完されるであろう。

A. 各図書の貸出とポアソン過程

まず，ある一冊の図書を注目した場合に，その貸出回数がポアソン分布によって「近似的に」記述されることを示す（これは第I章で述べた貸出頻度分布ではないことに注意)。これを最初に提唱したのは Morse[4] あるいは Buckland et al.[5] であるが，これらはオペレーションズ・リサーチ等の分野で知られている事例，特に，待ち行列の理論とのアナロジによって導入されたものである。しかし，図書の貸出という事象の特質とポアソン過程の定義とを照らし合わせれば，ポアソン分布が各図書の貸出回数を記述する理由は明確である。それを以下に述べる。

各図書の貸出の最も基本的な特徴は，「各図書の貸出回数は，確実に1回ずつ増加し，その累積回数が途中で減少するようなことは絶対に起こり得ない」ということである。一方，ポアソン過程は，“確率連続な加法過程 $X(t)$, $a<t<b$ の見本過程がほとんど高さ1の飛躍のみで増加する右連続階段関数である”[6] ような確率過程である。これらを比べればわかるように，貸出の基本的な特徴はポアソン過程の定義にほぼあてはまる。

しかし，このポアソン過程の定義中の「加法過程」に注意する必要がある。“(確率過程 $X(t), a<t<b$ に関して) $a<t_1<\cdots<t_n<b$ なる任意の $\{t_k\}$ に対して，$X(t_k)-X(t_{k-1}), k=2,3,\cdots,n$ が独立であるとき $X(t)$ は加法過程である”[6]から，図書の貸出がこの加法過程となる条件を満たしているかどうか調べなければならない。

加法過程の定義中の $X(t_k)-X(t_{k-1})$ が独立だという条件は，簡単に言えば，ある一冊の図書のある貸出がそれ以前のその図書の貸出と独立に生起するということである。例えば，他の事例として，貸出カウンターへの利用者の到着のような場合は，それぞれの到着はまったく無関係に起こると考えられるので，加法過程となるだろう。しかしここで問題としている一冊の図書の貸出の場合，前の貸出からの「返却」が行われなければ次の貸出は起こり得ないから，各貸出が完全に独立であるとは言いきれない。例えば仮に一度貸出するとその図書を一か月以上返却できないような図書館が存在するとすれば，この図書館においては，各貸出はその次の貸出の生起に著しい影響を与えることとなり，もはや各貸出が独立と仮定することは困難になろう。

しかしながら多くの場合，この例のような極端な場合はあり得ず，返却が無作為に（自由に）行われるという条件が満たされれば，「近似的に」各貸出が独立だと仮定してさしつかえないと考えられる。なぜなら，返却が無作為に行われれば，貸出の独立性に大きな影響は与え

ないと推測できるからである。

したがって，この各貸出の独立性さえ成り立てば，上記のポアソン過程および加法過程の定義との比較から，各図書の貸出が「近似的に」ポアソン過程になると結論できる。そこで，ポアソン過程の確率変数の分布はポアソン分布になる[6]から，各図書の貸出回数がポアソン分布で記述されることが演繹的に導出されたことになる。ただし，各貸出の独立性には上で述べた条件が必要となることに注意する必要がある。

B. 複合確率分布としての貸出頻度分布

1．複合確率分布

前節で導出したポアソン分布は，ある一冊の図書の貸出回数を記述するものであるから，貸出可能な図書全体の度数分布である貸出頻度分布を記述するための確率分布を導出するには，複合確率分布の理論を援用する必要がある。つまり，ポアソン分布のパラメター λ は各図書の貸出回数の期待値（あるいは潜在的な貸出回数の平均値）を意味するが，この各図書の λ が蔵書全体としてある確率分布にしたがって分布していると仮定して，下に述べる複合確率分布の定義中の数学的操作を行うのである。これによって貸出頻度分布が得られる。複合確率分布（あるいは複合分布）の定義は，“確率変数 X の確率密度関数が，母数 λ を用いて $f(x,\lambda)$ と表せるときに，λ それ自体がまた確率分布に従い，密度関数 $g(\lambda)$ をもつ［場合に］･･･複合分布は λ が離散変数の場合のとき，$\sum_{i=1}^{k} f(x;\ \lambda_i)g(\lambda_i)$ で表され，λ が連続的変数のとき $\int f(x;\ \lambda)g(\lambda)d\lambda$ で表される”[7] である。今回の場合は $f(x;\lambda)$ がポアソン分布になるわけである。

貸出頻度分布のモデル化にあたって，この複合確率分布を適用するということは，Burrell[2], Bagust[3] をはじめ，Gelman and Sichel[8], Hayse[9] などの多くの研究者が行ってきた。これは，各図書によって「有用性」あるいは「人気」に差異があり，その結果，各図書の貸出回数の分布が違っているという図書の貸出の状況が，複合確率分布の構造によく似ているためである。

2．ガンマ分布の演繹的な導出

そこで，貸出頻度分布を記述する確率分布を演繹的に導出するにあたっての次なる問題は，上の複合確率分布の定義中の $g(\lambda)$ の特定である。

そのために，まず $g(\lambda)$ としてどのような要件が必要か考えてみる。既に述べたように，λ は各図書の貸出回数の期待値として意味づけられる。一方，図書の実際の貸出頻度分布は，非常に「ゆがんだ(skew)」分布になることが経験的に知られている。この場合の「ゆがんだ」とは，度数分布の左極にほとんどの度数が集中する一方で，分布の右側の度数はなかなか0とならず，分布が長い尾を引くような形状になっていることを指す。つまり図書の貸出に関して言うならば，貸出回数が0回や1回しかないような，利用頻度の低い図書が多数存在する一方で，同時に，貸出回数が非常に多い図書もまた少数ながら存在するような状態が「ゆがんだ(skew)」と表現されるのである。このような貸出頻度分布の形状は，各図書の貸出回数の期待値である λ の分布から生じると考えられる。なぜならば，もし λ が全図書で一定であると仮定するならば，貸出頻度分布が，経験的に知られてるような極端な「ゆがんだ」形状にはなり得ないからである。したがって，$g(\lambda)$ はある程度「ゆがんだ」分布になっていると想定される。

ところで，このような「ゆがんだ」形状を示すデータは，Bradford の法則や Zipf の法則あるいは Lotka の法則の対象として従来図書館・情報学分野において研究が進められてきたものである。このいわゆるビブリオメトリックスの法則が成立する（すなわち「ゆがみ (skewness)」が生じる）原理については Simon[10] や Price[11] の「成功が成功を育む」という理論が知られている。これは確率変数が $x \rightarrow x+1$ に増える確率が，その増殖前の x が大きいほど高くなるということであり，確率過程論では一般に「線形増殖過程」と呼ばれている現象を示している（正確には，増殖の度合いを示すパラメターが x の一次式で表されるものが「線形増殖過程」である）。

したがって $g(\lambda)$ を考えるときにも，この理論を適用してみることが考えられる。ただし，λ は離散変数とは限らないので，Simon[10] や Price[11] のビブリオメトリックスの法則の導出法は直接利用できない。そこで，その理論を援用しながらも独自に $g(\lambda)$ の導出の理論を構成する必要がある。その $g(\lambda)$ の演繹的な導出の過程を以下に述べる。

さて，貸出頻度分布はある一定期間に利用された図書の回数の分布であるから，その成立までにはある時間的な経過を経ている。さらに具体的に言うならば，貸出回数の計数の開始時点から，1時間後の貸出頻度分布，1日後の貸出頻度分布，1か月後の貸出頻度分布がそれぞれ成立しているはずであり，つまり貸出頻度分布は時間に依存すると考えられる。

このように考えれば，その貸出頻度分布を決定づける

各図書の貸出回数のパラメターの分布 $g(\lambda)$ も，時々刻々と変化していくと考えられる。そこで，ある時点 t での $g(\lambda)$ を $g(\lambda;t)$ とする。この $g(\lambda;t)$ は確率過程とみなすことができる。

ここで，瞬間的な $g(\lambda;t)$ の変化，すなわち時間微分 $\partial g(\lambda;t)/\partial t$ を考える。そして，この微小時間 dt のあいだの $g(\lambda;t)$ の変化をモデル化し，それを t で積分することによって，$g(\lambda;t)$ を求めることとする。

ただし本論文は，観測の開始から1時間や2時間程度経過した後の貸出頻度分布ではなく，少なくとも1年間以上経過した時点での貸出頻度分布を問題にするため，$g(\lambda;t)$ を導出してから，$t\to\infty$ として定常分布 $g(\lambda)$ を導くことにする。つまり，本論文が対象とする貸出頻度分布は，あくまで一定の期間が経過して，その分布の形状の変化が落ち着いた時点の分布であり，この理由から $t\to\infty$ なる操作が必要となる。

この $\partial g(\lambda;t)/\partial t$ のモデル化にあたっては，物理学，生物学等でよく用いられるマスター方程式を利用する。なぜなら，このマスター方程式は確率的現象の時間発展を根本的・本質的に記述するモデルであって，非常に広い応用性を持っているからである。

このマスター方程式で記述するために，確率変数 λ の任意のある2つの状態を λ_1, λ_2 と書く。さらに微小時間 dt で図書の貸出回数の期待値が λ_2 から λ_1 へ遷移する確率密度を $W(\lambda_1, \lambda_2)$ とする。この遷移確率密度 $W(\lambda_1, \lambda_2)$ を用いると，$\partial g(\lambda;t)/\partial t$ は，

$$\frac{\partial g(\lambda;t)}{\partial t}=\int\Big\{W(\lambda_1,\lambda_2)g(\lambda_2,t)-W(\lambda_2,\lambda_1)g(\lambda_1,t)\Big\}d\lambda_2 \qquad (2.1)$$

とモデル化できる。これがマスター方程式である。$W(\lambda_1, \lambda_2)g(\lambda_2, t)$ は，時間 t に確率変数 λ が状態 λ_2 にあった確率密度が状態 λ_1 へと遷移する「流入」(λ_1 から見た場合，この遷移は確率密度の流入である）を示す。逆に $W(\lambda_2, \lambda_1)g(\lambda_1, t)$ は時間 t に状態 λ_1 にあった確率密度が状態 λ_2 に遷移する「流出」である。すなわち，状態 λ_2 から状態 λ_1 に流入する確率密度から，状態 λ_1 から状態 λ_2 へと流出していく確率密度を引き，その流入源・流出先としてすべての状態を考えるために，λ_2 に関して積分しているのが (2.1) 式である。積分は全区間すなわち全状態について行うから，(2.1) 式には確率密度が流入・流出せずに状態 λ_1 に留まっているという状況さえ含んでいる。したがって，この (2.1) 式は $\partial g(\lambda;t)/\partial t$，つまり微小時間での $g(\lambda;t)$ の変化を，非常に一般的に余すところなく記述していることになる。

さて，(2.1) 式は数学的には扱いにくいので，通常次のように近似される[12]。

$$\frac{\partial g(\lambda;t)}{\partial t}=-\frac{\partial}{\partial\lambda}[a_1(\lambda)g(\lambda;t)]+\frac{1}{2}\frac{\partial^2}{\partial\lambda^2}[a_2(\lambda)g(\lambda;t)] \qquad (2.2)$$

これはコルモゴロフの前向き微分方程式あるいはフォッカープランク方程式と呼ばれるものである。この変形は λ の微小時間での変化が小さい場合のみ有効であるが，図書の貸出回数は dt のような短時間でいきなり急増したり激減することは不可能であるから，(2.1)式を(2.2)式で近似することは，この場合特に問題ない。なお a_n, $n=1,2$ は，

$$a_n=\int(\lambda_1-\lambda_2)^n W(\lambda_1,\lambda_1-\lambda_2)d(\lambda_1-\lambda_2) \qquad (2.3)$$

である。すなわち，a_n は λ の変化の n 次のモーメント（平均，分散に相当する）である。この a_n は微小時間における $g(\lambda;t)$ のふるまいを決定する重要な要因であり，したがって最終的な $g(\lambda;t)$ の分布を決める要因でもある。つまり，この a_1, a_2 の具体的な関数形を決めることが，各図書の貸出回数のパラメター λ の分布 $g(\lambda)$ を導くのに，中心的な役割を果たす。

しかし，(2.2)式では，a_1, a_2 が λ の偏微分の項のなかに入っていて，数学的にまだ扱いにくいので，(2.2)式を

$$\frac{\partial g(\lambda:t)}{\partial t}=a_1(\lambda)\frac{\partial}{\partial\lambda}g(\lambda:t)+\frac{1}{2}a_2(\lambda)\frac{\partial^2}{\partial\lambda^2}g(\lambda;t) \qquad (2.4)$$

のように変形する。(2.4) 式はコルモゴロフの後向き微分方程式であり，ある条件のもとで[13]，その解は (2.2) 式の解に一致することが知られている。この条件は，注13) に示したように，貸出回数の期待値 λ が非負であるために問題なく成立する。そこで，今後は (2.2) 式ではなく，数学的に扱いやすい (2.4) 式を用いる。

次に実際に (2.4) 式の a_1, a_2 を具体的に決定するために，既に述べた Simon[10], Price[11] の「成功は成功を育む」という原理，すなわち線形増殖の原理を適用する。なぜならばこの a_1, a_2 に線形増殖性を組み込むことにより，$g(\lambda;t)$ はビブリオメトリックスの諸法則と同様に

「ゆがみ」を記述することができるようになると考えられるからである。そこで，

$$a_1(\lambda)=u-v\lambda \qquad (2.5)$$
$$a_2(\lambda)=2q\lambda \qquad (2.6)$$

とおく。$a_1(\lambda)$, $a_2(\lambda)$ ともに λ の一次式になっており，これを(2.2)あるいは(2.4)式に代入すると(2.2)，(2.4)式は，線形増殖モデルとなる。つまり，確率変数 λ の確率密度が，λ の大きさに比例して v の割合で増加していき，そしてその増加は λ が大きくなるほど q の割合で分散するモデルが与えられる。ただし，(2.5)式には定数項 u が加えられているので，λ が小さければ，u の値の大きさによっては逆に λ の確率密度が減少することもありうる。なお(2.6)式中で q に2を乗じてあるのは，(2.4)式の右辺第2項の1/2を消去するためである。

(2.5)，(2.6)式を使えば，(2.4)式は，

$$\frac{\partial g}{\partial t}=(u-v\lambda)\frac{\partial g}{\partial \lambda}+q\lambda\frac{\partial^2 g}{\partial \lambda^2} \qquad (2.4)'$$

となる。この微分方程式(2.4)′を解き，$t\to\infty$ とすると

$$g(\lambda)=\frac{(v/q)^{u/q}}{\Gamma(u/q)}\lambda^{u/q-1}e^{-(v/q)\lambda} \qquad (2.7)$$

となることが知られている[14)]。観測の最初の時点のごく短い期間での，パラメーター t に依存する貸出頻度分布のモデルも導くことも可能であるが，本論文では $t\to\infty$ として定常分布を導いた。

さて，(2.7)式において，u/q を k, v/q を b に置き換えると，(2.7)式は，

$$g(\lambda)=\frac{b^k}{\Gamma(k)}\lambda^{k-1}e^{-b\lambda} \qquad (2.8)$$

となる。これは，パラメーター k, b をもつガンマ分布である。したがって，線形増殖の仮定から $g(\lambda)$ としてガンマ分布が演繹的に導出されたことになる。

このガンマ分布は，各図書の貸出をポアソン過程とした Burrell[2)], Bagust[3)] が，そのパラメーター λ の確率分布として採用したものであるが，その根拠はあくまでデータから類推された経験的なものであった。しかし，ここでの議論より，λ がビブリオメトリックスの法則と同様に線形増殖の原理によって分布を変えていくという仮定だけでガンマ分布が理論的に導かれることが明らかになった。

3．負の二項分布の導出

上で述べた，複合確率分布 $\int f(x;\lambda)g(\lambda)d\lambda$ において $f(x;\lambda)$ がポアソン分布，$g(\lambda)$ がガンマ分布であることが導出されたので，次にこの積分を実行することにより，負の二項分布，

$$P(x)=\binom{x+k-1}{x}\left(\frac{b}{b+1}\right)^k\left(\frac{1}{b+1}\right)^x$$
$$=\binom{x+k-1}{x}p(1-p)^x \qquad (2.9)$$

ただし，$p=b/(b+1)$

が得られる[2)]。つまり，(1.1)式で表される貸出頻度分布モデルが負の二項分布をベースとしていることが，演繹的に導出された。

Ⅲ．貸出頻度分布の時間変化

A．貸出頻度分布モデルにおける時間変化

既に第Ⅰ章で述べたように，(1.1)式は貸出頻度分布の経年変化を記述することが可能である。

この「経年変化」あるいは「時間変化」は，貸出頻度分布のモデル化にとって非常に重要な要因である。なぜなら，図書の利用は経年的に減少することが知られているからである(いわゆる「obsolescence」)。このことは，貸出頻度分布が少なくとも一定期間は定常とはなりえないことを意味している。

したがって，もしモデルが時間変数 t を含まないものならば，そのモデルは各時点での貸出頻度分布を別個に「記述」することができても，将来を「予測」することは不可能であり，実際の図書館運営・図書館経営におけるそのモデルの価値は大幅に減少する。

そこで，貸出頻度分布モデルに時間変数 t を含めて，モデルが時間変化に対応できるようにする必要がある(ここでの t は，前章のガンマ分布の導出においてとった時間 t とは異なる。前章の t は非常に微小な時間を示すものであり，ガンマ分布を定常分布として導出したために，パラメーターとしてはモデル内には含まれない。一方この章で問題とする t は，1年間あるいは2年間，3年間のような非常に長い時間である)。

この時間変化を最初に貸出頻度分布モデル内に取り込んだのは，Morse[3)] である。彼のモデルは，基本的には時間的な変化を記述するマルコフ連鎖モデルであるが，Morseは，その遷移確率を定義するにあたって，その遷移確率の中に，obsolescence の要素を組み込んだ。

一方，ポアソン過程モデル（すなわち，負の二項分布モデル）も確率過程の一種であり，もともと時間変化を記述するモデルであるが，そこに図書の経年的な利用減少の要因を含めたのは Burrell[2] が最初である。彼は，このために，確率過程論の非定常ポアソン過程の理論を利用した。すなわち Burrell は期間 $[0, t]$ における各図書の平均貸出回数を記述するポアソン分布のパラメターを，

$$m(\lambda, t)=\int_0^t \lambda(s)ds$$

とおき，さらに，$\lambda(s)$ の具体的なかたちとして，

$$\lambda(s)=\lambda_0 e^{-as} \tag{3.1}$$

を仮定した[2]。これは，図書館・情報学分野においてよく知られた obsolescence の数式による表現であり，多くの経験的なデータから導出されたものである。この表現では自然対数の底である e のパラメターが負になっているので，(3.1) 式は時間変数 s の増加にしたがって，λ が減少していくようになっている。

しかし，この Burrell のモデルでは，現実の貸出頻度分布の時間変化には対応できないことが明らかになった[1]。この原因は，obsolescence の要因をポアソン分布のパラメター λ に含めるだけで，貸出頻度分布の時間変化に対応させようとしたことにある。

それに対してモデル (1.1) 式では，この欠点を補う工夫がなされている。本章では，まずその妥当性や問題点等を論じる。そしてさらに時間変化に関して拡張された一般的なモデルを提示する。

B. 時間の経過にしたがって変化するガンマ分布

1．ガンマ分布を時間変化させる妥当性

(1.1) 式のモデルを貸出頻度分布の時間変化に対応するようにするためには，図書の貸出回数の期待値の分布を示すガンマ分布のパラメターを時間関数とおいて，このガンマ分布自体を時間変化させることが考えられる。つまり，ガンマ分布 (2.8) 式中のパラメターを $k(t)$, $b(t)$ に置き換えて，

$$g_t(\lambda)=\frac{b(t)^{k(t)}}{\Gamma[k(t)]}\lambda^{k(t)-1}e^{b(t)\lambda}$$

としておいてから，複合確率分布 $\int f(x;\lambda)g_t(\lambda)d\lambda$ を計算するのである。パラメターを t の関数と置いてもこの積分には影響しないから，

$$\begin{aligned}P(x;t)&=\binom{x+k(t)-1}{x}\left(\frac{b(t)}{b(t)+1}\right)^{k(t)}\left(\frac{1}{b(t)+1}\right)^x\\&=\binom{x+k(t)-1}{x}p(t)^{k(t)}(1-p(t))^x\end{aligned} \tag{3.2}$$

ここで $p(t)=\dfrac{b(t)}{b(t)+1}$

のように，得られる確率分布は負の二項分布のままであり，(2.9) 式の k, b, p をそれぞれ $k(t)$, $b(t)$, $p(t)$ に置き換えたものとなる。

Burrell モデルでは，図書の貸出回数の期待値の蔵書全体での分布であるガンマ分布の形状は，時間の経過に対して全く変化せず，そのスケールだけが λ の減少に対応して縮小していく。したがって前節で述べたように，貸出頻度分布の時間変動を λ のみで対応せざるを得ず，現実のデータに適合しきれない。それに対して，ガンマ分布のパラメターを時間関数とするモデルは，より柔軟性がある。つまり，そのスケールだけでなく，平均貸出回数の分布の形状自体をも変化させることができるために，現実の貸出頻度分布の変化に柔軟に対応する能力を持っているのである。例えば，大学図書館においては，その利用者母集団は，入学，卒業あるいはキャンパスの移動などにより短期間でかなりの部分が入れ替わる場合が多い。このような利用者母集団自体が変わってしまうような，変化のはげしい状況に対しては，図書の貸出回数の期待値の分布の形状が変化しないモデルでは，十分対応できないと考えられる。したがって，図書館の現実を考えた場合，より高い柔軟性を得ることのできる，ガンマ分布を時間変化させるモデルのほうが望ましいと結論できる。

2．ガンマ分布のパラメターの関数形の特徴

そこで，次なる問題は $k(t)$, $b(t)$ の関数形を決めることである。ただし $b(t)$ に関しては，(3.2)式に示されているように $p(t)=b(t)/[b(t)+1]$ である。さらに第I章で述べたように，$p(t)$ は $k(t)$ が確定している場合の負の二項分布の最尤推定量として定義されている。そこで，$p(t)$ は $k(t)$ に大きく依存することになる（(1.3) 式参照。$p(t)$ は $k(t)$ と平均貸出数 $\bar{x}_t$ によって決まる)。したがって時間変化する貸出頻度分布モデルにおいて大きな役割を果たすのは，$k(t)$ の関数形ということになる。

(1.1) 式のモデルでは，Burrell と同様に負の指数関数を使って，$k(t)$ を (1.2)′ 式のようにおいた。これは，(1.1) 式を構築した時点では，この貸出頻度分布モデルにおけるガンマ分布の意味が十分明らかにされていなか

ったために，$k(t)$ がまったく未知の関数だったからである。そこでとりあえず Burrell と同じく，$k(t)$ に obsolescence の数式を用いたのである。

しかしながら，前章でガンマ分布を導出した際に，パラメター k, b の持つ意味がはっきりしたので，ここでは前章の成果を用いて，$k(t)$ の関数形を決定することが可能となった。

前章で示したように，$k=u/q$ および $b=v/q$ であった。この u, v はそれぞれ (2.5)，(2.6) 式で示されているとおり，u は変数 λ がある状態からある状態へ遷移する度合いを決める一次式中の定数項であり，v はその一次式において，λ の大きさの影響力を決めるパラメターである。したがって，貸出が減少するためには，v が小さくなる必要があることは明らかである。なぜならば v が小さくなればそれだけ大きな貸出回数の期待値を持つ図書が減るからである。そうすると，u も同様に小さくならなければならない。もしそうでなければ，(2.4) 式において，確率変数 λ のとる確率の変化が，常に λ が減少する方向に動くという状況になってしまうからである。したがって，貸出が減少するには k, b が共に減少する必要があるということが前章の結果から導かれる。

すなわち，$k(t)$ の関数形は，k の値が t の増加とともに減少するようなものを考えればよいということになる。

C. obsolescence の式の一般形

前節において $k(t)$ は減少関数になればよいことがわかった。しかし，具体的にどのような減少関数であるかを規定する条件は，前章のガンマ分布の導出過程からは特に見い出せない。

そこで，本節ではこの時間と共に減少する一般的な関数を構成する。つまり前章の議論中で，$k(t)$ を減少関数という以上に強く規定する要因がなかったことから，一般的な減少関数を構成しておいて，モデルの一般性を広げておくのが，モデルの応用という観点からは有用だと考えられるからである。

さて，$k(t)$ を減少する関数にするには，

$$\frac{dk}{dt}=-c(t)k-h(t) \tag{3.3}$$

とおいて，この微分方程式を解けばよい。なぜなら，(3.3) 式は k の微小時間における変化をモデル化したものであり，その右辺の各項はいずれも負になっているので，この微分方程式の解 $k(t)$ は減少関数となるのである。しかも，右辺第一項はその減少の単位時間あたりの割合が k の大きさに依存する部分，第二項は依存していない部分であり，この 2 つの部分を含む (3.3) 式は十分一般的であると考えられる（また，過度に一般的過ぎるということもない。例えば，(3.3) 式右辺第一項をより一般的に $c(t, k)$ とおくこともできるが，これは一般的すぎて逆に扱いが困難である）。

例えば (3.3) 式で $c(t)=0$，$h(t)=h$ とおき，それを代入して (3.3) 式を解けば，解は $k=-ht+C$ なる一次式となり，時間 t とともに一定の割合で k が減少していく関数が得られる（C は任意定数）。また，$c(t)=c$，$h(t)=0$ とするならば，(3.3) 式は，

$$\frac{dk}{dt}=-ck$$

となるが，この微分方程式を解くと，$k=Ce^{-kt}$，すなわち，Burrell の (3.1) 式，あるいは (1.2)′ 式で利用されている，従来の obsolescence の数学的な表現である負の指数関数が得られる（すなわち，(3.3) 式の解は，従来の obsolescence の式の一般形である）。

この他にも，上述した 2 つの例と同様に $c(t)$，$h(t)$ の設定次第によっては，さまざまなかたちの減少関数が導かれる。

このように，(3.3) 式は多くの減少関数を導く，非常に一般的な微分方程式であることがわかる。そこで，(3.3) 式を，$k(t)$ を表現するための一般的な方程式として採用することにすれば，貸出頻度分布モデルの適用範囲が非常に広くなる。そして，各図書館でこのモデルを利用する場合には，(3.3) 式中の最適な $c(t)$，$h(t)$ を実際の貸出頻度分布のデータから推測することにより，その図書館に適した $c(t)$，$h(t)$ を決めればよい。その環境・状況によって各図書館の実情は大きく異なる。したがって，このように柔軟な適応力を持ったモデルは有効であろう。

しかし，このモデルの利用のためには，(3.3) 式の解の一般形を求めておく必要がある。(3.3) 式は線形微分方程式であるから，

$$k(t)=e^{-\int c(t)dt}\int h(t)e^{-\int c(t)dt}dt+C' \tag{3.4}$$

となる（C' は任意定数）。したがって，本論では (3.4) 式を $k(t)$ の関数形とおくことにする。

なお，obsolescenceは図書・雑誌の利用あるいは引用

の経年的な「減少」に関する現象であるから，(3.3) 式の解である (3.4) 式は obsolescence の数式の一般形として捉えることもできる。

IV. 貸出頻度分布モデルの一般形とその応用

A. 貸出頻度分布モデルの一般形と特殊形

1. 一般的モデル

前章においては，貸出頻度分布の時間変化を記述するのに重要な役割を果たす，ガンマ分布のパラメター $k(t)$ の一般的な関数形 (3.4) 式を導出した。この (3.4) 式を用いた貸出頻度分布モデルは，その (3.4) 式の一般性によって，貸出の時間変化に対して，きわめて敏感に反応できるモデルとみなすことができる。その具体的な数式は次のようになる。

$$P(x_t=r)=\binom{r+k(t)-1}{r}p(t)^{k(n)}[1-p(t)]^r \quad (4.1)$$

ここで，

$$k(t)=e^{-\int c(t)dt}\int -h(t)e^{-\int c(t)dt}dt+C$$

$$p(t)=k(t)/[k(t)+\bar{x}_t]$$

$\bar{x}_t$ は第 t 期間における標本平均（平均貸出数）

これは時間変化に対して非常に一般的な貸出頻度分布モデルである。

ただし，$p(t)$ はモデル (1.1) 式と同様に $k(t)$ が既知の場合の最尤推定量となるように定義した。そのため，第I章で述べたように，$p(t)$ が $k(t)$ と $\bar{x}_t$ で決まることになる。この $\bar{x}_t$ は，第 t 期間（例えば第 t 年次）の標本平均を意味しており，貸出頻度分布モデルを予測に使おうという場合には，この $\bar{x}_t$ も何らかの方法で予測する必要がある。モデル (1.1) 式では，(1.3)′ 式から明らかなように，obsolescenceの数式である負の指数関数を $k(t)$ に関する (1.2)′ 式同様に利用して，任意の t における $\bar{x}_t$ を求めることができるようにしている。

この $\bar{x}_t$ の関数形の決め方にも，$k(t)$ と同様にさまざまな方法がありうる。しかし本論では，便宜的に各モデルの $\bar{x}_t$ としては，そのモデルの $k(t)$ と同じ関数形をそれぞれ用いることとする。ただし，この約束は本質的なものでなく，$k(t)$ と $\bar{x}_t$ とを全く別の関数にしても理論的な問題は生じない。

2. 一般的モデルから導かれるいくつかの特殊形

一般的モデル (4.1) 式において，$c(t), h(t)$ の具体的な関数を各図書館の実情に合わせて決定すれば，その図書館に合った貸出頻度分布モデルを得ることができる。

ここでは，(4.1) 式の適用の方法を提示するために，例として，いくつか $c(t)$, $h(t)$ の関数形を決め，(4.1) 式からの貸出頻度分布モデルの特殊形の導出を試みる。次の3つの場合を考える。

① $c(t)=c$， $h(t)=0$

② $c(t)=c/t$， $h(t)=0$

③ $c(t)=0$， $h(t)=h$

まず第一に，$c(t)=c, h(t)=0$ の場合である。この場合は，前章ですでに述べたように，$k(t)$ は負の指数関数，すなわち (1.2)′ 式になる。したがって，貸出頻度分布モデルは，モデル (1.1) 式に一致する。なお (1.1) 式のモデル中の (1.2)′ 式で $c=0$ とおいた，さらに特別の場合が Burrell[2] の各年次分布のモデルに等しくなる。したがって，Burrell の各年次分布のモデルは，(4.1) 式において $c(t)=0, h(t)=0$ とおいた場合となる。これらのモデルに関しては，第I章で詳しく説明してあるので，ここでは特に述べない。

第二の場合として $c(t)=c/t, h(t)=0$ とおく。これは，従来の obsolescence の表現である負の指数関数においてはその減少率が一定（(3.3) 式において $c(t)=c$ とおくから）であるのに対して，時間の経過と共に減少率が減っていくモデルを与える。すなわち，この場合従来の負の指数関数で表現される場合よりも，obsolescenceが緩やかになる。この場合の $k(t)$ は，(3.4) 式より，

$$\begin{aligned}k(t)&=\exp\left[-\int\frac{c}{t}dt\right]+C'\\&=\exp[-c\log t+C']\\&=C^*t^{-c}\end{aligned} \quad (4.2)$$

となる（ただし，$C^*=e^{c'}$）。(4.2) 式で $t=1$ とおけば，$k(1)=C^*$ を得る。そこで (1.2) 式と同様に $k(1)$ を k_1 とかけば，(4.2) 式は，

$$k(t)=k_1t^{-c} \quad (4.3)$$

となる。また，$\bar{x}_t$ も同様に，

$$\bar{x}_t=\bar{x}_1t^{-d} \quad (4.4)$$

であるから，これらを (4.1) 式に用いると，

$$P(x_t=r)=\binom{r+k(t)-1}{r}p(t)^{k(t)}[1-p(t)]^r \quad (4.5)$$

ここで，

$$k(t)=k_1t^{-c}$$

$$p(t)=\frac{k_1t^{-c}}{k_1t^{-c}+\bar{x}_1t^{-d}}$$

なる貸出頻度分布モデルが新たに得られる。

第三の場合として，$c(t)=0, h(t)=h$ とおく。これは前節ですでに述べたように，

$$k(t)=-ht+C \tag{4.6}$$

となる関数形を導く。(4.2) 式から (4.3) 式を得たときと同様に k_1 を $k(t)$ の初期値と考えると，(4.6) 式は，

$$k(t)=-h(t-1)+k_1 \tag{4.6$'$}$$

となる。$\bar{x}_t$ は (4.6)$'$ 式と同様に，

$$\bar{x}_t=-m(t-1)+\bar{x}_1 \tag{4.7}$$

となるから，貸出頻度分布モデルは，

$$P(x_t=r)=\binom{r+k(t)-1}{r}p(t)^{k(t)}[1-p(t)]^r \tag{4.8}$$

ここで，

$$k(t)=-h(t-1)+k_1$$

$$p(t)=\frac{-h(t-1)+k_1}{-h(t-1)+k_1-m(t-1)+\bar{x}_1}$$

である。

$c(t), h(t)$ の設定によっては，この他にも貸出頻度分布モデルが導かれうる。しかし，この3例によって，十分に一般モデル (4.1) 式からの実際の貸出頻度分布モデルの典型例が表されると考えられる。

B. 貸出頻度分布モデルにおけるパラメター推定法

(4.1) 式の一般的モデルから導出された貸出頻度分布モデルを実際に応用するためには，モデル中のパラメターをデータより推定しなければならない。

第Ⅰ章で述べたように，本論でのモデルは貸出頻度分布の時間変化を記述するモデルであるから，そのパラメターの推定には最低異なる2つの期間の貸出データ（例えば，1年間を単位とすれば，第1年次と第2年次の貸出データ）が必要になる。したがって，パラメターの推定は，通常の負の二項分布のそれよりもいくぶん複雑になる。ここでは，前節で新たに導出した貸出頻度分布モデル (4.5), (4.8) 式のパラメターを積率推定法を用いて推定する方法を示す。

積率推定法は，統計学でよく用いられる手法で，基本的には，モデルである確率分布のモーメント（積率）とそれに対応するデータの標本平均や標本分散等を等しいと仮定して，モデル中のパラメターを推定する方法である（(1.1) 式のモデルにおけるパラメター推定である，(1.4), (1.5) 式も積率推定量である)。

まずモデル (4.5) 式のパラメターの推定法を考える。第Ⅰ章と同様に，二つの時点 $t=1, t=2$ の標本平均をそれぞれ $\bar{x}_1$, $\bar{x}_2$, また標本分散を s_1^2, s_2^2 と表記する。負の二項分布 (2.9) 式の平均 μ, 分散 σ^2 は，

$$\mu=k(1-p)/p, \qquad \sigma^2=k(1-p)/p^2$$

であるが[3]，これを連立方程式と見なして k について解くと，

$$k=\mu^2/(\sigma^2-\mu) \tag{4.9}$$

が得られる。ここで μ を標本平均，σ^2 を標本分散に置き換えて，(4.9) 式を (4.3) 式に代入すると，

$$\frac{\bar{x}_2^2}{(s_2^2-\bar{x}_2)}=\frac{\bar{x}_1^2}{(s_1^2-\bar{x}_1)}\times 2^{-\hat{c}}$$

が得られる。そこで，これを c について解けば，

$$\hat{c}=-\log\left[\frac{\bar{x}_2^2(s_1^2-\bar{x}_1)}{\bar{x}_1^2(s_2^2-\bar{x}_2)}\right]\Big/\log 2 \tag{4.10}$$

となる。

一方，(4.4)式の $\bar{x}_t$ を決めるためのパラメター d に関しては，(4.4) 式で $t=2$ とおいたものを，$t=1$ とおいたもので割ると，$\bar{x}_2/\bar{x}_1=2^{-\hat{d}}$ が得られるから，これを解いて，

$$\hat{d}=-\log(\bar{x}_2/\bar{x}_1)\times\log 2 \tag{4.11}$$

を得る。

モデル (4.8) 式に対しても，全く同様の方法を用いれば，(4.6)$'$ 式の h は，

$$\frac{\bar{x}_2^2}{(s_2^2-\bar{x}_2)}=-\hat{h}+\frac{\bar{x}_1^2}{(s_1^2-\bar{x}_1)}$$

から，

$$\hat{h}=\frac{\bar{x}_1^2}{(s_1^2-\bar{x}_1)}-\frac{\bar{x}_2^2}{(s_2^2-\bar{x}_2)} \tag{4.12}$$

と推定でき，また (4.7) 式の m も，

$$\bar{x}_2/\bar{x}_1=(-\hat{m}+\bar{x}_1)/\bar{x}_1$$

から，

$$\hat{m}=\bar{x}_1-\bar{x}_2 \tag{4.13}$$

と推定できる。

C. 実際のデータへのモデルの適用

ここでは，慶應義塾大学日吉情報センターの貸出頻度分布のデータ[1)]をひとつの事例として，本章で導出された貸出頻度分布モデルを実際のデータに適用してみる。

各モデルのパラメターの推定に用いるための標本平均および標本分散は，$\bar{x}_1=0.995$, $s_1^2=3.263$, $\bar{x}_2=0.983$, $s_1^2=3.467$ である[1)]。したがって，前節で新たに導出した貸出頻度分布モデル (4.5) 式，(4.8) 式の各パラメターは，それぞれ (4.10), (4.11) 式，(4.12), (4.13) 式を用いると，$\hat{c}=0.1679$, $\hat{d}=0.018$, $\hat{h}=0.049$, $\hat{m}=0.0125$ と計算される。また，標本平均と $x=0$ の頻度を用いる方法によって，第1年次のデータより k_1 を求めると $\hat{k}_1=0.475$ であるから[1)]，これらのパラメターの値を (4.5) 式，(4.8) 式に代入すれば，モデルとしての貸出頻度分布を求めることができる。

実際にこの手順により，貸出頻度分布の理論的な値を求めたものを第1表，第2表に示す。第1表はモデル (4.5) 式であり，第2表はモデル (4.8) 式である。なお表中では，実際の貸出頻度分布のデータとの比較も行っており，その適合度のひとつの目安として，カイ二乗値も表示してある。また，同様のデータに対するモデル (1.1) 式の適合度も第3表に示した。

次に，これらの3つの表を比較する。第3年次（すなわち1985年度）のデータについては，モデル(4.5), (4.8) 式とモデル (1.1) 式とで自由度が異なるのでカイ二乗値での比較は難しいが，第2表のモデル (4.8) 式の適合度がわずかに良いことがわかる。したがって，慶應義塾大学日吉情報センターの1982年度受入の図書に関しては，一般モデル (4.1) 式において $c(t)=0$, $h(t)=h$ とおいたモデル (4.8) 式が，その貸出頻度分布をよく記述すると結論できる。

この結論は，貸出データがわずかに3年分しかなかったために，十分に確定的なものとはいえないが，その一方で，前論文において提示した (1.1) 式以外のモデルが

第1表　新しい貸出頻度分布モデル (4.5) 式の現実のデータへの適合
データ：慶應義塾大学日吉情報センターの1982年度受入図書の貸出記録[1)]

1983年度貸出				1984年度貸出				1985年度貸出			
	データ	理論値	χ^2 値		データ	理論値	χ^2 値		データ	理論値	χ^2 値
0回	5542	5542.0	0.00	0回	5661	5703.7	0.32	0回	5899	5798.6	1.73
1	1876	1782.8	4.86	1	1804	1687.0	8.10	1	1779	1631.1	13.39
2	826	890.0	4.60	2	843	839.1	0.01	2	710	809.8	12.30
3	475	496.9	0.97	3	442	473.7	2.13	3	423	460.1	2.99
4	283	292.2	0.29	4	253	283.4	3.26	4	234	277.9	6.94
5	170	177.0	0.27	5	155	175.2	2.34	5	119	173.8	17.31
6	92	109.3	2.74	6	94	110.7	2.52	6	81	111.2	8.23
7	72	68.4	0.18	7	66	71.0	0.35	7	77	72.3	0.29
8	44	43.2	0.02	8	33	46.0	3.70	8	53	47.5	0.61
9	24	27.5	0.46	9	33	30.1	0.27	9	39	31.5	1.73
10	36	17.6	18.97	10	27	19.8	2.57	10	22	21.1	0.03
11	16	11.3	1.86	11	29	13.1	19.10	11	13	14.2	0.10
12	9	7.3	0.35	12	19	8.7	12.00	12	10	9.5	0.01
13-14	7	7.9	0.10	13	5	5.8	0.12	13	4	6.5	0.96
15	8	5.6	1.01	14-	16	11.2	2.00	14-	17	12.6	1.52
計	9480	9480.0	36.73		9480	6480.0	58.86		9480	9480.0	68.24
平均			0.99	平均			0.98	平均			0.90
分散			3.26	分散			3.46	分散			3.21
パラメター $\hat{k}$			0.475	パラメター $\hat{k}$			0.423	パラメター $\hat{k}$			0.395
$\hat{p}$			0.323	$\hat{p}$			0.300	$\hat{p}$			0.288
自由度			14	自由度			14	自由度			14

第2表　新しい貸出頻度分布モデル (4.8) 式の現実のデータへの適合

データ：慶應義塾大学日吉情報センターの1982年度受入図書の貸出記録[1)]

	1984年度貸出				1985年度貸出		
	データ	理論値	χ^2値		データ	理論値	χ^2値
0回	5661	5696.0	0.21	0回	5899	5887.0	0.00
1	1804	1692.3	7.35	1	1779	1591.7	22.02
2	843	841.7	0.00	2	710	786.3	7.40
3	442	474.8	2.27	3	423	447.0	1.29
4	253	283.7	3.32	4	234	270.8	5.00
5	155	175.2	2.32	5	119	170.1	15.36
6	94	110.5	2.47	6	81	109.3	7.37
7	66	70.7	0.32	7	77	71.5	0.42
8	33	45.8	3.59	8	53	47.3	0.68
9	33	29.9	0.31	9	39	31.6	1.72
10	27	19.6	2.71	10	22	21.2	0.00
11	29	13.0	19.62	11	13	14.3	0.13
12	19	8.6	12.39	12	10	9.7	0.00
13	5	5.7	0.10	13	4	6.6	1.08
14-	16	11.0	2.21	14-	17	13.1	1.09
計	9480	9480.0	59.27		9480	9480.0	63.67
パラメター	$\hat{k}$		0.425	パラメター	$\hat{k}$		0.377
	$\hat{p}$		0.302		$\hat{p}$		0.282
自由度		14		自由度		14	

第3表　貸出頻度分布モデル (1.1) 式の現実のデータへの適合

データ：慶應義塾大学日吉情報センターの1982年度受入図書の貸出記録[1)]

	1984年度貸出				1985年度貸出		
	データ	理論値	χ^2値		データ	理論値	χ^2値
0回	5661	5703.6	0.31	0回	5899	5881.1	0.00
1	1804	1684.8	8.42	1	1779	1585.8	23.51
2	843	838.5	0.02	2	710	785.5	7.26
3	442	473.8	2.14	3	423	448.1	1.41
4	253	283.7	3.34	4	234	272.5	5.45
5	155	175.6	2.43	5	119	171.9	16.28
6	94	111.1	2.64	6	81	111.0	8.11
7	66	71.3	0.40	7	77	72.8	0.23
8	33	46.3	3.84	8	53	48.4	0.42
9	33	30.3	0.22	9	39	32.4	1.30
10	27	20.0	2.42	10	22	21.9	0.00
11	29	13.2	18.60	11	13	14.9	0.25
12	19	8.8	11.64	12	10	10.2	0.00
13	5	5.9	0.14	13	4	7.0	1.28
14-	16	11.4	1.82	14-15	11	8.1	0.99
				16-	6	5.8	0.00
計	9480	9480.0	58.45		9480	9480.0	66.61
パラメター	$\hat{k}$		0.422	パラメター	$\hat{k}$		0.374
	$\hat{p}$		0.300		$\hat{p}$		0.279
自由度		14		自由度		15	

現実のデータにかなり適合したということ，すなわち，本論において構築した一般的な貸出頻度分布モデル(4.1)式から導かれたモデルが現実のデータをよく説明したということは，本論で展開した議論の妥当性の裏付けとして捉えることができる。つまり，従来よりobsolescenceの数学モデルとして採用されてきた負の指数関数((1.2)′式あるいは(3.1)式)を貸出頻度分布モデルに無批判に採用することは否定され，(3.4) 式が示す obsolescence の数式の一般形より導かれた一般的な貸出頻度分布モデルからひとつの特殊形を選択するという手順をとることの有効性が確認されたのである。一般モデル(4.1)式で，$c(t)=0$, $h(t)=h$ とおくことにより導かれた (4.8) 式は，慶應義塾大学日吉情報センターの場合では妥当であったが，すべての図書館にあてはまるモデルであるかどうかまでは，ここではわからない。しかしいずれにせよ一般的貸出頻度分布モデル (4.1) 式の有効性は，ここで十分に確かめられたのである(もし(4.1)式がなければ，(4.8)式は導かれなかった)。

V. おわりに

本論文ではまず，筆者らが以前に提案した，負の二項分布をベースとする貸出頻度分布モデルの演繹的な導出を行い，それによって，その貸出頻度分布モデルの持つ意味を明確にした。具体的には，①ある一冊の図書に注目したときの，その貸出回数の確率分布は近似的にポアソン分布で記述できること，②その各図書のポアソン分布のパラメターの蔵書全体における分布として，ビブリオメトリックスの諸法則の導出において知られる線形増殖の原理を貸出という事象に応用することにより，ガンマ分布が得られること，の2つの点を示し，そしてこの①，②から負の二項分布が貸出頻度分布モデルとして演繹的に導出されることを提示した。

次に，貸出頻度分布モデルの時間変化の方法について

議論したあと，貸出頻度分布モデル中のパラメターの現実的な意味を考察し，そこから貸出頻度分布モデルの時間変化に関する一般化を行った。具体的には，パラメター $k(t)$ が一般的な減少関数であることを結論し，その $k(t)$ として従来の obsolescence の式を拡張した (3.4) 式を用いて，貸出頻度分布の一般的なモデル (4.1) 式を得た。

さらに，その一般モデル (4.1) 式の現実のデータへの適用の方法を慶應義塾大学日吉情報センターの貸出データをひとつの事例として用いて，提示した。今回は典型例として，① $c(t)=c$, $h(t)=0$, ② $c(t)=c/t$, $h(t)=0$, ③ $c(t)=0$, $h(t)=h$ の3例を示したのみであるが，(3.4) 式中の積分が実行可能な限り，ここで述べた手順によってあらゆる $c(t)$, $h(t)$ に対する貸出頻度分布モデルを現実に応用できる。

以上の本論文の議論・考察により，負の二項分布をベースとする貸出頻度分布モデルの正当性に対する理論的な裏付けが得られ，さらに応用範囲の広い一般的なモデルが構築された。この貸出頻度分布モデルは，現実の図書館の運営・管理に対して，さまざまな応用可能性を持っている。例えば，既存の貸出データからパラメターを推定しさえすれば，t を任意にとって将来の貸出頻度分布のシミュレーションを行うことができる。これは，図書の廃棄や別置の問題に対して有力な情報を提供する。従来は，図書の貸出の将来予測の技法が不十分であったために，将来の利用を常に予期しながら実行することが必要な廃棄・別置という作業に，定量的な示唆（例えば出版後10年経った図書の x% を別置せよ，など）を与えることが困難であった。しかし，本論文で構築した貸出頻度分布モデルはかなりの精度で将来予測を行うことができるので，このような種類の意思決定に対して十分応用できる。

また，各図書の貸出回数がポアソン分布にしたがうときのパラメター λ は，Morse の「circulation interference」の理論[4]と密接に結び付いているため，この理論を貸出頻度分布モデルに組み込むことにより，複本購入の問題にもこのモデルを利用することも可能である。

一方，モデルの構成に関して残された課題は，なるべく数多くの，さまざまなタイプの図書館の，多年度にわたる貸出データを用いて，それへの貸出頻度分布モデルの適用を試してみるということである。この試みによって，さらに貸出頻度分布モデルやその理論の妥当性・応用可能性が詳しく議論され，モデルがより洗練あるいは修正されていくことになろう。

なお，本論文の執筆にあたりまして，慶應義塾大学文学部図書館・情報学科高山正也教授からは多くの貴重な御助言をいただきました。ここに感謝の意を表します。

1) 岸田和明，原田隆史，高山正也，小川治之，逸村裕．大学図書館における図書の貸出頻度についての確率過程モデルの検討—象の二項分布を中心として．Library and Information Science. No. 25, p. 25-39 (1987).
2) Burrell, Quentin. A note on ageing in a library circulation model. Journal of Documentation. Vol. 41, No. 2, p. 100-115 (1985).
3) Bagust, A. A circulation model for busy public libraries. Journal of Documentation. Vol. 39, No. 1, p. 24-37 (1983).
4) Morse, Philip M. Library effectiveness: a system approach. Cambridge, M.I.T. Press, 1968. 207p.
5) Buckland, M.K. et al. System analysis of a university library. Lancaster, University of Lancaster, 1970. p. 35-55.
6) 伊藤　清．確率論．東京，岩波書店，1953. 405p.
7) 芝　祐順ほか編．統計用語辞典．東京，新曜社，1984. p. 222.
8) Gelman, E.; Sichel, H.S. Library book circulation and beta-binomial distribution. Journal of the American Society for Information Science. Vol. 38, No. 1, p. 4-12 (1987).
9) Hayes, Robert M. The distribution of use of library materials: analysis of data from the university of Pittsburgh. Library Research. Vol. 3, No. 3, p. 215-260 (1981).
10) Simon, Herbert A. On a class of skew distribution function. Biometrika. Vol. 42, No. 3/4, p. 425-440 (1955).
11) Price, Derek de Solla. A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science. Vol. 27, No. 5, p. 292-306 (1976).
12) Reichl, L.E. 現代統計物理 上．鈴木増雄監訳．東京，丸善，1983, 334p.
13) (2.4)式は(2.2)式の随伴偏微分方程式であるので，これらの解が同等となる条件は，グリーン関数の相反定理が成立する条件に等しい。これは，(2.2), (2.4) 式の境界条件を反射壁，あるいは吸収壁とすることである。貸出回数の期待値は非負であるからこの条件は満たされている。つまり，$\lambda=0$ を境界として，確率は $\lambda<0$ の領域に流出しない。

14) 初期条件は $P(\lambda_2, s|\lambda_1, t)=\delta(\lambda_2-\lambda_1), t=s$ である（$\delta(x)$ はディラックのデルタ関数）。境界は，正則境界あるいは流入境界である必要があるので，$2v/q>0$ でなければならない。これらの条件の下に，固有関数展開を利用する方法に用いて偏微分方程式 (2.4)′ 式を解き，さらに $t\to\infty$ としてその定常分布をとれば，ガンマ分布が得られる。詳しくは，小倉久直．続物理・工学のための確率過程論．東京，コロナ社，1985．260p．などを参照。