Library and Information Science No. 25 1987

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
——負の二項分布を中心として——

An Examination of Stochastic Models for the Number of Book Circulations in Academic Libraries
—Concerning to a Negative Binomial Distribution—

岸田和明 *Kazuaki Kishida*
原田隆史 *Takashi Harada*
高山正也 *Masaya Takayama*
小川治之 *Haruyuki Ogawa*
逸村裕 *Hiroshi Itsumura*

Résumé

Recently, some library circulation models have been developed, and it has been tried to make use of them for decision making in libraries. One of the main approaches is to use a negative binomial distribution as a book circulation frequency distribution.

In this paper, firstly, we try to examine the validity of a negative binomial distribution as a book circulation frequency distribution in comparison to other mixture distributions—beta binomial, Neiman's type A, and logarithmic series distributions. As a result, the examination indicates the superiority of the negative binomial distribution.

Secondly, we examine a nonstationary Poisson process model which Q. E. Burrell has developed. It's result clarifies that Burrell's model doesn't be enough to fit the year-by-year change of real circulation frequency distributions of Hiyoshi Library and Information Center in Keio University. So, we modify this model. Our modified model gives a better fitness to real deta than Burrell's model.

岸田和明：慶應義塾大学大学院文学研究科図書館・情報学科専攻修士課程，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.
原田隆史：慶應義塾大学大学院文学研究科図書館・情報学科専攻博士課程，東京都港区三田 2-15-45.
Takashi Harada: Graduate School of Library and Information Science, Keio University, 2-15-45, Mita, Minato-ku, Tokyo.
高山正也：慶應義塾大学文学部図書館・情報学科教授，東京都港区三田 2-15-45.
Masaya Takayama: Professer, School of Library and Information Science, Keio University, 2-15-45, Mita, Minato-ku, Tokyo.
小山治之：慶應義塾大学三田情報センター選書課長，東京都港区三田 2-15-45.
Haruyuki Ogawa: Mita Library and Information Center, Keio University, 2-15-45, Mita, Minato-ku, Tokyo.
逸村　裕：上智大学図書館，東京都千代田区紀尾井町 7-1.
Hiroshi Itsumura: Sophia University Library, 7-1, Kioicho, Chiyoda-ku, Tokyo.

I. は　じ　め　に

最近の出版物の増加，諸物価の上昇，地価の高騰に伴う用地の取得難といった状況のなかで，図書館がその予算やスペースの大幅な増加を期待することが困難になりつつある。そこで，大学図書館においても，効果的かつ効率的な蔵書構築・蔵書管理がますます必要になってくると考えられる。それは，実際には，選書のほかにも，例えば，図書・雑誌の保存書庫への別置や，廃棄といった手段によって行われているが，現在までに，その決定をサポートするような，さまざまな数学的モデルの開発が進められている。

その主流のひとつは，図書の貸出頻度分布（ある一定期間に各図書が何回貸し出されたかを計数し，各貸出回数ごとにその冊数を合計したものを，本論文では貸出頻度分布と称する。図示すれば，横軸に貸出回数，縦軸にその図書の 冊数をとったグラフとなる。）を確率分布としてモデル化し，貸出の予測や，図書の廃棄・別置の方針に役立てようとするもので，その源流は，Morse[1] に求められる。

Morse は，図書の利用の分析のために統計学の理論を導入し，図書の貸出の予測を試みた。まず，彼は，図書の貸出について，以下の 3 つの仮定を置いた。

①図書の貸出は，ランダムに生じ，ポアソン過程にしたがう。

②平均的には，図書の貸出は，時間の経過にともなって，指数的に減少する。

③図書の貸出事象は，マルコフ性をもつ。

Morse は，これらの仮定に 基づいて，図書の 貸出についてのマルコフモデルを構築し，予測の方法を開発したが，その際，その第 1 年の貸出回数の分布（すなわち，貸出頻度分布）に対して，幾何分布を適用した。これは，図書が r 回貸し出される確率を $P(X=r)$ と書くことにすれば，

$$P(X=r)=(1-\omega)\omega^r$$

となる（ω は幾何分布のパラメター）。

Morse のこのような貸出データの分析から，その後，図書の貸出頻度分布に対するさまざまな統計学的な研究が生まれ，理論的にはかなり高度に洗練されつつある。

しかし，そこで貸出頻度分布として提案された確率分布に対しての，実際のデータを用いての検証は，いまだ不十分であり，モデルの実用化には至っていない。実用化のためには，さらに，研究とデータの積み重ねをおこなっていく必要がある。そこで，本論文では，まず，この貸出頻度分布の研究についてのレビューと考察を行いさらに，実際のデータとして，慶應義塾大学日吉情報センターの貸出記録を用い，それらのモデルの有効性等を検証する。そして，最終的には，既存のモデルを修正した新たなモデルも提出し，その優位性を示す。

II. 貸出頻度分布とその確率的構造

A. 確率過程モデルと Burrell の理論

Morse の研究以後，同様な試みがいくつか行われたがそのなかでも，Burrell[2] は，図書の 貸出事象を 確率過程として捉え，その確率的構造を考慮した上で，Morse が用いた幾何分布を，貸出頻度分布として改めて演繹的に導出した。そのモデルは以下の仮定，すなわち，

①貸出事象はポアソン過程である，

②各図書の “desirability（望ましさ)” は，ある確率分布に従う，

に基づいて構築される。これは，各図書の貸出が，ポアソン分布に従って起こる一方，その貸出が起こる回数の単位時間あたりの 平均 が，各図書の人気 や 有用性など

(すなわち "desirability") に影響されて，各図書で異なっているということを意味する。そして，その結果として，図書によって貸出の頻度に違いが生じるということから，Burrell は，貸出頻度分布の形状を説明しようとした。実際に，彼は，この "desirability" が，負の指数分布に従うと仮定して，ポアソン分布の平均値 λ に負の指数分布を組み込んで，幾何分布を導出し，これを貸出頻度分布として実際のデータにあてはめた。よってこの幾何分布は，いわゆる複合ポアソン分布である。

しかし，同様の理論を展開した Burrell and Cane[3] に対する批判が，Journal of Royal Statistical Society 誌上でいくつか行われている。C. Chatfield らは，ポアソン分布のパラメーター λ にガンマ分布を適用することによって導かれる負の二項分布のほうが，貸出頻度分布として幾何分布よりも優れていると指摘した[4]。

負の二項分布を図書館の貸出頻度分布に適用する方法は，その後1983年に，A. Bagust によって改めて発表された[5]。Bagust は, Burrell の理論を改良するかたちで, "desirability" の分布にガンマ分布をあてはめて，演繹的にこの負の二項分布を導出し，公共図書館のデータによってそれを検証した。

そして，さらに，この Bagust のモデルに，obsolescence（蔵書の陳腐化）の要素を組み込み，負の二項分布モデルの改良をおこなったのが，Burrell の1985年の論文である[6]。本論文では, この1985年以降, Q.L. Burrell が理論的に展開してきたモデルを，便宜上，「Burrell モデル」と称する。そして，以下，この Burrell モデルについて，やや詳しく述べる。

この Burrell モデルは，前述の①, ②で示した，1980年の Burrell[2] の仮定に，

③ 1年間の平均貸出数は年を追って，指数的に減少する，

という仮定を加えたものである。これによって，ポアソン過程は，非定常ポアソン過程となる。すなわち，ポアソン分布のパラメータ λ は時間 t の関数 $\lambda(t)$ となり，ポアソン分布の式は，平均値関数，

$$m(\lambda, t)=\int_0^t \lambda(s)ds \qquad (1)$$

を用いて，

$$P(X_t=r)=\frac{\exp[-m(\lambda, t)]m(\lambda, t)^r}{r!} \qquad (2)$$

となる。ここで，Burrell は，この $\lambda(t)$ について，具体的に，

$$\lambda(t)=\lambda e^{-at},\ \ t\geqq 0 \qquad (3)$$

とおいた。従って，平均値関数 $m(\lambda, t)$ は，(1)式の積分を実行して，

$$m(\lambda, t)=\frac{\lambda}{a}(1-e^{-at})=\lambda m(t) \qquad (4)$$

と計算される。ここで λ の分布 $f(\lambda)$ を，

$$f(\lambda)=\frac{b^k}{\Gamma(k)}\lambda^k e^{-b\lambda} \qquad (5)$$

というガンマ分布とする。そして，複合分布の定義に (2), (4)を用いた，

$$P(X_t=r)=\int_0^\infty \frac{\exp[-\lambda m(t)]\lambda m(t)^r}{r!}f(\lambda)d\lambda \qquad (6)$$

に，(5)を代入して，積分を実行すれば，

$$P(X_t=r)=\binom{r+k-1}{r}p(t)^k q(t)^r,\ \ r=0, 1, 2, \cdots$$

ここで，

$$p(t)=\frac{1}{1+bm(t)}=1-q(t) \qquad (7)$$

と計算でき, 最終的に負の二項分布が得られる。これは，例えば，t を年単位でとれば，t 年が経過したのちの，貸出頻度の分布を表している。さらに，Burrell は，各年ごとの貸出頻度の分布を求めている。

以上のように，この Burrell モデルは，基本的には負の二項分布であるが, obsolescence を組み込むことにより，貸出頻度分布の時間的変動に対応しうるように，拡張されたモデルである。

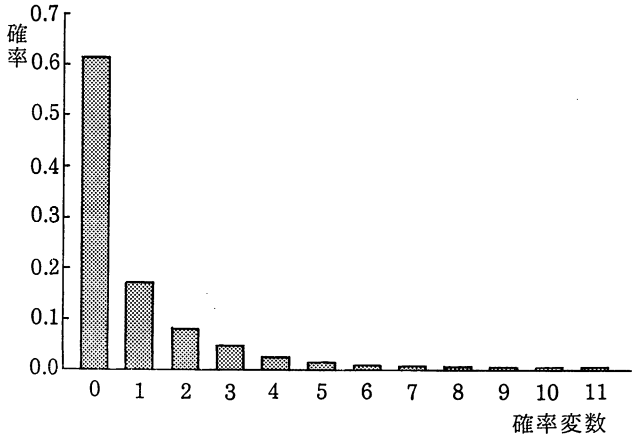
さらにまた，Burrell[7)8)] では，このモデルが図書の平均貸出数の予測や，別置の問題に応用されている。

B. 貸出頻度分布の確率的構造

ここでは，Burrell モデルの理論的構造や，さらに，貸出頻度分布の確率的構造に関する考察を行なう。

1. 負の二項分布の構造

1920年に，Greenwood and Yule[9] は，大工場において各作業者が事故に合う回数がパラメーター λ のポアソン分布で記述され，さらにその λ 自体がガンマ分布にしたがうという仮定から，その工場の事故件数が，負の二項分布にしたがうという結果を得た。これは，Bagust[5] も指摘しているように，前節で述べた Burrell の貸出に関する理論に類似している。さらに，負の二項分布が，工



第１図　負の二項分布の形状（パラメター $k=0.4$, $p=0.3$）

業や商業などの全く異なる状況において，これと同様の理論から導出された例がいくつかある[10]。

このように，負の二項分布は，ガンマ分布を組み込む複合ポアソン分布から導かれるが，似たような状況は他にも多く存在し，様々な現象が負の二項分布で近似されると考えられる。これは，ランダムな現象の多くがポアソン分布で記述され，さらに，ガンマ分布が非常に柔軟性に富んでいることに起因している。

さらに，この負の二項分布の汎用性は，別の事実によっても示される。例えば，生物学において，下等生物の増殖数は，Yule 過程として知られているが，これは，負の二項分布となる。この場合の負の二項分布の導出は，複合ポアソン分布からではなく，別の仮定から行われる[11]。

一方，負の二項分布は，グラフ化すると，右すそに長くなることが知られている。例として，第１図に，典型的な負の二項分布の形状を示してある。このように，負の二項分布は，いわゆる“J”形（裏返しの）のかたちをしており（ただし，k がある程度大きいと単峰形となる），さらに，右すその確率はなかなか０とならない。このかたちは，従来報告されている貸出頻度分布の形状によく適合しており，この点からも，負の二項分布の採用の妥当性が裏付けられる。

２．貸出頻度分布の確率的構造

これまでの議論では，負の二項分布の優位性を論じてきた。しかし，その現実への妥当性の検証をおこなう前に，貸出頻度分布として用いうる可能性のある，他の確率分布についても考察を行っておく必要がある。また，その分布を対立する仮説として考えれば，のちに負の二項分布の妥当性を論じるときの助けともなる。

そこでまず，図書館の貸出事象の確率的構造についてもうすこし一般的な記述をおこなってみると，

①貸出はある確率分布 f_A にしたがって生じる，

②その f_A のパラメターは各図書によって異なり，それ自体もある確率分布 f_B にしたがう，

③その結果，貸出頻度分布に対して複合分布 f_C が導かれる，

となる。そこで，問題は，この f_A, f_B の特定とそれらからの f_C の算出となる。Burrell モデルでは，f_A はポアソン分布，f_B はガンマ分布と仮定し，これから，f_C として負の二項分布を導出した。これは，前節で述べたように，かなりの汎用性がある。しかし，図書館には多くの種類やそれぞれに固有の状況があり，すべてが，必ずしも，Burrell が採用した確率分布にしたがうと考える必要はない。例えば，ある図書館においては，f_B はガンマ分布よりも，ポアソン分布による近似のほうがよいかもしれないし，またそれほど図書の数が多くなく，貸出も少ない図書館では，f_A は二項分布にしたがうかもしれない。

このように考えると，各 f_A, f_B の組み合わせに対して，さまざまな f_C が導かれ，貸出頻度分布の候補となりうる可能性が出てくる。そこで，便利なように，これらを第１表にまとめる。他にもさまざまな複合分布が存在するが，ここでは，実用性等を考慮したうえで，その

第1表 おもな複合分布

f_A	f_B：パラメターの分布	f_C：複合分布	適用例
ポアソン分布	負の指数分布	幾何分布	Burrell (1980)[2]
ポアソン分布	対数級数分布	負の二項分布	
ポアソン分布	ガンマ分布	負の二項分布	Bagust[5], Burrell モデル
ポアソン分布	ガンマ分布	対数級数分布（負の二項分布の成功回数を0に近づけたときの極限分布）	
ポアソン分布	ポアソン分布	ネイマンのA型伝播性分布	
二項分布	ポアソン分布	ポアソン二項分布	
二項分布	ベータ分布	ベータ二項分布（負の超幾何分布）	Gelman and Sichel[12]

うち主要なものをあげるにとどめた。このうち，幾何分布と，負の二項分布についてのレビューはすでにおこなったので，それ以外の分布について簡単に述べる。

a．ベータ二項分布（負の超幾何分布）

この分布は，最近，Gelman and Sichel[12] によって，貸出頻度分布として用いられた。彼らは，図書の館外貸出の頻度が館内利用の頻度に比べて低いことと，貸出が各図書館の貸出条件の制限を受けるために有限であることの2点を考慮して，f_A の分布として，二項分布を採用した。そして，f_B の分布としては，柔軟性に富むベータ分布を導入して，複合分布として，以下のようなベータ二項分布，

$$P(X=r)=\binom{n}{r}\frac{B(a+r,b+n-r)}{B(a,b)},\qquad r=0,1,2,\cdots \tag{8}$$

を導出した（$B(a,b)$ はベータ関数)。ただし，Gelman and Sichel は，$X=0$（すなわち未貸出の図書数）を切断 (truncate) している。このベータ二項分布は負の二項分布の一般化でもあり，ある極限をとれば，負の二項分布に近づく[11]。

b．対数級数分布

対数級数分布は，

$$P(X=r)=-\frac{1}{\ln(1-\theta)}\frac{\theta^r}{r},\qquad 0<\theta<1,\quad r=1,2,\cdots \tag{9}$$

であたえられる。この分布は，ポアソン分布のパラメターがガンマ分布にしたがうと仮定し，そして，ガンマ分布のパラメター k（(5) 式参照）を無限に0に近づけると得られる分布である[13]。すなわち，対数級数分布は，負の二項分布の極限分布である。ただし，この分布では，$X=0$（すなわち，未貸出図書）は考慮されない。

c．ネイマンのA型伝播性分布

f_A がポアソン分布，さらに，f_B も，ポアソン分布，

$$f(\lambda)=\frac{h^\lambda}{\lambda!}e^{-h}$$

で記述できるとすれば，複合分布の定義により，

$$\begin{aligned}P(X=r)&=\sum_{j=1}^{\infty}e^{-h}\frac{h^j}{j!}e^{-jd}\frac{(jd)^r}{r!}\\&=\frac{d^r}{r!}e^{-h}\sum_{k=1}^{r}(he^{-d})S_r^{(k)}\exp[de^{-d}],\\&\qquad r=0,1,2,\cdots\end{aligned} \tag{10}$$

と計算できる[13]。ただし，$S_r^{(k)}$ は，第2種のスターリング数，d はパラメターである。この分布はいわゆる，ネイマンのA型伝播性分布である。

d．ポアソン二項分布

複合分布の最後として，ポアソン二項分布をあげる。この分布は，f_A が二項分布にしたがい，f_B がポアソン分布で記述されるときに導かれる分布である。しかし，この分布は，その計算およびパラメターの推定がかなり複雑で，現段階では，その実用性がかなり低い。そこで本論文では，その可能性の示唆だけにとどめ，この以上の考察は行わない。

3．切断分布の利用

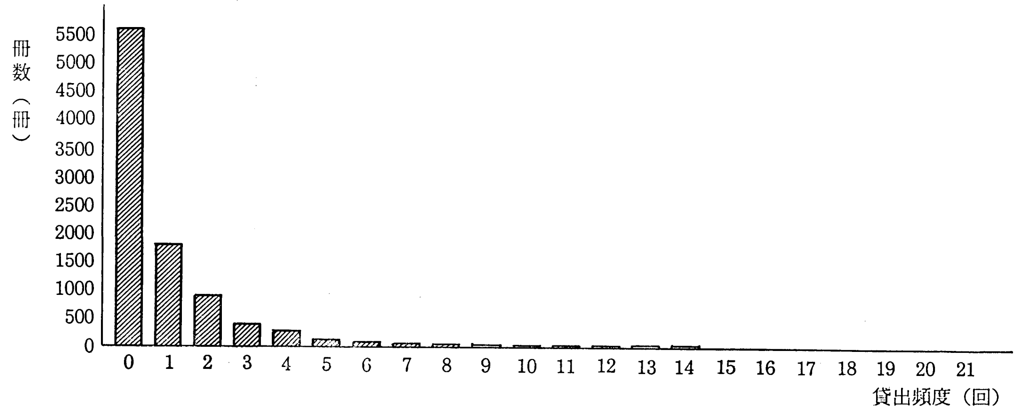
1度も貸し出されない図書（いわゆる未貸出図書）の冊数には，紛失や盗難などに起因するものが含まれるために，未貸出図書数はモデルでは記述できないとして，確率分布 $X=0$ の部分を切断する (truncate) という考え方がある。この考えに立つ切断分布としては，Morse[14] や Burrell and Cane[3] の切断幾何分布があり，さらに前述したように，Gelman and Sichel[12] のベータ二項

分布モデルは切断分布である。

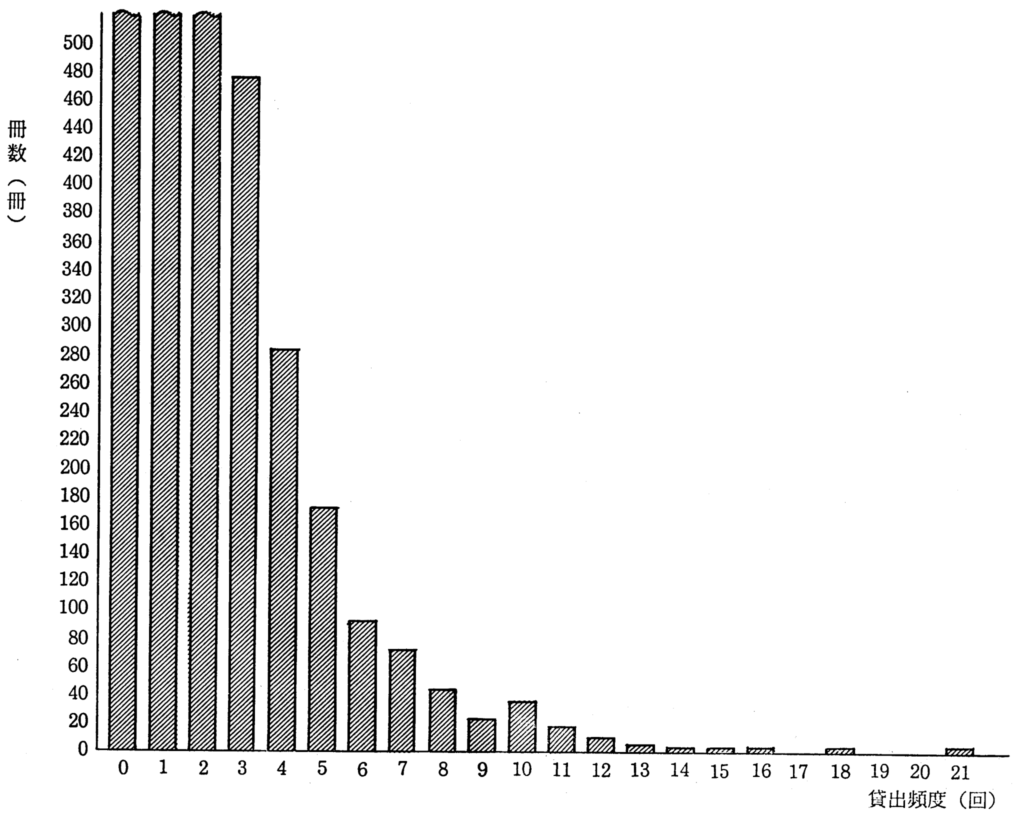
この切断分布は，たとえば，何らかの理由により図書の紛失が多いとか，参考図書やリザーブ図書の数と貸出可能の図書の数とがいっしょになってしまっている場合などに，有効であると考えられる。

III. 実際のデータを用いた検証

著者らは，Burrell モデルや，前章で議論した各複合



全 体 図



拡 大 図

第２図　日吉情報センターにおける1982年出版図書の83年度における貸出頻度分布（全体図と拡大図）

分布を，実際の図書館の貸出データに適用し，その有効性を検証することを試みた。

利用可能なデータは，慶應義塾大学日吉キャンパス内にある日吉情報センターの1983年度から1985年度までの貸出記録である。その日吉キャンパスには，教養課程が置かれている。したがって，日吉情報センターは，おもに，各学部の1,2年生を奉仕対象とする，いわゆる学習図書館である。その蔵書規模は，1986年の時点で，和書約21万冊，洋書約10万冊であり，その奉仕対象者数は約1万人である。

A. 各複合分布についての分析

まず，1983年度の貸出記録のなかから，1982年出版の

第2表 各分布の観測値へ適合度の比較（その理論値とカイ2乗統計量）

頻度	観測値[1]	負の二項分布[2]		ベータ二項分布[3]		ネイマンのA型伝播性分布[4]		幾何分布[5]		対数級数分布[6]	
0回	5542	5542.2[7]	0.00[8]	5610.5	0.84	5539.8	0.09	4751.5	131.51	—	—
1	1876	1782.2	4.93	1680.9	22.63	1376.1	189.43	2370.0	102.96	2705.0	254.04
2	826	889.8	4.58	870.3	2.25	1122.9	78.52	1182.2	107.28	747.7	8.19
3	475	497.0	0.97	508.1	2.14	693.4	68.79	589.6	22.28	275.6	144.29
4	283	292.3	0.30	310.4	2.41	376.6	23.28	294.1	0.42	114.3	249.15
5	170	177.2	0.29	193.1	2.77	194.2	97.31	146.7	3.70	50.5	282.37
6	92	109.4	2.77	120.7	6.81	97.3	0.29	73.2	4.85	23.3	202.80
7	72	68.5	0.18	75.1	0.13	47.4	12.75	36.5	34.54	11.0	336.86
8	44	43.3	0.01	46.2	0.11	22.5	20.66	18.2	36.56	5.3	280.07
9	24	27.6	0.48	28.0	0.56	10.4	17.93	9.1	24.52	2.6	1760.92
10	36	17.7	18.85	16.6	22.78	4.7	551.59	4.5	497.21	1.3	—
11	16	11.4	1.83	9.6	4.35	2.1	—	2.3	—	0.7	—
12	9	7.4	0.34	5.3	2.54	0.9	—	1.1	—	0.3	—
13	5	4.8	0.18	2.8	16.54	0.4	—	0.6	—	0.2	—
14	2	3.1	—	1.4	—	0.2	—	0.3	—	0.09	—
15	2	2.0	—	0.7	—	0.1	—	0.1	—	0.04	—
16	3	1.3	—	0.3	—	0.03	—	0.06	—	0.02	—
17	0	0.9	—	0.1	—	0.01	—	0.03	—	0.01	—
18	1	0.6	—	0.04	—	0.00	—	0.01	—	0.01	—
19	0	0.4	—	0.01	—	0.00	—	0.01	—	0.00	—
20	0	0.3	—	0.00	—	0.00	—	0.00	—	0.00	—
21	2	0.2	—	0.00	—	0.00	—	0.00	—	0.00	—
合計	9480	9479.8	35.7	9479.9	86.86	9480.0	966.3	9480.0	965.8	3938.0[9]	3464.69

注：1．データは日吉情報センターの1982年出版の図書についての83年度の貸出記録。平均貸出数は，0.995，分散は3.26。

2．パラメターは，$\hat{k}=0.475$，$\hat{p}=0.323$。$X=0$ の頻度と標本平均から推定。

3．パラメタ I は，$\hat{a}=0.400$，$\hat{b}=8.042$，$n=21$。積率推定法による。

4．パラメターは，$\hat{n}=0.714$，$\hat{d}=1.349$。$X=0$ の頻度と標本平均から推定。

5．パラメターは，$\hat{p}=0.501$。最尤推定量。

6．パラメターは，$\hat{\theta}=0.552$。最尤推定量。

7．各分布による理論値。

8．カイ2乗統計量。各分布とも，理論値が5以下の図書数はまとめて，カイ2乗値を算出した。

9．対数級数分布は，$X=0$ で定義されないので，合計冊数は，貸し出された図書の総数に等しくなるように規格化されている。

図書についての貸出のみを計数し，そのデータに各複合分布をあてはめることを試みた。そのデータをグラフ化したもの第2図に，そしてあてはめた結果を，第2表にまとめる。

第2図からは，この1982年出版の図書についての貸出頻度分布が，右方向にかなりなめらかに減衰し，右すそを長くひいていることがわかる。ただし，貸出頻度10回の図書数が，そのまわりに比べてやや大きく，特異な値を示している。しかし，全体としては，これは典型的な貸出頻度分布の形状である。

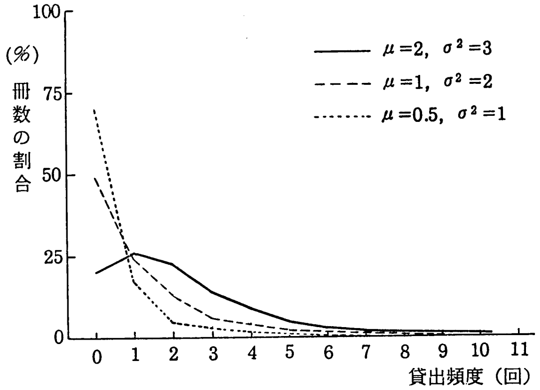
このデータに，前章で考察した各分布をあてはめた結果が，第2表である。この表にはデータと理論値との適合の目安としてカイ2乗統計量を示してあるが，これから，次のことが読み取れる。

①負の二項分布とベータ二項分布の適合度が他の分布に比べてかなり良好である。特に，負の二項分布は，特異値である頻度10を除いて考えれば，全体的に非常によく適合しており，カイ2乗適合度検定を行っても，かなりよい結果が得られる。また，ベータ二項分布も，パラメターの推定を，さらに丹念におこなえば，より精度があがる可能性はある。

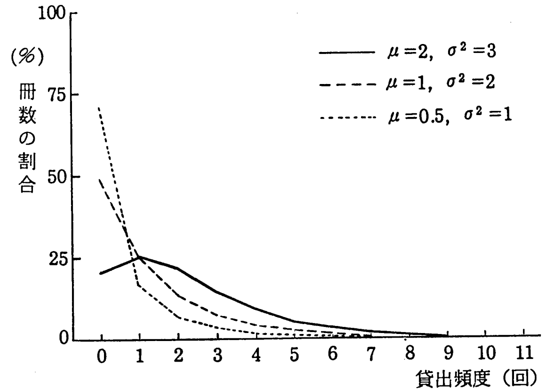
②幾何分布の適合性は先行研究の結果と同じく，不十分であった。また，対数級数分布はもっとも適合していない。もっとも，これらの分布は，負の二項分布のパラメター k を，それぞれ，$k=1$, $k\to 0$ としたものであるが，第2表の注2より，負の二項分布のパラメター $\hat{k}$ は 0.475 なので，このような結果になったのは，当然であろう。

③ネイマンのA型伝播性分布の結果は，それほどよくなかった。

このように，第2表の結果からは，負の二項分布とベータ二項分布の優位性が示されたが，つぎに，この2つの分布の比較を試みる。第3図は，平均と分散の3種類の組み合わせ（① $\mu=2$, $\sigma^2=3$, ② $\mu=1$, $\sigma^2=2$, ③ $\mu=0.5$, $\sigma^2=1$）に対して，両分布のパラメターを積率推定で算出し，それによる分布を図示したものである。この図は，両分布の形状にほとんど違いのないことを示している。ただし，右すそに注目すると，ほんのわずかながら，負の二項分布の確率のほうが大きい。この差は全体からみれば，僅少であるが，図書館の蔵書数は一般にかなり大きいので，このような差が，実際に効いてくる可能性がある。例えば，第2表においても，頻度の大きいものの確率が，ベータ二項分布では過小評価され，その分，負の二項分布よりもカイ2乗統計量が大きくなっている（ただし前述したように，ベータ二項分布は，負の二項分布の一般形であり，パラメター n を大きくすれば，右すその確率は大きくなる。第2表のデータに対しても，実際に n を適当に大きくすれば，負の二項分布よりも，右すそを長くひく。しかし，その分，頻度の小さい部分への適合は悪くなる）。



負の二項分布



ベータ二項分布

第3図　負の二項分布とベータ二項分布との比較

以上のように，貸出頻度分布としては，負の二項分布が優れていることが，日吉情報センターのデータによって検証された。ただし，ベータ二項分布の可能性も同時に示された。前述の Gelman and Sichel[12] の仮定も考慮に合わせると，貸出がそれほど活発でなく，それほど貸出頻度の大きい図書が存在しないような図書館においては，ベータ二項分布のほうがよい結果がえられるということが考えられる。

B. Burrell モデルの検証

次に，現実のデータへの Burrell モデルの適用を試みる。

前節のさまざまな分布に比べ，Burrell モデルの優れている点のひとつは，obsolescence がモデル自体に組み込まれていることである。そのため，第Ⅱ章で述べたように，Burrell モデルは，非定常ポアソン過程となるが，これは，貸出数の時間的な変動を記述する上で，定常ポアソン過程モデルよりも，より現実に即している。例えば，定常ポアソン過程モデルで記述をおこなうと，各図書の貸出は時間が経ってもいっこうに衰えず，いつかはすべての図書が必ず貸し出されるという，現実に即さない結論が得られることになる。そのため，前節で扱ったような複合分布では，ある一定期間のみの貸出頻度分布は記述することができても，その時間的変動を扱うのには適していない。

さらに，Burrell[6] では，図書館の実際の現場で便利なように，各年次ごとの分布 $P(Y_n)$ が与えられている。これにより，各年次ごとの貸出頻度分布の変動の記述が可能となる。

ここでは，その Burrell の各年次別分布モデルが，貸出数の年ごとの変動をどの程度記述できるかを，日吉情報センターの1982年出版の図書についての，83年度から85年度までの貸出記録によって，検証する。ただし，この1982年出版の図書数は，1983年の時点で固定されるものとする。

1．パラメターの推定法

現実のデータに適用する際，パラメターの推定は重要である。負の二項分布のパラメターは，基本的には p と k のふたつであるが，Burrell モデルでは，p のかたちが，(7) 式に示されているように複雑なので，いくつかの工夫を行わなければ，パラメター $p(t)$ を決めることはできない。Burrell[6] による，詳しい手順は，ここでは省略するが，結果的には，(7) 式のパラメターは以下のように求められる[6]。

①まず，一般の統計理論を用いてパラメター k を推定する。この負の二項分布の k の推定については，付録にまとめてある。

②第 n 年のパラメター p を p_n と書くことにする。いくつかの工夫をおこなうと，これは，

$$p_n=\frac{k}{k+\mu_1\theta^{n-1}} \tag{11}$$

と変形できる。ただし，μ_1 は，第1年の貸出数の期待値，$\theta=e^{-a}$（(3)式参照）である。ここで，p_n を推定するには，k の推定はすでに①で行われているから，μ_1 と，θ を推定すれば p_n が求められることになる。μ_1 は，$E(Y_n)$ であるから，実際のデータから第1年の標本平均（平均貸出回数＝総貸出数／総図書数）を計算して，これを用いればよい。また，θ についても，

$$\frac{E(Y_n)}{E(Y_{n-1})}=\theta \tag{12}$$

となることから，実際のデータの連続する2年間の平均貸出数を用いて計算し，それを θ の推定値とおけばよいことになる。

2．実際のデータによる検証

次に，実際に日吉情報センターの1982年出版の蔵書についての貸出データから，これらのパラメターの推定を試みる。

k については，第2表の注に示すとおり，$\hat{k}=0.475$ である。これは，$X=0$ の度数と標本平均を用いて推定したものであるが，積率推定で推定した $\hat{k}=0.44$ よりも，適合が良好であったために，そちらを採用した。Anscombe[15] の研究によれば，k が大きく，μ が小さい場合には，積率法が有効であり，k も μ も小さい場合は $X=0$ の度数の方法が望ましい。

次に，パラメター θ を，83年度の総貸出数と，84年度の総貸出数から推定すると，$\hat{\theta}=0.987$ と計算された。そこで，この $\hat{k}$ と $\hat{\theta}$ とをもちいて，82年出版の図書についての83，84，85年度の貸出数を理論的に求めた。その結果と，その実際の値とのずれをみるためのカイ2乗値とを，第3表に示す。さらに，その適合度を評価する目安として，同様のデータについて，Burrell モデルではなく，各貸出年度ごとに，0の度数と標本平均を用いて k と p とを推定したものを第4表として示す。

これらは，Burrell モデルによる理論値のデータへの適合が，84，85年度と年が経つにつれて悪化していることを示している。84年度は，第4表の結果とくらべて，まだ許容範囲内としても，85年度はかなりずれがある。すなわち，Burrell モデルは，貸出数の年次変化にそれほど対応していないことがわかる。

この原因は，ふたつ考えられる。ひとつは，パラメター θ の推定値がばらつくことである。Burrell モデルでは，(3) 式のかたちから，θ の値が一定であることが必要である。ところが，その θ は連続する2つの年度の平

第3表　82年出版の図書の貸出記録とその理論値（年度別）

総図書数：9480冊

頻度	観測値			理論値[1]			カイ2乗統計量		
	83[2]	84	85	83	84	85	83	84	85
0回	5542	5661	5899	5542.2	5566.6	5590.0	0.00	1.60	16.99
1	1876	1804	1779	1782.2	1782.1	1781.9	4.93	0.26	0.04
2	826	843	710	889.8	885.9	881.8	4.58	2.07	33.48
3	475	442	423	497.0	492.6	488.1	0.97	5.19	8.70
4	283	253	234	292.3	288.4	284.6	0.30	4.35	8.98
5	170	155	119	177.2	174.0	170.9	0.29	2.07	15.76
6	92	94	81	109.4	107.0	104.6	2.77	1.58	5.33
7	72	66	77	68.5	66.7	64.9	0.18	0.01	2.23
8	44	33	53	43.3	42.0	40.7	0.01	1.93	3.70
9	24	33	39	27.6	26.7	25.7	0.48	1.50	6.84
10	36	27	22	17.7	17.0	16.4	18.85	5.84	1.95
11	16	29	13	11.4	10.9	10.5	1.83	29.88	0.62
12	9	19	10	7.4	7.0	6.7	0.34	20.29	1.61
13	5	5	4	4.8	4.6	4.3	0.18[3]	7.42	11.27
14	2	11	8	3.1	3.0	2.8	—	—	—
15	2	0	3	2.0	1.9	1.8	—	—	—
16	3	1	1	1.3	1.3	1.2	—	—	—
17	0	2	4	0.9	0.8	0.7	—	—	—
18	1	0	0	0.6	0.5	0.5	—	—	—
19	0	1	1	0.4	0.4	0.3	—	—	—
20	0	1	0	0.3	0.2	0.2	—	—	—
21	2	0	0	0.2	0.1	0.1	—	—	—
合計	9480			9479.8	9479.7	9479.7	35.72	84.02	117.49
平均	1.00	0.98	0.91						
分散	3.26	3.47	3.22						

注：1．パラメターは，$\hat{k}=0.475$，$\hat{\theta}=0.98$。
2．貸出の行われた年度。
3．カイ2乗適合度検定のため，貸出頻度13回以上については，その図書数を合計し，そのカイ2乗値をもとめた。

第4表　0の度数と標本平均によるのパラメター推定によるあてはめの結果

年度	パラメター $\hat{k}$	パラメター $\hat{p}$	カイ2乗統計量（計）
84	0.435	0.309	62.2
85	0.402	0.307	61.0

均貸出数から推定するのであるが，第5表が示すとおり推定値 $\hat{\theta}$ は厳密には一定ではなく，平均値を算出する貸出年度によって変動がある。もともと，この種の事象は不確定の誤差をともなうので，厳密な一致は不可能である。この点からは，第5表の $\hat{\theta}$ の値は1.00から0.90の範囲に収まっており，大局的には，(3) 式が成り立つとしても，誤差を取り除き，真の θ を推定するのは困難である。

もうひとつの原因は，パラメター k の問題である。この負の二項分布の k は，複合分布である Burrell モデルにおいては，“desirability” の分布であるガンマ分布 (5)式のパラメター k である。ガンマ分布において，パラメター b が尺度的な因子として作用するのに対して，パ

第5表 日吉情報センターにおける82年から77年までの各出版年の図書の貸出記録によるパラメーターθの推定値 $\hat{\theta}$[1]

出版年	82	81	80	79	78	77
貸出年度						
83-84[2]	0.987	0.917	0.913	0.905	0.912	0.965
84-85[3]	0.922	0.919	0.964	0.930	1.000	0.928

注：1. $\hat{\theta}=n+1$年度の総貸出数／n年度の総貸出数。
2. 上式で，$n=83$。すなわち，83年度の貸出総数／84年度の貸出総数。
3. 上式で，$n=84$。

ラメーターkは，その質的な影響を与える。Burrell モデルでは，このkは年次経過を通して変化しないということが，暗黙のうちに仮定されている。しかし，第4表が示すように，k は年が経つにつれて，変化する可能性がある。つまり，"desirability" の形状は年次的に変化すると仮定できる。つまり，大学図書館においては，その利用者が入学・卒業等によって，毎年変化するので，その図書に対する "desirability" もその影響を受け，その形状が微妙に変化すると考えることができる。しかし，Burrell モデルでは，k が固定されているので，この変化に対して，結果的に，パラメーターpのみで対応しなければならず，ここに無理が生じることになる。そして，これによって，年が経つにつれて適合度が落ちるという結果になったと推察できる。

そこで，このパラメーターkの問題の解決策として，著者らは，Burrell モデルを一部修正したモデルを次章にて提案する。パラメーターθについては，特に，抜本的な解決は得られなかったが，最小2乗法等を用いるなどして，精度をあげることが考えられる。

IV. 修正モデルの導出と検証

A. 修正モデルの導出

Burrell モデルの改良の最も直接的な方法は，ガンマ分布のパラメーターk自体に，時間的要素を組み込むことである。それには，(5)式の k を $k(n)$ と置き換えて，

$$f(\lambda)=\frac{b^{k(n)}}{\Gamma[k(n)]}\lambda^{k(n)-1}e^{-b\lambda} \tag{13}$$

とする。ここで，$k(n)$ は，第 n 年におけるガンマ分布のパラメーターkをあらわす。すなわち，修正モデルでは "desirability" の確率分布の形状が年次的に変化すると仮定する。修正モデルは，このパラメーター $k(n)$ の変化によって，貸出頻度分布の時間的変動に対応するため Burrell モデルのように，非定常ポアソン過程となる必要はなく，普通の定常ポアソン過程である。つまり，n 年から $n+1$ 年に年次が変わると，パラメーター $k(n)$ が $k(n+1)$ に置き換わり，その結果，貸出数の変動が起きるものとする。

そこで，まず，ポアソン分布は，

$$p(Y=r)=\frac{e^{-\lambda t}(\lambda t)^r}{r!}$$

となるが，単位時間を1年とするので，$t=1$ である。次に，このポアソン分布のパラメーター λ がガンマ分布(13)にしたがうと仮定する。そこで，(6)式に対応する積分，

$$p(Y_n=r)=\int_0^\infty \frac{e^{-\lambda}\lambda^r}{r!}f(\lambda)d\lambda \tag{14}$$

に，(13)を代入して，積分を実行すれば，

$$\begin{aligned}p(Y_n=r)&=\frac{b^{k(n)}}{r!\Gamma[k(n)]}\int_0^\infty \lambda^r e^{-\lambda}e^{-b\lambda}\lambda^{k(n)-1}d\lambda\\&=\frac{b^{k(n)}}{r!\Gamma[k(n)]}\int_0^\infty \lambda^{r+k(n)-1}e^{-\lambda(b+1)}d\lambda\\&=\frac{b^{k(n)}}{r!\Gamma[k(n)]}\frac{\Gamma[r+k(n)]}{(b+1)^{r+k(n)}}\\&=\frac{[r+k(n)-1]!}{r!}\left(\frac{b}{b+1}\right)^{k(n)}\left(\frac{1}{b+1}\right)^r\end{aligned}$$

と計算できる。ここで，$p=b/(b+1)$ とおけば，この修正モデルは，$k(n)$ と p をパラメーターとしてもつ，負の二項分布となる。

ところで，負の二項分布のパラメーターkが既知の場合のpの推定値は，

$$\hat{p}=k/(k+\bar{x})$$

である（付録参照）。そこで，第 n 年の平均貸出数を $\bar{x}_n$ と書くことにすれば，第 n 年のパラメーター p の推定量 $\hat{p}_n$ は，

$$\hat{p}_n=k(n)/[k(n)+\bar{x}_n] \tag{15}$$

と求めることができる。

さて，この修正モデルを実際に適用するとすれば，$k(n)$ を n の関数と見なすのが便利である。その関数 $k(n)$ の具体的なかたちとして，今回は，

$$k(n)=k_1e^{-c(n-1)} \tag{16}$$

と定めることにする。ただし，k_1は，第１年の負の二項分布のパラメターkである。この $k(n)$ については，今回は(16)式を採用するが，今後，さらに考察が必要である。さて，ここで，$e^{-c}=\gamma$ とおけば，(16)式は，

$$k(n)=k_1\gamma^{n-1} \tag{17}$$

とかける。そこで，このγの値がなんらかの方法で求めることができたならば，k_1から，それ以降の $k(n)$ の値を任意に求めることができる。これは，実際に貸出数の予測などに役立つであろう。ここでは，γ を求める最も簡単な方法として，第１年目と第２年目の実際のデータから推定する方法を採用する。それは，第１年の平均貸出数と分散を $\bar{x}_1$, s_1^2, 第２年のそれを $\bar{x}_2$, s_2^2 と書けば，負の二項分布の平均，分散の定義と(17)式を用いて，

$$\bar{x}_1^2/(s_1^2-\bar{x}_1)*\hat{\gamma}=\bar{x}_2^2/(s_2^2-\bar{x}_2)$$

という関係がえられるので，これを解いた，

$$\hat{\gamma}=\frac{(s_1^2-\bar{x}_1)\bar{x}_2^2}{(s_2^2-\bar{x}_2)\bar{x}_1^2} \tag{18}$$

から γ を推定する方法である。もちろん，これも $k(n)$ の場合と同じく，将来の改良が期待できる。

この $\hat{\gamma}$ を(17)式に代入して，$k(n)$ を決定することができれば，この修正モデルによって，年次別の貸出頻度分布を記述することが可能となる。しかし，$k(n)$ と同様に $\bar{x}_n$ についても，

$$\bar{x}_n=f(n)$$

のかたちを定めて，x_n を関数的に決定できるようにすれば，予測等に便利となる。この $f(n)$ については，さまざまなものが可能であり，たとえば，過去の値から，最小２乗法によって定める方法などが考えられる。ここでは，$k(n)$ と同じく，

$$\bar{x}_n=\mu_1 e^{-d(n-1)}$$

と定めることにする（μ_1 は第１年の分布の期待値)。ここで，やはり，$e^{-d}=\theta$ とおけば，$\bar{x}_n=\mu_1\theta^{n-1}$ となる。これを，(15)に代入すれば，p_n が定まることになるが，実はこの p_n は，Burrell モデルの(11)式の k を $k(n)$ に置き換えたものと一致する。すなわち，この修正モデルは，$\gamma=1$ とおくと，Burrell モデルの年次別分布に一致することになる。

以上の結果から，第n年の貸出頻度分布予測をおこなう場合の，修正モデルは，

$$P(Y_n=r)=\binom{r+k(n)-1}{r}p(n)^{k(n)}[1-p(n)]^r,$$
$$r=0,1,2,\cdots$$

ただし，

$$k(n)=k_1\gamma^{n-1},\quad p(n)=\frac{k_1\gamma^{n-1}}{k_1\gamma^{n-1}+\mu_1\theta^{n-1}} \tag{19}$$

となる。

各年の平均と，分散は，負の二項分布のそれらの定義により，

$$\left.\begin{aligned} E(Y_n)&=k(n)\frac{1-p(n)}{p(n)}=\mu_1\theta^{n-1},\\ V(Y_n)&=k(n)\frac{1-p(n)}{p(n)^2}\\ &=\frac{k_1\gamma^{n-1}\mu_1\theta^{n-1}+\mu_1^2\theta^{2n-2}}{k_1\gamma^{n-1}}\end{aligned}\right\} \tag{20}$$

と計算できる。

B. 修正モデルの検証

修正モデルを第３表の実際のデータにあてはめるために，いくつかのパラメターの推定をおこなう。

まず，k_1については，83年度の貸出データより，０の度数と標本平均を用いて，推定するが，これは，既に求めており，$\hat{k}_1=0.475$ である。θ については，(20)より，

$$\frac{E(Y_{n+1})}{E(Y_n)}=\theta$$

が成り立つから，Burrell モデルと同様に推定できる（(12)式参照)。ただし，ここでも，前章でのべたθの推定の問題がふたたび起こる。$\hat{\theta}=0.987$, $\hat{\theta}=0.992$, $\hat{\theta}=(0.987+0.922)/2$ などが考えられるが，ここでは，第１年次と第２年次のみのデータから求めることができる，$\hat{\theta}=0.987$ を採用する。次に，γ については，(18) を用いる。データより，$\bar{x}_1=0.99515$, $\bar{x}_2=0.98259$, $s_1^2=3.26$, $s_2^2=3.47$ であるから，計算を実行すると，$\hat{\gamma}=0.88769$ となる。

これらのパラメターを用いて，データへのあてはめをおこなった結果を第６表に示す。修正モデルのカイ２乗統計量は，58.5と66.64であり，これは，Burrellモデルによる理論値よりも，修正モデルのそれのほうが実際のデータにより適合していることを示している。さらに，これを第４表の数字と比べても，この修正モデルの適合がかなり良好であることがわかる。これらのことから，修正モデルの妥当性が示された。

第6表 修正モデルによる理論値（82年出版の図書の貸出について）

頻度	84年度[1]理論値		85年度理論値	
	$\hat{k}=0.422$ $\hat{p}=0.300$	カイ2乗統計量	$\hat{k}=0.374$ $\hat{p}=0.279$	カイ2乗統計量
0回	5708.5	0.39	5875.4	0.09
1	1648.2	8.53	1586.6	23.34
2	837.6	0.03	786.5	7.44
3	473.1	2.05	449.1	1.51
4	283.2	3.21	273.3	5.66
5	175.2	2.33	172.5	16.59
6	110.8	2.54	111.5	8.33
7	71.1	0.37	73.2	0.19
8	46.2	3.75	48.7	0.38
9	30.2	0.26	32.7	1.21
10	19.9	2.51	22.1	0.00
11	13.2	18.89	15.0	0.27
12	8.7	11.84	10.3	0.08
13	5.9	0.13	7.1	1.33
14	3.9	1.66	4.9	0.25
15	2.7	—	3.4	—
16	1.8	—	2.3	—
17	1.2	—	1.6	—
18	0.8	—	1.1	—
19	0.6	—	0.8	—
20	0.4	—	0.6	—
21	0.4	—	0.4	—
計	9479.6	58.50	9479.1	66.64

注：1. 貸出年度。

しかし，$k(n)$ の性質や，パラメター θ, γ の推定の問題は残る。今後は，$k(n)$, $f(n)$ や，γ の決定方法の研究・改良がさらに必要である。

V. 結　論

以上の分析から，貸出頻度分布としての負の二項分布の妥当性が検証され，また，貸出頻度分布の時間的変動を記述・予測するモデルとして，著者らの案出したモデルが有効であることが明らかとなった。

しかし，すでに述べたように，著者らの修正モデルに関しては，$k(n)$ の改良や，パラメター θ, γ の推定など，問題がいくつか残っている。また，負の二項分布の妥当性についても，すべての図書館の貸出頻度分布が負の二項分布にしたがうとは，断言できない（第Ⅲ章ではベータ二項分布の有効性も示唆された)。さらに，この貸出モデルの，実際の図書館政策への具体的な応用については，今回は特に触れなかった。そこで，今後，さらに，さまざまなデータの分析・モデルの改良を進める必要がある。

ところで，一方，貸出が図書館の蔵書の「有効性」を測定する唯一の尺度であるかどうかという疑問も生じる。この点に関して，貸出データが，図書館の「有効性」をどの程度示しているかということも，今後，探究されるべき問題である。

1) Morse, Philip M. Library Effectiveness: a systems approach. Cambridge, M.I.T. Press, 1968, 207p.
2) Burrell, Quentin. A simple stochastic model for library loans. Journal of Documentation. Vol. 36, No. 2, p. 115-132 (1980).
2) Burrell, Quentin and Cane, Violet R. The analysis of library data. Journal of the Royal Statistical Society, Series A. Vol. 145, p. 439-463 (1982).
4) Discussion of the paper by Mr. Burrell and Professor Cane. Journal of the Royal Statistical Society, Series A. Vol. 145, p. 463-471 (1982).
さらに，貸出頻度分布が負の二項分布にしたがうという点にのみについては，Wall, T. Letter to the editor. Journal of Documentation. Vol. 36, No. 4, p. 343-344 (1980). が指摘している。
5) Bagust, A. A circulation model for busy public libraries. Journal of Documentation. Vol. 39, No. 1, p. 24-37 (1983).
6) Burrell, Quentin. A note on ageing in a library circulation model. Journal of Documentation. Vol. 41, No. 2, p. 100-115 (1985).
7) Burrell, Quentin. A second note on ageing in a library circulation model: the correlation structure. Journal of Documentation. Vol. 42, No. 2, p. 114-128 (1986).
8) Burrell, Quentin. A third note on ageing in a library circulation model: applications to future use and relegation. Journal of Documentation. Vol. 43, No. 1, p. 24-45 (1987).
9) Greenwood, Major and Yule, G. Udny. An inquiry into the nature of frequency distributions of multiple happening with particular reference to the occurence of multiple attacks of disease or repeated accidents. Journal of the Royal Statistical Society, Series A. Vol. 83, p. 255-279 (1920).
10) 工業関係では，Wise, M. E. The use of the neg-

ative binominal distribution in an industrial sampling probrem. Journal of the Royal Statistical Society, Series B. Vol. 8, p. 202-211 (1946).
商業関係では，Chatfield, C. et al. Progress on a simplified models of stationary purchasing behavior. Journal of the Royal Statistical Society, Series B. Vol. 129, p. 317-360 (1966).
11) 竹内　啓，藤野和健．二項分布とポアソン分布．東京，東京大学出版会，1981，262p.
12) Gelman, E. and Sichel, H. S. Library book circulation and the beta-binominal distribution. Journal of the American Society for Information Science. Vol. 38, No. 1, p. 4-12 (1987).
13) Johnson, Norman L. and Kotz, Samuel. Discrete distribution: distributions in statistics. New York, John Wiley & Sons, 1969, 328p.
14) Morse, Philip M. Measure of library effectiveness. Library Quarterly. Vol. 42, p. 15-30 (1972).
15) Anscombe, F. L. Sampling theory of negative binomial and logarithmic series distribution. Biometrika. Vol. 37, p. 358-382 (1950).

付　録

複合分布とそのパラメターの推定法

本論文で扱った確率分布は，一般的な統計学の教科書ではそれほど解説されていないものが多い。そこで，便利なように，そのパラメターの推定法を，付録としてまとめておく。ただし，その多くは，文献の 13) を参考にした。

①負の二項分布

一般的なかたちは，

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

である。このとき，x は非負の整数であるが，k は任意の実数である。そこで，k が整数以外の場合は，二項係数を拡張する必要がある。それには，

$$(a)_r=a(a-1)(a-2)\cdots(a-r+1)$$

と定義して，

$$\binom{a}{r}=\frac{(a)_r}{r!}$$

とする。なお，$k=1$ の場合が，幾何分布である。

パラメターの推定については，まず k が前もって知られている場合は，$\bar{x}$ を標本平均として，

$$\hat{p}=\frac{k}{\bar{x}+k}$$

が，最尤推定量である[11]。次に，p と k の両方を推定する場合であるが，これには 3 つの方法がある。それは，

①積率推定法
②$x=0$ の度数と標本平均を用いる方法
③最尤推定法

である。しかし，この場合，最尤推定法は計算が煩雑であり，実用的ではないので，ここでは省略する。そこで，まず，積率推定法であるが，s^2 を標本分散として，

$$\hat{p}=\bar{x}/s^2, \quad \hat{k}=x^2/(s^2-\bar{x})$$

である。次に，$x=0$ の度数 f_0 を用いる方法は，

$$\hat{p}\bar{x}\ln\hat{p}=(1-\hat{p})\ln(f_0/n)$$
$$\hat{k}=\ln(f_0/n)/\ln\hat{p}$$

という連立方程式の解を $\hat{p}$, $\hat{k}$ とする[11]。これは，第 1 式を，$\hat{p}=f(\hat{p})$ のかたちに解き，そして，反復計算で $\hat{p}$ を求め，その $\hat{p}$ を第 2 式に代入すればよい。

②ベータ二項分布

二項分布とベータ分布の複合から導かれる式は，本文中の第 (8) 式であるが，ベータ関数の計算が困難な場合は，前に拡張した二項係数を用いて，

$$P(x)=\binom{-a}{x}\binom{-b}{n-x}\Big/\binom{-a-b}{n}$$

というかたちを利用すればよい。この分布の平均と分散は，

$$\mu=na/(a+b)$$

$$\sigma^2=\frac{na}{a+b}\cdot\frac{b}{a+b}\cdot\frac{a+b+n}{a+b+1}$$

であるので，μ, σ^2 のかわりに，標本平均 $\bar{x}$，標本分散 s^2 をもちいて，これを解けば，a, b が推定できる。したがって，

$$\hat{a}=\frac{\hat{a}^2n\bar{x}(n-\bar{x})(\hat{a}+\bar{x})}{(\hat{a}^2n^2+\bar{x})s^2}$$

から，反復計算で $\hat{a}$ を算出し，

$$\hat{b}=\hat{a}(n-\bar{x})/\bar{x}$$

で，$\hat{b}$ を求めればよい。

③ネイマンのA型伝播性分布

(10)において，$S_k^{(r)}$ は，第2種のスターリング数で，

$$S_k^{(r)}=\frac{1}{k!}\sum_{i=0}^{n-1}(-1)^i{}_kC_i(k-i)^r$$

と定義される。この分布の平均と分散は，

$$\mu=hd \tag{21}$$

$$\sigma^2=hd(1+d) \tag{22}$$

である。

パラメターの推定については，まず，(21), (22)から，積率推定量は，

$$\hat{h}=\bar{x}/\hat{d}$$
$$\hat{d}=(s^2-\bar{x})/\bar{x}$$

となる。ただし，$\bar{x}$, s^2 は，それぞれ標本平均，標本分散である。

つぎに，頻度0の割合，

$$f_0=\exp[-h(1-e^{-\hat{d}})] \tag{23}$$

と(21)を用いる方法がある。これは，まず，(23)の両辺の対数をとって，

$$\log f_0=-\hat{h}(1-e^{-\hat{d}})$$
$$h=-\log f_0/(1-e^{-\hat{d}})$$

と変形してから，μ を $\bar{x}$ に置き換えた(21)を代入すれば，

$$\hat{h}=-\log f_0/(1-e^{-\bar{x}/\hat{h}})$$

となるので，これを反復法でとけばよい。

④対数級数分布

対数級数分布のパラメター θ の最尤推定量は，

$$\bar{x}=\frac{\hat{\theta}}{-(1-\hat{\theta})\ln(1-\hat{\theta})} \tag{24}$$

を満たす。ただし，これは，一意には解けないので，かわりの推定量として，

a) 1－[$x=1$ の割合]/$\bar{x}$

b) 1－[原点回りの2次の積率]/$\bar{x}$

などがある。第2表では，a)の方法による推定量をもとに，近似的に(24)を解いてある。