# 大学図書館における図書資料除架基準策定問題に対する数理モデルを用いた解法の研究

M2004MM004 花里　真吾

指導教員　鈴木　敦夫

## 1　はじめに

図書館における情報の活用は 1960 年代から 80 年代にかけて盛んに行われた [1][2][4][5]．そのなかでも Philip M.Morse によって行われたマサチューセッツ工科大学の大学図書館を対象とした研究はよく知られている．Morse は図書館の運営に関するさまざまな問題をオペレーションズ・リサーチの手法を用いて考察している．Morse が研究を行った 1960 年代はデータを取得することが大変な労力であったが, 現在では日常の図書館運営において自然にコンピュータにデータが蓄積されるようになっている．またコンピュータの処理性能が飛躍的に高まっている今日において, さまざまな計算結果を実業務に直接利用できる場面は多いと考えられる．一方, 図書館におけるデータは年間貸出冊数や予約件数, ベストリーダ, 蔵書回転率, などに集計され, 図書館運営に反映されている．最近は出版点数の増加が顕著であるにもかかわらず図書館のスペースは一定の大きさであるために, 書誌の選択だけではなく除架の問題も重要になってきた [6].

## 2　背景

新刊書籍の 1 年間の出版点数は近年 30 年間において約 3.69 倍にもなっている．その伸びは 1990 年ごろから特に大きく伸びている．インターネットが急速に普及した 1995 年以降でも従来と同様に出版点数は増加傾向にある．編集作業が以前より効率化し, 小ロットでも製本コストを抑えた出版が可能になってきたこと, 多様化している読書ニーズを探るように小ロットでも多種の本を出版するようになってきたことなどがその要因として考えられる．

図書館は物理的なスペースが限られているため, 出版点数の増加はスペース不足の問題を引き起こす．限られた予算, 限られたスペースで一定の蔵書回転率を保つためには図書の選択の問題と併せ, 除架の問題が重要となる．特に私立大学の図書館には, 厳しい経営環境を反映した図書関連予算の減少により年間受入冊数を増やせないという問題もある．その上で従来どおり学生, 教職員のニーズを満たしていくという使命がある．『2004 年度出版年鑑』および『2003 年度日本の図書館 2004』によれば 1998 年から 2002 年の大学図書館における蔵書冊数については 1998 年に比べ 2002 年では和書, 洋書ともにどの大学区分でも増加している．年間受入冊数においては和書ではどの大学区分でも経年とともに増加傾向にあるが, 洋書については減少傾向にある．また和書と洋書の合計では 2000 年を境に増加傾向に歯止めがかかっている．図書資料費については 1999 年から 2002 年までの 4 年間ですべての大学区分で減少している．これらのことから, 限られた予算で利用者の要求を満たそうと努力している図書館の姿が浮かび上がる．

これらの背景を考えると不必要な本を早く除架し, スペースを確保することが重要性を増していることがわかる．

## 3　非定常ポアソン過程

非定常ポアソン過程とは偶然に発生するイベントの発生率が時間の変化により変動する現象をあらわす過程である．地震の発生などの自然現象などに適用されている．図書館の書籍においても時間当たりの貸出回数は時間経過とともに変化すると考えられるため, 非定常ポアソン過程でモデルの作成を行う [3].

**定義**

$$N(\ 0\ ) = 0 \tag{1}$$

$$\text{過程は独立増分を持つ} \tag{2}$$

$$P[\ N(t+h) - N(t) = 1\ ] = \lambda(t)\ h + \ o(h) \tag{3}$$

$$P[\ N(t+h) - N(t) \geq 2\ ] = o(h) \tag{4}$$

非定常ポアソン過程は, 独立増分を持つが定常増分を持たない．なお $\lambda(t)$ は, 強度関数であり, 書籍の貸出においては時間当たりの貸出回数に相当する．

**平均値関数**　時刻 $t$ までの累積貸出回数は次のようになる．また期待値にもなっている．

$$\Lambda(t) = \int_0^t \lambda(s)\ ds \tag{5}$$

定常ポアソン過程における時刻 $t$ までにおける累積回数の期待値が $\lambda t$ であることに対応している．

**貸出間隔**　分布関数の定義は満たさないが, 瞬間 $MTBF(t)$ , 累積 $MTBF(t)$ を以下のように定義して代替的な尺度と考える．
なお $MTBF(Mean\ Time\ Between\ Failure)$ は信頼性工学分野で使用される用語である．なお定常ポアソン過程に従う場合イベント発生の時間間隔の分布は指数分布になり, 期待値は $\frac{1}{\lambda}$ となる [7].

$$\text{瞬間 } MTBF(t) : \frac{1}{\lambda(t)} \tag{6}$$

$$\text{累積 } MTBF(t) : \frac{t}{\Lambda(t)} = \frac{t}{\int_0^t \lambda(s)\ ds} \tag{7}$$

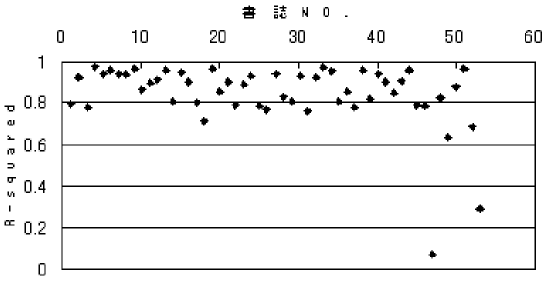


図 1　平均値関数の推定結果 (指数)

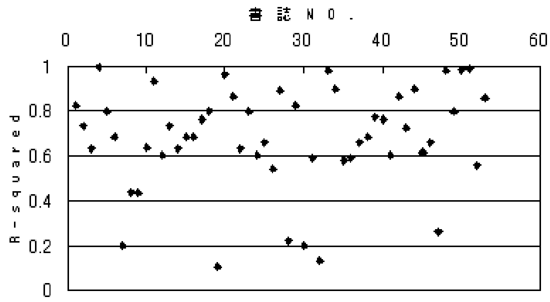


図 2　強度関数の推定結果 (指数)

## 4　指数モデル

書籍は経年により, 内容が陳腐化したり物理的に劣化したり, または本の存在が認知されたりするなど, 時間当たりの貸出回数は変化していくと考えられる. そこで瞬間の貸出率は時間の経過に比例して変化すると仮定する.

書誌の時間当たりの貸出回数が, ある定数に比例して小さくなっていると仮定すると以下の式が成り立つ.

$$\frac{d}{dt}\lambda(t) = -b\ \lambda(t) \quad (\ 0 \le b < 1\ ) \tag{8}$$

強度関数は次のようになる.

$$\int \frac{1}{\lambda(t)} d\lambda(t) = -\int b\ dt \tag{9}$$

$$\lambda(t) = c_1 e^{-bt} \quad (\ c_1 > 0,\ 0 \le b < 1\ ) \tag{10}$$

次に, 平均値関数は, 次のようになる.

$$\Lambda_E(t) = \int_0^t \lambda(s)\ ds = \int_0^t ce^{-bs}\ ds = \frac{c_1}{b}(\ 1 - e^{-bt}\ ) \tag{11}$$

ここで, $A = \frac{c_1}{b},\ B = b$ とおくと, 式 (11) は, 以下の式になる.

$$\Lambda_E(t) = A(\ 1 - e^{-Bt}\ ) \tag{12}$$

次に単位時間を半年にし, 2000 年 11 月 1 日から 12 月 31 日までに受け入れられた書籍のうち貸出回数が 10 回の書籍 53 タイトルについて非線形最小 2 乗法の適用を試みた. $R^2$ は図 1 のとおりとなった.

結果は $R^2$ の値が 53 タイトル中 14 タイトル以外はすべて 0.8 を超えており全体的に良好だった. 受入からの時間経過において初期に時間当たりの貸出率が増えていくようなケースのいくつかは近似した結果が直線的になった. 次に貸出間隔から直接強度関数を推定することも試みる. (6) 式と (7) 式を用いて以下の式を仮定する.

$$\frac{1}{\lambda(t)} = \frac{t}{\Lambda(t)} \tag{13}$$

両辺の逆数を考える.

$$\lambda(t) = \frac{\Lambda(t)}{t} = \frac{N(t)}{t} \tag{14}$$

$\lambda(t) = AB^t$ とおいて両辺の対数をとり, 最小 2 乗法で近似を行った. $R^2$ は図 2 のとおりである.

$R^2$ の値が 0.8 を超えたものは 53 タイトル中 16 タイトルにとどまり, モデルの適合がよいとはいえないことがわかった. 平均値関数を推定するより適合がよくなかったことについては, $\frac{1}{\lambda(t)} = \frac{t}{\Lambda(t)}$ という仮定をおいたこと, 分布関数の和が 1 にならないこと等によると思われる.

## 5　ロジスティック曲線モデル

図書館が蔵書として受入れたことが, 時間経過とともに認知されることにより, 書誌の時間当たり貸出回数が貸出に応じて増え, 内容の陳腐化などにより時間が経過するにつれ 2 乗に比例して減るモデルを考える. 貸出回数はある定数に比例して大きくなり, 貸出回数の 2 乗はある定数に比例して減るものと仮定すると, 以下の式が成り立つ.

$$\frac{d}{dt}\Lambda(t) = (\ a - b\Lambda(t)\ )\ \Lambda(t) \quad (\ a \ge 0,\ b > 0\ ) \tag{15}$$

式 (15) はベルヌーイ型の微分方程式となる. しかし部分分数に分けることで変数分離形になる.

$a > 0,\ b > 0,\ \Lambda(\ 0\ ) = \frac{a}{b}$のとき

$$\frac{1}{(\ a - b\Lambda(t)\ )\ \Lambda(t)} d\Lambda(t) = dt \tag{16}$$

$$\Lambda(t) = \frac{\frac{a}{b}}{1 + \frac{1}{bc_2}\exp[-at]} \tag{17}$$

ここで, $A = \frac{a}{b}, B = \frac{1}{bc_2}, C = a$ とおくと平均値関数は次のようになる. また考察の範囲を $A \ge 0,\ B > 0,\ C > 0$ とする.

$$\Lambda_L(t) = \frac{A}{1 + B\exp[-Ct]} \quad (\ A \ge 0,\ B > 0,\ C > 0\ ) \tag{18}$$

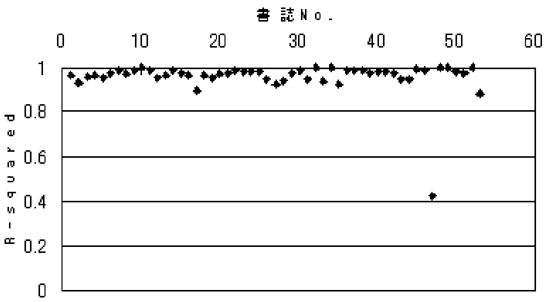


図 3　平均値関数の推定結果 (ロジスティック曲線)

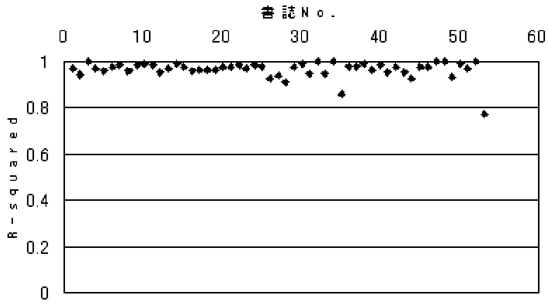


図 4　平均値関数の推定結果 (ゴンペルツ曲線)

ロジスティック関数は初期に貸出が増加傾向にある場合があることをうまく説明している．強度関数は, 次のようになる．

$$\frac{d}{dt}\Lambda_L(t) = \frac{d}{dt}\left(\frac{A}{1+B\exp[-Ct]}\right) = \frac{ABC\exp[-Ct]}{(1+B\exp[-Ct]\ )^2} \tag{19}$$

また $\infty$ 時間経過後の漸近値は次のようになる．

$$\lim_{t\to\infty}\Lambda_L(t) = \lim_{t\to\infty}\frac{A}{1+B\exp[-Ct]} = A \tag{20}$$

図 3 はロジスティック関数を平均値関数としてデータのあてはめを行ったものである．指数モデルと同様に単位時間を半年にし, 2000 年 11 月 1 日から 12 月 31 日までに受け入れられた書籍のうち貸出回数が 10 回の書籍 53 タイトルについて非線形最小 2 乗法を用いて適用を試みた． $R^2$ の値は非常に良好であった．

# 6　ゴンペルツ曲線モデル

書誌の時間当たり貸出回数が, ユーザに認知されることにより時間経過とともに増えたり, 書誌の内容の陳腐化などにより貸出回数が減るモデルにおいて, 任意の時間間隔で借りられる回数が指数的に変化していくとする仮定をおき, 以下の式で表す．

$$\frac{d}{dt}\Lambda_G(t) = (\ \log a\ )(\ \log b\ )\ b^t \times \Lambda_G(t) \quad (\ a>0,\ b>0,\ t>0\ ) \tag{21}$$

よって

$$\Lambda_G(t) = c_1 a^{b^t} \quad (c_1 > 0 : \text{parameter}\ ) \tag{22}$$

あらためて $A = c_1,\ B^{e^{-Ct}} = a^{b^t}$ とおくことで, 平均値関数は次のようになる．

$$\Lambda_G(t) = AB^{e^{-Ct}} \quad (\ A>0,\ B>0,\ C>0\ ) \tag{23}$$

また, 強度関数は, 次のようになる．

$$\frac{d}{dt}\Lambda_G(t) = \frac{d}{dt}(\ AB^{e^{-Ct}}\ ) = -AB^{e^{-Ct}}Ce^{-Ct}\log B \tag{24}$$

また $\infty$ 時間経過後の漸近値は次のようになる．

$$\lim_{t\to\infty}\Lambda_G(t) = \lim_{t\to\infty}AB^{e^{-Ct}} = A \tag{25}$$

ゴンペルツ関数も初期に貸出が増加傾向にある場合があることをうまく説明している．また, ロジスティック曲線に比べると漸近値に緩やかに近づいていく傾向がある．

図 4 はゴンペルツ関数を平均値関数としてデータのあてはめを行ったものである．指数モデル, ロジスティック曲線モデルと同じデータを用いて非線形最小 2 乗法により, あてはめを試みた．また $R^2$ の値も求めた．その結果 $R^2$ の値は非常に良好であった．しかしロジスティック曲線モデルと同様パラメータ数が指数モデルに比べ 1 つ多いためその結果あてはまりがよくなっている可能性がある．

# 7　AIC によるモデル選択

ロジスティック曲線モデル, ゴンペルツ曲線モデルは, $R^2$ の値も非常によかったがモデルのパラメータ数が指数モデルに比べ多いため AIC を用いて評価を行った．

$$AIC = -2\times L + 2\times P \tag{26}$$

$L$ : 最大対数尤度

$P$ : パラメータ数

AIC の第 1 項は最大対数尤度が大きければ小さくなり第 2 項はパラメータ数が多くなれば大きくなる．同じデータセットに対しては AIC の値が小さいほどよいモデルであるといえる．

結果はほとんどの書誌で指数モデルよりロジスティック曲線モデル, ゴンペルツ曲線モデルのほうが小さい値となった (図 5) ．またロジスティック曲線モデルとゴンペルツ曲線モデルでは差は軽微だがロジスティック曲線

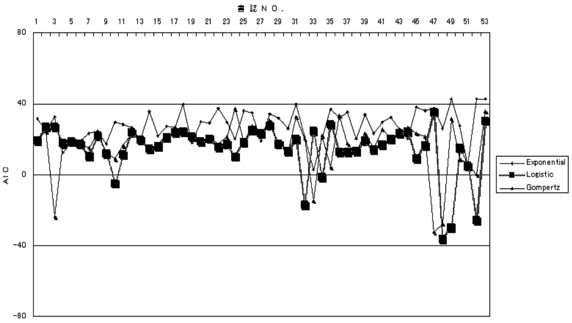


図 5　AIC によるモデル評価

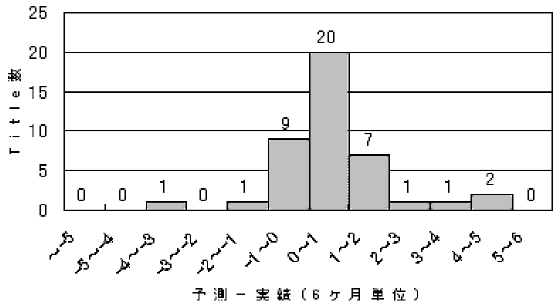


図 6　予測値－実績値の頻度分布

モデルのほうが全体的には AIC が小さい値を示す傾向があった．また, ロジスティック曲線モデルへの当てはまりがよく, ゴンペルツ曲線に当てはまりが悪いもの, ロジスティック曲線モデルへの当てはまりが悪く, ゴンペルツ曲線に当てはまりがよいものがあった．

結果的に受入から初期には貸し出しが増加傾向にあり, 一定期間経過した後に貸し出しが減少に転じるというモデルが適していることがわかった．モデルとしては初期には貸出は図書館が受け入れてから認知される過程に関連すると思われる増加要因が一次関数的もしくは指数関数的に関係し, 後期には内容の陳腐化などによる書籍の老化要因により 2 次関数的, もしくは指数関数的に関係するモデルとなる．

次に AIC によるモデルの評価で良好な結果を示した, ロジスティック曲線モデルを用いて次回の貸出日を予測した．10 回貸出があった書籍について 9 回目までのデータを用いて 10 回目の貸出日を予測する．まず単位期間を半年として, 半年毎に貸出回数を数え累積した．そのデータを用いてロジスティック曲線を推定した．モデルで推定した 10 回目の貸出日と実際の 9 回目の貸出日の差を次回貸出日までの日数とする．なお同一日に貸出があった場合については対象外とした．その上で実際の貸出間隔と予測による貸出間隔を比べると, 結果は図 6 のようになり, 29 タイトルが実際の貸出日から半年以内に入っていた．

## 8　考察

モデルにより得られた予測関数の漸近値を見ることによって, その時点における潜在的な本の寿命が貸出回数の意味でわかる．そのため実用にあたり漸近値の持つ意味は非常に大きい．例えばモデルからはすでに寿命が尽きていると想定されている書籍でも貸出があったものがあった．そのため慎重に判断する必要がある．今後はマルチスタート等により, グローバルな解に少しでも近づけるようプログラムの改良を図る必要がある．

## 9　おわりに

今回は 10 回目までの貸出が 4 年以内に起こっているケースのみを対象に検証したが, 仮に一律に 10 年経過していることを除架, 廃棄する際のリストアップ要件にすると, 4 年経過時点において, 次回貸出の予測が 6 年以上の本は廃棄の対象としてもよいと考えられる．そのことにより書架スペースの有効な利用が可能になる．また, どのような書籍の寿命が短いかを知ることは書誌の選択に際しても重要な意味を持つ．実際の除架, 廃棄にあたっては, 書誌自体に価値があるものや, あまり借りられてはいないがいろいろな書籍に引用されている書籍, 常備しておくべき基本的な書籍, 等は図書館に存在していることが望ましいため, 除架, 廃棄の判断には, 引用数等の情報も考慮したうえで最終的には司書の専門性に委ねられる．また, このことを考慮した上で, 実際に現場の書誌の廃棄プロセスにおいて次回貸出日を参考に廃棄を行えるようアプリケーションのフロントエンドも整えたい．

## 参考文献


[1] Ching-chih Chen, *Quantitative Measurement and Dynamic Library Service*, ORYX PRESS, Phoenix, 1978.

[2] 原田勝, 海野敏, 鈴木志元, 谷口祥一, 石井啓豊, 図書館情報学における数学的方法, 日外アソシエーツ, 1994.

[3] 尾崎俊治, 確率モデル入門, 朝倉書店, 1996.

[4] Philip M. Morse, *Library Effectiveness -A Systems Approach-*, THE M.I.T PRESS, Cambridge, 1969.

[5] Quentin L. Burrell and Violet R. Cane, The Analysis of Library Data, *J.R.Statist.Soc.*, **A,145**(1982) , Part 4, pp.439-471.

[6] Stanley J. Slote, *Weeding Library Collections -Library Weeding Methods- Forth Edition*, LIBRARIES UNLIMITED, INC., Colorado, 1997.

[7] 山田茂, ソフトウェア信頼性モデル-基礎と応用, 日科技連, 1994.