Studies on frequency distributions of recorded use for students using academic library collections

STUDIES ON

FREQUENCY DISTRIBUTIONS OF RECORDED USE

FOR STUDENTS USING

ACADEMIC LIBRARY COLLECTIONS

by

Terry Keith Wall

A Doctoral Thesis
Submitted in partial fulfilment of the requirements
for the award of the degree of
Doctor of Philosophy
of the Loughborough University of Technology
October 1987

CONTENTS page

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

ACKNOWLEDGEMENTS

My sincere thanks are due to the following:

Professor P. Havard-Williams and Professor A.J. Meadows, successively Heads of the Department of Library and Information Studies of the Loughborough University of Technology, for guidance during the course of this work;

Mr. D.M. Ellis, formerly Lecturer in the Department of Mathematics, UWIST, for advice and for pointing out the simplified expression for $p(0)$ used in Figure 8.1;

Dr. J.M. Brittain, Senior Lecturer in the Department of Library and Information Studies of the Loughborough University of Technology, for commenting on a draft of Chapters 1 to 6;

Mr. R.J. Horrill, Librarian of University College Cardiff, and Ms. A. Moloney for making available circulation data;

Mr. J.C. Hartas, formerly Librarian of Bristol Polytechnic, and Mr. C. Pinder for making available circulation data;

The Principal, Librarian and Academic Staff Committee of UWIST for supporting a request for study leave;

My wife, Gill, for typing and retyping the manuscript.

ABSTRACT

Frequency distributions of recorded use for students using academic libraries were analysed using statistical models not previously employed for the purpose. The suitability of the data for such analysis is discussed. Evidence suggested that frequency distributions of recorded library use reflected real differences in amounts of library use by users. A computer simulation of library use by students was used to investigate the effects of competition among users upon distributions of use.

Negative binomial probability distributions were found to reproduce some of the observed patterns of user activity, but were rejected on grounds of fit and applicability. Other two and three-parameter probability distributions were considered. A novel modification of the negative binomial distribution (being a Neyman Type A-gamma distribution instead of a Poisson-gamma distribution) gave good fit to frequency distributions of recorded use from various libraries. The fitted parameters appeared to be related to statistics of use for the observed populations, but the diversity observed in reality among users was clearly simplified in a stochastic model with only three parameters.

In the second part of the study, methods of using the model were explored. Given stability in two of the three parameters, the model could be scaled with time to predict future frequency distributions. The extrapolation of numbers of non-users from one set of data is described. The effect upon the uptake of titles from a library collection of distributions of activity among students was also considered. By simplifying the model, relationships between the mean use by a group of users and maximum amounts of use by individuals, and between numbers of uses and numbers of titles used are suggested. A key factor in relating user activity to uptake is the extent to which users diversify in their use of titles.

SYMBOLS

Explanations of mathematical symbols are included in the text. The symbolism is intended to convey a clear meaning within a given context, but does not approach the rigour or sophistication of the mathematician. Some letters have duplicate meanings and a list of all the symbols employed (excepting those used in the computer simulation programs) is therefore given below.

a	1) scale parameter of the modified negative binomial and generalized inverse Gaussian-Poisson distributions 2) exponent used in the approximation converting numbers of uses to numbers of title uses 3) a variable
b	scale parameter of the gamma, negative binomial and modified negative binomial distributions
c	a constant
d	the denominator of a fraction
e	a constant, equal to 2.718
e_i	the expected frequency of the ith class or outcome in a contingency table
$E[X]$	mean or expected value of the variate named x
$f(x)$	the frequency with which a variable assumes the value x
$\Gamma(n)$	the gamma function of n
$g(x)$	the proportion of observations in which a variable assumes the value x
h	the highest observed value of a variable, r
k	shape parameter of the gamma, negative binomial, modified negative binomial and generalized inverse Gaussian-Poisson distributions
λ	mean of a Poisson variate
μ	mean of a lognormal variate
μ_r	the rth moment about the mean
m	1) the mean of a variate 2) the name or value of a continuous gamma variate representing mean rates of occurrence
m', m''	parameters of Neyman's Type A distribution

n	1) size of sample or component population 2) exponent used in the approximation for $b \ln[b/(b + 1)]$
N	size of population
o_i	the observed frequency of the ith class or outcome in a contingency table
p	1) the probability that an event occurs in a Bernouilli trial 2) scale parameter of the geometric, negative binomial and modified negative binomial distributions
$p(x)$	the probability or relative frequency with which a variable assumes the value x
$p_t(x)$	the probability or relative frequency with which a variable assumes the value x in a period of observation of t time periods
$\underline{P}$	1) the probability that an observed frequency distribution could have arisen in random sampling from a hypothetical population 2) the probability that an observed chi-squared statistic would be exceeded in random sampling
$P(x\|y)$	the conditional probability of x given y
q	1) scale parameter of the geometric and generalized inverse Gaussian-Poisson distributions 2) $1 - p$
ρ	population correlation coefficient
r	1) product-moment correlation coefficient 2) a variable
σ	standard deviation
s	1) standard deviation 2) a variable
s^2	variance
t	1) a statistic conforming to Student's t distribution 2) number of time periods
u	a variable
var$[X]$	the variance of the variable named x
x	a variable
$\bar{x}$	sample mean
y	a variable
χ^2	observed chi-squared statistic

CHAPTER 1

INTRODUCTION

1.1 PREAMBLE AND SCOPE OF THE STUDY

Library users differ in the amount of library material they use and in their frequency of library use. Even those with similar tasks to perform can differ widely. Fellow students, for example, rely on libraries to differing degrees (1) and use information with differing levels of sophistication (2). When numbers of users are tabulated against numbers of recorded uses for a library collection, a skewed frequency distribution will often result. Most users record the smaller numbers of uses or no uses at all, and relatively few users record the larger numbers of uses. This pattern, moreover, persists even when users or uses are subdivided. Knapp (3) found that the 'same pattern occurred no matter how the students were grouped,...by sex...scholastic aptitude...achievement...class level,...whether loans or titles,...reserve or general collection withdrawals, course borrowing or non-course borrowing'.

Of course, frequency distributions of recorded use are not inexorably skewed or of large variance; but where students are unconstrained in their method of seeking information and in the amount of information they use, it appears that it is often so. Thus, when a large class tackles, for example, a programme of essays, information may be derived from a variety of sources and processed in different ways. As a result, it seems, the frequency distribution of amounts of library use has a great range. But when a small group of students is assigned some reading and is closely monitored, then most perform to expectation and record similar amounts of use (4).

Frequency distributions of recorded use by potential library users have been noted by various writers, but nowhere in any analytical detail. Quite often, the writer is only concerned to relate differences in library use to the possession of particular attributes. Sometimes only the proportion of users and non-users in the potential user population needs to be calculated. Thus Lubans (5), investigating the non-use of academic libraries, compared the attitudes and background of potential users who recorded at least one circulation use in a period of twelve months with those who did not. For other writers, differences between users who use extensively and those who use only a little are also relevant. Wills and Oldman (6) and Harrop (1,9), for example, discuss factors which may cause such differences. A few writers, among them Knapp (3), Ritter (7), Clayton (8), Schnaitter (24), Maxted (25) and Wall (20), present more fully graduated distributions but none except Wall appears to have sought to generalise their form. Morse, using questionnaire data collected from visitors to the MIT Science Library found that the frequency distributions of numbers of books consulted per visit by homogeneous groups of users produced 'fairly good straight-line semilog plots' (10:31). He did not, however, consider how this geometric probability distribution might be modified as visits cumulated. The aim of this investigation is to describe and model such long-term frequency distributions of recorded library use.

1.1.1 Diversity in amount of library use by students.

It is not surprising that librarians have not analysed distributions of library use more fully. Faced, like Knapp, with evidence that students vary widely in their amount of library use, their attention has been drawn to the cause of the variation rather than the symptom itself. Frequently, it seems, they conclude that the variation is an abnormal state, that light users are underusing a valuable educational resource and therefore require instruction or motivation to bring them up to the level of the heavy users. To be sure, there have been those who point out other possible causes. Line (11), for example, stresses the barriers to information seeking in libraries, and the persistence needed to overcome them: the contrast between the librarian's sedulity and the user's expediency in such matters is therefore inevitable. Harrop demonstrates exactly how the teacher and the nature of the course can influence the pattern of library use (9). Wills and Oldman (6) point out that heavy users may not be more mature in their use of the library than light users: it is certainly well known that they may not be academically more successful (13:13,14:57-58). Furthermore, users may depend more than librarians realise upon convenient access and the ready availability of material. Studies such as that of Buckland (15) show how much the librarian's management of his resource can facilitate and promote wider library use.

Clearly, students do need instruction, or at least practice, in information seeking. Their bibliographical immaturity does not, however, decide the functional purposes to which they put libraries and the natural individual variation in style or amount of library use which results. Of course, librarians of academic libraries need to stress the centrality of their resource to the educational process if they are to maintain or improve it. It may not be easy, however, to demonstrate that the performance of undergraduate students at least is improved by access to a good library (9,14:57).

A recent collection of essays (16) on student reading needs seems likely, from its provenance, to reflect some commonly held views in the UK. The contributions of the librarians stress the need to educate students towards a more independent style of learning involving wider and more critical use of library resources; but the viewpoint of the student is also well recorded. His approach is likely to be practical:

> 'library use is for most students purely a means towards an end. The majority are motivated most directly by the demands of their course. All the evidence suggests that...the ultimate motivation towards library use comes from the setting of tasks by the tutor, and...determines whether...and also how it will be used' (17:5).

The evidence comes especially from Harrop (1,9,18). It is interesting to compare Knapp's summary of the American experience. The

> 'average undergraduate uses the library for course-related materials. There seems to be some correlation between [mean] borrowing and sex, scholastic achievement, and academic class. The most significant differences seem to relate to the instructors and their requirements' (19:301).

As Harris (13) notes, however, the material used by the student may not be that recommended by the lecturer. Sometimes, for example, the recommendation will not suit the abilities of all students. A humanities lecturer acknowledges that the

> 'significant variation in students' ability to assimilate, process, and above all to <u>use</u> information in a creative manner, is known to all teachers...'(29:171),

- especially to those in the humanities (9,see also 2). Further variation can be expected from differences in motivation. According to Mann:

> 'Many students today, not only in the sciences, consider their lecture notes the most important degree-getting aids they have. 'Reading round' a topic may be for the high-fliers or budding 'academics', but

for the person satisfied with a reasonable lower second not much 'extra' work is needed' (30:185,see also 9:28).

Similarly, Mays concluded that supplementary reading by the Australian undergraduates which he surveyed was prompted more by personal reading habits than by expectations of academic profit (14:59).

It seems reasonable to assume that library use will be directly influenced by these variations among students in propensity to read or ability to read effectively. (Oppenheim (31) concluded that students who disliked LSE library seemed to 'dislike reading generally'.) In this study, such differences within groups of users are taken to be universal and inevitable.

Attempts have naturally been made to relate observed differences between individuals or groups of users to their attributes. Some generalisations are easily made for averaged use by groups which are homogeneous in some respect. As early as 1934, McDiamid (32), using circulation records from seven colleges in the central US, concluded that, within wide local variations, more recorded use tended to be made by: humanities and social science students; women; senior students; and academically good students. (It may be, of course, that some of these variables are themselves correlated.) Harrop (1) and Wills and Oldman (6) note however, that academically poor students may also make much use of the library. The evidence is similarly conflicting from the American and Australian studies: correlation between grade-point averages and numbers of books borrowed (or used) is very low, but all investigators agree that students with higher scores tend to record the use of more library material (7,14,19,33-36,38). Of course, the teachers are highly influential in determining amounts of library use, as in the UK. Indeed, Waples (40) was able to show a correlation between amounts of recorded use by students and amounts of recorded use by their teachers.

Other attributes are more difficult to relate to amount of recorded library use. Lubans' results indicated differences in the characteristics of users and non-users, but these differences were not confirmed in a second study (12). McDowell (41) and Musavi (43) conducted similar surveys of academic library users. The non-engagement of the non-users in Musavi's sample appeared to extend beyond the library into academic life more generally. For a public library (where library use was, of course, optional) Madden has drawn quite a convincing picture of the differences in character and attitude between users and non-users from market research (44): the users claimed to be more active in other social settings. Even in a context where library use should not have been optional, Musavi's finding echoes that of Madden, and it seems, therefore, that to explain the diversity in amount of library use among similar students, an analysis of the actual or perceived benefits or rewards accruing to each individual from library use would be necessary (2,6,46,47:43,48). Dunn (49) performed such an analysis for supplementary information gathering by American undergraduates and successfully correlated the information sources they used (the college library was ranked second in importance behind the teachers) with broad categories of psychological motivation for seeking information.

1.1.2 Proportions of non-users

The proportion of non-users in the potential user population may be difficult to determine absolutely. Slater (51) quotes proportions of users and non-users in constructing profiles of typical non-users in industry and commerce. Unfortunately, the periods of observation are not quoted so that the figures are particularly meaningless. For a similar reason, Whitlatch (52) is able to quote reported proportions from 11% to 63% for non-users in academic libraries. The proportion of non-users will almost certainly decline as the period of observation is extended. The time period is therefore a necessary qualification in citing non-user data.

Strain (53) and Blagden (54) in surveying the use of technical libraries over known periods of time are also concerned with the proportion of users within the population of potential users. In terms of the 'penetration' of library service to potential users the results are disappointing. Oseasohn is similarly disappointed by the amount of use recorded by local practitioners of a modern medical library (55). Only one quarter of those eligible borrowed during a two year period, although two thirds enrolled as members. Unfortunately, the author does not discuss unrecorded use. It seems possible that in special library services of this sort, borrowing will represent a smaller proportion of total collection use than in public or academic libraries (c.f. 56:43). Even in academic medical libraries, it seems, students spend less time in the library, and buy or photocopy more material than in other disciplines (57:92). In these cases it is not clear whether the potential user population has been insufficiently defined, or whether, as Slater suggests (51), libraries are just not addressing the requirements of many potential users.

In academic libraries, the potential user population is easy to define and lists of those who are eligible to use the services of the library are usually available to the investigator. If the proportion which uses the library is to be gauged accurately, of course, the use of all aspects of the library service must be surveyed. Many investigators survey only use of the circulation service and risk being misled (see below, Section 1.3). In public libraries, the potential user population is impossible to define. Even lists of members constitute poor sampling frames (58,59). Nonetheless, differences in amounts of use among recorded users may be compared. Clark, for example, groups users according to frequency of visit and amount of use (62). Neither sex nor distance of residence seemed to account for these differences.

Strain's study indicated that only a small core within the potential user population were regular users of her technical library. The majority of users seemed to be infrequent users emerging only irregularly from the population. Lubans also concludes that his non-users were often intermittent users who happened not to have recorded use during the period of observation (12). Even after a year, some who claimed to be infrequent users of the library had still not been observed recording use. Potential users can be classed as non-users only in respect of the period of observation, therefore. Nonetheless, the relative amounts of use recorded by even arbitrarily dichotomised groups of 'heavy' and 'light' or 'frequent' or 'infrequent' users will be clearly distinguishable, whatever the time period.

1.1.3 Other measures of diversity

It is revealing to calculate the proportion of total uses generated by the most active users. In academic libraries a large amount of recorded use is often generated by only a small proportion of all potential users. Table 1.1 shows examples using data from various sources. In all cases the users are college or university students enrolled on taught courses. There are noticeable similarities between libraries. It appears that all the libraries represented receive uneven use from their potential users. This does not necessarily imply that non-users or infrequent users fail

TABLE 1.1

Proportions of recorded collection use generated by proportions of most active potential users in academic libraries.

Period of observation and source of data	% of potential users	% of total uses
Four weeks (20)	9	49
Nine weeks (7)	13	51
"	24	68
Quarter (21)	18	50
16 weeks (20)	11	51
Semester (8)	19	50
Semester (22)	25	67
Semester (3)	14	48
"*	9	50+
Semester (23)	10	50-
Semester (24)	9	50
Two terms (25)	10	49
36 weeks (26)	8	34
- (27)	33	75
"	50	86
- (28)	50-	95-

*Use of reserve collection only.
'+' = 'more than'; '-' = 'less than'. Some original data given with less precision than shown; e.g. 'about half' is shown as 50%. Where possible the proportion of users generating about 50% of use has been calculated.

to use the library in other ways or utilize less information in completing their assignments. The data relate simply to those uses which were recorded.

If non-users are disregarded, the disproportion in the use of library collections by potential users is more markedly similar. Figure 1.1 shows a plot for various sets of data (20). This form of plot, though confirming the 'heavy half' theory from market research (46) and suggesting a rule of thumb for the observed disproportion, does not, however, display clearly the differences between the distributions of use. The plot is constrained at each end and can occupy only the area above the diagonal between these corners. Two very different distributions are summarised in Table 1.2. They represent recorded use of short-loan collections in a UK academic library by second-year undergraduates and are constructed from data collected by the author. The distribution for the pharmacy students is reversed J-shaped, while that for the economics students is bell-shaped with a mode close to the mean. Although the plot of disproportionate use in Figure 1.2 indicates the greater disproportion among the pharmacists, it gives little indication of the scale of the difference.

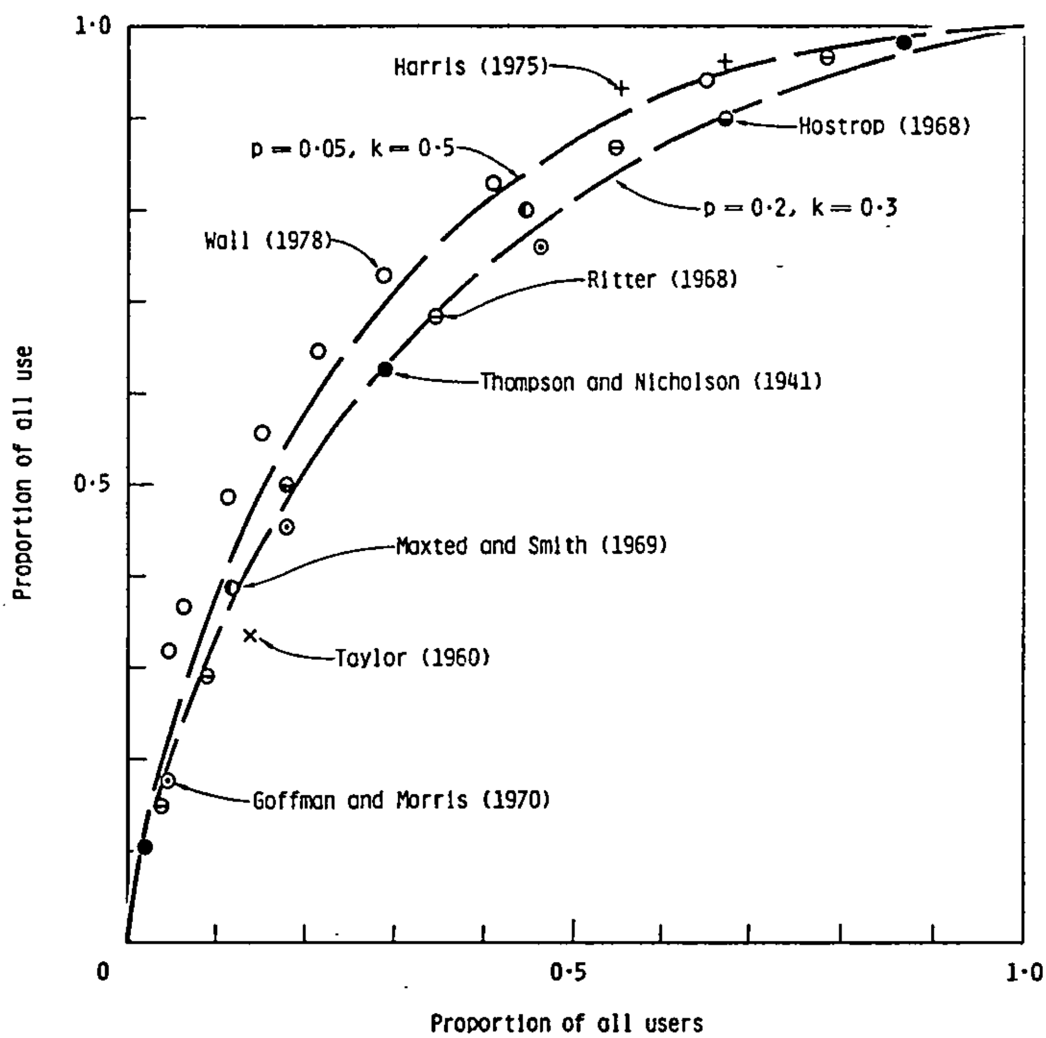
Clearly too, the proportion of non-users in the potential user population and the share of total use generated by proportions of active users will vary with time, making comparisons between different sets of data difficult when they are expressed only in these terms. For example, increasing the period of observation is likely to result in relatively more infrequent users being observed and this will accentuate the disproportion among users.

1.1.4 Scope of the study

A more rigorous analysis of some frequency distributions of use by library users is attempted in the following chapters, with the aim of finding a more satisfactory quantitative method for describing them. The frequency distributions chosen for analysis are taken from academic libraries. It is in these libraries that populations of potential users can be identified most precisely and where groups of users engaged on similar tasks can be readily found and compared. A mathematical model will be sought to represent what are felt to be the important features of the phenomena observed. Although the model will be arbitrary and not

FIGURE 1.1

Disproportionate use of library collections by users: proportions of total uses generated by proportions of most active users plotted from data in Goffman (60), Harris (61), Hostrop (21), Maxted (25), Ritter (7), Taylor (26), Thompson (22) and Wall (42). Lines represent proportions for negative binomial probability distributions fitted by Wall to frequency distributions of recorded use for students using a short-loan collection in a UK university library for time periods of differing length.



Source: redrawn from Wall (42).

TABLE 1.2

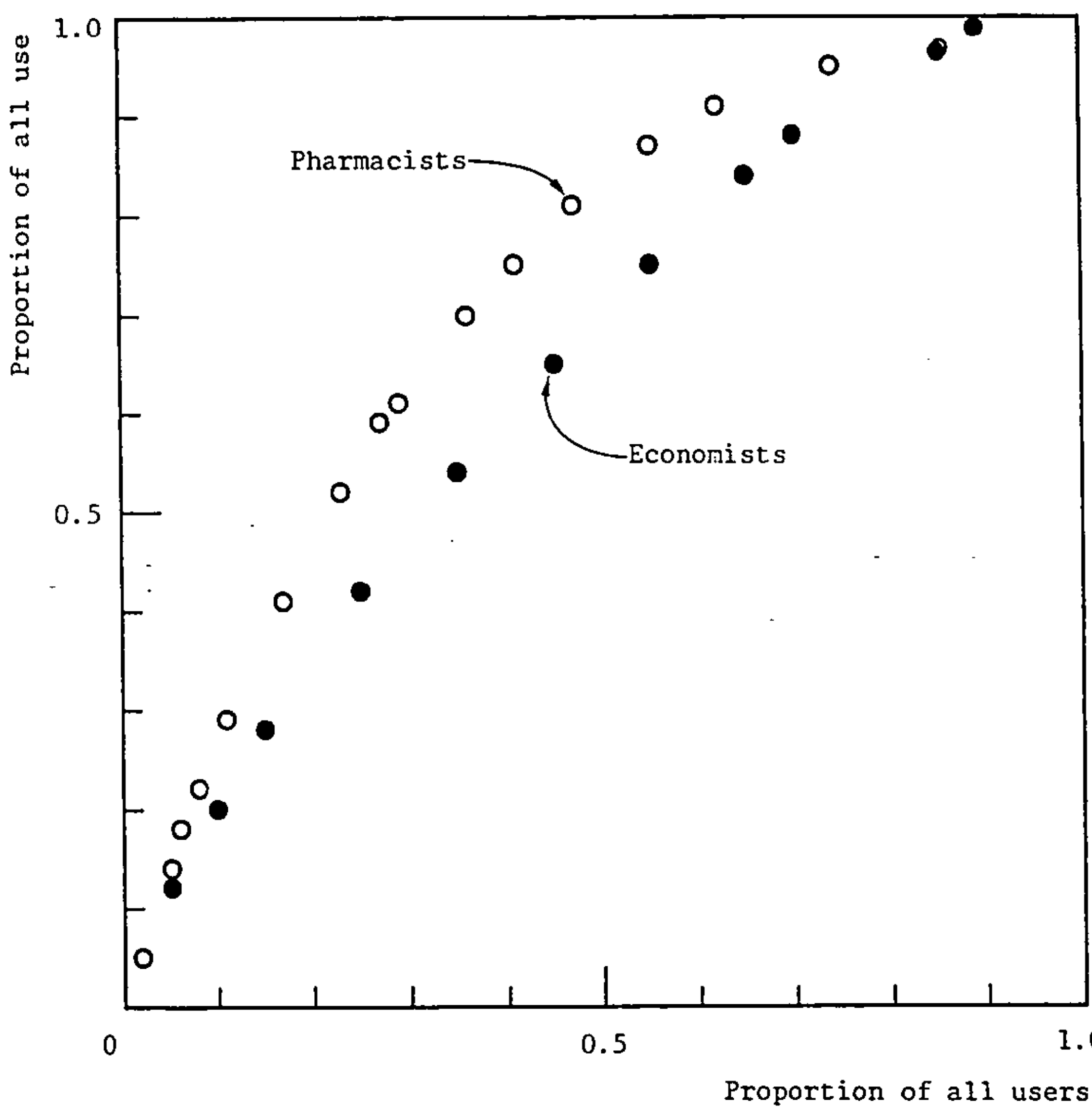
Summary of frequency distributions of numbers of recorded uses for two groups of second-year undergraduates using short-loan collections in a UK academic library over similar periods of time.

Number of uses	Pharmacy students	Economics students
0	13	0
1-4	29	2
5-9	13	1
10-14	12	3
15-24	11	7
25+	1	7
Period	8 weeks	10 weeks
Potential users	79	20*
Total uses	535	425
Mean use	6.8	21.25
Variance	47	135

*First 20 names from class of 86 persons.

FIGURE 1.2

Proportions of total uses generated by proportions of most active users from classes of second-year undergraduate economists and pharmacists using short-loan collections in a UK university library.



necessarily unique, it will confer the advantage, otherwise denied where phenomena are difficult to control and impossible to repeat, of allowing the manipulation of the variables observed. Only a statistical model will be supported: the data are not qualitatively or quantitatively distinguishable (the intensity, duration or value of each use is unknown) and they give no clue as to how use is generated. The representation will not necessarily be more than a statistically plausible approximation to the observed outcomes of the phenomena under discussion, therefore; a probability function relating numbers of users to given amounts of use. The use of a statistical model requires of the investigator a familiarity not only with the phenomena under investigation, but also with the application of statistical techniques. Though not without pitfalls, statistical methods are fortunately readily accessible to the layman who will find from the literature many examples of analyses to guide him.

1.2 DATA

As we have seen, a freely available library service seems rarely to be taken up at the same rate by all potential users: the provision of a loan collection in an academic library, for example, does not attract identical amounts of recorded use from each individual student - nor even from those enrolled on the same course. The data required to quantify accurately these differences among users determine to a large extent the services which can be studied. The amount of data needed will be considerable and the data must therefore be easy to collect.

The number of users studied must be large enough to yield a coherent pattern of diversity and a sample taken from a large population must be of sufficient size to represent the population to within given limits of statistical expectation. (The sampling unit would, of course, need to be the user: samples of uses would result in an underestimation of the numbers of infrequent users and so would bias the frequency distribution of use.) The period of observation must be sufficiently long to collect data for intermittent users, but not longer than the time period for which the potential user population remains unchanged. Sample sizes of the order of hundreds and periods of observation of the order of months will therefore be required.

Another consideration argues against the use of large samples, however. It is clear that the potential user populations of academic libraries can be subdivided into smaller, more homogeneous, groups some of which will differ in their mean amounts of collection use. Thus second-year economics students differ in their mean use from second-year pharmacy students, and would be likely to differ also from lecturers in engineering, part-time sociology students, and so on. Examples of these differences are given by Saunders (63), Whitlatch (52) and Harrop (18,64). It would seem therefore that, for taught-course students, a single class or course/year would provide the most suitable group of users for study. By definition, such a group is homogeneous in that its members have the same or similar tasks to perform.

Whether the group is a sample or an entire population, however, it must be large enough to ensure that, in analysing frequency distributions of use, the values of the variate are represented with sufficient frequency to enable tests of hypotheses about the form of the distribution to be conducted. Tests conducted using the chi-squared statistic will require a minimum of about five values in each cell. To avoid undue aggregation of the values in the tail or tails of a frequency distribution, therefore, samples of many tens or some hundreds will probably be required.

Large, less homogeneous, groups of users may demonstrate coherent patterns of use, but samples taken from them will also need to be large if they are to be representative. For example, data from a random sample of about 100 potential users would be required in order to estimate mean use to ±10% at the 95% level of confidence assuming that user scores were normally distributed in an infinite population with a standard deviation of half the mean. For values exponentially distributed, the sample size would need to be closer to 400.

Questionnaire, interview or diary methods of data collection are unlikely to yield sufficiently precise data to support the envisaged statistical analysis of frequency distributions and would in any case require an immense amount of labour. There is evidence too that respondents may wittingly or unwittingly contribute inaccurate information (14,62:6,65). In practice, therefore, only data generated as a consistent by-product of the service used will be suitable for investigating quantitative differences among library users. Circulation data provide the most common example. These data can be collected easily and unobtrusively, but they suffer from shortcomings which are considered in the next two sections.

1.3 VALIDITY OF THE DATA

Although complete sets of data may be readily available for some services, it is unlikely that a record of use will be available for every aspect of library service. It cannot be assumed that non-users of the recorded services are also non-users of every other service. Lubans (12), in particular, has drawn attention to this point. He compares structured interview responses from students on the Boulder campus of the University of Colorado who borrowed at least one book for home use from the Norlin Library to responses from students for whom there was no record of borrowing in twelve months. The survey was restricted to users and potential users of humanities and social sciences books. 69 users and 73 non-users were interviewed, but responses for a further 139 non-users were also available from interviews one year before. These latter responses showed differences in reported amounts of library use and library orientation between borrowers and non-borrowers, but little difference was detected in the later study.

Table 1.3 shows the percentages of each of these groups of respondents making each of the four permissible responses to the question, "How often do you use the CU Library?". (It appears from the interviewers' questionnaire reproduced by Lubans in Appendix I of his report, that the term 'use' was not qualified in any way and could presumably have denoted use of the library as a study hall to some non-borrowers.) Clearly the non-borrowers are not always non-users of the library, although the differences in numbers of responses between borrowers and non-borrowers are statistically significant. Even the differences between the first two columns would be exceeded in random sampling from a population making aggregate responses with a probability of only 0.05 (chi-squared test with 3 degrees of freedom).

In response to another question, over three-quarters of the borrowers replied that the assignments they were set required the use of more library material than just the books placed on reserve (not surveyed by Lubans). Curiously, however, almost two-thirds of the 69 non-borrowers and almost one half of the 139 non-borrowers also made this assertion. Whether they used material within the library or spoke without reference to their own practice is not revealed. Clearly, in failing to include reserve collection users among his borrowers, Lubans runs considerable risk of wrongly categorising library users as non-users, but his study

TABLE 1.3

Interview responses from borrowing and non-borrowing students to the question, "How often do you use CU Library?".

	Percentages responding among 73 borrowers	69 non-borrowers	139 non-borrowers
	%	%	%
2+ times a week	58	47.5	0.5
9+ times a semester	23	32.5	0.5
Few times a semester	16	9.5	64.5
Very seldom or never	3	9.5	35.5

Scores are read to the nearest 0.5% from bar charts in Lubans (12) and do not sum to 100%.

serves to cast doubt nonetheless on how good an index of library use circulation records will constitute.

If the record of use includes all tangible issues or circulations, it seems reasonable to assume that differences between users will approximately reflect their differences in library-related activity and will therefore be worth studying. But it is easy to think of two types of use which will still not be recorded and which it might be desirable to include in a tally of the total use made of library material; namely, material reached down from the open shelves and consulted in the library, and material shared among groups of users while issued to only one of them.

In a survey of US university students, Meier (66) finds reported use to be 84% greater than recorded use and attributes the difference to the unrecorded exchange among students of books in high demand. In seeking to quantify the differences between users, however, we are less concerned with obtaining a complete tally of library use than with assessing the representativeness of the available record as an index of differences. If all users reach down books in the library or share books with colleagues in proportion to their recorded use, then the record would still be worth analysing for differences between users.

It would be difficult to measure the actual rates of these unrecorded activities, however, unless large numbers of users agreed to keep accurate diaries. But if unrecorded use predominated for some users, then we might expect to record their use only on those occasions when use was unavoidably recorded or when individuals were obliged to use on their own behalf: borrowing for home use in the vacation, for example, or completing reading assignments from a closed-access reserve collection.

Two sets of data are presented which show recorded issues retained for vacation use compared to all other recorded issues over a period of one academic year (October to June) for groups of second-year undergraduate students using UK university libraries. Table 1.4 shows issues to a systematic sample of 34 arts or social science students. There are no students who record use only for vacations. Nonetheless, the proportions of vacation use compared to other use are very varied. Figure 1.3 shows the data plotted as a scattergram. The numbers are small and the trend is not pronounced, but on the assumption that the trend is linear and that the data are drawn from a single population, the product-moment correlation coefficient was calculated for the relationship between

TABLE 1.4

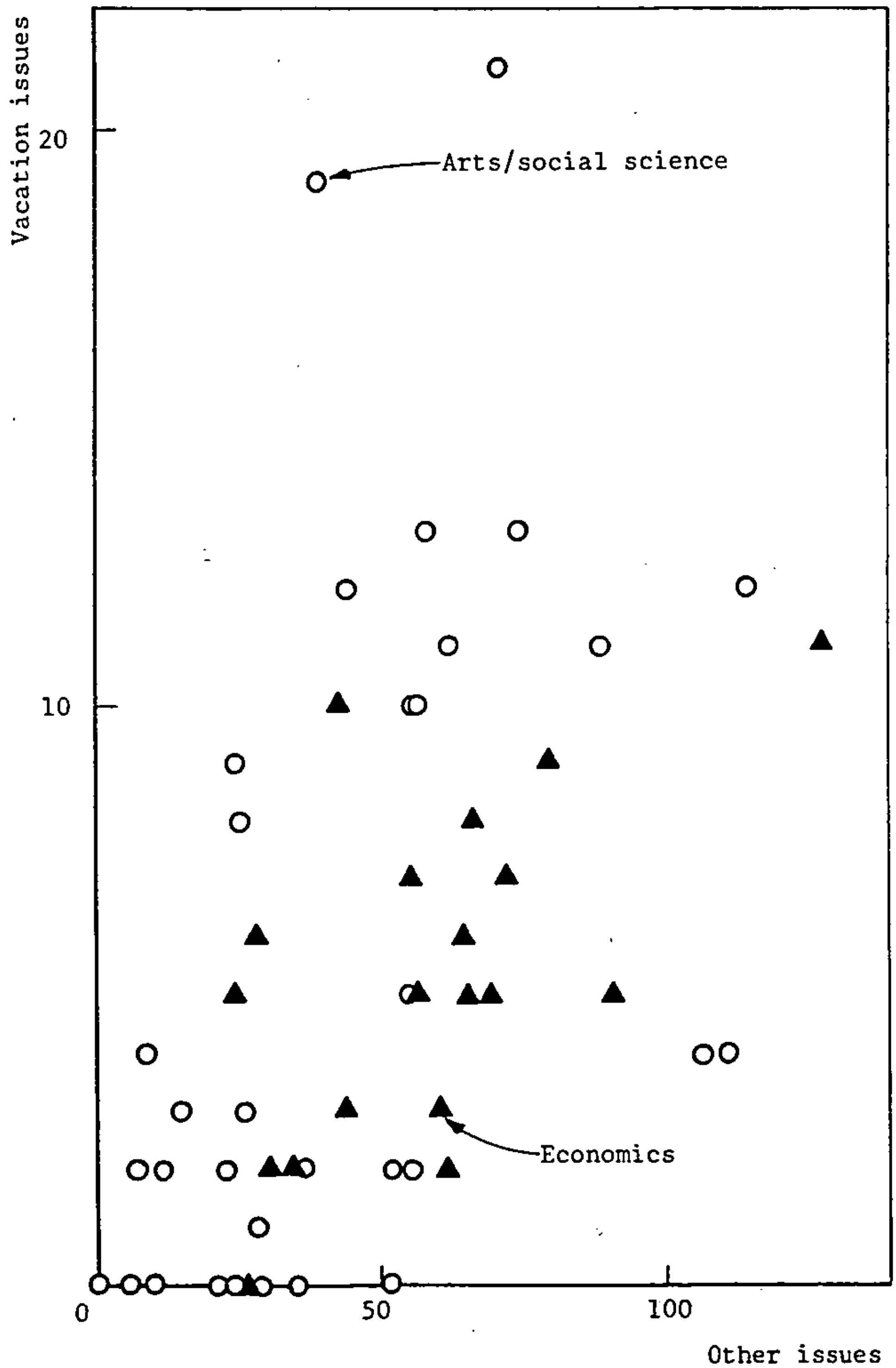
Numbers of books issued before and retained throughout two vacations and numbers of all other issues to a sample of 34 second-year undergraduate arts or social science students during the academic year.

Vacation issues	Other issues	Vacation issues	Other issues
0	1	2	36
0	1	19	39
0	6	12	44
2	7	0	51
4	9	2	52
0	10	5	55
2	12	2	55
3	15	10	55
0	21	10	56
2	23	13	58
0	24	11	62
9	24	21	71
8	25	13	74
3	26	11	88
1	28	4	106
0	29	4	110
0	35	12	114

Total vacation issues = 185; total other issues = 1422.

FIGURE 1.3

Numbers of recorded issues retained for vacation use plotted against numbers of all other issues for samples of 34 second-year arts or social science undergraduates and 20 second-year economics undergraduates using UK university libraries.



numbers of vacation issues and numbers of all other issues. Its value is 0.5, which in random sampling from a population in which there was no correlation between the two types of issues (i.e. $\rho = 0$) would be exceeded with a probability of less than 0.01. Thus although the students were not enrolled in the same course and would have had, therefore, differing requirements for library material, their amounts of vacation and other use appear not unconnected.

The second set of data relates to the systematic sample of 20 second-year economics students previously referred to in Section 1.1. Recorded issues from a short-loan collection during one academic year are shown in Table 1.5. Again, no users appear solely in the vacation issues column. The sample is from an academically homogeneous population and the relationship between numbers of vacation issues and numbers of other issues is roughly linear (Figure 1.3). The product-moment correlation coefficient was in this case 0.65, which value would be exceeded in random sampling from the null-hypothesis population (i.e. $\rho = 0$) with a probability of less than 0.01.

Neither of these sets of data provide evidence for unrecorded non-vacation users. The use of reserve material was also recorded for the first group of students. No student used reserve material without also recording the use of ordinary loan material. Again, therefore, there is no evidence of unrecorded users of borrowed material.

Even if there is no evidence for unrecorded users in these two cases, however, it would be impossible to claim that unrecorded use would be indexed faithfully by recorded use. Almost certainly an otherwise homogeneous student group would contain users who collaborate to differing degrees or who vary in their use of the library as a study hall (1). Away from extremes of behaviour, however, it seems reasonable to expect some correlation between individual amounts of recorded and unrecorded use. This is indeed the conclusion of Mays (14:58) on the basis of survey responses.

Whether or not the position is defensible, there is an overriding justification for analysing recorded use data. Despite its shortcomings, recorded use is widely used as an indicator of library performance. It is the simplest measure of output to establish and is therefore likely to remain important whenever it is necessary to quantify library 'productivity'.

TABLE 1.5

Numbers of books issued before and retained throughout two vacations and numbers of all other issues to a sample of 20 second-year undergraduate economics students using a short-loan collection during the academic year.

Vacation issues	Other issues
0	8
5	24
0	26
6	28
2	30
2	34
10	42
3	43
7	55
5	56
3	60
2	61
6	64
5	65
8	66
5	69
7	72
9	79
5	90
11	127

Total vacation issues = 101; total other issues =1099.

Thus in the US, charged circulation is prescribed by the national standard for library statistics (67) as a measure of the utilization of local resources. (Academic libraries can subdivide this into general and reserve circulation, if appropriate.) It is true that other measures are obtained by sampling: during a sample week, numbers of visitors and borrowers, and numbers of uncharged uses are also counted. By this means, the circulation total can be put into context. But clearly, ratios of output to input will depend to a large extent on the circulation total, both as a measure of output in its own right and as an index for the other, sampled, activities.

In the UK, a pilot study of the calculation of business ratios for academic libraries by the Centre for Interfirm Comparison (68) and the recent creation of a database of academic library statistics (69) have been partly stimulated by the belief that the comparison of such ratios might reveal differences in the levels of efficiency attained by academic libraries (70). The Centre for Interfirm Comparison collected data for resources, activity and costs relating to the year 1980 from 12 university libraries and 8 polytechnic libraries. Wide differences were apparent within each group. Table 1.6 shows the recorded use figures per unit of population. Clearly it would be unwise to interpret such data without local knowledge. The statistics now collected annually by the Standing Conference of National and University Libraries (69) include not only circulation data but also data from a sample-day survey of unrecorded use (71). But again, it seems clear that the sampled data will largely be utilized in conjunction with the circulation data.

Data relating to recorded use is freely employed, therefore, in quantifying the output of academic libraries and the performance of their collections. It appears to be assumed that it is a valid indicator for this purpose. The relative extents to which potential library users participate in generating recorded library use is unlikely to be included among commonly-collected library-use statistics because they are a difficult set of quantities to express in a single measure. Nonetheless, an investigation of frequency distributions of recorded use by users does seem worthwhile in view of the general use made of recorded use data.

TABLE 1.6

Loans per member of the potential user population (full-time equivalent academic staff, researchers and students) and short-term loans per student for 12 universities and 8 polytechnics in the UK for the academic year centred on 1980.

Universities		Polytechnics	
Loans	Short loans	Loans	Short loans
17	5	22	3
20	-	26	9
24	1	29	6
29	20	30	10
35	15	32	9
38	16	40	3
40	12	42	15
47	15	55	12
52	34		
60	13		
82	32		
82	59		

Source: Centre for Interfirm Comparison (68)

1.4 UNITS

Counts of recorded uses can be made easily and unobtrusively. The unit counted, however, may not be uniform. Recorded circulations, for example, may relate to a whole range of activities, from deep and complete study of an item to no consultation at all. This is a severe limitation in the data and one which has not been successfully overcome.

Meier (66) and Hamburg (72) have proposed more discriminating measures of document use than recorded uses. Both authors were concerned with developing performance measures for libraries. They thus required a measure of output to set against measured inputs. Meier proposed a unit of 'item-use days' to be applied to all uses of library materials. The numbers of units produced in unrecorded activities and the weightings applied to recorded activities were derived from the results of sample surveys. Table 1.7 shows the weightings proposed by Meier for converting recorded or observed use to item-use days. He acknowledges that the surveys would need to be regularly repeated. Hamburg proposed an hour of 'document exposure' as the unit of measurement. Again, the total number of units scored for each library was derived from a mixture of existing information and sample surveys.

Although both units are easy to define and apply, and permit comparisons between libraries, they are designed for use with aggregate measures which average out individual diversities in library use. In constructing an aggregate measure, averaged weightings or estimated averages are quite adequate to the purpose and simple to obtain. It should be possible, for example, to obtain estimates to within ±10% of a mean from sample sizes in hundreds and the cumulative error in an aggregate of these estimates need not be greater. (Nonetheless, both authors are concerned with advocating and employing the measures rather than testing them, and pay little attention to the problems of sampling from large and very diverse populations.)

In comparing individuals in a population, however, an averaged weighting which converts one unit into another advances our understanding very little, although it may allow us to reflect differences rather more sensitively. There is no guarantee, however, that additional spurious differences are not also introduced.

Figure 1.4 shows the result of weighting the use recorded by the sample of 20 undergraduate economists. Books were issued either for one day

TABLE 1.7

Weightings proposed by Meier for converting recorded library use into item-use days.

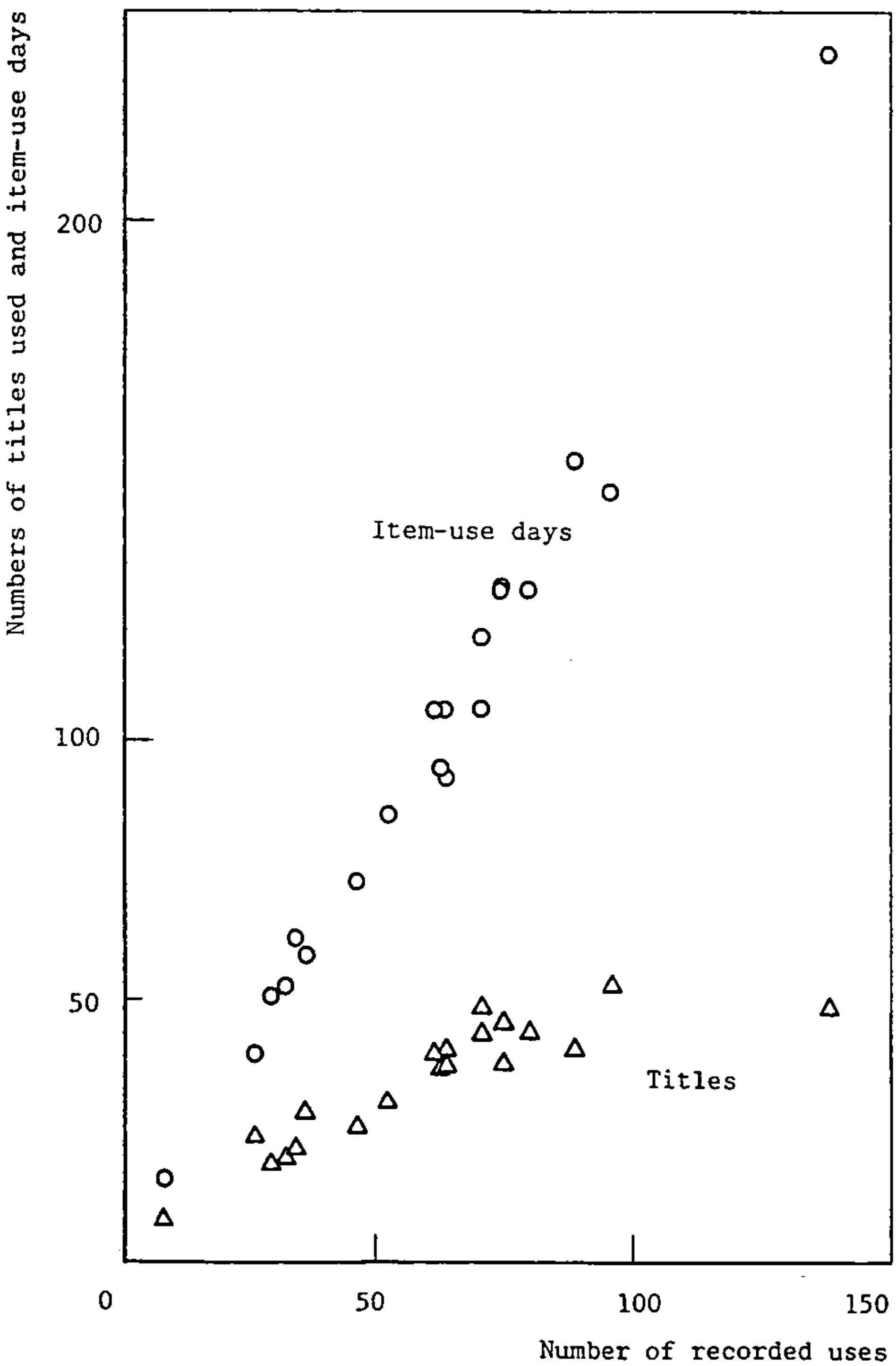
Recorded activity	Weight
Circulation of book	
with two-week loan period or longer	3
with one-week loan period	2
for overnight loan	1
from closed reserve	3
Book reshelved after library use	2
Use of reference book	1
Book found out of place*	1

*Presumed misplaced by user after use.

Source: Meier (66)

FIGURE 1.4

Numbers of recorded uses weighted as item-use days and numbers of titles used plotted against numbers of recorded issues for a sample of 20 second-year economics undergraduates using a UK university library.



loan or one week loan. In the manner of Meier, one day loan books were counted as one use and one week loan books as two uses. The plot remains linear throughout its range. In this case, therefore, the proportions of each kind of book used were similar for all users. Conversion of the data would alter statistics relating to the scale of the frequency distribution of use but not those relating to the shape of the distribution.

The work of Knapp (3) contains an interesting approach to the measurement of use. As well as counting numbers of circulations recorded by college students, she also counts numbers of titles used. For academic libraries this suggests a unit which will suppress some of the effects of the differences in the way individuals choose to use similar library materials. Whether a user records several uses in reading a particular piece, or consumes it at one go, the result will be one title use. The data then record only the differences among users with respect to numbers of titles used. Differences in the amount of a particular title used and in the nature of that use will not, of course, be discriminated.

In general, the effect of recording titles used is to decrease the range of amounts of recorded use as well as reducing absolute numbers of uses (see below, Section 9.2). Figure 1.4 shows the effect for the 20 undergraduate economists. The student who recorded most issues and most weighted usage now shares joint second place.

For academic libraries, there seems some justification in analysing the use of course-related material in terms of titles used. Especially in the case of recommended or required reading, the range of titles used and the distribution of this use over the users are perhaps of more importance both to the teachers and to those managing the provision of library resources than absolute numbers of issues to each user. In these cases at least, counts of recorded uses are likely to be reliable, particularly when the material is issued from a closed collection. In other cases, however, (where different groups of library users or different types of library material are represented within the population or collection, for example) it seems wise to assume that care and local knowledge will be needed to prevent the analysis of recorded use degenerating into a meaningless numerology aggregating incompatible units of measurement.

1.5 AIMS OF THE STUDY

Despite their shortcomings, it seems likely that recorded-use data will continue to be used to quantify the output of libraries. An investigation into the distributions of recorded use over potential user populations in academic libraries was therefore considered worthwhile. It appeared well-known that potential users rarely participated equally in generating recorded use, yet the form of the frequency distributions of use among those users appeared to have been little considered.

The aims of the study were therefore set down as follows:

i) To investigate, using available data, the distribution of recorded uses over potential users in academic libraries.

ii) To consider the nature of factors, apart from the propensity of the users for library use, which may cause the distribution to arise.

iii) To examine the effect of the distribution of use among users upon the uptake of material from the library collection.

1.6 LIMITATIONS IN THE MODEL

The limitations inherent in the available data make necessary, as we have seen, assumptions about the capacity of the data to reflect such differences in library use as exist among users. In describing and comparing frequency distributions of recorded use, the information available for fitting and employing the statistical model is similarly limited by the nature of the data.

Many variables could be proposed as factors influencing the recorded library use of individual users. For example, to quantify the propensity of users to use the library and their success in using it, such variables as: rates of visiting; rates of searching for individual items; success rates per search; recorded use rates per success; and so on, could be included in a description of the process of library use. Users' reactions to success or failure, and to competition and the various restrictions and regulations imposed upon them would also need to be taken into account. A range of individual, group and institutional factors could thus be relevant to the analysis.

The data available support no quantifications of these sorts of variables, however. Only the simplest descriptive model of the outcome of the process is therefore appropriate. It needs to be capable of furnishing a probability distribution (preferably a discrete probability distribution) which recreates an observed frequency distribution for numbers of uses to within the limits of statistical expectation (assuming random sampling from a process with fixed parameters). The model will thus need to be based on as few parameters as possible, since the data will yield too few statistics to sustain stable estimates of large numbers of parameters in a more complex model. The model implies no understanding of the process producing the outcome (indeed, the outcome may be described by more than one model) and it may be that, without this understanding, no model can be found which holds exactly. (Nonetheless, local knowledge about the library collection and its users may allow the data to be fitted with more insight, even if alternative models cannot be distinguished on statistical evidence.)

Only the variable represented in the data is described or predicted by the model and the predictive success of the model is related, of course, to the capacity of the data (73). Improving the fit of a model to a set of data by increasing the number of fitted parameters necessarily decreases its generality and its predictive power unless additional data are available from which to test for and reduce the effects of chance (or error) in the original data. A model that serves for a wide range of sets of data gains in usefulness, however, and a probability distribution which models the outcome of the process of use for various groups of users may allow the range of outcomes and their change over time to be predicted. Even then, little may be disclosed about the factors which determine the outcome.

1.7 TESTING THE MODEL

To assess the suitability of a model of user activity, it is necessary to test the ability of the model to furnish a probability distribution similar to the observed distribution of relative frequencies of use. To investigate models which do furnish such a distribution, the known characteristics of the actual library users and their patterns of library use could also then be compared to the characteristics which the model assigns to the users and the method by which it generates the distribution of probabilities.

The goodness of fit of the expected frequency distribution to the observed distribution of use is tested after the limits are set within which fits are to be accepted. Within these limits, differences between the two frequency distributions are explained as chance variations such as would occur in random sampling from a known population. In some cases it was to prove sufficient to compare only a few terms of the distributions, or even some simple statistics, in order to reject a fit. But elsewhere, the chi-squared test was used.

1.7.1 Test of goodness of fit

In testing the goodness of fit of the expected frequency distributions, the conventional chi-squared test was used throughout. The result of this test provides a measure of the probability that an observed frequency distribution could have been generated by the random sampling of a given number of observations from a population distributed according to the hypothesis being tested. The frequencies (in this case, numbers of users) of each outcome or class (in this case, numbers of uses) would be expected, under random sampling, to vary within known statistical limits. For large samples, the sampling distribution of any class frequency becomes approximately normal and the sampling distribution of the sum of the differences (see below) between the actual class frequencies and the expected class frequencies under random sampling is approximately that of the chi-squared probability distribution. The sum of the differences between an observed and a hypothetical distribution (the chi-squared statistic) is therefore compared to the chi-squared probability distribution in order to assess goodness of fit. Reference to a table showing probabilities of the values of the chi-squared statistic for the given number of observations (less the number of constraints introduced in estimating the parameters of the hypothetical distribution) yields the probability, P, of that value of the sum of differences being exceeded in random sampling from the hypothetical distribution.

If the probability, P, is small (that is, if the sum and therefore the value of the statistic are large), then the hypothesis (usually the null hypothesis) is open to rejection. If not, the hypothesis, although in no way being proved correct, at least is not shown to be incorrect. Where a null hypothesis is being tested, the values of P taken to indicate rejection are conventionally 0.05 or 0.01; the so-called 5% or 1% levels of significance. For testing the fits of expected frequency distributions of recorded library use, a 20% significance level was adopted, as explained below in Section 5.3, because a failure to reject the hypothesis would support the candidature of the model being tested. This result was to be accepted with more caution than in the case of a null hypothesis, therefore.

The sum of the differences between the observed and expected frequencies, o_i and e_i, for all i classes or outcomes is calculated as $\Sigma[(o_i - e_i)^2/e_i]$. It can be shown that this variate conforms approximately to the chi-squared distribution if the null hypothesis is true. The approximation does not hold, however, for small class frequencies, because the binomial sampling distribution is not then approximately normal. It is usual, therefore, to pool classes with small frequencies to form classes with frequencies of at least five.

Wall (20) used the Kolmogorov-Smirnov test to compare hypothetical distributions with the observed distributions of recorded use for the population of undergraduates described in Section 3.1. Although this test is designed to compare continuous rather than discrete distributions, it was adopted in the belief that it would give a better test of cumulating discrepancies of the same sign. Although some runs of such discrepancies were apparent, the results of the Kolmogorov-Smirnov tests differed little from those of chi-squared tests on the same data. The Kolmogorov-Smirnov test was not used, therefore, in this study.

1.8 SUMMARY

Most librarians are interested in the proportion of potential users who use their services and many have conducted and reported surveys. Less attention has been paid to quantifying the differences among users with respect to amounts of use, although there appears to be evidence that differences are real and characteristic of the users. Potential user populations are most readily enumerated for academic libraries and students in the same class will provide good subjects for study since they will have similar tasks to perform and, in respect to course-related use, therefore, will constitute an homogeneous population.

To quantify and compare frequency distributions of use accurately, a great deal of exact data are required. Periods of data collection must extend over some weeks or months. Diary, questionnaire or interview methods of data collection are cumbersome and inaccurate for producing large amounts of data relating to long periods of time, and data reported or recorded as a by-product of library housekeeping therefore need to be employed. A number of factors make these data unsatisfactory for consistent analysis, however. Recorded use will relate indiscriminately to a whole range of types and lengths of use and many valid uses may not be recorded. The interpretation of the data requires considerable care, therefore, although the counting of numbers of titles used rather than uses could avoid some ambiguities in the data.

In practice, many factors will determine the amount of use made of the library by individual users. Similar library activities are performed by users with very different academic attainments and critical abilities, yet their recorded use will not be quantitatively distinguishable. The model used to describe frequency distributions of amounts of library use among users will relate to the statistical outcome of the process of library use rather than the process itself, therefore, and will be based on as few parameters as possible.

CHAPTER 2

PREVIOUS WORK

Some previous work by Wall (20) forms the starting point of the present study. Wall collected data for a group of about 1550 undergraduate students using a short-loan textbook collection in a UK university library. Negative binomial probability distributions were fitted to frequency distributions of recorded use constructed from these data. The fits were generally good (see Figure 2.1), but the fitted parameters did not vary predictably over time.

2.1 NEGATIVE BINOMIAL DISTRIBUTION

The negative binomial distribution was used by Greenwood and Yule (74) in 1920 to describe the frequency distribution of accidents among industrial workers. It has found a number of subsequent applications. The distribution is derived by Greenwood and Yule as a mixture of Poisson distributions, the means of which are continuously distributed according to a gamma distribution governed by two parameters. Figure 2.2 shows this derivation. For applications to recorded library use by potential users, the assumptions therefore are that:

i) Potential users have constant mean individual rates of library use. These can be expressed as a continuous probability distribution of expected rates for any time period. Only rates greater than zero are possible; there are no potential users with zero expectations of use.

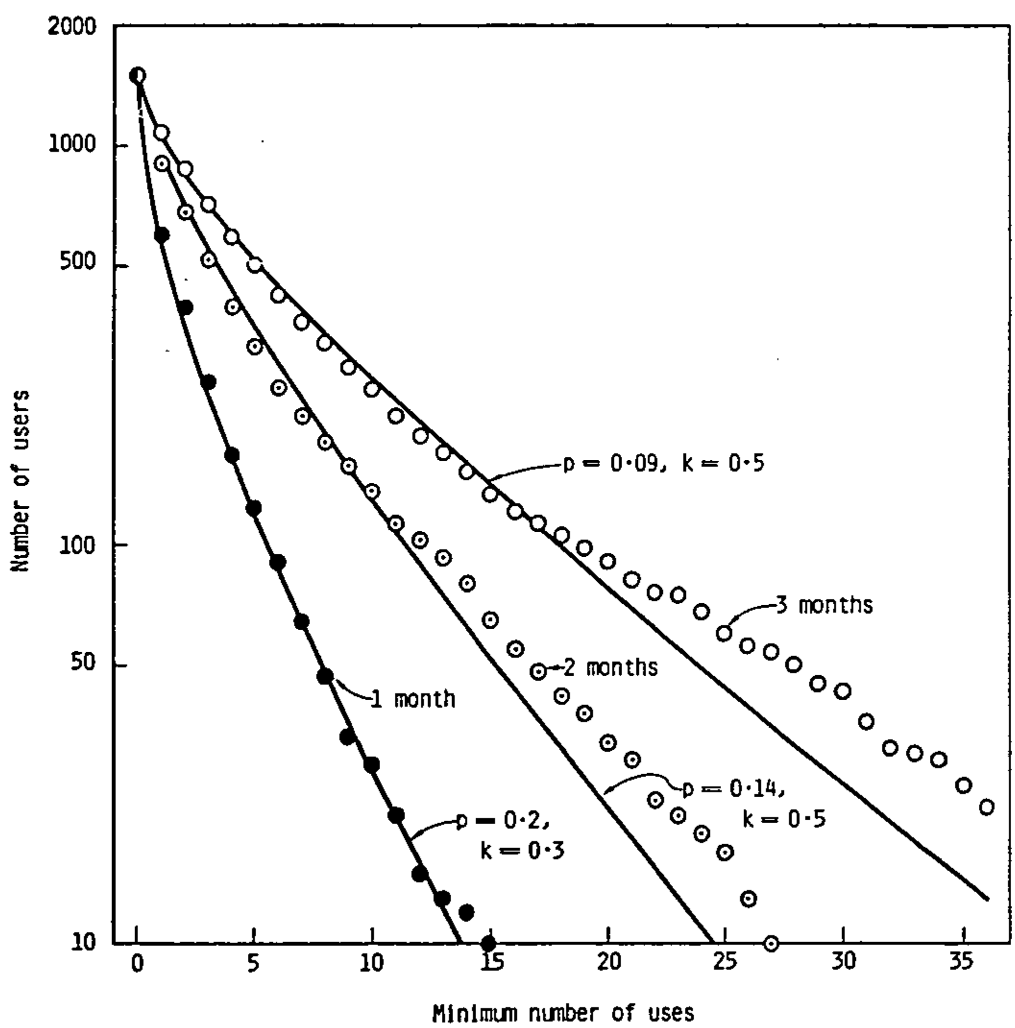
ii) Actual amounts of use observed for each individual user in time periods of similar length have a Poisson distribution with constant mean.

The assumption that observed numbers of uses are Poisson distributed about a constant mean can be tested for given users and time periods. The form of the distribution of means over the population is fitted from the data and cannot be directly tested.

The nature of the negative binomial distribution is discussed by Williamson and Bretherton (37). There are two parameters. The shape parameter, k, can be estimated from the sample mean and the proportion of zeros in the observed distribution (see Figure 2.3) or from the maximum

FIGURE 2.1

Numbers of users plotted against their minimum number of recorded uses for 1550 science or engineering taught-course students using a short-loan textbook collection for three periods of observation. Negative binomial distributions representing best fit from tabulated examples (37) are shown for clarity as continuous lines with parameter values indicated. Logarithmic scale on the ordinate.



Source: redrawn from Wall (42)

FIGURE 2.2

Negative binomial probability distribution

A continuous variable, m, is a gamma variate with parameters b and k. It represents mean numbers of events per time period and is distributed with probability density:

$$P(m) = \frac{b^k}{\Gamma(k)} m^{k-1} \varepsilon^{-bm}, \quad 0 \leqslant m \leq \infty$$

where $e = 2.718$, k is a shape parameter, b is a scale parameter and the gamma function $\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} \, dx$.

Observed integer numbers of events in any time period, given a mean of m events per time period, are distributed about this mean with Poisson probabilities. Thus the probability of s events, given m, is:

$$P(s \mid m) = \frac{\varepsilon^{-m} m^s}{s!}, \quad s = 0, 1, 2, \ldots$$

For all m, the proportion of observations in which s events are expected, $p(s)$, is then:

$$p(s) = \int_0^\infty \frac{b^k}{\Gamma(k)} m^{k-1} \varepsilon^{-bm} \frac{\varepsilon^{-m} m^s}{s!} \, dm$$

$$= \frac{b^k}{\Gamma(k) s!} \int_0^\infty m^{k+s-1} \varepsilon^{-m(b+1)} \, dm$$

Since $\int_0^\infty x^{n-1} e^{-ax} \, dx = \Gamma(n)/a^n$,

$$p(s) = b^k \frac{\Gamma(k+s)}{\Gamma(k) s!} \frac{1}{(b+1)^{k+s}}.$$

FIGURE 2.3

Proportion of zeros and mean of negative binomial distribution

---

From Figure 2.2, the proportion of zeros, $p(0)$, is:

$$p(0) = \left(\frac{b}{b+1}\right)^k .$$

For the mean of m, $E[M]$, we can write:

$$E[M] = \int_0^\infty \frac{b^k m^{k-1} \varepsilon^{-bm} m}{\Gamma(k)} dm$$

$$= \int_0^\infty \frac{b^k m^k \varepsilon^{-bm}}{\Gamma(k)} dm$$

$$= \frac{b^k \Gamma(k+1)}{b^{(k+1)} \Gamma(k)} = \frac{k}{b}$$

Similarly, the variance of m, $var[M] = E[M^2] - (E[M])^2$, can be written:

$$s^2 = \int_0^\infty \frac{b^k m^{k-1} \varepsilon^{-bm} m^2}{\Gamma(k)} dm - \left(\frac{k}{b}\right)^2$$

$$= \frac{k}{b^2} .$$

The variance of the Poisson distribution equals its mean, $E[M] = k/b$, so that the total variance of m in the negative binomial distribution is:

$$\frac{k}{b^2} + \frac{k}{b} .$$

likelihood equation (see Figure 2.4). The scale parameter, b, is then estimated from the values of the sample mean and the shape parameter, k.

Other notations may be adopted for convenience. Figure 2.5 shows that of Williamson and Bretherton, which is used from now on. The mean of the distribution is kq/p and the variance kq/p^2. The number of users recording s uses, $f(s)$, out of a total of N potential users is:

$$f(s) = N \binom{k + s - 1}{k - 1} p^k q^s, \quad s = 0,1,2,\ldots$$

Each sth term is more easily evaluated as the $(s - 1)$th term multiplied by $q(k + s - 1)/s$. The starter value for this recurrence relationship, the zero term, is Np^k.

2.2 APPLICATION OF THE NEGATIVE BINOMIAL DISTRIBUTION

The discrete negative binomial probability distribution, derived as shown above, can be thought of as a more general form of the Poisson distribution. Events are not restricted to occurrence at a single mean rate, as with the Poisson distribution, but rather occur at different mean rates which vary over the population with considerable freedom. If members of a human population are not homogeneously disposed to the event observed, or if the occurrence of one event predisposes individuals to further events, then the variance of the data exceeds the mean and a fit of the Poisson distribution becomes unlikely. Many such situations appear to be well modelled by the negative binomial distribution, for example, where mean expectations vary over the population with respect to accidents, arrivals, absences, errors, sickness, and so on. Three examples are presented below.

Descriptions of real patterns of human behaviour are unlikely, however, to suggest models based on random and independent processes governed by fixed parameters. The events observed are likely to involve and be conditional on many unknown factors, not least past behaviour. Nevertheless, as a descriptive model, a mixture of Poisson distributions serves as a convenient first approximation, condensing the effect of the large number of factors in a complex observed activity to the simplicity of random variation governed by a single parameter. The use of the negative binomial distribution thus implies the assumption that numbers of events are distributed with Poisson probabilities about means which

FIGURE 2.4

Maximum likelihood equation for estimating parameter k of the negative binomial probability distribution

---

Let $f(r)$ be the observed frequency of r events out of a total number of N observations, let h be the highest value of r observed and let m be the mean value of r.

The solution of the equation

$$N \log\left(1 + \frac{m}{K}\right) - \frac{f(1) + f(2) + \ldots + f(h)}{k} - \frac{f(2) + f(3) + \ldots + f(h)}{(k+1)} - \ldots - \frac{f(h)}{k+h-1} = 0$$

gives the maximum likelihood estimate of k.
[From Williamson and Bretherton (37:12-13)]

FIGURE 2.5

Notation of Williamson and Bretherton

Let $p = b/(b + 1)$ and for $\Gamma(k)$ write $(k - 1)!$. Further, let $q = (1 - p) = 1/(b + 1)$. Then $b = p/q$, the proportion of zeros $p(0) = p^k$ and, in general, the probability, $p(s)$, of observing s events is:

$$p(s) = p^k \binom{k+s-1}{k-1} q^s.$$

We verify that the sum of the probabilities of all numbers of events, $s = 0,1,2,\ldots$, is unity and that the mean or expected value of s, $E[S]$, is kq/p.

We have

$$\sum_{s=0}^{\infty} p(s) = p^k + p^k kq + p^k \frac{k(k+1)q^2}{2!} + \ldots$$

$$= p^k \left[1 + kq + \frac{k(k+1)q^2}{2!} + \ldots \right.$$

$$= p^k (1-q)^{-k} = 1,$$

using the binomial expansion, $(1 - x)^{-n} = 1 + nx + n(n + 1)x^2/2! + \ldots$

Similarly,

$$E[s] = 0 \times p^k + 1 \times p^k kq + 2 \times p^k \frac{k(k+1)q^2}{2!} + \ldots$$

$$= p^k kq \left[1 + (k+1)q + \frac{(k+1)(k+2)q^2}{2!} + \ldots \right.$$

$$= p^k kq (1-q)^{-(k+1)} = kq/p.$$

are gamma distributed. This gamma distribution of means is then the result of the many causes which differentially dispose individuals in the population to record events. If this form of distribution is suggested by the data then it may be of little importance whether the generating process in the model corresponds to what is known of reality. If the sole requirement is simple description or prediction of the gross aspects of a situation. then a simple and rather superficial model may be adequate. Of course, an empirical test of how well the Poisson assumption is reflected in the data provides one line for investigating the patterns which underlie the observed frequency distribution of events.

The gamma distribution of means allows a wide range of unimodal observed distributions to be accommodated. When the negative binomial parameter, k, is small, especially when less than unity, a reversed J-shaped distribution results, with frequencies decreasing monotonically after a mode at zero. When $k = 1$, a geometric distribution is generated. As k increases, the mode moves away from zero and the distribution forms a positively-skewed hump. Clearly, the frequencies of the zero and first terms are equal when $kq = 1$.

In the derivation given above, k tends to be independent of time when the heterogeneity in a given population is constant. When k remains constant, b (Figure 2.2) or p (Figure 2.5) are inversely proportional to the length of the period of observation (or the mean) and expressions for the probability of use in time periods of differing length can be derived.

2.2.1 Purchasing behaviour

Ehrenberg (39), Chatfield and Goodhardt fitted negative binomial distributions to observed frequency distributions of numbers of 'purchase occasions' reported by members of consumer panels maintained by market research organisations. Their purpose was to describe and predict consumption and brand loyalty among the buyers of regularly-purchased branded consumer goods such as breakfast cereals. Time periods were chosen for which there was little change in aggregate sales rate or market shares and in which purchasing behaviour was largely independent of behaviour in a previous time period. Despite the diversity of buyers, brands and outlets, a simple short-term pattern in aggregate purchasing emerged for various brands. Table 2.1 shows the fit to some data for

TABLE 2.1

Observed numbers of purchases of a household product by 2000 households in 26 weeks and expected frequencies for the negative binomial distribution fitted by Ehrenberg (39).

	Numbers of households	
Number of purchases	Observed	Expected
0	1612	1612
1	164	157
2	71	74
3	47	44
4	28	29
5	17	20
6	12	15
7	12	11
8	5	8
9	7	6
10	6	5
11-15	11	12
16+	8	7

TABLE 2.1 (continued)

Statistics of the frequency distribution of household purchases and fitted parameters ($\underline{k}$ and p) and chi-squared test statistics for the negative binomial distribution.

	Observed	Expected
Total households	2000	
Mean	0.64	
Variance	2.12	
$\underline{k}$		0.115
p		0.153
Chi-squared		3.53
Number of cells		13
$\underline{P}$		>0.95

$\underline{P}$ is the approximate probability of the observed chi-squared value being exceeded in random sampling.

incidences of purchase (that is, purchase occasions, rather than amount purchased) from one panel over a period of 26 weeks.

When data were collected from similar panels for time periods of differing length, numbers of buyers were found to increase predictably over time. The parameter *k* remained roughly constant. Where sales rate and market share were stable, the assumption that observed numbers of purchase occasions were distributed with Poisson probabilities about the mean could be used to estimate that proportion of users in one time period who would not be observed buying in another time period of similar length. Because an equal number of users are expected to be observed in each time period, those not observed in one are replaced by an equal number not observed in the other. Subtracting these numbers from the total expected to be observed in either time period yields an estimate for the number of repeat buyers who are expected to buy in both periods. Summing the expected frequency of purchase (gamma-distributed) over all single period buyers gives an estimate of the contribution of these buyers to aggregate sales and, by subtraction, an estimate of the contribution of repeat buyers.

Thus Ehrenberg is able to use the model as a base for prediction over time; for comparing different sets of data (e.g. for different brands); and for quantifying in a limited way the patterns of behaviour which generate the data. It is true, however, that stationary conditions may not last long; that independence may not apply for short time periods; that many types of purchasing may not be amenable to modelling in this way; that discrepencies in fit may occur even for those that do; that total sales must be derived by using an averaged multiplier to convert incidences of purchase to amount purchased; that data are subject to reporting and sampling errors; and that the homogeneity of the population is not established.

To avoid some of these difficulties whilst retaining the benefits of the model, subsequent writers have proposed alternative components, varying either the gamma distribution of long run means or the Poisson distribution of observed events per time period. Thus Sichel (75), for example, generalises the gamma distribution by the addition of a third parameter and improves the fit to several sets of data.

2.2.3 Surgery consultations

Froggatt, Dudgeon and Merrett (76) present data on surgery consultations by 2810 female patients registered in a group medical practice. Table 2.2 shows the frequency distribution of numbers of consultations in a single year. A negative binomial distribution fits the data well.

Three other theoretical distributions were considered by Froggatt but were rejected on grounds of fit. Each distribution represented a hypothesis relating to the generation of the observed frequency distribution. The negative binomial distribution represented the hypothesis that the population was homogeneous but differentially prone to consult their doctor, this proneness being unchanging. Changes in the environment were assumed to affect the whole population equally, therefore. This hypothesis was then further tested by considering data for succeeding years. Correlation coefficients for numbers of consultations by individuals in each pair of three years were positive and thought significant, but they varied excessively and fell below the level expected if proneness was to account for all the variation in the data. The linear regression of consultations in the second of a pair of years was well predicted and negative binomial fits to two-years data were good, but fits to single year distributions using parameters derived from two-year parameters were less so.

Symmetrical bivariate negative binomial distributions did not therefore model the data sufficiently well to suggest that constant proneness entirely explained the variation among the women. Other influences were thought to confound the operation of simple proneness, albeit to a modest extent: distributions similar to those observed could have arisen if sections of the population were unequally exposed to risk; were recorded unevenly or prompted to consult unequally; if the population was liable to attend only in random spells; or if consultation or a threshold number of consultations, altered the liability for subsequent consultation. The constancy vouchsafed in the parameters of the distributions was, it seemed, the net effect of a large number of what could be continually changing individual factors, both personal and environmental.

Nonetheless, the predictive capacity of the model based on limited data was equal to a regression based on extra data.

TABLE 2.2

Observed numbers of surgery consultations in one year by 2810 female patients and expected frequencies for the negative binomial distribution.

Number of consultations	Numbers of women Observed	Expected
0	820	819.0
1	535	533.8
2	369	378.0
3	283	274.8
4	201	202.3
5	149	150.1
6	106	112.0
7	76	83.8
8	77	62.9
9	54	47.3
10	32	35.6
11	31	26.9
12	27	20.3
13	14	15.3
14	3	11.6
15	6	8.8
16	8	6.7
17	3	5.0
18	2	3.8
19+	14	12.0
Total	2810	
Mean	2.77	
Variance	11.8	

Source: Froggatt (76)

TABLE 2.2 (continued)

Statistics of the frequency distribution of surgery consultations and fitted parameters (k and p) and chi-squared test statistics for the negative binomial distribution.

	Observed	Expected
Total women	2810	
Mean	2.77	
Variance	11.8	
k		0.8525
p		0.2355
Chi-squared		19.0
No.of cells		19
P		0.25

Chi-squared test: expected frequencies were pooled to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

2.2.3 Library book use

Wall (20) fitted negative binomial distributions to some frequency distributions of the recorded use of short-loan textbooks by students enrolled on taught-courses in a UK university. The number of potential users (mainly scientists and engineers) was initially estimated at 1500 but is now taken to be 1550 after further inspection of class lists and course calendars. About 50 separate classes (course/years) were represented. Most of these were wholly served by the short-loan collection. Other classes were excluded from consideration, except for a few classes, a proportion of whose members could be expected to use the collection as a result of choosing particular options in their courses.

The use of any book from the collection (either within the library or taken away on loan) resulted in a transaction record. These records were saved after cancellation for a period of almost ten weeks in the first term of an academic year. Quick reference consultations and browsing of the material in the collection were not recorded, but constituted only a minimal part of the total use made of the collection. Thus, although the unit was not standardised, the record was thought to be complete.

Frequency distributions of numbers of recorded uses were constructed for the first three and six weeks of the term as well as for the full ten week period of data collection. Table 2.3 shows a summary of the distributions. The parameter k of these distributions was estimated using the maximum-likelihood equation shown in Figure 2.4. Values of the chi-squared statistic were calculated from the full data and indicate that the observed frequency distributions were adequately fitted by the negative binomial distributions shown.

The fitted parameters k appear to change progressively as the time period of observation lengthens. Prediction based on parameters fitted to the data would be difficult, therefore, because neither of the parameters would be either constant or indexed by the mean. One solution to this problem would be to assume that, at the beginning of the academic session, not all of the 1550 possible users were yet potential users, but that users were gradually recruited to the potential user population as they received assignments or reading recommendations. As the potential user population was progressively augmented, k would approach an upper limit representing its value for a steady state after full recruitment.

Table 2.3

Observed numbers of recorded uses from a textbook collection in cumulating periods of three, six and ten weeks by 1550 potential library users and expected frequencies for the fitted negative binomial distribution.

Number of uses	Observed and expected numbers of users					
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	955	950.4	648	650.9	470	479.5
1	204	230.2	221	234.7	204	210.2
2	137	120.2	164	145.5	165	142.0
3	86	74.0	123	102.8	118	107.2
4	44	49.0	81	77.1	89	85.2
5	34	33.8	66	59.8	83	69.7
6	26	23.9	37	47.4	61	58.1
7	17	17.3	29	38.2	40	49.1
8	14	12.7	23	31.1	42	42.0
9	5	9.4	22	25.6	34	36.1
10	7	7.0	22	21.2	35	31.3
11	6	5.2	11	17.6	21	27.2
12	2	4.0	10	14.7	18	23.8
13	1	3.0	13	12.4	17	20.8
14	2	2.3	15	10.4	19	18.3
15	2	1.8	10	8.8	13	16.1
16	2	1.3	7	7.5	8	14.3
17	1	1.0	6	6.3	7	12.6
18	1	0.8	4	5.4	8	11.2
19	1	0.6	6	4.6	7	9.9
20	1	0.5	3	3.9	9	8.8
21	0	0.4	6	3.3	6	7.9
22	2	0.3	2	2.9	1	7.0
23	0	0.2	2	2.5	7	6.3
24	0	0.2	2	2.1	8	5.6
25+	0	0.5	17	13.1	60	49.8

TABLE 2.3 (continued)

Statistics of the frequency distribution of textbook uses and fitted parameters (k and p) and chi-squared test statistics for the negative binomial distribution.

	Three weeks	Six weeks	Ten weeks
Total users	1550	1550	1550
Mean use	1.223	2.993	5.048
Variance	6.076	27.40	69.38
k	0.302	0.41	0.48
p	0.198	0.1205	0.0868
Chi-squared	13.17	25.86	40.37
No.of cells	15	25	33
P	0.35	0.25	0.1

Chi-squared test: expected frequencies were pooled to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

Evidence that k would stabilize was available from data collected for a sample of 309 users over almost two terms. The sampling method is described later. Table 2.4 shows cumulating use by this sample over a period of sixteen weeks and fitted values of k. By holding k constant at its estimated steady state value (say, k = 0.54), estimates of the size of the potential user population can be made which result in the best fit of negative binomial distributions to the data for each time period. The final column in Table 2.4 shows these estimates. Fits were as good as those obtained when k was allowed to vary.

Unfortunately, k was less than 0.54 for each of the sixteen single-week frequency distributions represented in the sample data. With two exceptions, values were between 0.2 and 0.3. From any starting point (not just the beginning of session) k progressively changed. It seemed, therefore, that the distribution for any period of observation less than about one term in length would show either k less than its steady value or the potential user population less than expected.

2.3 FURTHER WORK

The work described above suggested a further line of investigation to set alongside the general aims set down in Section 1.5:

iv) To find a method of predicting frequency distributions for time periods exceeding periods of observation.

The results of the investigations are reported below. First, the data for the use of the short-loan collection are examined in order to compare observed patterns of use with the assumptions inherent in the negative binomial model. The effects of other factors, such as competition among users, are also assessed. Second, other sets of data from the literature are tested for the fit of negative binomial distributions. A novel, modified model is then suggested and fitted to all sets of data. Third, using this modified model, some extrapolations are suggested, yielding information unavailable from the original data.

2.4 SUMMARY

Examples have been quoted in which negative binomial distributions have been used to model frequency distributions resulting (in part at least) from the uneven disposition of the members of a population to particular

TABLE 2.4

Numbers of observed users, numbers of recorded uses, estimated negative binomial parameter, *k*, and estimated potential user population size for *k* = 0.54 for a sample of 309 potential short-loan collection users over periods from one to 16 weeks.

Weeks	Users	Uses	*k*	Est. size
1	63	122	0.23	175
1-2	103	267	0.30	216
1-3	125	394	0.33	235
1-4	148	572	0.37	251
1-5	165	713	0.41	267
1-6	178	888	0.43	275
1-7	190	1055	0.44	280
1-8	201	1212	0.48	291
1-9	211	1372	0.51	300
1-10	218	1608	0.51	300
1-11	223	1748	0.52	303
1-12	231	1912	0.55	310
1-13	234	2087	0.54	309
1-14	237	2230	0.54	309
1-15	238	2386	0.53	306
1-16	239	2509	0.53	305

actions or events. The model was derived as a gamma distribution of long-run individual means with observed numbers of events per time period being distributed about these means according to the Poisson law. Numbers of events and their contributory causes are usually large, but the model can provide a summary of the outcomes of complex observed activities which is tolerant of discrepancies between reality and the assumptions underlying its derivation.

Some of the dependence of the parameters of the model on underlying real factors can be investigated by testing the assumptions in the model against reality. Examples of such analyses are described. An example in which negative binomial distributions provide equivocal fits to some library use data is taken as a starting point in the present study.

CHAPTER 3

COMPARISON OF ASSUMPTIONS IN THE NEGATIVE BINOMIAL MODEL WITH OBSERVED PATTERNS OF USE

3.1 DATA

The data described above (Section 2.2.3) for the short-loan collection users were re-examined in order to test the assumptions implicit in the negative binomial model. Data were available for almost 16 weeks of term and represented recorded use by a sample of 309 users out of the 1550 in the original study. With the records of use for an intervening vacation, therefore, more than half of the use recorded in the academic year could be analysed.

Some trouble had been taken to find a method of choosing a representative sample from the population of 1550 users (42). The sample consisted of users with surnames beginning with the letters B, S or T. This sample represented well the proportions of users enrolled in each of the classes wholly served by the library which housed the short-loan collection. Table 3.1 compares the statistics of the use of the collection by the population and by the sample for a period of observation of almost 10 weeks. The sample mean is just within 95% confidence limits for a random sample of similar size. For the present investigation, however, the sample did not need to be particularly representative of the population. It was the usage patterns of individuals which were to be examined rather than the statistics of their aggregate use.

The testable assumption in the negative binomial model (Section 2.1) relates to the representation of the expected numbers of uses by each individual in similar time periods as a Poisson series with constant mean. Use in one time period should therefore be independent of use in any other time period and the relative frequencies of the observed numbers of uses should correspond to the Poisson probabilities for the observed mean.

The period of observation was divided into units of one week, this being the shortest period of time for which adjacent units could conceivably exhibit similar levels of academic activity. (The stipulated loan period for many of the books in the collection was also one week

TABLE 3.1

Statistics of the use of a short-loan collection for a period of 10 weeks by a potential user population of 1550 students compared to statistics for a non-random sample of 309 students taken from the population.

	Population	Sample
Mean use	5.05	4.44
Variance	69	58.8
k*	0.478	0.505
p*	0.0865	0.1021
Expected variance	58.4	43.5

*Negative binomial parameters fitted by the maximum likelihood method.

and it is possible that the observed pattern of use would thereby be influenced.) The weekly units of time were numbered from 1 to 17. Week 1 included a small number of uses recorded in the days before the term began as well as those recorded in the first week of term. Data collection ended during Week 17 which is therefore incomplete. Week 10 is a fabrication and is included only where the analysis requires. It consists of uses recorded towards the end of the last week of the first term and all uses during the vacation. While the number of users recorded in Week 10 was similar to the numbers recorded in other weeks, the amount of use was half as much again.

Table 3.2 shows the cumulating number of uses by potential users during the period of observation. At the end of the period 78% of potential users had recorded use, 241 out of 309 potential users.

An array was now constructed from the raw data. Each cell contained the number of uses recorded for a particular user in a particular week. The array is shown in Appendix A. From this array, the frequency distribution of the weekly amount of recorded use for each user was constructed.

3.2 FREQUENCY DISTRIBUTIONS OF WEEKLY AMOUNTS OF RECORDED USE

During term (Weeks 1 to 9 and 11 to 17), all potential users would have been engaged in academic work and therefore could be expected to have occasion to use the short-loan collection. The collection contained recommended or well-used textbooks serving the core syllabus of each course. The mean weekly use for each user is estimated from the total number of recorded uses divided by the total number of weeks in the period of observation. Frequency distributions of weekly use which conformed to Poisson series would contain counts of weekly amounts of use which fell within ranges prescribed by confidence limits calculated for a Poisson distribution with the given mean.

3.2.1 Initial test

A rough test of the agreement of each frequency distribution with the expected Poisson series was performed by noting the number of counts for each user which exceeded critical values in each tail of the expected distribution. Critical values embracing the 10% significance levels were

TABLE 3.2

Cumulating use of the short-loan collection by the sample of 309 potential users during the period of observation.

Weeks	Number of uses	Mean use	Variance	Number of users
1	122	0.395	0.986	63
1-2	267	0.864	3.28	103
1-3	394	1.275	5.91	125
1-4	572	1.85	11.6	148
1-5	713	2.31	17.0	165
1-6	888	2.87	26.2	178
1-7	1055	3.41	38.9	190
1-8	1212	3.92	47.0	201
1-9	1372	4.44	58.8	211
1-10	1608	5.20	77.7	218
1-11	1748	5.65	89.2	223
1-12	1912	6.19	103	231
1-13	2087	6.75	122	234
1-14	2230	7.22	140	237
1-15	2386	7.72	170	238
1-16	2509	8.12	192	239
1-17	2559	8.28	200	241

adopted. Because the expected distribution was discrete, the actual significance level tested fell on the division between the integer value embracing the 10% level and the next integer closer to the tail. The actual proportion of the distribution, the probability mass, which exceeded each critical value varied therefore, although it was always less than 10%. For means less than 2.33, the lower critical value was zero. For these means, therefore, only the upper tail could be tested, Very infrequent users were excluded from consideration because their frequency distributions could contain too few extreme values to be testable.

Critical values were calculated for each mean and are shown in Table 3.3. Strictly, the mean itself is an estimate and subject to error: the 15 weeks of the period of observation (Weeks 1 to 9 and 11 to 16) are sampled from a longer series of weeks. The population of weeks is less than double the size of the sample, however, and the sampling is not randomised. It is not appropriate therefore to calculate conventional confidence limits for the estimate of the mean. The range of the estimate would, however, not be large even for a random sample from a large population. In the case of the largest means, 5.3 and 9.3, it would be less than ±0.5 at the 95% level of confidence for a sample of similar size and variance. The estimate of each mean was therefore taken as a single point-value derived from the total recorded use for the weeks sampled.

Out of 139 user samples with 15-week totals from 4 to 34, 61 had one count exceeding a critical value, 48 had counts in more than one week and 30 had none. The actual level of significance was less than 10% and only one tail could be tested. About one excess count in each sample was therefore to be expected. Of the nine users with means of 2.33 or greater, for whom about two excess counts could be expected, five had 4 or more excess counts, one had 3, one had 2 and two had 1 excess count.

These results were taken as preliminary evidence for a greater than expected variation in weekly counts of recorded use.

3.2.2 Extreme values for users recording two and three uses

For the 27 users recording only two uses during the 15 weeks, the probability of these two uses being recorded in a single week would be around 0.008 if weekly amounts of use were Poisson distributed. A '2' could therefore be expected in only 3.3 out of the 27 × 15 = 405 weeks

TABLE 3.3

Critical values of weekly use count which if exceeded indicated rejection at the 10% significance level of the hypothesis that weekly amounts of use were distributed as Poisson series.

Total recorded use	Mean weekly use*	Lower critical value	Upper critical value
3-7	0.20-	0	1
8-16	0.53-	0	2
17-26	1.13-	0	3
27-34	1.80-	0	4
35-36	2.33-	1	4
37-47	2.53-	1	5
48-58	3.20-	1	6
59-69	3.93-	2	7
70-79	4.67-	2	8
80-81	5.33-	3	8
--	--	-	-
130-141	8.67-	5	13

*The lower value in the range is shown; the upper value is that immediately preceding the value in the following line.

sampled. In fact, ten 2's were observed. The null hypothesis that the proportion in the observed population was 0.008 was clearly to be rejected: the number of 2's would exceed 7.5 on only 1% of occasions in random sampling, assuming a normal distribution about the mean of 3.3 with binomial variance of 405 × 0.008 × 0.992 = 3.2.

Similarly for the 30 users making three uses during the 15 weeks, the probability of two or three uses being recorded in a single week would be about 0.018. 17 such extreme values (all 2's) were observed for these 30 users, more than double the expected number (450 × 0.018 = 8.1). The null hypothesis that the proportion of 2's or 3's in the observed population was 0.018 was again clearly rejected: the number of extreme values would exceed 14.7 on only 1% of occasions in random sampling.

3.2.3 Test of the equality of sample mean and variance

As a check on the initial test of Section 3.2.1, 44 sets of data for users picked arbitrarily from the 169 users with means between 0.2 and 2.27 were examined. Twenty-two sets with two or more excess counts and 22 without excesses were tested for the equality of mean and variance in the data.

None of the sets were considered suitable for testing with the chi-squared test, since at best only three aggregated cells could be created from the data. Necessarily, therefore, proportions of extreme values in the tails of the distributions would not be discriminated.

The test of the equality of mean and variance is described by Elliott (77). The statistic $s^2(n - 1)/\bar{x}$ is calculated where s^2 is the sample variance, $\bar{x}$ is the sample mean taken as an estimator of the variance of the expected distribution and n is the size of the sample. The statistic is then compared to tabulated values of the chi-squared statistic for $(n - 1)$ degrees of freedom. The probability of the calculated value of the statistic being observed for a Poisson distribution (where mean and variance are equal) is approximately the level of significance associated with the chi-squared value.

Of the 22 sets of data with excess counts, test statistics for 14 had values which would have been observed in random sampling from Poisson distributions with probabilities of less than 0.05. A further four had probabilities between 0.1 and 0.05. Table 3.4 shows the results of the test.

TABLE 3.4

Test of the equality of mean and variance for 22 sets of recorded use data: probabilities (P) of chi-squared values equal to the test statistics being observed in random sampling.

User *	Total uses	Mean	Vari- ance	Test stat- istic	P **
1/3	6	0.40	0.83	29.0	++
1/5	16	1.07	2.35	30.9	++
1/11	139	9.27	24.5	37.0	++
1/16	13	0.87	1.27	20.5	
1/26	5	0.33	0.52	22.0	+
1/30	27	1.80	6.17	48.0	++
1/34	13	0.87	1.55	25.1	++
1/51	8	0.53	1.27	33.3	++
1/53	17	1.13	2.41	29.8	++
1/58	19	1.27	2.07	22.8	+
1/60	19	1.27	2.21	24.4	++
3/1	7	0.47	0.55	16.5	
3/6	26	1.73	3.78	30.5	++
3/7	47	3.13	8.27	36.9	++
3/9	42	2.80	4.60	23.0	+
3/13	6	0.40	0.54	19.0	
3/20	6	0.40	0.83	29.0	++
3/21	7	0.47	0.83	25.	++
3/23	7	0.47	0.55	16.5	
3/26	8	0.53	1.27	33.3	++
3/32	5	0.33	0.52	22.0	+
3/45	11	0.73	3.07	58.6	++

*Appendix A: sheet number/row number.
**Probability: '++' denotes 'less than 0.05', '+' denotes 'between 0.1 and 0.05'.

Of the 22 sets of data without excess values, the test statisics for all but one sample had values which would have been observed with probabilities of greater than 0.1.

The test was also conducted for all nine users with means of 2.33 or greater. Table 3.5 shows the results. The values of the test statistics for five users were associated with probabilities of less than 0.05. A further one had a probability between 0.1 and 0.05.

3.2.4 Results

The results of Section 3.2 suggest that at least one third of the users tested recorded a greater range in weekly use count than would be expected for frequency distributions approximated by Poisson series. Very few seemed to show a smaller range than expected, but many of the samples contained too few uses for such a result to occur anyway.

3.3 INDEPENDENCE OF WEEKLY INCIDENCE OF USE

3.3.1 Runs test

A runs test from Sokal and Rohlf (78:624) was employed as an initial test of the independence of the weekly incidence of use for each user. A run is defined as one or more consecutive weeks in which the same event (either use or no use) is recorded. Week 10 data were included for this test so that the sequence of weeks should not be broken. If, as assumed in the model, use in one time period occurs independently of use in any other time period for each user, then abnormally high or abnormally low numbers of runs would occur with predictable frequency.

The test does not support a consistant level of significance for all sets of data, even though, as before, a maximum level can be established. Thus, for example, if use is observed in two or 14 out of the 16 weeks, 120 different combinations of these use and no-use weeks could be observed. The highest number of runs is five which could occur in 78 ways. The lowest is two which could occur in two ways. Between them, three or four runs could occur in 40 ways. Two runs occur in less than 10% of possible outcomes. In random sampling it should occur in 1.7% of outcomes. The highest number of runs, five, occurs too frequently to provide a small enough level of significance. Thus, for users with two

TABLE 3.5

Test of the equality of mean and variance for 9 sets of recorded use data for users recording more than 34 uses in 15 weeks.

User *	Total uses	Mean	Vari-ance	Test stat-istic	P **
1/11	139	9.27	24.5	37.0	++
3/7	47	3.13	8.27	36.9	++
3/9	42	2.80	4.60	23.0	+
3/28	57	3.80	4.60	16.9	
3/53	80	5.33	4.24	11.1	
3/60	52	3.47	7.98	32.2	++
4/2	42	2.80	4.89	24.4	++
4/6	57	3.80	3.03	11.2	
4/12	38	2.53	10.7	59.1	++

*Appendix A: sheet number/row number.
**Probability: '++' denotes 'less than 0.05', '+' denotes 'between 0.1 and 0.05'.

or 14 use/weeks, the test is one-tailed with a critical value for the 1.7 percentage point. For three or 13 use/weeks the percentage point is 2.7, and so on. For six or ten, seven or nine, and eight use/weeks, both tails have a critical value which defines roughly 95% confidence limits for numbers of runs. Table 3.6 shows the critical values for the (notional) 5% significance level in either the one-tailed or the two-tailed tests of numbers of runs.

48 out of the 239 users showed use in only one week and so could not be tested. Of the 191 remaining samples, 12 had run counts equal to or less than a lower critical value in Table 3.6. None of the 63 samples for which a two-tailed test could be conducted had a run count in the upper critical region, although two were among the 12 samples already identified. Table 3.7 shows the Week numbers in which use was recorded for each of the 12 samples.

Two of the users represented in these 12 samples were known not to have been present for the latter part of the period of observation. Discounting these two, the proportion of abnormal run counts observed corresponds to the notional significance level of the test. As we have seen, however, the actual level of the test is stricter than this notional level. Rather fewer abnormal samples could have been expected, therefore.

3.3.2 Re-use in succeeding weeks

The apparent independence in the activity of each user in each week can be tested in other ways. If the rate of use remains constant for each user (as assumed in the model) and if use in a time period of the particular length we choose to observe is for each user independent of use in any other time period of similar length, then for a group of users recording use in any given time period, the proportion observed to record in any other time period should be constant. For, if use is a Poisson-distributed random variate, the probability for any user of recording one or more uses in any time period remains constant at $(1 - e^{-m})$, where m is the mean use per time period.

Data for the 239 users recorded in Weeks 1 to 16 were tested in this respect. The numbers of users who recorded use in each Week and who also recorded use in arbitrarily chosen Weeks after this Week were counted. The numbers are set out in Table 3.8. Although the 95% confidence intervals for the mean numbers of subsequent users (re-users)

TABLE 3.6

Critical values for observed numbers of runs of use weeks or non-use weeks which indicate rejection at the 0.05 level of significance of the hypothesis of random activity for a period of activity of 16 weeks.

Observed numbers of use weeks	Tails tested	Critical values
2 or 14	One	2
3 or 13	One	3 or less
4 or 12	One	4 or less
5 or 11	One	4 or less
6 or 10	Two	4 or less; 13 or more
7 or 9	Two	4 or less; 14 or more
8	Two	4 or less; 14 or more

Critical values taken from Rohlf and Sokal (45) Table 28, p.175.

TABLE 3.7

Week numbers in which use was recorded for 12 users with less than five runs of use or non-use in the 16 week period.

Two runs	Three runs	Four runs
1 and 2	7 to 9	1 to 3; 12
1 and 2	10 to 12	1 to 3; 9 and 10
1 to 4		1 to 3; 5 to 10
1 to 5		4 to 6; 12 to 16
4 to 16		4 to 8; 10 to 16

TABLE 3.8

Numbers of users recorded in the first, second, third, eighth or tenth weeks after use in the week indicated in column one.

	Numbers of users who recorded use in the					
Week number	Week indic-ated	1st week after	2nd week after	3rd week after	8th week after	10th week after
1	63	33	30	32	28	29
2	73	34	36	28	37	25
3	66	42	33	32	29	31
4	84	46	38	39	38	38
5	79	40	38	44	33	33
6	77	40	36	33	32	38
7	77	41	34	35	33	
8	84	38	37	35	33	
9	80	39	27	33		
10	86	39	39	38		
11	70	36	37	31		
12	79	47	45	42		
13	80	42	39	36		
14	73	45	34			
15	71	34				
16	64					
Mean	75.4	39.73	35.93	35.23	32.88	32.33
Var.	51.05	18.64	18.69	20.53	11.84	26.27
SE	1.786	1.115	1.155	1.257	1.217	2.092
CL	±3.81	±2.39	±2.496	±2.74	±2.88	±5.38

Var.:variance. SE: standard error of the mean. CL: 95% confidence limits for the mean.

overlap except in the case of the first and eighth weeks, it seems clear that a gentle decrease occurs in the numbers of re-users as time goes on. The extent to which independence is approximated is noticeable, however. The stability in the numbers of re-users throughout the period of observation does not appear to result from the activity of a core of regular users. From Table 3.9 we see that very few users use regularly enough to make up such a core. Of the 63 users first using in Week 1, for example, 33 re-use in Week 2, but only four re-use in every succeeding week (i.e. all 16 weeks).

It is also clear from the totals shown in Table 3.9 that new users are recruited throughout the period of observation but in gradually declining numbers as would be expected under conditions of independence where the probability of recording use remains constant.

3.3.3 Use in adjacent weeks

The independence of user activity in adjacent pairs of weeks was also tested. As we have seen, for the assumptions in the model, the probability for any user of no use in any week is e^{-m}, where m is that user's expected rate of weekly use. The probability that one or more uses are recorded is $(1 - e^{-m})$ therefore. For any two adjacent weeks taken in order of occurrence, four outcomes are possible for each user: (no use, no use); (use, no use); (no use, use) and (use, use). The probabilities associated with these four outcomes are, respectively: e^{-2m}; $e^{-m}(1 - e^{-m})$; $e^{-m}(1 - e^{-m})$ and $(1 - e^{-m})^2$.

The frequencies of these outcomes were compared to expectation for an arbitrary sample of users. Frequencies in 8 separate pairs of weeks in the period of observation (i.e. for Weeks 1 and 2; Weeks 3 and 4; and so on) were counted for all users recording 5, 12, 17 to 19, 29 to 30 and 56 to 60 uses. The correspondence of the observed and expected frequencies was tested with the chi-squared test. The result is shown in Table 3.10.

In general, (use, use) weeks appear less frequently than expected and outcomes with no use in one or both weeks more frequently than expected. This appears to bear out the result of the tests in Sections 3.2.1 and 3.2.2. Some users concentrate an unexpectedly high proportion of their recorded use into some weeks. In other weeks, therefore, their count will be at the lower extreme and for many will result in an excess of zero-use weeks. For the highest users in Table 3.10 it could appear that use in

TABLE 3.9

Numbers of users first recorded in each week of the period of observation (columns) arranged according to the number of weeks in which they record use.

Weeks of use	Numbers of users first observed in week number 1	2	3	4	5	6	7	8	9
1	2	2	4	5	1		2	4	6
2	5	8	1	1	4		2	2	1
3	6	3	1	3	6	3	5	3	1
4	7	5		2	5	6		2	1
5	3	8		2		2			1
6	1	1	4	3		1	2		
7	6	2	3	1			1		
8	4	2	4	3	1	1			
9	5	3	1	1					
10	6	4	3						
11	4	2							
12	2			1					
13	6			1					
14	1								
15	1								
16	4								
Totals	63	40	22	23	17	13	12	11	10

Total users = 241.

TABLE 3.9 (continued)

Numbers of users first recorded in each week of the period of observation (columns) arranged according to the number of weeks in which they record use.

Weeks of use	Numbers of users first observed in week number 10	11	12	13	14	15	16	17
1	4	2	5	2	2	1	1	2
2		2	2		1			
3	2	1	1	1				
4	1							
Totals	7	5	8	3	3	1	1	2

TABLE 3.10

Incidence of use in eight adjacent pairs of weeks: observed (O) and expected (E) frequencies of four possible outcomes for users recording 5, 12, 17 to 19, 29 to 30 and 56 to 60 uses in total.

Out-come	Mean use* 0.313		0.75		1.13		1.88		3.63	
	O	E	O	E	O	E	O	E	O	E
(N.N)	53	47.1	25	17.9	28	10.1	3	0.75	0	0.02
(N.U)	15	17.3	19	19.9	17	21.1	6	4.26	5	0.83
(U.N)	17	17.3	16	19.9	22	21.1	10	4.26	3	0.83
(U.U)	3	6.3	20	22.3	29	43.8	13	22.7	24	30.3
Users	11		10		12		4		4	
χ^2	2.8		3.9		37.4		14.9		25.0	
$\underline{P}$	0.4		0.3		0		0.001		0	

$\underline{P}$ is the approximate probability of the observed chi-squared value being exceeded in random sampling. Degrees of freedom: 3 except col. five (2) and col. six (1).
*Respectively, 18/16; 30/16 and 58/16 for last three columns.

one week reduced the probability of use in another, but the results of Sections 3.3.1 and 3.3.2 do little to support this view. The explanation above involving erratic use rather than dependence seems preferable. Of course, this conclusion holds only for the units of time (one week) adopted here. It is possible that inter-period dependence could be demonstrated more strongly for time periods of different length.

3.4 CONCLUSION

The weekly time period into which the data were divided coincides not only with a recurring cycle of academic activity but also with the stipulated loan period of five-sixths of the issues recorded. (The rest were issued for one day. Only about one quarter of the issues of either type of loan would have been returned precisely on the date due, however (42:Appendix 12).) Despite these potential coincidences, it has been impossible to demonstrate any marked regularity in either the frequency or the incidence of recorded use. For a large minority of the users represented in the array of Appendix A, recorded use appears to proceed more erratically than would a random process. In employing the negative binomial distribution for describing frequency distributions of recorded use, it is assumed for simplicity that recorded use is approximately randomly distributed about a constant mean for each user in each time period. It seems that this hypothesis will often fail and that an alternative would require either that the means vary over time (sometimes reducing to zero, perhaps) or that a different theoretical distribution which is capable of accommodating a variance greater than the mean is substituted for the Poisson component of the negative binomial distribution. The assumption of the independence of the weekly incidence of use is less markedly challenged by the data. This hypothesis could be accepted for the time period analysed, especially if mean rates of use were allowed to vary.

Clearly these hypotheses (inter-period independence in the incidence of use; constancy of individual mean rates of use; equality of mean and variance) can only be tested with reference to a given unit of time. For very short time periods (days or hours) one or more would almost certainly be rejected. For example, the overall mean rate of use per user for the period of observation is about 0.13 uses per working day. The Poisson probability of recording more than one use in any day is then about 0.005. Experience tells us however, that such an outcome occurs in real life much more frequently.

3.5 AGGREGATE STABILITY IN PATTERN OF USE

In Table 3.8 there is a noticeable equilibrium in the aggregate amounts of use recorded in each week. It is what we should expect if mean rates of use for individual users remained constant over time. In Section 3.4 we have doubted this hypothesis without being able to prove an alternative, and it may be that some of the success of the negative binomial model in fitting the recorded use data owes to this equilibrium in aggregate rather than individual amounts of use. How the equilibrium comes about is not clear: few users used regularly enough week by week to sustain it. Figure 3.1 shows the average pattern of users' transitions from use to no use for all sequences of four successive weeks. Equilibrium clearly depends upon a stable pattern of intermittent use involving most users sooner or later.

In this section, some aspects of this equilibrium are reported, and in the next chapter the availability of library material is examined to see whether the supply of material could have regulated aggregate use.

3.5.1 Weekly numbers of users and uses

Table 3.11 shows recorded numbers of users and uses for each of the Weeks 1 to 16. No particular trend is evident, although the first and last weeks have low numbers. The pattern of use in the first weeks of the first term (Weeks 1 to 6) is clearly different to the first weeks of the second term (Weeks 11 to 16). Normality in the distribution of weekly values about sample means with estimated variances could not be rejected (Chi-squared test: expected distribution divided into six intervals, P(users) = 0.25; P(uses) = 0.4). The 95% confidence intervals for single-week values sampled from the supposed populations are shown in Table 3.11. (Similar tests applied to numbers of re-users observed in successive weeks (c.f. Table 3.8) gave similar results and are not further described).

Correlation between pairs of user and use counts was tested. On the assumption that each set of values was normally distributed and that a linear relationship was expected between them, the product-moment

FIGURE 3.1

Average weekly transition between use and no use in all sequences of four successive weeks from a period of observation of 16 weeks for a sample of 309 students: mean numbers of users [U] and non-users [N] with 95% confidence limits where calculated.

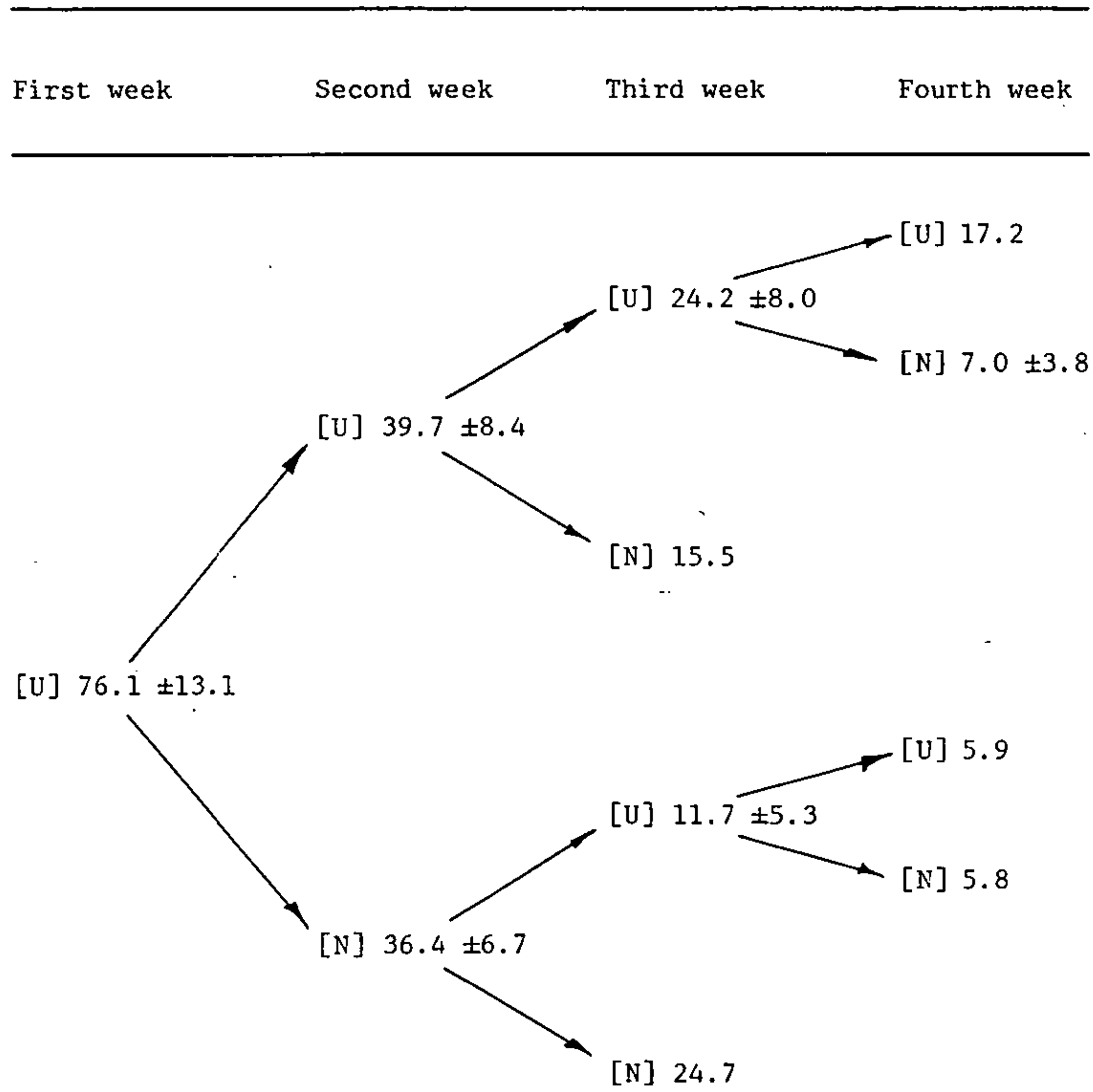


TABLE 3.11

Numbers of users recording use and amount of use for Weeks 1 to 16.

Week number	Number of users	Number of uses
1	63	122
2	73	145
3	66	127
4	84	178
5	79	141
6	77	175
7	77	167
8	84	157
9	80	160
[10	86	236]
11	70	140
12	79	164
13	80	175
14	73	143
15	71	156
16	64	123
Mean	74.7	151.5
Var.	46.1	354.7
CL	61.4-88.0	114.6-188.4

Means calculated for Weeks 1 to 9 and 11 to 16. Var.: variance. CL: 95% confidence limits for single week values assuming random sampling from normally distributed population with means and variances shown.

correlation coefficient, r, was calculated in order to test the null hypothesis that no correlation existed between pairs of values (i.e. $\rho = 0$). The value of the statistic, r, gave a highly significant rejection of the hypothesis ($r = 0.78$; $t = 4.46$ for $(n - 2) = 13$ degrees of freedom where $t = r\sqrt{(n-2)}/\sqrt{(1-r^2)}$; $P(\rho = 0) < 0.001$).

3.5.2 Cumulating use

The pattern of the cumulation in numbers of users and amounts of use (Table 3.12) was roughly similar for time periods beginning in Week 1, Week 5 and Week 11, again indicating that patterns of use were relatively stable irrespective of time. (Chi-squared test of the comparison of the numbers of users first observed in each week of each period: for $(3-1) \times (5-1) = 8$ degrees of freedom, $P = 0.2$; chi-squared test of the comparison weekly amounts of use (not cumulated): for $(3-1) \times (5-1) = 8$ degrees of freedom, $P = 0.1$)

3.5.3 Frequency distributions of weekly use by users

A composite frequency distribution of weekly amounts of use by users was formed from the means of the terms in the 15 single week distributions (Week 10 omitted). Table 3.13 shows this composite distribution and Table 3.14 the array of individual distributions. Each term in the array was tested against the corresponding composite term in order to test the hypothesis that the array could have been obtained in random sampling from a hypothetical population of users whose use was distributed as in the composite distribution (the terms of which, being row means, represented the best estimates of these population values). In order to perform the chi-squared test, frequencies were pooled for four and five uses and for six or more uses. The expected values in the composite distribution were then 5.6 and 3.47 respectively. The hypothesis that the array was sampled from the composite population could not be rejected (Chi-squared test for $(15 - 1) \times (6 - 1) = 70$ degrees of freedom, $P > 0.9$). A similar test was performed for frequency distributions of fortnightly amounts of use with similar results and is not further described.

TABLE 3.12

Cumulating numbers of users and total amounts of use over periods of five weeks beginning week 1, week 5 and week 11.

Period in weeks	Numbers of users and amounts of use in weeks: 1 to 5		5 to 9		11 to 15	
	Users	Use	Users	Use	User	Use
1	63	122	79	141	70	140
2	103	267	116	316	113	304
3	125	394	140	483	134	479
4	148	572	158	640	151	622
5	165	713	177	800	156	778

TABLE 3.13

Composite frequency distribution formed by averaging distributions of recorded use for 15 single weeks.

Number of recorded uses	Observed number of users
0	234.3
1	39.7
2	17.2
3	8.73
4	3.47
5	2.13
6	1.67
7	0.40
8	0.80
9	0.13
10	0.067
11	0.20
12	0
13+	0.20

Total users: 309. Total uses: 151.5

TABLE 3.14

Frequency distributions of recorded use for single weeks.

Number of recorded uses	Observed numbers of users in week:								
	1	2	3	4	5	6	7	8	9
0	246	236	243	225	230	232	232	225	229
1	33	39	31	48	48	36	41	42	47
2	17	14	17	16	16	21	21	24	13
3	5	12	11	7	10	8	7	10	10
4	3	4	6	4	2	3		5	4
5	2	2	1	3		4	3	1	2
6	3	1		2		2	1	2	2
7					1	2			
8				2	2		2		1
9				2					
Others		10				15	*		11

*11;15.

TABLE 3.14 (continued)

Frequency distributions for single weeks.

Number of recorded uses	Observed numbers of users in week: 11	12	13	14	15	16
0	239	230	229	236	238	245
1	35	43	41	38	36	37
2	19	15	19	21	15	10
3	7	7	7	6	12	12
4	3	7	4	1	4	6
5	3	2	3	3	2	1
6	2	4	1	3	1	1
7	1		2			
8		1	3	1		
9						
Others					22	11

3.5.4 Conclusion

Throughout the period of observation, the number of users who had recorded use of the short-loan collection increased. Large changes in aggregate rates of use might have been expected as students tackled the various aspects of their courses and received varying amounts of tuition and exposure to the collection. Yet the pattern of user activity, as measured above, appears to have changed little during the period. The amount and distribution of use in each week and each fortnight remained similar even though those generating the use changed continually. Use cumulated, and new users appeared, at similar rates from three starting points within the period of observation.

In view of the diversity of activities supported by the collection, the pattern of use is less variable than expected therefore, and a model which assumes that the users' propensity to use is fixed over time and independent of their recent activity may clearly be adequate. Individual amounts (or perhaps rates) of use do appear to vary more widely than expected, but in aggregate it seems, some of these variations cancel out giving an equilibrium which simplifies the description of user activity and encourages the adoption of a simple stochastic model. There is an extra justification, therefore, for using the negative binomial distribution to approximate the observed frequency distributions of use in addition to the initial agreement between the observed and expected shapes of these distributions.

3.6 SUMMARY

Data for recorded library use by 309 students over 16 weeks were examined to see if individual weekly use totals were Poisson distributed about the sample mean and if use occurred independently of the activity preceding or succeeding it.

A large minority of users showed a greater range in their weekly totals of use than would have been expected from the random sampling of a Poisson variate. The assumption of independence in the weekly incidence of use appeared a workable hypothesis for most users, however.

Even if, as individuals, real users were more erratic in their amounts of weekly use than those in the model, in aggregate they presented a surprising equilibrium in their weekly pattern of use, To this extent, therefore, the model was not an unreasonable simplification of reality.

CHAPTER 4

EFFECT OF THE AVAILABILITY OF LIBRARY MATERIAL ON USE

In Section 3.5 an equilibrium was observed in numbers of users and amounts of use throughout the period of observation. It is possible that the amount of useful material available in the short-loan collection was insufficient to sustain a greater level of use, thus constraining the weekly amount of use observed and possibly determining the observed distribution of use. This possibility is investigated below; first, by assessing the amount of material in the collection during the period of observation and comparing it to the amount taken up by the users, and second, by simulating the effect upon use of varying the amount of useful material available from a collection in an attempt to gauge the extent to which the observed distributions of use were attributable to the amount of material available.

The findings of Sections 3.2 and 3.3 were important in this respect. It appeared reasonable to suppose that amounts of use for any individual user in similar time periods could, as a first approximation, be estimated by calculating the probabilities of each amount in any time period as a Poisson series. The distribution of mean rates of use per time period among users could then be assigned arbitrarily in order to test hypotheses or could be derived from recorded use data in order to simulate non-deterministically the use of actual collections. For these reasons, and because it was thought difficult to control extraneous variables in observing or experimenting with live users, it seemed that simulation would provide the best tool for investigating the effect of availability on use.

4.1 AVAILABILITY OF MATERIAL IN THE SHORT-LOAN COLLECTION

The short-loan collection comprised about 3200 books. Assuming that the sample of users described in Section 3.1 comprised about one fifth of all potential users and that it was representative of them, then the average number of issues per week to potential users was $151 \times 5 = 755$. In the year of the survey each book received an average of about eight issues (books receiving less than four issues were relegated annually to the main collection). Some books, especially the one-day loans, might be used more than once per week, so that somewhat less than 755/3200 = 24% of the collection would have been used in any week. Assuming that the average retention time was about one week (Section 3.4), we can say that between one fifth and one quarter of the collection would have been absent from the shelves at any time.

Although many of the books in the collection were standard texts for the syllabuses of science or engineering courses, undoubtedly some of the books would have been useful for less than the whole period of observation. Many titles were available in multiple copies, so that the 3200 books represented less than 1500 titles. It is possible, therefore, that at any moment of time only a minority of the apparent resource was judged useful by the users and, this having been taken up, further use was inhibited.

In the year following this survey, however, while the collection and the number of users remained largely unchanged, the number of issues from the collection increased markedly, from 25964 to 30946. It appears, therefore, that had their rates of use been higher, the users observed in the previous year could also have recorded substantially more use. Thus the constraint imposed by the amount of useful material in the collection should not have been severe enough to maintain the aggregate rate of use at its observed level.

4.2 EFFECT OF AVAILABILITY ON FREQUENCY DISTRIBUTIONS OF RECORDED USE

Even if the shortage of useful material in the collection was not severe enough to create the equilibrium observed in Section 3.5, it would certainly have caused the demands of potential users to interfere with each other. There was by no means enough potentially useful library material to satisfy every user on demand. If competition among the users altered the scale or ranking of their long-run rates of use, then an observed distribution of use would result not only (as assumed up till now) from the users' propensity to use but also from the level of availability of useful material. For reasons outlined at the beginning of this chapter, a computer simulation of library use by students was employed to test the effects of competition on patterns of use. The procedure and its result are briefly described below: further details are presented in Appendix B.

4.2.1 Description of library use simulation

It was assumed that 100 students from the same class used a collection of useful library material for a period of 50 working days (say, 10 weeks). During this period, items from the collection gained and then lost their potential usefulness to the users. Useful items could be removed from the collection and used by the students for varying lengths of time. Only potentially useful books were included in the simulation: it was assumed that they could be derived from all sections of the library or even different libraries but, for convenience, they were treated as a single 'collection'.

Various parameters were assigned to describe the behaviour of the users and the availability of the material. Where possible their values were calculated from data for actual users or libraries. Failing this, values were estimated initially and then adjusted during calibration runs of the simulation program to yield two required outcomes. First, a total of about 800 uses was required. This was thought reasonable for course-related use by science or technology students and was estimated from published reviews (15,68) supplemented by local data available to the author. Second, the frequency distribution of numbers of recorded uses per observed (i.e. successful) visit was required to approximate to one of three sets of data available to the author from local libraries (Table 6.18). It could not be claimed that adjusting the parameters in the simulation program to achieve these requirements would guarantee the assignment of realistic values, but the approach was taken to be superior to pure guesswork.

4.2.1.1 User variables

Potential users were deemed to differ not only in their rate of visiting the collection and in the number of items they sought to use per visit, but also in their characteristic reaction to failure. Some users regularly returned on up to two occasions to attempt the use of an item not available at a previous attempt; a few made no further attempts on that visit; a few attempted to use substitute material; and one half were unaffected by failure. These rates and characteristics were an arbitrary and crude attempt to quantify some of the patterns of behaviour reported in surveys (e.g. 1,79) and some of the variables identified from user studies, for example (following Wilson): individual perception of information need; resort to the information channel under investigation; patterns of information seeking and evaluation; and the acceptability of the information channel within the individual context of need and the general environment of information use (48). The values finally assigned are shown in Table 4.1.

These rates and characteristics were combined over the users so that all possible permutations were represented in their appropriate frequencies. The mean rate of visiting was determined during calibration and was about one third that reported by Harrop for social science students (18). About one half of Harrop's visits, however, appear not to have been made with the primary purpose of using library material. The distribution over the users of the different probabilities of a further attempt at use (Table 4.1) was estimated indirectly from local data and adjusted during calibration.

The distribution among the users of rates of visit in 50 days was used as the experimental variable. Two arbitrary distributions of rate of visit were adopted: a Poisson distribution roughly symmetrical about the mean and a geometric distribution. The lowest expectation in each case was one visit.

All users, it was assumed, would gain experience and encouragement from a successful attempt to use material in the collection. A crude mechanism increasing the probability of a further attempt after early success was therefore included in the simulation. It was adjusted during calibration to produce the required frequency distribution of observed numbers of uses per visit referred to in Section 4.2.1.

4.2.1.2 Collection variables

A total of 400 items was assumed to be potentially useful during the period represented by the simulation. No differentiation of the collection into titles and copies was made. Each day, new items became available and others, having received a number of uses, lost their usefulness. Initial estimates of the size of the collection (Appendix B) and the daily increment of useful stock were adjusted during calibration runs of the simulation program to produce the required total of uses for 100 users sharing a single mean rate of use and visit. Table 4.2 shows the values adopted for the collection parameters.

TABLE 4.1

Values of variables used in simulation: users.

	Value
Number of users	100
Time period	50 'days'
Mean number of expected visits	10
Mean number of attempts per visit	1.45
Probabilities of a further attempt	0.1; 0.25; 0.45
Multiplier after success	$2/s^2$*
Users revisiting after failure	30
Maximum number of revisits	2
Users substituting after failure	10
Probability of making substitute use	0.5
Users abandoning visit after failure	10

*Where s is the number of successes so far on that visit. Thus a user with probability of 0.25 of making a further attempt has a probability of 0.5 after one success and 0.125 after two successes of making a further attempt.

TABLE 4.2

Values of variables used in simulation: initial collection.

	Value
Total number of useful items	400
Maximum possible uses in 50 'days'	1046
Initial number of useful items	0
Daily addition of useful items	
20 day use	3
5 day use	3
1 day use	2
Probability that return still useful	
20 day use	0.67
5 day use	0.67
1 day use	0.8
Probability attempt to use succeeds	0.8a/c*

*Where a is the number of currently useful books not in use and c is the number of currently useful books.

The probability that a user would succeed in an attempt to use was determined by the proportion of currently useful material which was not in use. When use occurred, the length of time for which the item was to be retained was allocated at random, but according to the proportions of material available at the time in each of the three retention categories.

During the simulation of each day's use, potential users were taken in order of their expected frequency of visit. It was assumed that the most frequent visitors would also be the first to the library at any opportunity. On average, 20 visits occurred per day and in general would result in less than 25 uses (the expected mean being 16). Prior users would not have greatly depleted the stock of useful items therefore.

The maximum probability of a successful attempt, given that all currently useful items were available, was set at 0.8 to take account of known failure rates in academic library use (80 and Appendix B). These failures were assumed to be associated with errors in taking references or searching catalogues or shelves, or with errors or inefficiencies in the guiding of the library, the indication of locations or in the prompt and correct replacement of items returned from use.

4.2.1.3 Calibration and simulation runs

Figure 4.1 shows in chart form the method of calibrating the simulation program. Parameter values for rate variables were represented in the program as probabilities of the occurrence of each event. Thus, a user expected to visit ten times during the period had a daily probability of visit of 0.2 and a book expected to have a useful life of five uses had a probability of 0.8 of retaining its usefulness when returned after each use. When the program was run, the outcome for each event was determined by calling a fresh random number, greater than zero but less than unity. If the probability value of the parameter exceeded the random number, the event occurred. This method was expected to result in numbers of events distributed about individual mean rates of occurrence with approximately Poisson probabilities (Figure 4.2). The distribution of the random numbers was tested and shown to be acceptably uniform (Appendix B).

Each run of the simulation consisted of 50 repetitions of the following procedure. The probability of visit of each user was tested against a fresh random number in order to determine whether a visit occurred on

FIGURE 4.1

Sequence of adjustment to simulation parameters during calibration

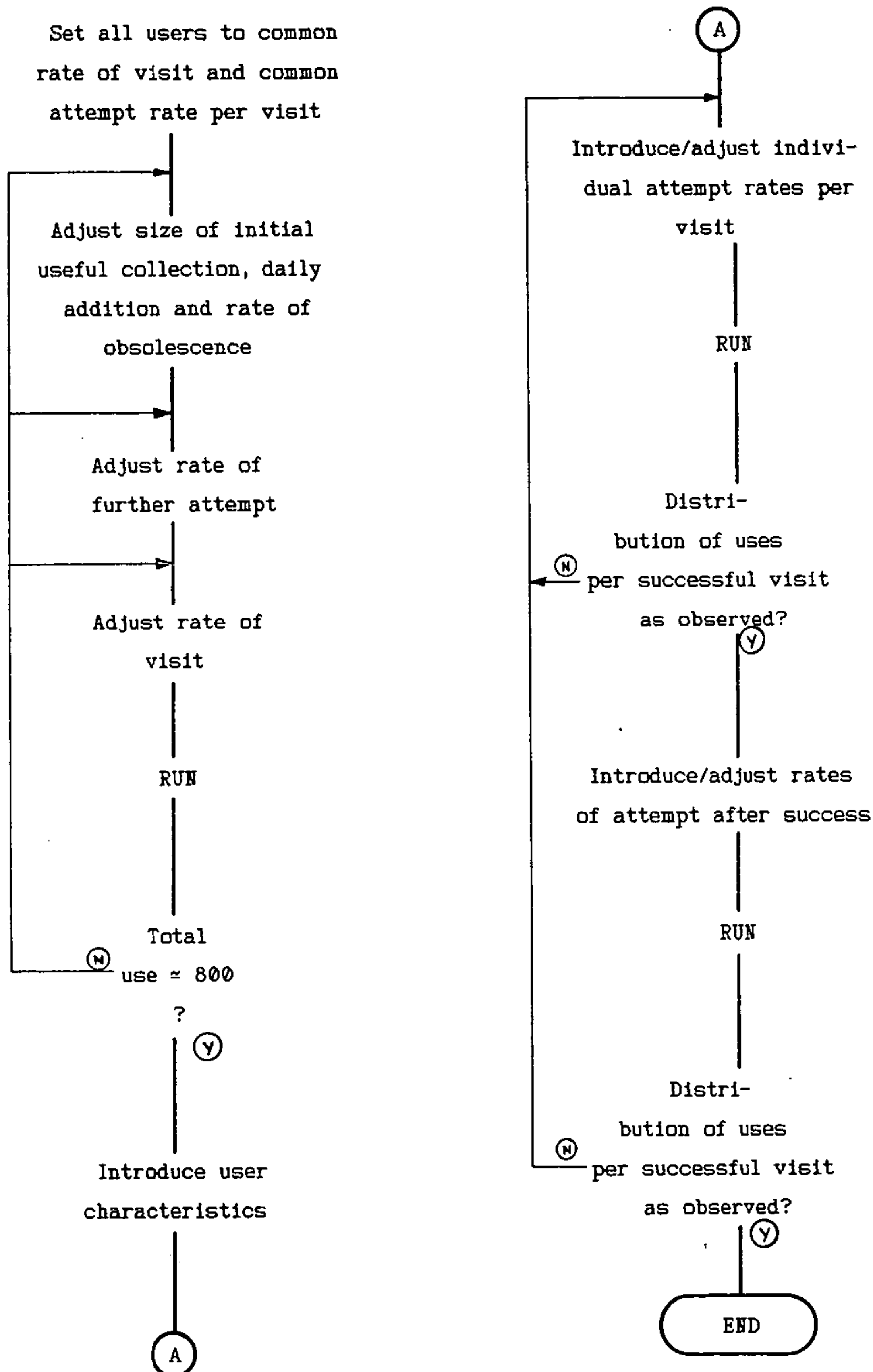


FIGURE 4.2

Distributions of numbers of events as observed and as simulated

---

In the simulation, the occurrence of an event results from a Bernouilli trial, that is, a trial with only two outcomes not necessarily equally likely (for example: visit; no visit). If the trials are independent, the numbers of events expected to be recorded at the end of the period of simulation will be distributed with binomial probabilities about the mean rate of occurrence.

Observed numbers of uses per time period were distributed with approximately Poisson probabilities about observed means (Section 3.2.4), and it is assumed that numbers of events for other activities (for example, visits) are also distributed in this fashion. The incidence of use appeared not to be greatly dependent upon incidences in preceding or succeeding time periods. Because sums of independent Poisson variates are also Poisson variates, it is possible to assume that a process which results in a Poisson distribution of observed numbers of events (such as recorded uses) can be divided into short time periods where the mean rate of occurrence is so very much smaller than unity that, in practice, only zero or one events can occur. The Poisson probabilities of the outcome in each of these time periods will then be similar to the probabilities associated with the equivalent Bernouilli trial. Thus, if p is the probability of an event occurring in the Bernouilli trial and λ is the mean rate of occurrence in observed time periods, then, for the probability of no event, $e^{-\lambda} \simeq (1 - p)$ and, for the probability of one event, $\lambda e^{-\lambda} \simeq p$, where λ and p are equal and much less than unity.

When compared over longer time periods, the binomial and Poisson distributions of numbers of events will show similar means, but the variance of the binomial distribution will be rather less than that of the Poisson distribution.

that 'day'. If the outcome was a visit, then further trials were made as shown in the flow-chart of Figure 4.3 to determine the amount of use. If no visit occurred, then the probability of visit of the next user was tested, and so on for all 100 users. At the end of the run, frequency distributions of total use, and daily amounts of use were printed out. Figure 4.4 shows an example for a run using a uniform distribution of expected numbers of visits ranging from 1 to 19.

A maximum of 1046 uses was possible during the useful life of the material in the collection. About 730 to 740 uses occurred during simulation runs, about 70% of those possible. (Further uses during each simulation run resulted from the use of substitute material considered to be derived from outside the collection).

Further simulation runs were made with reduced and enlarged collections so that the effect upon the frequency distribution of use could be observed. Collections of 250 and 550 items were assumed, with daily additions of 2, 2 and 1, and 4, 4 and 3 items respectively for 20-day, five-day and one-day material. For the reduced collection, a maximum of 620 uses was possible. About 570 to 590 uses, more than 90% of those possible, occurred in simulation runs. This collection was taken as the lower realistic level of availability. More than 70% of visits (excluding revisits for which the proportion was higher) failed to record any use, compared to about 60% for the initial collection. Towards the end of a run, the probability of succeeding in an attempt began each day at about 0.3 and declined to about 0.2. For the initial collection, these figures were about 0.4 and 0.3 respectively. The maximum possible number of uses for the enlarged collection was 1471, of which about 60% were taken up in simulation runs.

It was assumed that the availability of material in the short-loan collection was similar or better than that represented in the reduced simulation collection. Even if the 30964 uses (Section 4.1) constituted a maximum for the collection, use in the year in question was then only 84% of the maximum possible, which was less than the proportion generated for the reduced collection in the simulation.

4.2.2 Result and conclusions

Table 4.3 shows the aggregated frequency distributions of use for three simulation runs with the initial collection and three runs with the

FIGURE 4.3

Flow chart for simulation program: daily iteration for each user

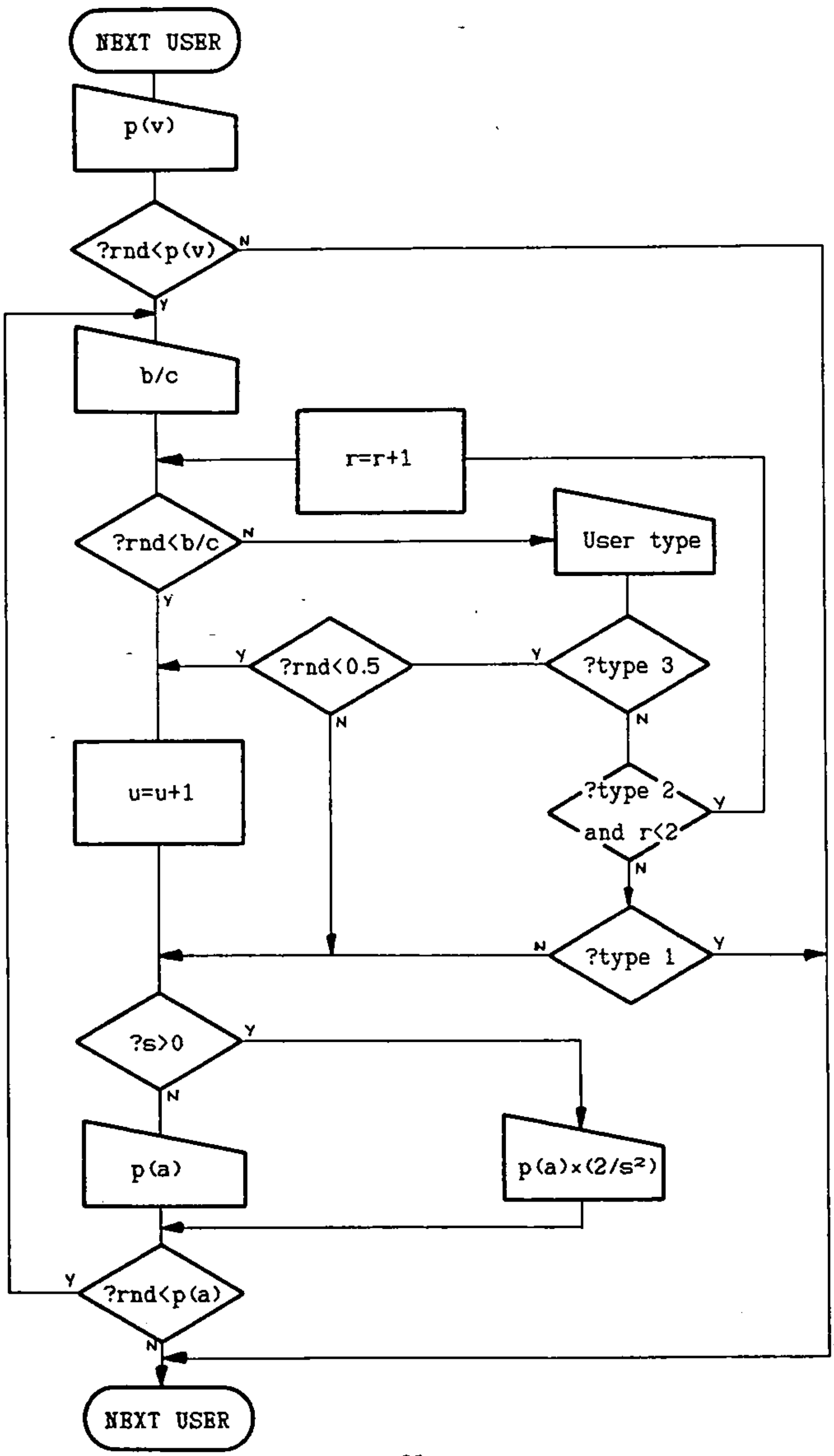
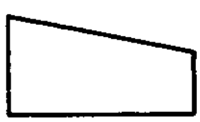


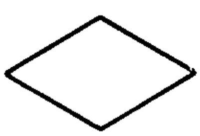
FIGURE 4.3 (continued)

Key to flow chart for simulation program

---

b	Number of currently useful books not in use
c	Number of currently useful books
p(a)	Individual probability of making another attempt
p(v)	Individual probability of daily visit
r	Number of revisits on this visit: r=0 initially
rnd	Random number
s	Number of uses on this visit
u	Individual number of uses so far


Input of value of parameter or variable


Decision

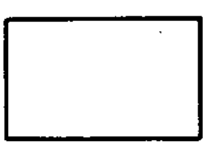

Operation

FIGURE 4.4

Example of output from simulation run with uniform distribution of rates of visit.

4.4.1 Frequency distribution of numbers of users (Usrs) for each number of uses.

4.4.2 Numbers of visits (Vsts) and numbers of uses recorded for each user (No. 1 to 100), with type of user from 1 to 4 (see 4.4.4) and individual probability of further attempt (.1; .25; or .45) shown under 'A+Ty'. The number of revisits made by type 2 users is also shown.

4.4.3 Daily totals (Days 1 to 50) for numbers of useful books in the collection at the start of the day (Bks), numbers of useful books available at the start of the day (Avb) and numbers of uses recorded during the day. The frequency distribution of numbers of visits (Vsts) and revisits (Rvts) for numbers of recorded uses is also shown, along with the total number of substitute uses made from outside the collection by type 3 users, the total numbers of uses made for each retention category of book and their daily rate of addition to the collection.

4.4.4 Expected number of visits (EVt), probability of further attempt (EPrA) and user type for each user (No. 1 to 100). The proportions of each type of user in the population are also shown. A uniform distribution of visits, designated the second type of distribution investigated ('Dist'n of visits = 2'), is shown.

DISTRIBUTION FIGURE 4.4.1

Total uses = 792
Mean use = 7.92

Uses	Usrs	Uses	Usrs
0	14	50	
1	10	51	
2	9	52	
3	8	53	
4	8	54	1
5	7	55	
6	6	56	
7	1	57	
8	5	58	
9	3	59	
10	3	60	
11	3	61	
12	6	62	
13	1	63	
14	2	64	
15	1	65	
16		66	
17		67	
18	1	68	
19	1	69	
20	2	70	
21	1	71	
22		72	
23	1	73	
24		74	
25		75	
26		76	
27	1	77	
28		78	
29	1	79	
30	1	80	
31		81	
32		82	
33	1	83	
34		84	
35	1	85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43	1	93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

USAGE

FIGURE 4.4.2

Total visits = 1683, including 713 revisits
Total uses = 792

No.	Vsts	Uses	A+Ty	No.	Vsts	Uses	A+Ty
1	4	5	4.25	51	9	5	4.1
2	19	21	4.45	52	7	2	4.25
3	21	6	4.1	53	10	3	4.45
4	15	5	4.25	54	11	2	4.1
5	19	20	4.45	55	8	1	4.25
6	39	14	2.1	56	26	8	2.45
7	50	19	2.25	57	22	8	2.1
8	70	54	2.45	58	26	12	2.25
9	16	8	1.1	59	8	2	1.45
10	11	12	3.25	60	7	9	3.1
11	23	27	4.45	61	8	1	4.25
12	15	3	4.1	62	2	0	4.45
13	17	10	4.25	63	9	0	4.1
14	20	18	4.45	64	7	2	4.25
15	11	4	4.1	65	3	2	4.45
16	49	20	2.25	66	13	3	2.1
17	90	43	2.45	67	11	4	2.25
18	34	10	2.1	68	17	12	2.45
19	17	3	1.25	69	7	1	1.1
20	17	33	3.45	70	6	6	3.25
21	17	6	4.1	71	10	9	4.45
22	17	6	4.25	72	3	0	4.1
23	14	12	4.45	73	3	0	4.25
24	11	4	4.1	74	7	2	4.45
25	16	11	4.25	75	7	4	4.1
26	105	35	2.45	76	18	5	2.25
27	45	15	2.1	77	29	9	2.45
28	43	23	2.25	78	14	4	2.1
29	14	4	1.45	79	1	0	1.25
30	14	10	3.1	80	3	6	3.45
31	14	6	4.25	81	6	0	4.1
32	10	3	4.45	82	3	1	4.25
33	7	2	4.1	83	4	2	4.45
34	9	4	4.25	84	4	1	4.1
35	13	3	4.45	85	5	0	4.25
36	31	8	2.1	86	13	5	2.45
37	32	11	2.25	87	10	1	2.1
38	85	30	2.45	88	17	11	2.25
39	10	5	1.1	89	5	0	1.45
40	10	12	3.25	90	6	4	3.1
41	13	13	4.45	91	0	0	4.25
42	15	5	4.1	92	1	0	4.45
43	8	1	4.25	93	3	0	4.1
44	12	12	4.45	94	2	1	4.25
45	15	1	4.1	95	3	3	4.45
46	26	7	2.25	96	6	0	2.1
47	73	29	2.45	97	3	1	2.25
48	17	8	2.1	98	9	2	2.45
49	9	3	1.25	99	1	0	1.1
50	8	14	3.45	100	0	0	3.31

Initial no. useful books = 0 Daily addition = 8
Prop'n day = 0 , add'n = 2 , uses = 284
Prop'n week = 0 , add'n = 3 , uses = 262
Prop'n month = 0 , add'n = 3 , uses = 185
Max success base rate = 0.8, prob of substitute = 0.5
Total of daily uses = 792, substitute uses = 61

Day	Bks	Avb	Uses
1	8	8	13
2	16	10	8
3	23	12	9
4	30	15	12
5	37	16	12
6	43	18	11
7	50	20	11
8	56	24	13
9	62	25	10
10	67	31	17
11	73	32	14
12	79	31	17
13	84	31	14
14	90	31	15
15	95	39	17
16	100	42	14
17	105	44	19
18	109	48	16
19	116	47	15
20	122	46	17
21	126	49	11
22	131	55	14
23	136	57	17
24	138	62	19
25	141	64	19
26	145	65	23
27	149	63	16
28	153	65	18
29	158	67	13
30	161	73	13
31	165	78	13
32	169	83	12
33	173	89	23
34	178	83	13
35	182	89	13
36	186	93	22
37	191	91	12
38	194	97	9
39	199	106	16
40	202	111	24
41	205	110	32
42	210	102	23
43	214	98	15
44	218	102	21
45	221	103	23
46	220	111	13
47	222	119	23
48	226	119	12
49	228	129	18
50	230	134	18

Uses	Vsts	Rvts
0	600	494
1	219	219
2	112	
3	30	
4	5	
5	4	
6		
7		
8		
9		

Dist'n of visits = 2
Mean no. of visits = 10, attempts per visit = 1.45distributed
Prop'n who renege, [type 1] = 0.1, who revisit, [type 2] = 0.3
Prop'n who substitute, [type 3] = 0.1, balance, [type 4] = 0.5
Max no. of revisits = 2

No.	EVt	EPrA	Type	No.	EVt	EPrA	Type
1	9	.25	4	51	10	.1	4
2	19	.45	4	52	10	.25	4
3	19	.1	4	53	10	.45	4
4	19	.25	4	54	9	.1	4
5	19	.45	4	55	9	.25	4
6	19	.1	2	56	9	.45	2
7	18	.25	2	57	9	.1	2
8	18	.45	2	58	9	.25	2
9	18	.1	1	59	8	.45	1
10	18	.25	3	60	8	.1	3
11	18	.45	4	61	8	.25	4
12	17	.1	4	62	8	.45	4
13	17	.25	4	63	8	.1	4
14	17	.45	4	64	8	.25	4
15	17	.1	4	65	7	.45	4
16	17	.25	2	66	7	.1	2
17	16	.45	2	67	7	.25	2
18	16	.1	2	68	7	.45	2
19	16	.25	1	69	7	.1	1
20	16	.45	3	70	6	.25	3
21	16	.1	4	71	6	.45	4
22	16	.25	4	72	6	.1	4
23	15	.45	4	73	6	.25	4
24	15	.1	4	74	6	.45	4
25	15	.25	4	75	5	.1	4
26	15	.45	2	76	5	.25	2
27	15	.1	2	77	5	.45	2
28	14	.25	2	78	5	.1	2
29	14	.45	1	79	5	.25	1
30	14	.1	3	80	4	.45	3
31	14	.25	4	81	4	.1	4
32	14	.45	4	82	4	.25	4
33	13	.1	4	83	4	.45	4
34	13	.25	4	84	4	.1	4
35	13	.45	4	85	4	.25	4
36	13	.1	2	86	3	.45	2
37	13	.25	2	87	3	.1	2
38	12	.45	2	88	3	.25	2
39	12	.1	1	89	3	.45	1
40	12	.25	3	90	3	.1	3
41	12	.45	4	91	2	.25	4
42	12	.1	4	92	2	.45	4
43	12	.25	4	93	2	.1	4
44	11	.45	4	94	2	.25	4
45	11	.1	4	95	2	.45	4
46	11	.25	2	96	1	.1	2
47	11	.45	2	97	1	.25	2
48	11	.1	2	98	1	.45	2
49	10	.25	1	99	1	.1	1
50	10	.45	3	100	1	.31	3

TABLE 4.3

Aggregate frequency distributions of use after three simulations of the use of the initial and reduced collections by 100 users with Poisson-type distribution of rate of visit and expected frequencies for the fitted negative binomial distribution.

	Simulated and expected numbers of users			
Number of	Initial		Reduced	
uses	Sim.	Exp.	Sim.	Exp.
0	12	15.5	29	31.2
1	18	21.6	33	33.8
2	27	24.0	39	32.1
3	32	24.4	27	29.1
4	23	23.6	30	25.8
5	28	22.3	24	22.5
6	17	20.5	19	19.5
7	22	18.7	14	16.7
8	20	16.8	15	14.2
9	16	14.9	9	12.2
10	8	13.2	4	10.2
11	11	11.6	6	8.6
12	6	10.2	10	7.3
13	10	8.9	8	6.1
14	5	7.7	3	5.1
15	6	6.7	3	25.6*
16	5	5.8	5	
17	5	5.0	5	
18	1	28.6*	5	
19	3		1	
20+	25		11	

*Expected frequencies less than 5 were pooled for the test of fit (see over).

TABLE 4.3 (continued)

Mean of the frequency distribution of simulated uses for the Poisson-type distribution of rates of visit and fitted parameters (k and p) and chi-squared test statistics for the negative binomial distribution.

	Initial collection	Reduced collection
Mean use	8.10	5.99
k	1.685	1.32
p	0.172	0.180
Chi-squared	22.56	11.69
No.of cells	19	16
P	0.1	0.5

Chi-squared test: expected frequencies less than 5 were pooled. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

reduced collection for Poisson-distributed rates of visit. Table 4.4 shows the results for the geometrically distributed rates of visit. Table 4.5 shows the expected frequencies of visit for the Poisson-type and geometric-type distributions.

The differences assigned to users as failure characteristics and as differential rates of attempt rendered their frequency distributions of use positively skewed even when their rates of visit were symmetrically (Poisson) distributed. Negative binomial distributions could be fitted to all the frequency distributions of use without test values of the chi-squared statistic exceeding the 90% level of confidence. In the case of the geometric distributions of visit this is not surprising; the distribution is itself negative binomial with the shape parameter set to unity. For the Poisson-type distribution of visit, the shape parameters of the fitted negative binomial distributions of use exceeded unity, even for the use of the reduced collection. Observing these distributions in reality, therefore, we could not mistake the influence of the underlying symmetrical distribution of visits. The value of the shape parameter certainly falls with a reduction in the size of collection (for the enlarged collection the fitted value was 2.15), but a much more drastic reduction would be necessary before a reversed J-shaped distribution resulted. For the geometric distribution of visit, the shape parameters remained more stable and indeed were identical for the initial and enlarged collections.

From the evidence of the simulation results, it seems reasonable to conclude that, unless the distribution of propensity among users (the product of all their rates and characteristics) is itself reversed J-shaped, competition among users is unlikely to produce such a frequency distribution of use at realistic levels of availability.

4.3 SUMMARY

The equilibrium in aggregate rates of use observed for the users of the short-loan collection (Section 3.5) was unlikely to have been the result of a constraint imposed by a shortage of useful material in the collection. In the following year, under largely unchanged conditions, the number of uses sustained by the collection rose by nearly 20%.

The simulation of use for a collection with similar or lower levels of availability appeared to show that reversed J-shaped frequency

TABLE 4.4

Aggregate frequency distributions of use after three simulations of the use of the initial and reduced collections by 100 users with geometric-type distribution of rate of visit and expected frequencies for the fitted negative binomial distribution.

Number of uses	Simulated and expected numbers of users Initial Sim.	Exp.	Reduced Sim.	Exp.
0	45	49.4	68	69.9
1	33	32.6	45	39.8
2	33	25.7	28	29.2
3	23	21.4	20	23.1
4	16	18.3	15	18.9
5	21	15.9	22	15.7
6	14	13.9	15	13.3
7	10	12.3	12	11.3
8	14	10.9	8	9.7
9	9	9.7	11	8.4
10	15	8.6	5	7.3
11	3	7.7	7	6.3
12	5	6.9	3	5.5
13	3	6.2	1	41.6*
14	2	5.6	5	
15	1	5.1	3	
16	4	49.8*	5	
17	5		3	
18	4		2	
19	3		3	
20+	37		19	

*Expected frequencies less than 5 were pooled for the test of fit (see over).

TABLE 4.4 (continued)

Mean of the frequency distribution of simulated uses for the geometric-type distribution of rates of visit and fitted parameters ($\underline{k}$ and $\underline{p}$) and chi-squared test statistics for the negative binomial distribution.

	Initial collection	Reduced collection
Mean use	8.13	5.71
$\underline{k}$	0.718	0.632
p	0.081	0.0997
Chi-squared	21.50	7.84
No.of cells	17	14
$\underline{P}$	0.1	0.75

Chi-squared test: expected frequencies less than 5 were pooled. $\underline{P}$ is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 4.5

Expected frequencies of visit for Poisson-type and geometric-type distributions of rate of visit.

Number of visits	Expected number of visitors Poisson-type	Geometric-type
0	0	0
1	0	10
2	0	9
3	1	8
4	1	7
5	4	6
6	7	6
7	9	6
8	11	4
9	12	5
10	13	4
11	11	3
12	11	3
13	7	3
14	5	3
15	4	2
16	2	2
17	1	2
18	1	1
19	0	2
20	0	1
21+	0	13*

*Namely: 21, 21, 22, 23, 24, 26, 27, 29, 31, 33, 34, 38, 44.

distributions of use would not be observed unless the users underlying rate of recourse to the collection was itself of that form.

It appears therefore reasonable to assume that patterns of use observed for users of the short-loan collection (Table 2.3 and Chapter 3) reflect the propensity of those users towards library use.

CHAPTER 5

FIT OF NEGATIVE BINOMIAL DISTRIBUTIONS TO LIBRARY USE DATA FROM THE LITERATURE

In Section 2.2.3, negative binomial distributions were shown to fit observed frequency distributions of use for the users of a short-loan collection. Although the fit to the data was reasonable, the model was unsatisfactory. The fitted parameters varied unpredictably with time, at least in the short term, thus robbing the model of its potential usefulness. In Chapter 3, the data were shown to exhibit some of the properties assumed in the negative binomial model and in Chapter 4 the observed distributions of use were judged to reflect real differences among the users.

In this chapter, the model is tested against sets of library-use data reported in the literature. Unfortunately, no examples could be found representing the use of the same collection over varying time periods: the test of the model against these new sets of data represents only a data-fitting exercise, therefore. On the assumption that the findings of Chapters 3 and 4 applied to the new sets of data, reasonable fits would confirm the negative binomial distribution as a useful approximation to frequency distributions observed for library users; but unless the context of the data was known, the doubts expressed in Section 1.3 (about the coherence of the record) and in Section 1.4 (about the integrity of the unit) would discourage generalisation from the fitted model.

5.1 THE USE OF LIBRARY DATA FROM THE LITERATURE

Even data fitting posed some problems with these new data. Author's estimates of the size of potential user populations could not be checked, and estimates of the numbers of potential users for particular collections (which were possibly smaller than for all collections) were not available. Data for the use of particular collections (such as withdrawals from a reserve collection) were not analysed, therefore, even though an accurate estimate of the proportion of zeros would not always have been critical to fit (81). In one case, for example, a poorer fit was achieved for the aggregate distribution for all types of use than for the use of a reserve collection known to be serving only a part of the potential user population. (It is possible that in this case (and perhaps in others) the aggregate record of use comprised sets of data with incompatible distributions which arose under differing conditions and for different sections of the overall population.)

For one set of data, a proportion of the withdrawals from the collection probably went unrecorded; for others, the importance of unrecorded use (within the library, say, or from subsidiary collections) remains unknown.

For convenience, most authors grouped together some of the frequencies of use, especially in the tails of their distributions. The data were always used as presented when the fits of negative binomial distributions were tested. But in order to estimate the parameters of the distribution by using the maximum-likelihood equation (Figure 2.4) individual frequencies had to be reconstructed. This was done in an arbitrary way, so that the reconstruction should not appear too smooth, but, of course, within the constraints of group totals or known statistics, such as the mean or standard deviation.

5.2 SETS OF DATA FROM THE LITERATURE

Sets of data for recorded library use by users of academic libraries were reported by Ritter (7), Maxted (25), Knapp (3), Clayton (8) and Schnaitter (24).

Ritter gives circulation totals for 468 students in a US liberal arts college for a period of nine weeks. The students were enrolled on various courses, especially in education. Frequencies were not grouped and 11 students, unlikely to use, were discounted by the author from the potential user population.

Maxted records borrowings over two successive terms from an unsupervised library in a UK hall of residence by 342 resident students. All disciplines within the university were represented but use was largely recreational. Even if all the residents were, by definition, potential borrowers, there must be doubt about the completeness of a voluntary record of borrowing.

Knapp records course-related reserve and general collection withdrawals from a US college library by 738 students during one semester. Before estimation of the negative binomial parameters, the distribution was reconstructed around a smooth curve sketched through the mean values of grouped frequencies.

Clayton presents data similar to that of Knapp for 545 students in another US college during one semester. Clayton acknowledges that some students would not have had items placed on reserve for them during the period of his study.

Schnaitter records issues from a main library to 3755 junior students at a US university during one semester. Issues of loans, short loans, journals and reserve material were recorded but six divisional branch libraries on the campus were not surveyed.

All these writers were concerned to assess differences among students before or after library use and to relate them to observed distributions of use. Although marked differences in amounts of use were always observed, no equally marked differences in purpose or benefit were found. In many ways, Wilson's criticism of user studies is exemplified (48). Even though, as students, these users share similar tasks and roles, a more sophisticated concept of information need and use would appear to be be required before the use of particular information channels could be explained.

5.3 FIT OF NEGATIVE BINOMIAL DISTRIBUTIONS

The frequency distributions of recorded library use taken from the literature are shown in Tables 5.1 to 5.3. Negative binomial distributions are fitted alongside. The parameters were estimated using the maximum likelihood equation set out in Figure 2.4. Frequencies are grouped for convenience, but fit was tested using as many individual frequencies as were presented by the authors.

The null hypothesis predicted that the observed frequency distributions were sampled from a negative-binomial distributed population with parameters estimated from the data. This hypothesis was tested using the chi-squared test of goodness of fit. The hypothesis was to be rejected for observed values of the chi-squared statistic exceeding the 20% level of significance for a chi-squared distribution with three less degrees of freedom than the number of cells into which the frequencies were divided. (The expected distribution depended upon two parameters estimated from the data as well as the total of the frequencies). A majority of the sets of data were required to have values of the test statistic outside the

TABLE 5.1

Frequency distributions and statistics of recorded library use from Ritter (7) and Maxted (25) with expected frequencies, parameters (k and p) and chi-squared test statistics for fitted negative binomial distributions.

Number of recorded uses	Observed (Obs) and expected (Exp) numbers of users			
	Ritter Obs.	Exp.	Maxted Obs.	Exp.
0	150	135.4	149	143.7
1	27	56.5	35	45.3
2	41	38.3	20	28.0
3	23	29.3	25	20.1
4	23	23.7	13	15.5
5-9	84	75.0	58	44.2
10-14	41	40.1	19	20.1
15-24	50	39.2	18	16.5
25-49	26	25.6	4	7.8
50+	3	4.9	1	0.8
Mean use	6.8		3.97	
Variance	84.1		45.1	
k		0.445		0.342
p		0.0614		0.0793
Chi-squared		48.7		20.9
No.of cells		27		16
P		0.003		0.08

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 5.2

Frequency distribution and statistics of recorded library use from Knapp (3) with expected frequencies, parameters (k and p) and chi-squared test statistics for the fitted negative binomial distribution.

Number of recorded uses	Numbers of users Observed	Expected
0	111	107.8
1	58	74.2
2	61	59.9
3	54	50.6
4	49	43.8
5-10	192	176.1
11-15	90	82.1
16-25	62	83.4
26+	61	60.1
Mean use	9.12	
Variance	120	
k		0.745
p		0.0756
Chi-squared		12.8
No.of cells		11
P		0.13

P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 5.3

Frequency distributions and statistics of recorded library use from Clayton (8) and Schnaitter (24) with expected frequencies, parameters (k and p) and chi-squared test statistics for fitted negative binomial distributions.

Number of recorded uses	Observed (Obs) and expected (Exp) numbers of users			
	Clayton Obs.	Exp.	Schnaitter Obs.	Exp.
0	48	34.6	1308	1266.4
1	28	31.8	298	421.1
2	20	29.5	291	271.3
3	23	27.6	209	203.0
4	22	25.8	168	162.4
5-9	102	106.4	527	515.1
10-14	73	77.4	349	286.1
15-24	114	97.9	306	304.2
25-49	94	89.7	226	248.2
50+	21	24.3	73	77.2
Mean use	15.5		7.57	
Variance	216		172	
k		0.976		0.348
p		0.0593		0.044
Chi-squared		26.8		82.8
No.of cells		23		30
P		0.15		0

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

critical region for the model to be accepted. The conservative significance level was adopted in order to give some protection against Type II errors, that is, an acceptance of the negative binomial model when an alternative might be preferable. The test was one-tailed: the critical region being assumed to lie only in the upper tail.

The results of the tests indicated a rejection of the negative binomial model for these sets of data. All the values of the test statistic fell within the critical region, although in two cases (Knapp and Clayton) with a level of significance between 10% and 20%.

Poor fit to the zero or early terms of the distributions seemed to be responsible for the rejection. It is possible that potential user populations were not well estimated, although this could not explain the general lack of smoothness in the early part of most distributions which was perhaps due to the aggregation of separate distributions for the use of different components of the collections. If this were the case, the distributions would be difficult to model with as few as two parameters.

5.4 SUMMARY

Negative binomial distributions were fitted to five frequency distributions of recorded library use reported in the literature but gave poor results. It is possible that the data came from incomplete surveys or were mixtures of incompatible records of use.

CHAPTER 6

ALTERNATIVE MODELS FOR FREQUENCY DISTRIBUTIONS OF LIBRARY USE

On the evidence of Section 2.2.3 and Chapter 5, the negative binomial probability distribution has to be rejected as a model for frequency distributions of library use. Other probability distributions were therefore reviewed. Of these, the lognormal distribution, Neyman's Type A distribution and the generalized inverse Gaussian-Poisson distribution were considered candidates capable of fitting the data so far presented, as well as being justifiable *a priori* as models, albeit superficial models, of the process observed.

The fitted distribution needed to be reversed J-shaped with zeros in any proportion (although usually less than one half) and a standard deviation in the range between one and two times the mean. Table 6.1 shows the behaviour of the negative binomial distribution within this range. Values of k need to be rather less than unity to achieve the skewness of the observed distributions so that much freedom with the scale parameter q is lost. Nonetheless, the general shape of the observed distributions was approximated.

It was possible to reject all three candidate distributions on the evidence of fit to one or more sets of data.

6.1 LOGNORMAL DISTRIBUTION

The lognormal distribution is described by Aitchison and Brown (50). It is generated when the logarithm of the variate is normally distributed. Now a normal variate comprises the sum of a constant and an error value. If the constant is zero and the error is itself the sum of many errors, then the distribution is similarly normal. For the sampling distribution of the means (and therefore the sums) of large random samples from a population of errors (however distributed) will be approximately normally distributed. But if the errors are combined in some multiplicative process, as a product rather than a sum, then a lognormal distribution will result. The logarithm of the variate will comprise the sum of the logarithms of the error values.

This is an attractive model for the process leading to recorded library use, for a number of the variables have already been characterised as

TABLE 6.1

Values of the negative binomial shape parameter, k, and proportion of zeros, p(0), for various values of the scale parameter, $q = (1 - p)$, when the standard deviation, s, is equal to or greater than the mean, m.

	$s=m$		$s=\sqrt{2}m$		$s=2m$	
q	k	$p(0)$	k	$p(0)$	k	$p(0)$
0.25	4.0	0.32	2.0	0.56	1.0	0.75
0.5	2.0	0.25	1.0	0.5	0.5	0.71
0.75	1.33	0.16	0.67	0.4	0.33	0.63
0.99	1.01	0.01	0.5	0.1	0.25	0.32

Calculated from: mean = kq/p; variance = kq/p^2; proportion of zeros = p^k.

operating serially and thus multiplicatively. In Section 4.2, the number of recorded uses simulated for each user depended, among other factors, upon the number of visits multiplied by the number of attempts per visit multiplied by the rate of success per attempt and so on. On the other hand, the number of factors may not be large enough to achieve the normality in the model. The process, too, may branch into parallel paths, such as the revisits or substitutions of the simulation.

6.1.1 Fit of the lognormal distribution

For each of the sets of data presented in Chapter 5, the cumulative proportions of users were plotted on a probability scale against the logarithm of the maximum number of uses. The resulting plots were variously concave to the probability axis (lognormally-distributed data would produce a straight line), suggesting that the logarithmic transformation was overcorrecting for the positive skewness of the data. It is true that some curvature in the plot of the lowest proportions would result for a discrete form of the lognormal distribution (for example, the Poisson-lognormal distribution (82) described by Cassie). The model would certainly need to be used in such a discrete form. But the observed curvature persisted throughout the plot for most sets of data.

A test of the skewness in relation to the height of the observed and expected distributions was therefore performed. The ratio of the excess of kurtosis to the square of skewness was calculated for each set of data and compared to values for the lognormal distribution.

It is customary to define the skewness of a distribution (its departure from symmetry) as the ratio of the third moment about the mean to the cube of the standard deviation (see Hines (83:302-304), for example). Kurtosis is defined as the ratio of the fourth moment about the mean to the square of the variance. The kurtosis of a normal distribution takes the value 3 and is constant. Hence the excess or coefficient of kurtosis is defined as the value of the kurtosis reduced by three. A ratio of the excess of kurtosis to the square of the skewness thus describes a particular relationship between symmetry and peakedness for a distribution. It was used to compare lognormal distributions against the observed distributions of Chaper 5.

Table 6.2 shows the means and central moments for these data together with calculated values of the ratio. By trial and error, lognormal means and standard deviations were found which represented the transformed means and variances of the observed distributions. These are shown in Table 6.3. Values of the skewness and excess of kurtosis for the lognormal distribution were calculated from formulae presented by Aitchison and Brown (50:8) and are shown in Table 6.4 for the range of standard deviations represented in Table 6.3. The ratios of the excess of kurtosis to the square of skewness shown in Table 6.4 are clearly consistently larger than the ratios calculated for the data (Table 6.2).

By comparison, the ratio for the gamma distribution is closer to observed values, being constant at 1.5. On this evidence, therefore, fits of a discrete lognormal distribution to the observed distributions are likely to be poorer than those obtained for the discrete gamma distribution, the negative binomial distribution. Values of the ratio for the negative binomial distribution are shown in Table 6.5.

6.2 NEYMAN'S TYPE A DISTRIBUTION

Neyman's Type A or contagious distribution mixes two Poisson distributions and shows similarities to forms of the negative binomial distribution (77,84,85). Applied to library use, the model can be described, following Froggatt (76), as representing use occurring in short, infrequent spells; the parameters of the two Poisson distributions determine the mean number of spells per time period and the mean number of uses per spell for all users. In Figure 6.1, the distribution function is set out together with the methods of estimating the two parameters and the expected proportion of zeros. Using the means and variances of the five sets of data described in Chapter 5 (Table 6.2) the expected proportion of zeros was calculated and compared to the observed proportion. The result is shown in Table 6.6.

Clearly the Neyman Type A distribution would give a poor fit to the observed frequency distributions. The proportion of zeros will lie within a narrow range for any combination of mean and variance. If the mean and standard deviation are related in the manner shown in Table 6.1 and as required by the library use data, then, as Table 6.7 shows, the zero term is closely defined also.

TABLE 6.2

Means and central moments for five frequency distributions of recorded library use and ratio of the excess of kurtosis to the square of skewness.

Origin of data*	Mean	μ_2	μ_3	μ_4	Ratio
Ritter	6.80	84	1547	55647	1.21
Maxted	3.97	45	1029	40867	1.48
Knapp	9.12	130	3434	151608	1.11
Clayton	15.5	221	5004	278642	1.17
Schnaitter	7.57	176	8708	733742	1.49

*see Chapter 5.

μ_r represents the rth moment about the mean: for grouped distributions the values are approximate. Ratio calculated from:

$$[\mu_4/(\mu_2^2)-3]/[\mu_3^2/\mu_2^3]$$

TABLE 6.3

Means and standard deviations (SD) for lognormal distributions fitted to frequency distributions of recorded library use.

Origin of data*	Lognormal mean	Lognormal SD	Expected mean	Expected variance
Ritter	1.40	1.02	6.82	85.2
Maxted	0.71	1.16	3.99	45.1
Knapp	1.74	0.97	9.12	130
Clayton	2.41	0.81	15.5	222
Schnaitter	1.31	1.19	7.52	177

*see Chapter 5.

Expected mean = $\varepsilon^{\mu+\sigma^2/2}$; expected variance = $\varepsilon^{2\mu+\sigma^2}(\varepsilon^{\sigma^2}-1)$ where μ is the lognormal mean and σ is the lognormal SD.

TABLE 6.4

Skewness, excess of kurtosis and ratio of the excess of kurtosis to the square of skewness for lognormal distributions with standard deviations, σ

σ	Skewness	Excess of kurtosis	Ratio
0.5	1.75	5.90	1.93
1.0	6.18	111	2.90
2.0	414	9220557	53.7

Data from Aitchison and Brown (50:8)
Skewness = $c^3 + 3c$; excess of kurtosis = $c^8 + 6c^6 + 15c^4 + 16c^2$ where $c = \sqrt{(e^{\sigma^2} - 1)}$.

TABLE 6.5

Values of the ratio of the excess of kurtosis to the square of skewness for negative binomial distributions with scale parameters $q = (1 - p)$.

q	Ratio*	q	Ratio*
0.001	1.0	0.6	1.47
0.1	1.17	0.7	1.48
0.2	1.28	0.8	1.49
0.3	1.36	0.9	1.5
0.4	1.41	0.99	1.5
0.5	1.44		

*Kurtosis = $3 + (1 + 4q + q^2)/kq$; square of skewness = $(1 + q)^2/kq$.
Source: Williamson and Bretherton (37).

FIGURE 6.1

Neyman's Type A probability distribution

Users have a mean number of spells per time period, m', and a mean number of uses per spell, m''. The probability of observing u uses, $p(u)$, is summed over all numbers of spells, s. Thus,

$$p(u) = \sum_{s} \frac{\varepsilon^{-sm''}(sm'')^{u}}{u!} \, \frac{\varepsilon^{-m'} m'^{s}}{s!}, \quad s = 0,1,2,\ldots;\ u = 0,1,2,\ldots$$

$$= \frac{\varepsilon^{-m'} m''^{u}}{u!} \sum_{s} \frac{(m'\varepsilon^{-m''})^{s} s^{u}}{s!},$$

where $0^0 = 1$ and $e = 2.718$.

The parameters, m' and m'', can be expressed in terms of the mean, $m = m'm''$, and the variance, $s^2 = (1 + m'')m'm''$, so that,

$m'' = (s^2 - m)/m,$

and $m' = m/m'' = m^2/(s^2 - m)$.

The proportion of zeros, $p(0)$, may be written,

$$p(0) = \varepsilon^{-m'}\left[1 + m'\varepsilon^{-m''} + \frac{(m'\varepsilon^{-m''})^2}{2!} + \ldots\right]$$

$$= \varepsilon^{m'(\varepsilon^{-m''} - 1)}.$$

The expressions for m' and m'' may then be sustituted.

TABLE 6.6

Observed proportions of users recording zero uses in five frequency distributions of recorded library use and expected proportions of zeros for Neyman's Type A distribution.

Origin of data*	Observed proportion of zeros	Expected proportion of zeros
Ritter	0.32	0.55
Knapp	0.15	0.50
Maxted	0.44	0.68
Clayton	0.088	0.31
Schnaitter	0.35	0.71

*see Chapter 5.
The expected proportion of zeros is calculated from the expression given in Figure 6.1 by substituting observed means and variances for each set of data.

TABLE 6.7

Proportion of zeros, $p(0)$, in Neyman's Type A distribution for various values of the mean, m, when the standard deviation, s, is equal to or greater than the mean.

Mean	Proportion of zeros, $p(0)$, when $s = m$	$s = \sqrt{2m}$	$s = 2m$
4	0.28	0.57	0.77
7	0.31	0.58	0.77
10	0.33	0.59	0.77
13	0.34	0.59	0.77
16	0.34	0.60	0.78

$p(0)$ calculated from $e^{-m'}(e^{-m''} - 1)$ where $m'' = (s^2 - m)/m$ and $m' = m/m''$.

6.3 THREE-PARAMETER MODELS

The two-parameter alternative models considered above did not seem likely to improve upon the fit of the negative binomial distribution. It became necessary therefore to consider three-parameter distributions in order to find a frequency curve with a more adaptable shape. In this way it was hoped to account for anomalies in the negative binomial fits by the introduction of an extra variable. Nonetheless, the step was taken with some reluctance. The nature and scale of the new variable would have to be assigned in a largely arbitrary fashion. No extra data were available to estimate the new parameter; the existing data disclosed only central tendency and range for the observed distributions. While it is true that under these conditions a model can in any case only be judged by fit (the probability of the occurrence of an observed set of data in random sampling from the model population) and while it is true that a model with improved fit was indeed required to satisfy the test of Section 5.3, it was nonetheless to be expected that within the sets of data requiring to be modelled there were not only the effects of unknown variables but also random fluctuations and, very probably, sampling errors. Since errors and genuine diversity were both unknowns, there was a real risk of adjusting the model to fit errors as well as diversity. Nonetheless, the availability of five different sets of observations gave some protection against modelling spurious effects, and it seemed reasonable therefore to fit one extra parameter.

It would have been possible to assume that all the populations represented by the data were to some degree heterogeneous. Fits could have been improved, therefore, by partitioning populations into two or more subsets to be modelled with separate parameters. An example of such an exercise is given by Brownsey (86). There was evidence, however, that heterogeneity did not always cause poor fit. Data collected by the author for four months library use by each of three classes (years) of undergraduate pharmacists were indeed well fitted by negative binomial distributions with $P > 0.25$ for chi-squared tests. But the fit of a negative binomial distribution to the frequency distribution of use for the aggregate of all three classes was even better ($P > 0.75$).

For most sets of data, the natural subdivision of the population would have been into classes, or at least disciplines, of which there would usually have been many in each population. No information about the sizes of classes or disciplines was available to the author except with respect to one set of data. Even in this case, it would have been beyond the ability of the author to construct an algorithm to fit many negative binomial distributions simultaneously. Again, therefore, it seemed reasonable to adopt the most parsimonious modelling approach by fitting only a single extra parameter.

6.4 GENERALIZED INVERSE GAUSSIAN-POISSON DISTRIBUTION

Sichel (87) has employed a discrete probability distribution with up to three parameters to model many sets of bibliometric data. In its most general form, this Generalised Inverse Gaussian-Poisson (GIGP) distribution is tedious to calculate, although a recurrence relationship based on two previous terms in the series is available (Figure 6.2). The negative binomial distribution is that special case of the GIGP distribution in which one of the parameters, a, is set to zero and another, k, is greater than zero.

When all three parameters are free, calculation is simplified if $k = -0.5, 0, 0.5, 1,\ldots$ According to Sichel (75:196), the shape of the distribution changes only gradually with the change in k. It was for these half-integer values of k, therefore, that the distribution was tested against the data.

6.4.1 $k = -0.5$

Sichel (75) has fitted the GIGP distribution with this parameter to sets of purchasing data for which the negative binomial distribution gave poor fits. The observed distributions had modes greater than zero, or were reversed J-shaped with large zero terms.

Table 6.8 shows the result of estimating the GIGP parameter a from the values of the zero and ones terms in Clayton's data for various permissible values of the parameter q. It is clear that the two estimates are incompatible.

6.4.2 $k = 0, 0.5, 1,\ldots$

For other values of k, use was made of the recurrence relationship shown in Figure 6.2. If the fit of the GIGP distribution was to improve

FIGURE 6.2

Generalized inverse Gaussian-Poisson distribution

The probability of r events, $p(r)$, is,

$$p(r) = \frac{(1-q)^{k/2}}{K_k(a[1-q]^{\frac{1}{2}})} \frac{(aq/2)^r}{r!} K_{r+k}(a),$$

where a, k and q are the parameters of the distribution, $a > 0$, $0 < k < \infty$, $0 \leq q \leq 1$, and $K_v(z)$ is the modified Bessel function of the second kind of order v with argument z. The parameter k is invariant with time while a and q change predictably and tend to their upper limits.

A recurrence relationship links successive terms of the distribution,

$$p(r) = \left(\frac{r+k-1}{r}\right) q\, p(r-1) + \frac{a^2 q^2}{4r(r-1)} p(r-2).$$

For $k = -0.5$, the start-up probabilities,

$$p(0) = e^{-a(1-(1-q)^{\frac{1}{2}})},$$

and $p(1) = p(0)aq/2$,

are easily calculated.

TABLE 6.8

Estimates of parameter a of the GIGP distribution which satisfy the observed proportions of zeros and ones reported by Clayton (8) for various values of the parameter q.

	Estimates of parameter a:	
q	For observed proportion of zeros	For observed proportion of ones
0.1	47	12.0
0.2	23	5.8
0.3	15	3.9
0.4	11	2.9
0.5	8.3	2.3
0.6	6.6	1.9
0.7	5.4	1.7
0.8	4.4	1.5
0.9	3.6	1.3

$k = -0.5$. Estimate of a calculated from $p(0) = e^{-a(1-(1-q)^{1/2})}$ and $p(1) = p(0)aq/2$ where $p(0) = 0.088$ and $p(1) = 0.583.p(0)$

on that of the negative binomial distribution, it was in the earlier terms of the series that a closer fit was required. In particular, the GIGP would need to reproduce the observed 'shelving' where adjacent terms took similar values instead of declining smoothly (Table 6.9). This shelf may have been easier to model by assuming heterogeneous subsets within the population of potential users, but this assumption was not thought justified until simpler models assuming homogeneous populations had been discredited.

On the assumption that the proportions of users making one and two uses were to be equal (the 'shelf'), an expression for the parameter *a* in terms of the other variables was derived from the recurrence relationship as shown in Figure 6.3. Using the observed relative frequencies of zero and one uses from two sets of data as start-up probabilities in the recurrence relationship, the distribution was then graduated for trial values of *k* and *q* until best fit to the terms shown in Figure 6.4 were obtained. The optimal values of *a*, *k* and *q* were then used in the probability function of Figure 6.2 in order to calculate expected proportions of zeros and ones.

The table in Figure 6.4 shows values calculated from the data of Knapp and Schnaitter. The correspondence to observed values is poor in three cases out of four and the GIGP was therefore rejected as a model.

6.5 MODIFIED NEGATIVE BINOMIAL DISTRIBUTION

Observed ratios of the excess of kurtosis to the square of skewness (Table 6.5) indicate a gamma-like distribution for the sets of data from the literature. Ratios of this order would also indicate a beta distribution, however. A Poisson-beta distribution (which would have three parameters) was therefore considered worthy of future investigation if a modification of the negative binomial (Poisson-gamma) distribution failed to produce good fit.

6.5.1 Modification of the negative binomial distribution

In considering the lognormal distribution (Section 6.1), it was noted that library uses accrue as the product of several variables, numbers of visits to the collection, attempts per visit, and so on. Following Froggatt's application (76), the Neyman Type A distribution would also

TABLE 6.9

Shelving (departure from smooth monotonic decline) in the early terms of frequency distributions of recorded library use.

Number of uses	Numbers of users Ritter*	Maxted*	Knapp*	Clayton*	Schnaitter*
1	27		58		298
2	41	20	61	20	291
3	23	25	54	23	
4	23			22	

* See Chapter 5.

FIGURE 6.3

Calculation of parameters for the generalized inverse Gaussian-Poisson distribution which would give fit to data with 'shelf' at $p(1)$ and $p(2)$.

From the recurrence relationship in Figure 6.2, we have,

$$p(2) = q\left(\frac{1+k}{2}\right)p(1) + \left(\frac{a^2q^2}{8}\right)p(0).$$

If $p(1) = p(2)$ and $p(0) = cp(1)$, where c is greater than unity and is calculated from an observed distribution, then,

$$c\left(\frac{a^2q^2}{8}\right) + q\left(\frac{1+k}{2}\right) = 1$$

and thus,

$$a = \sqrt{\frac{8-4q(1+k)}{cq^2}}.$$

For $k = 0, 0.5, 1, \ldots$ and $0 < q < 1$, if $p(0)$ and $p(1)$ are supplied from an observed distribution, an expected distribution of best fit may be graduated by trial and error from the recurrence relationship of Figure 6.2.

FIGURE 6.4

Fit of generalized inverse Gaussian-Poisson distribution to data from Knapp (3) and Schnaitter (24)

Expected numbers of users were calculated for trial values of k and q in the ranges $k = 0, 0.5, 1, \ldots$ and $0 < q < 1$ using $a = \sqrt{[(8 - 4q(1 + k))/cq^2]}$ and $p(0)$, $p(1)$ and $c = p(0)/p(1)$ supplied from the data. Best fit was obtained for the values of k and q shown. These values were inserted in to the probability function of Figure 6.2 to provide the estimates of $p(0)$ and $p(1)$ shown at the foot of the table.

Observed (Obs.) and expected (Exp.) numbers of users, observed and estimated proportions of zeros and ones, and fitted values of the parameters $\underline{k}$ and $\underline{q}$.

Number of uses	Knapp Obs.	Exp.	Schnaitter Obs.	Exp.
0	111		1308	
1	58		298	
2	61	60.0	291	295.0
3	54	54.1	209	205.9
4	49	48.3	168	158.5
5	40	42.8	126	127.9
-	-	-	-	-
18	8	8.3	31	32.6
$\underline{p}(0)$	0.15		0.35	
$\underline{p}(1)$	0.08		0.08	
$\underline{k}$		1.0		0
$\underline{q}$		0.88		0.99
$\underline{p}(0)$		0.06		0.18
$\underline{p}(1)$		0.07		0.13

reproduce such a serial process. For, the probability of a given number of uses would depend not only upon the probabilities of that number of uses occurring in all possible numbers of spells, but also upon the probability associated with each possible number of spells. It is therefore the sum of the products of these two probabilities over all numbers of spells.

The negative binomial distribution was accordingly modified from a Poisson-gamma distribution to a Neyman-gamma distribution. The extra sophistication accorded to the model was modest enough: potential users shared only a single common rate of use per spell. Nonetheless, the modified model could be adjusted to increase the zero term of the distribution without significantly altering the shape of the tail. For, depending upon the value assigned to the extra parameter, some individuals with few spells, as well as all those with none at all, would record zero uses.

In Section 3.4, it was noted that if a unit of time of one week was employed then the observed distributions of amounts of use per time period were reasonably well approximated by the negative binomial distribution. If the unit of time were one day, however, the model would grossly underestimate the observed range of amounts of use. The hypothesis that use occurs in spells of undefined length within any time period resolves this difficulty. The amount of use per spell is then independent of the unit of time chosen for the analysis: it is the expected number of spells which is dependent.

Clearly, the new model still represents only a little of the diversity assumed in the simulation model discussed in Section 4.2. The limited amount of flexibility derived from the extra parameter is applied, however, in a way which intuitively seems to correspond to reality.

6.5.2 Modified negative binomial distribution

The probability function for the modified negative binomial distribution is shown in Figure 6.5. The gamma variate, m, which is distributed with probability density as shown in Figure 2.2, represents now a mean number of spells per time period. Actual numbers of spells are distributed about this mean with Poisson probabilities. Uses are recorded only in these spells and occur for all users at a mean rate of a per spell. Thus the actual numbers of uses recorded by each user in each

FIGURE 6.5

Modified negative binomial distribution

Let the events defined in Figure 2.2 be spells (Section 6.2). Observed integer numbers of events occur only in these spells and are Poisson distributed about a mean rate, a, per spell. For all numbers, s, of spells, the proportion of observations in which r events are expected is then

$$p(r) = \sum_{s=0}^{\infty} \frac{\varepsilon^{-as}\,(as)^r}{r!}\, p(s)$$

and therefore from Figure 2.4

$$p(r) = \frac{p^k}{r!} \sum (as)^r \, \varepsilon^{-as} \binom{k+s-1}{k-1} q^s .$$

The sum of the probabilities approaches unity for any valid combination of parameter values. It can be shown (88) that the mean of r is akq/p and that the variance is $akq(a + p)/p^2$. Using the direct scale parameter, $b = p/(1 - p) = p/q$, we have, therefore

DISTRIBUTION	MEAN	VARIANCE
Gamma	$\frac{k}{b}$	$\frac{k}{b^2}$
Negative binomial	$\frac{k}{b}$	$\frac{k}{b^2} + \frac{k}{b}$
Modified negative binomial	$a\frac{k}{b}$	$a^2\frac{k}{b^2} + a^2\frac{k}{b} + a\frac{k}{b}$.

In the application to recorded library use, the number of users recording r uses, $f(r)$, out of a total of N potential users is,

$$f(r) = N\frac{p^k}{r!} \sum_{s} \binom{k+s-1}{k-1} \varepsilon^{-as} (as)^r q^s, \quad s = 0,1,2,\ldots$$

time period are Poisson distributed about a mean which is the product of the actual number of spells and the parameter value, *a*.

6.5.2.1 Fit of the modified negative binomial distribution

The frequency distributions of recorded library use described in Chapter 5 are shown again in Tables 6.10 to 6.12. Modified negative binomial distributions are fitted alongside. Parameters k and q were fitted by trial and error, and for each combination, a value of a was chosen so as to reproduce the sample mean. The parameter values which gave best fit to the proportions of zeros and ones were then finally adjusted to give a minimum chi-squared value when up to 13 terms of the observed and expected distributions were compared.

As before, chi-squared test statistics were calculated in order to assess the goodness of fit of the full distribution. Ungrouped frequencies were used where available. The null hypothesis was deemed rejected if chi-squared test statistics exceeded the value for the 20% level of significance of the chi-squared distribution having four less degrees of freedom than the number of cells into which the frequencies were distributed. The fit is improved compared to the original model (Section 5.3); only two of the five results now indicate rejection. For one of these, for Schnaitter's data, the fit was improved when the population was partitioned into sexes (Table 6.13). Frequency distributions for each of the three Pharmacy classes (Tables 6.14 and 6.15) were not fitted so well as with the original model. Although the chi-squared statistics were no worse (Table 6.16), the loss of one degree of freedom (for the extra parameter) reduced the probability of their values being exceeded in random sampling.

6.5.3 Conclusion

With the extra parameter, the modified model does cope better with large differences between proportions of zeros and ones. In some cases it does go a little way towards reproducing the 'shelves' described earlier (Table 6.9). These shelves are evident even in the frequency distributions for the classes of pharmacists and they are therefore not necessarily caused by heterogeneity of discipline. It may be that, on the evidence from Schnaitter's data, differences between the sexes are also

TABLE 6.10

Frequency distributions and statistics of recorded library use from Ritter (7) and Maxted (25) with expected frequencies, parameters (*k*, *p* and *a*) and chi-squared test statistics for fitted modified negative binomial distributions.

Number of recorded uses	Observed (Obs) and expected (Exp) numbers of users			
	Ritter Obs.	Exp.	Maxted Obs.	Exp.
0	150	151.1	149	145.3
1	27	30.5	35	31.8
2	41	33.1	20	28.0
3	23	29.3	25	22.1
4	23	24.9	13	17.5
5-9	84	82.9	58	51.0
10-14	41	45.4	19	22.6
15-24	50	42.9	18	17.0
25-49	26	25.1	4	6.7
50+	3	2.8	1	0
Mean use	6.8		3.97	
Variance	84.1		45.1	
k		0.60		0.47
p		0.126		0.113
a		1.63		1.08
Chi-squared		30.4		13.5
No.of cells		26		16
P		0.1		0.3

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. *P* is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 6.11

Frequency distribution and statistics of recorded library use from Knapp (3) with expected frequencies, parameters (*k*, *p* and *a*) and chi-squared test statistics for the fitted modified negative binomial distribution.

Number of recorded uses	Numbers of users Observed	Expected
0	111	108.7
1	58	60.0
2	61	55.7
3	54	49.6
4	49	44.3
5-10	192	186.5
11-15	90	88.7
16-25	62	87.9
26+	60	55.6
Mean use	9.12	
Variance	120	
k		0.91
p		0.0654
a		0.70
Chi-squared		10.0
No.of cells		11
P		0.2

P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 6.12

Frequency distributions and statistics of recorded library use from Clayton (8) and Schnaitter (24) with expected frequencies, parameters (k, p and a) and chi-squared test statistics for fitted modified negative binomial distributions.

Number of recorded uses	Observed (Obs) and expected (Exp) numbers of users			
	Clayton Obs.	Exp.	Schnaitter Obs.	Exp.
0	48	48.0	1308	1285.3
1	28	24.6	298	301.4
2	20	26.4	291	260.7
3	23	25.4	209	209.6
4	22	24.2	168	172.3
5-9	102	103.3	527	563.2
10-14	73	77.6	349	316.9
15-24	114	100.1	306	331.3
25-49	94	91.5	226	251.3
50+	21	23.9	73	63.0
Mean use	15.5		7.57	
Variance	216		172	
k		1.08		0.423
p		0.0749		0.0499
a		1.16		0.941
Chi-squared		17.9		46.4
No.of cells		23		29
P		0.5		0.005

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 6.13

Frequency distributions and statistics of recorded library use from Schnaitter (24) with expected frequencies, parameters (*k*, *p* and *a*) and chi-squared test statistics for fitted modified negative binomial distributions.

Number of recorded uses	Observed (Obs) and expected (Exp) numbers of users			
	Women Obs.	Exp.	Men Obs.	Exp.
0	392	394.6	916	915.5
1	111	111.4	187	186.6
2	112	102.5	179	157.7
3	92	86.4	116	123.0
4	79	73.7	89	98.2
5-9	242	259.4	284	299.4
10-14	185	161.0	164	150.4
15-24	177	185.8	131	136.4
25-49	147	169.5	79	79.2
50+	61	53.7	12	10.6
Mean use	11.0		5.0	
k		0.53		0.38
p		0.047		0.067
a		1.025		0.95
Chi-squared		35.6		19.8
No.of cells		29		28
P		0.1		0.7

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. *P* is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 6.14

Frequency distributions and statistics of recorded library use for pharmacy undergraduates with expected frequencies, parameters (k, p and a) and chi-squared test statistics for fitted modified negative binomial distributions.

Number of recorded uses	Observed (Obs) and expected (Exp) numbers of users			
	Class A		Class B	
	Obs.	Exp.	Obs.	Exp.
0	31	30.7	18	18.0
1	8	8.0	9	9.0
2	7	6.6	6	7.4
3	4	5.1	8	6.0
4	6	4.0	3	4.9
5-9	14	11.8	22	15.1
10-14	2	5.3	2	6.7
15-24	3	3.9	1	4.5
25+	2	1.6	4	1.4
Mean use	4.1		5.0	
Variance	58		56	
k		0.48		0.77
p		0.09		0.07
a		0.85		0.49
Chi-squared		4.7		4.0
No.of cells		8		8
P		0.3		0.4

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 6.15

Frequency distribution and statistics of recorded library use for pharmacy undergraduates with expected frequencies, parameters (k, p and a) and chi-squared test statistics for fitted modified negative binomial distribution.

Number of recorded uses	Numbers of users Class C Observed	Expected
0	13	12.8
1	9	8.8
2	8	7.4
3	8	6.3
4	4	5.5
5-9	13	18.2
10-14	12	9.3
15-24	11	7.5
25+	1	3.2
Mean use	6.8	
Variance	47	
k		0.88
p		0.028
a		0.22
Chi-squared		9.7
No.of cells		11
P		0.2

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 6.16

Chi-squared statistic, degrees of freedom and level of significance for the fit of negative binomial (NB) and modified negative binomial (MNB) distributions to frequency distributions of library use for three classes of undergraduate pharmacists.

	Distribution	Class A	Class B	Class C
Chi-squared statistic	NB	6.0	5.7	9.5
	MNB	4.7	4.0	9.7
Degrees of freedom	NB	5	6	8
	MNB	4	4	7
Level of significance	NB	0.32	0.47	0.31
	MNB	0.34	0.42	0.22

implicated. On the other hand, there may be more dynamic causes. For example, using Nosik's model of consumer-purchasing behaviour (89), perhaps users diverge by being selectively converted into repeat users or latent (infrequent) users or lapsed users at critical incidents in their affairs, in Nosik's case by entering a trial or experimentalist phase of library use. For some potential users, the mere contemplation of library use may possibly be critical enough, of course.

Although further elaboration of the model would probably improve some of the fits, it was thought that a reasonable compromise between approximation and simplicity had been reached with this single modification. Clearly, an efficient method of estimating the parameters needs to be sought for the modified distribution to be serviceable.

6.5.4 Distribution of recorded uses per spell

The processes underlying library use are unavoidably complex. In the reduction of their description to a simple generalisation of the outcomes much accuracy is undoubtedly lost. It is possible, however, that the modified model does represent an extra observable aspect of the real process.

The distribution of spells in the model is, as before, inferred from the general shape of the observed frequency distribution of use and is assumed to be gamma. If uses constitute the events of the model, then the spells in which these events take place can only be purposeful visits to the library or periods during these visits when attempts to use library material are made. In the model, numbers of events per spell are distributed with Poisson probabilities about a constant mean which is assigned as the third parameter, a. Data collected from three UK academic libraries suggested that such a distribution roughly approximated the type of distribution actually observed, although a number of assumptions were necessarily made in analysing the data.

6.5.4.1 Data on number of uses per spell

Records of use or borrowing by random samples of 30 to 40 students were examined. Data for three academic libraries were collected; the period of observation was in one case four months, in the others nine months. Uses which appeared to have been made in the same visit or spell were identified for each user from the dates and times when use was recorded. Usually each period of activity was short and clearly defined, but in a very few cases arbitrary decisions based on circumstantial evidence had to be made in order to assign uses to spells.

6.5.4.2 Frequency distributions of numbers of uses per spell

The frequency distributions of numbers of uses per spell approximated geometric probability distributions for the majority of individual users in each sample (c.f. Morse, 10:31, Section 1.1), although for small individual totals of use these fits were trivial. The mean rates of use per spell varied among users as shown in Table 6.17. There appeared little correlation between numbers of spells observed and rate of use per spell (Figure 6.6). When frequencies were pooled for those users from each sample with similar rates of use per spell, all the resulting frequency distributions were approximated by geometric distributions. The aggregate frequency distributions for all the users in each sample were also approximated by geometric distributions (Table 6.18). Some anomalies resulted from abnormal amounts of use per spell prior to vacations; the proportion of fours in Sample C of Table 6.18 is the clearest example. The mean number of uses per spell for users who subsequently retained their borrowings for the vacation was 2.40 compared to 1.76 uses per spell overall for the users in this sample.

6.5.4.3 Zero uses

It was assumed that some spells occurred in which use was attempted but without success. Naturally, such spells were not observed. In the model, the proportion of these failed visits or spells is estimated as e^{-a}, where a is the third parameter and $e = 2.718$. If the proportion occurring in reality were similar, then to this extent the modification to the model reflects reality. Unfortunately, no data were available for the rate of complete failure in purposeful visits to academic libraries, either for individuals or in aggregate.

Typical proportions of failures for single known-item searches are between 20% and 50% for academic libraries (79,80,90-94). Proportions for browsing and subject searches may well be smaller because of the

TABLE 6.17

Mean rates of use per spell for students sampled from three academic libraries.

	Numbers of students with mean use per spell from				
Sample	1.0 to 1.49	1.5 to 1.99	2.0 to 2.49	2.5 to 2.99	3.0+
A	9	16	7	2	0
B	28	8	1	1	1
C	7	20	5	3	0

A: 34 arts or social science students; all collections; 9 months; 1 to 56 spells

B: 39 science or technology students; short-loan books; 4 months; 1 to 48 spells

C: 35 social science students; ordinary loan books; 9 months; 2 to 65 spells.

FIGURE 6.6

Numbers of uses per spell plotted against numbers of spells observed for the users in Sample A of Table 6.17.

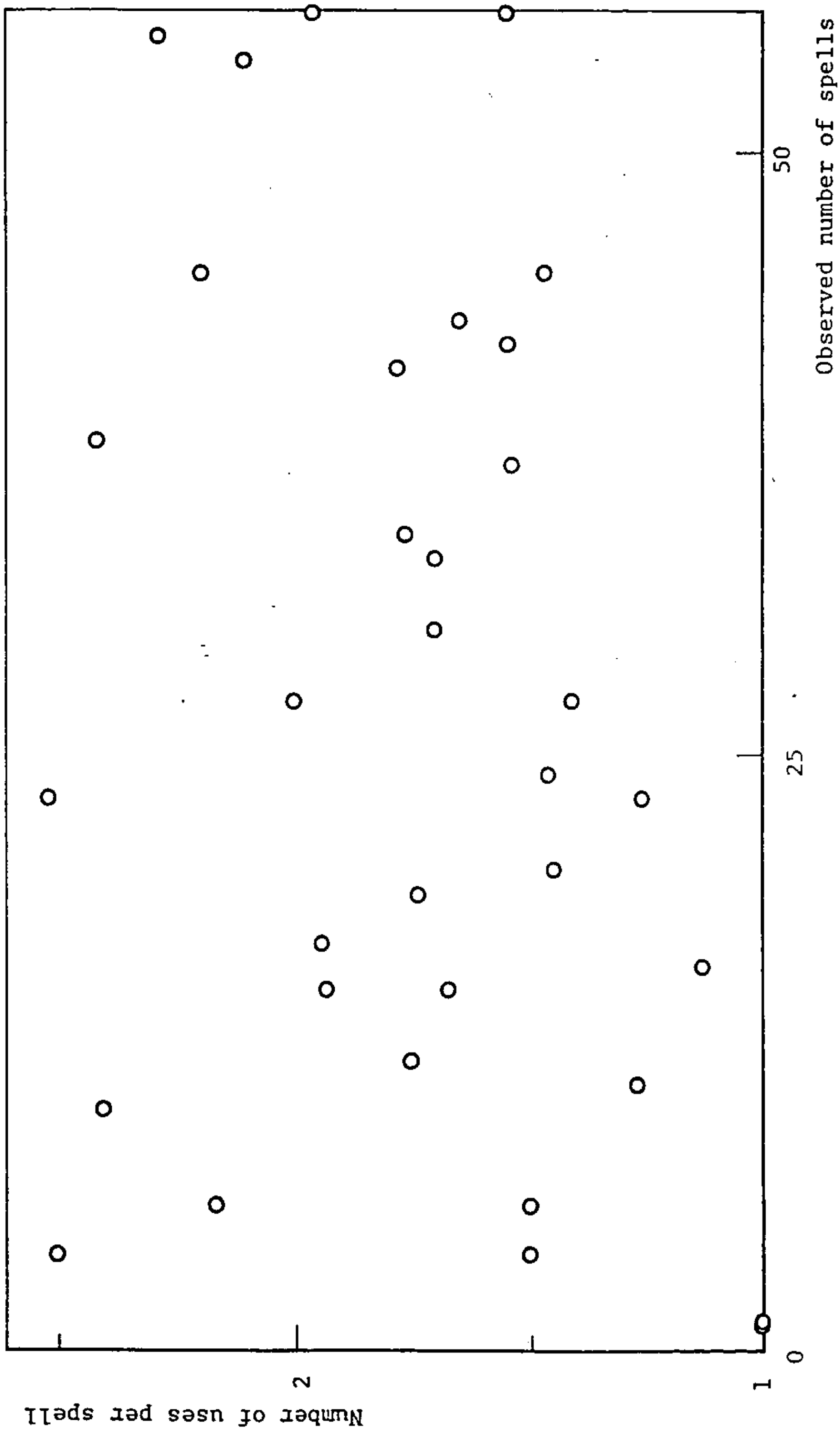


TABLE 6.18

Observed (Obs.) frequency distributions of amount of use per spell for students sampled from three academic libraries and expected (Exp.) numbers of uses for the geometric distribution.

Uses per spell	Numbers of spells Sample A		Sample B		Sample C	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
1	481	493.1	263	259.3	546	550.6
2	234	219.9	71	74.1	230	236.8
3	97	98.1	17	21.2	103	101.8
4	43	43.7	9	6.0	61	43.8
5	20	19.5	2	1.7	20	18.8
6	11	8.7	1	0.5	5	8.1
7+	4	7.0	-	0.2	1	6.1
Non-users	16		11		15	
Mean	1.81		1.40		1.76	
p*		0.554		0.714		0.570
P**		0.5		0.25		0.025

*Parameter of geometric distribution: relative frequency of r uses per spell is p(1 - p) where p is the reciprocal of the mean.
**Approximate probability of observing chi-squared test value in random sampling. Samples as in Table 6.17.

substitutability of the material. Even failed known-item searches might result in the recorded use of other relevant material found during the search. Overall, therefore, for the students in the samples, the proportion of attempts at use which failed was likely to be less than 50%. And, if more than one attempt was made, on average, in each spell, then the proportion of failed or zero-use spells would have been less again. In a specialised resource centre for undergraduate surveyors, less than 10% of visitors who returned questionnaires reported complete failure on that day's visit (95). It seems likely, however, that only a proportion of all visitors actually participated in the survey.

Proportions of zeros of 0.2, 0.3, 0.4 and 0.5 were added to the aggregate frequency distributions of Table 6.18. One value in each case yielded a distribution which was roughly approximated by a Poisson distribution of similar mean (Table 6.19). To this extent therefore, it was thought that the model might be found to agree with what was observed. As we have noted, however, individual mean rates of use per spell varied considerably. The behaviour of individual users was much more complex, therefore, than the single rate of use in the model would allow.

6.5.4.4 Constraints on the value of *a*

Some of the fitted values of *a* shown in Tables 6.10 to 6.15 have values less than 0.7. Following the argument above, it seems unlikely that such low mean rates of use per spell would occur in reality. For, if the actual distributions of uses per spell were roughly Poisson, the zero terms would then be greater than 0.5. A proportion of zeros this large would, it seems, be unusual in an observed distribution of numbers of uses per spell or visit. Applying a lower limit to *a* in estimating the parameters of the modified negative binomial distribution does not sacrifice goodness of fit, however. Values of *p* and *a* can be adjusted quite freely without altering the value of *k* to any large degree.

A further constraint on the value of *a* would be imposed if the same population was observed over different time periods. For, it would be reasonable to expect that the users' aggregate mean rate of use per spell would change little over time, except in the period just before vacations.

TABLE 6.19

Observed (Obs.) frequencies of use per spell from Table 6.18 expressed as relative frequencies with arbitrary zero term compared to Poisson (Exp.) distributions with similar mean.

Uses per spell	Sample A		Sample B		Sample C	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	0.3*	0.301	0.5*	0.497	0.3*	0.301
1	0.378	0.361	0.353	0.348	0.396	0.361
2	0.184	0.217	0.095	0.122	0.167	0.217
3	0.076	0.087	0.023	0.028	0.075	0.087
4	0.033	0.026	0.012	0.005	0.044	0.026
5+	0.029	0.008	0.017	0.001	0.018	0.008
Mean	1.24		0.68		1.23	
$\underline{m}$**		1.2		0.7		1.2

*Arbitrarily chosen zero term. **Mean of Poisson distribution.

If the Poisson distribution introduced into the modified model is considered an approximation to observed numbers of uses per visit, therefore, the estimation of its parameter, a, is simplified. The possible range can be estimated independently or constrained by averaging over different time periods. The argument from failure rates would suggest that values of a around or greater than unity should be common. There is evidence of such a rate from one polytechnic (79).

6.6 SUMMARY

Alternative two-parameter models could not be shown to approximate observed frequency distributions of library use more successfully than the negative binomial distribution. Accordingly, two three-parameter models were tested, one being a generalisation, the other a modification of the negative binomial distribution. The modified negative binomial distribution represented a gamma mixture of Neyman Type A (rather than Poisson) distributions. The parameters of the modified distribution were crudely estimated by minimizing a partial chi-squared statistic. The fit was acceptable in most cases and the model was adopted as a reasonable compromise between accuracy and convenience.

On the analogy of Froggatt's application of the Neyman Type A distribution, the third parameter could, it seemed, be taken to represent the mean rate of use by users during each spell of library use. If this correspondence were shown to exist, then independent evidence could be used to supplement goodness of fit in estimating the third parameter.

CHAPTER 7

APPLICATION OF THE MODIFIED NEGATIVE BINOMIAL DISTRIBUTION

The utility of the model adopted in Section 6.5 is now investigated. First, the discrepancies in the fits of the negative binomial distributions are reviewed, then the improvement obtained from the modified distribution is assessed, and finally, the potential usefulness of the model in quantifying information about the behaviour of library users is considered. In the following two chapters, two applications are described.

7.1 FAILURE OF THE NEGATIVE BINOMIAL DISTRIBUTION

When tested for fit to frequency distributions of library use, the negative binomial distribution was found to fail in two respects. In Section 5.3, the fit for five sets of published data was rejected because all values of the test statistic fell within the critical region. In Section 2.2.3, negative binomial distributions showed reasonable fit for two out of three sets of related data, but the model was rejected because the shape parameter, k, increased progressively with time.

7.1.1 Expected and observed behaviour of the negative binomial shape parameter, k.

If data are available only for a single period of time, then both parameters of the negative binomial distribution may be freely estimated in order to obtain best fit. A constraint is placed upon the permissible value of the shape parameter, k, however, if data relating to the same population are available for more than one time period. For, the model requires that individuals in the population have unchanging mean rates of use over time. Thus k must remain constant over time; otherwise a third parameter would need to be estimated to govern its change.

Figure 7.1 shows the probability function of the gamma distribution. The effect of varying the two parameters is shown in the examples sketched in Figure 7.2 where the shape and extent of the frequency curves may be compared. The shape parameter, k, determines the relative distribution of the population over the range of the abscissa. If k remains constant, then the area under the curve between points on the

FIGURE 7.1

Gamma probability distribution

The integral

$$\Gamma(k) = \int_0^\infty y^{k-1}\,\varepsilon^{-y}\,dy, \quad y > 0$$

is called the gamma function. By substitution

$$\Gamma(k) = \int_0^\infty (bx)^{k-1}\,\varepsilon^{-bx}\,d(bx)$$

and

$$\frac{\Gamma(k)}{b} = \int_0^\infty (bx)^{k-1}\,\varepsilon^{-bx}\,dx.$$

Rearranging, we have

$$\frac{b^k}{\Gamma(k)}\int_0^\infty x^{k-1}\,\varepsilon^{-bx}\,dx = 1.$$

Hence, the integral of the gamma variate

$$p(x) = \frac{b^k}{\Gamma(k)}\,x^{k-1}\,\varepsilon^{-bx}, \quad x \geqslant 0$$

is unity. There are two parameters, an inverse scale parameter, b, and a shape parameter, k, each greater than zero. Curves for some values of b and k are sketched in Figure 7.2.

As shown in Figure 2.2, the mean and variance of x are

$E[X] = k/b$

and $\mathrm{var}[X] = k/b^2$.

FIGURE 7.2

Probability curves for gamma distributions

Probability curves for gamma distributions with the following parameters are shown overleaf:

i)	$k = 1$	$b = 0.5$
	$k = 3$	$b = 0.5$
	$k = 5$	$b = 0.5$
ii)	$k = 0.5$	$b = 0.5$
	$k = 0.5$	$b = 0.25$
	$k = 0.5$	$b = 0.125$
iii)	$k = 2$	$b = 0.5$
	$k = 2$	$b = 0.25$
	$k = 2$	$b = 0.125$

Figure 7.2 (i) Graph of the gamma probability density function $p(x) = (b^k x^{k-1} e^{-bx})/\Gamma(k)$ for $b = 0.5$ and $k = 1$; $k = 3$; and $k = 5$ as indicated.

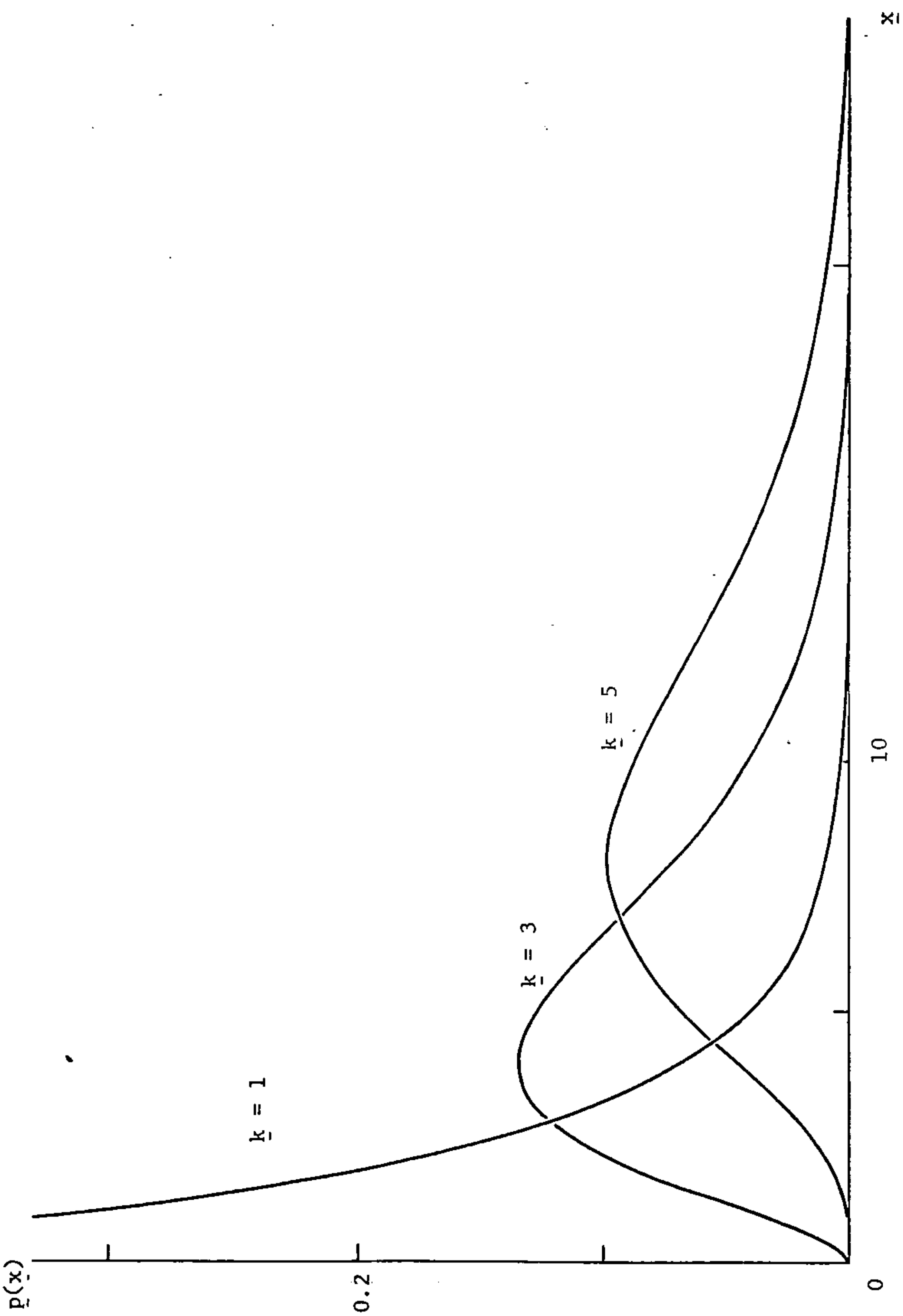


Figure 7.2 (ii) Graph of the gamma probability density function $p(x) = (b^k x^{k-1} e^{-bx})/\Gamma(k)$ for $k = 0.5$ and $b = 0.5$; $b = 0.25$; and $b = 0.125$ as indicated.

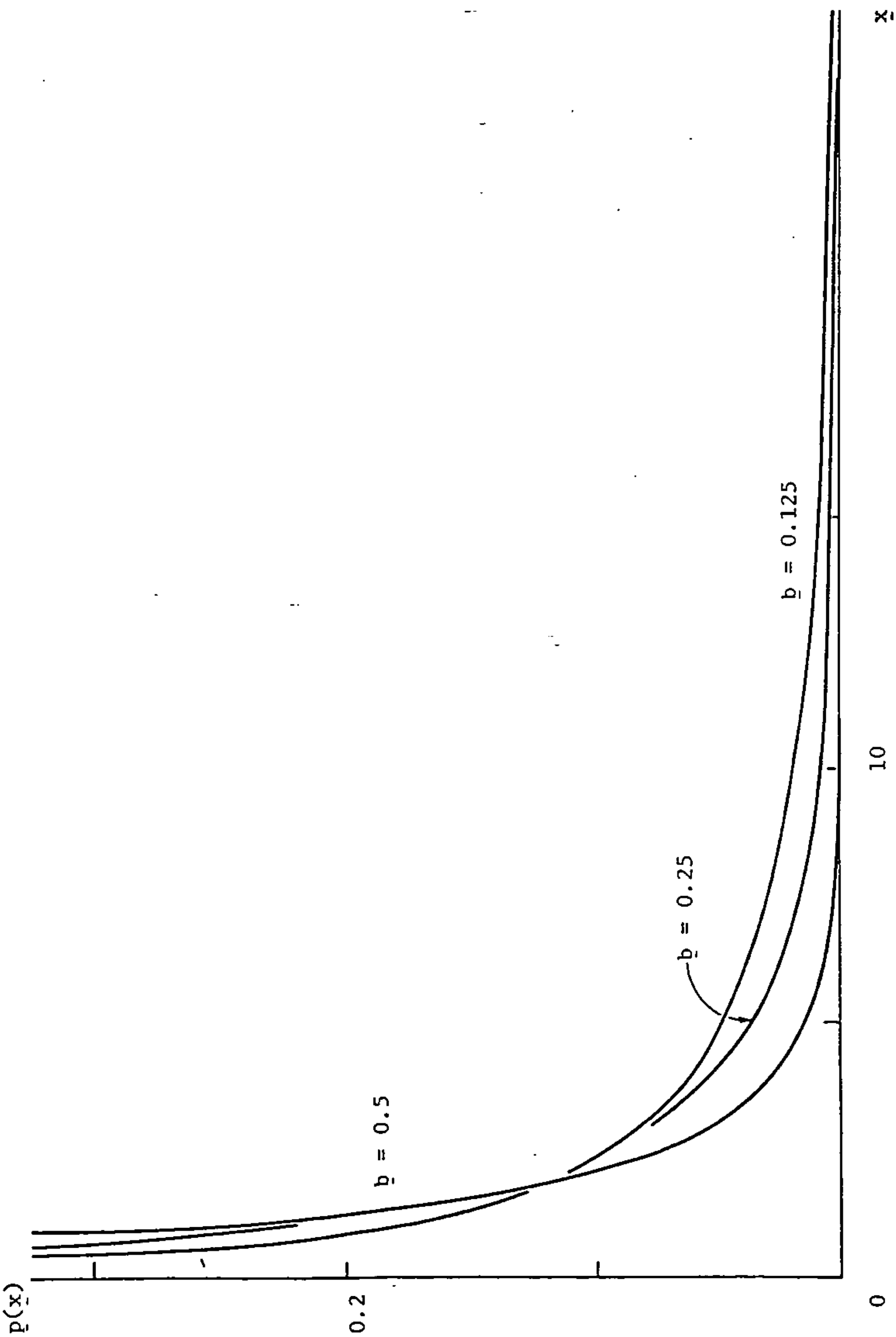
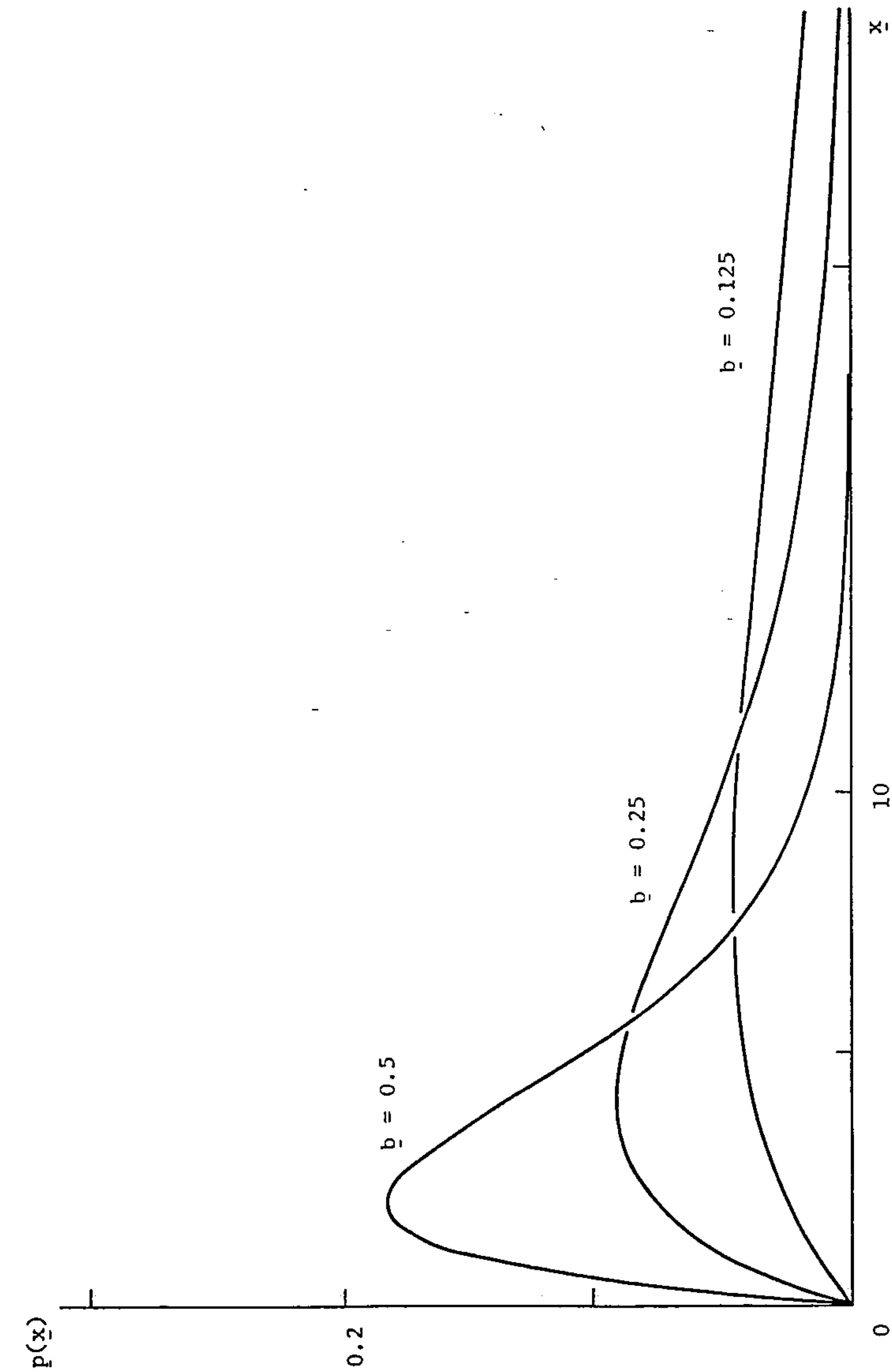


Figure 7.2 (iii) Graph of the gamma probability density function $p(x) = (b^k x^{k-1} e^{-bx})/\Gamma(k)$ for $k = 2$ and $b = 0.5$; $b = 0.25$; and $b = 0.125$ as indicated.



abscissa will also remain constant so long as the points retain their relationship to the mean. A doubling of the mean doubles also the width of any given interval on the abscissa; but the heights of the ordinates within this interval, being functions of their abscissa values, are reduced by half. The area beneath the curve thus remains the same (Figure 7.3).

7.1.2 Possible causes of failure of negative binomial distribution.

Some hypotheses offering possible explanations for the failure of the negative binomial fits have already been noted.

7.1.2.1 Multiple populations.

It is possible that the observed frequency distributions of use were made up of not one but two or more dissimilar distributions representing use by separate groups within the potential user population. In Section 6.5.2.1, for example, it was found that Schnaitter's data could be better fitted by using modified negative binomial distributions for separate male and female populations instead of the combined population (Tables 6.12 and 6.13). It seems likely that any large student population will contain separate homogeneous groups of users (whether classified, for example, by sex; discipline, faculty or class; type or level of course; or country of origin). There seems no reliable way, however, of detecting the criterion of classification or estimating the size of the groups without collecting extra information. Even then, an improvement in fit is not guaranteed: negative binomial distributions give better fit for an aggregate population comprising three classes of undergraduate pharmacists than for each class separately (Section 6.5.3).

Without extra information, the fitting of additional parameters (such as subpopulation sizes) may yield equivocal results when modelling observed distributions. Although a population may be partitioned by trial and error, and distributions fitted which, in sum, minimize a chi-squared test statistic, it is likely that a range of similar fits will be demonstrated for different combinations of parameter values and population sizes. To illustrate the limits of such data fitting, a further attempt was made to model Schnaitter's data and, in particular, the distribution of use recorded by the female members of the population. Despite the improvement in fit achieved for these data, first by using

FIGURE 7.3

Areas under the probability curves of gamma distributions

Consider two gamma distributions of differing means but with shape parameters of equal value. We determine the ratio of the heights of two ordinates, one for each distribution, which are proportionately the same distance from their means.

For the distribution with mean k/b, we have

$$p(x) = \frac{b^k}{\Gamma(k)} x^{k-1} \varepsilon^{-bx}$$

and for the distribution with a mean c times as large, $c(k/b)$

$$p(cx) = \frac{\left(\frac{b}{c}\right)^k}{\Gamma(k)} (cx)^{k-1} \varepsilon^{-\frac{bcx}{c}}, \quad c > 0.$$

The ratio between the height, $p(cx)$, of the ordinate raised at cx on the abscissa of the second distribution and the height, $p(x)$, of the ordinate at x on the abscissa of the first distribution is then

$$\frac{p_{cx}(cx)}{p_x(x)} = \frac{\dfrac{\left(\frac{b}{c}\right)^k}{\Gamma(k)} (cx)^{k-1} \varepsilon^{-\frac{bcx}{c}}}{\dfrac{b^k}{\Gamma(k)} x^{k-1} \varepsilon^{-bx}}$$

$$= \frac{c^{k-1}}{c^k} = \frac{1}{c}.$$

Thus when intervals on the abscissa are increased c-fold, ordinates are reduced to $1/c$th. The area beneath the probability curve for corresponding intervals on the abscissa is therefore constant.

the modified negative binomial distribution (Table 6.12), and then by partitioning the population into sexes (Table 6.13), the goodness of fit statistic for the women's distribution remained within the critical region.

Both the men's and the women's frequency distributions exhibit a 'shelf' (Section 6.4.2) at the twos term. In order to partition the women's population, it was assumed that, at this shelf, a hump representing the mode of a subpopulation interrupted the expected smooth, monotonic decline of the reversed J-shaped frequency curve. This mode gave an indication of the value of the shape parameter, k, which was needed to fit the distribution for the subpopulation. In the unscaled gamma distribution, the value on the abscissa below the mode is equal to $(k - 1)$ when $k \geq 1$; and for the scaled distribution (with inverse scale parameter, b) it is equal to $(k - 1)/b$ (Figure 7.4; see also Figure 7.2).

Trial and error fits were made for different values of the parameters of the distributions and for different sizes of the two populations. Three parameters and one population size needed to be assigned; the fourth parameter value and the second population size could then be calculated from the overall mean and overall population size respectively. Aggregate distributions which reproduced the observed zeros and ones terms to ±5 were compared for fit to the observed distribution by calculating the sum of the squared deviations from the twos to the thirteens terms.

The smallest sum was obtained, not for a subpopulation with a mode greater than zero as suggested above, but for the combination of two reversed J-shaped distributions. Table 7.1 shows the observed frequency distribution and the expected distribution made up of two components. Sums of squared deviations for some other subpopulation sizes were only slightly greater, however, so that other selections would have been possible. Table 7.2 shows one example with a larger subpopulation and a mode greater than that initially envisaged.

Using an irregularity (the 'shelf') in the shape of an observed frequency distribution to suggest the shapes of two component distributions does not, therefore, in this case, result in a well-defined best fit. Nor, indeed, does the best fit improve on the fit of a single negative binomial distribution. Table 7.3 shows the results of chi-squared tests carried out for the distributions of Tables 7.1 and 7.2 and for negative binomial and modified negative binomial distributions. As before, the data were tested in the original groupings presented by

FIGURE 7.4

Modal value of the gamma distribution

Consider the gamma distribution function

$$p(y) = \frac{1}{\Gamma(k)}\, y^{k-1} \varepsilon^{-y}.$$

Differentiating with respect to *y*, we have

$$\frac{d\,p(y)}{dy} = \frac{y^{k-1}}{\Gamma(k)} \frac{d(\varepsilon^{-y})}{dy} + \varepsilon^{-y} \frac{d\left(\frac{y^{k-1}}{\Gamma(k)}\right)}{dy}$$

$$= \frac{y^{k-1}\,(-\varepsilon^{-y})}{\Gamma(k)} + \frac{\varepsilon^{-y}\,(k-1)\,y^{k-2}}{\Gamma(k)}.$$

A single mode exists ~~(Figure 7.2)~~ at which

$$\frac{d\,p(y)}{dy} = 0$$

and therefore

$$y = k-1$$

Similarly, if

$$p(y) = \frac{b^k}{\Gamma(k)}\, y^{k-1}\, \varepsilon^{-by},$$

the mode occurs at

$$y = (k-1)/b. \qquad \text{(Figure 7.2)}$$

TABLE 7.1

Frequency distribution and mean of recorded library use by 1598 women from Schnaitter (24) with population sizes (n), parameters (k and p) and expected individual and aggregate frequencies for two fitted negative binomial distributions.

Number of recorded uses	Observed and expected numbers of users			
	Observed	Expected	Components n = 1310	n = 288
0	392	395.8	142.0	253.8
1	111	115.9	110.8	5.1
2	112	97.3	94.7	2.6
3	92	85.0	83.3	1.7
4	79	75.6	74.3	1.3
5-9	242	281.5	277.5	4.0
10-14	185	178.3	176.1	2.2
15-24	177	194.8	192.0	2.8
25-50	147	141.2	137.6	3.6
51+	61	32.6	21.7	10.9
Mean use	11.0			
k			0.84	0.020
p			0.071	0.0018

TABLE 7.2

Frequency distribution and mean of recorded library use by 1598 women from Schnaitter (24) with population sizes (n), parameters (k and p) and expected individual and aggregate frequencies for two fitted negative binomial distributions.

Number of recorded uses	Observed and expected numbers of users		Components	
	Observed	Expected	n = 1023	n = 575
0	392	390.4	382.9	7.5
1	111	113.4	97.3	16.1
2	112	83.7	59.9	23.8
3	92	74.0	44.1	29.9
4	79	69.2	35.1	34.1
5-9	242	294.7	112.9	181.8
10-14	185	202.2	66.6	135.6
15-24	177	195.2	79.3	115.9
25-50	147	117.8	87.7	30.1
51+	61	57.4	57.2	0.2
Mean use	11.0			
k			0.26	2.7
p			0.0228	0.2

TABLE 7.3

Chi-squared test statistics for the fit of four distributions (see below) to the frequency distribution from Schnaitter (24) for 1598 women, with partial sums of the statistic for the first five (0 to 4) and the last two (51 to 75, and 76+) cells.

Distribution:	NB	MNB	T1	T2
Chi-squared statistic	52.42	37.68	53.78	53.74
Degrees of freedom	27	26	24	24
Number of cells	30	30	29	29
Chi-squared, 0 to 4	17.38	1.64	3.19	15.39
Chi-squared, 51+	3.82	6.85	25.83	0.86

NB: negative binomial distribution; MNB: modified negative binomial distribution; T1: composite distribution of Table 7.1; T2: composite distribution of Table 7.2.

Schnaitter. Only the modified negative binomial distribution appears to give reasonable fit at both extremes of the observed distribution.

7.1.2.2 Effects of competition among users.

In Section 4.2, the effect of competition among users upon observed distributions of use was considered. The availability of material would, it seems, need to have been much lower than has been commonly reported by academic libraries for differences in the competitive success of users to have concealed differences in their rates of recourse. Thus it does not appear that the poor fits of the negative binomial distribution could be improved by incorporating some kind of differential failure rate into the model. (The modified negative binomial distribution (Section 6.5) could be regarded as applying an undifferentiated failure rate to all users.) It is possible , however, that the effects of competition may have influenced to a minor extent the final form of the distributions of use, since in simulation severe competition certainly produced distributions which were well modelled by negative binomial distributions (Tables 4.3 and 4.4).

In the longer term, of course, competition may alter the distribution of rates of recourse to the collection. But these kinds of dynamic variation in behaviour, generated by individual reactions to the library environment, would be difficult to incorporate into a simple model.

7.1.2.3 Excessive variance in observed distributions of the amount of weekly use.

In Section 3.2, it was seen that a large minority of the users of the short-loan collection used more erratically than the Poisson distribution in the negative binomial model would represent. It is possible:

i) that their rates of use varied over time or; that they were potential users of the collection only intermittently (phenomena not accommodated in the model, see Section 7.1.1), or;

ii) that their amounts of use per time period were more variable than if Poisson distributed for similar mean.

In order to account for this excessive variation and to explain a changing value of the shape parameter, k, over time, it was suggested in Section 2.2.3 that the short-loan collection users might have been recruited to the potential user population only gradually. An extra parameter, the potential user population size, was therefore estimated (Table 2.4). It became clear, however, that the estimated value of k changed as use cumulated from any starting point, not just from the beginning of the academic year. The poor fit of the negative binomial model would have to be explained, therefore, by intermittent membership or changing rates of use rather than by slow recruitment to the potential user population.

In Section 6.5.1 the negative binomial distribution was modified to provide a slightly more realistic (and successful) representation of the process by which use was thought to be generated. This model allowed amounts of use per time period to vary more widely than if Poisson distributed. The modification entailed replacing the gamma mixture of Poisson distributions (the negative binomial distribution) with a gamma mixture of Neyman Type A distributions, the Neyman Type A distribution being itself a Poisson mixture of Poisson distributions. If m' and m'' are respectively the means of the mixing and mixed Poisson distributions, then the aggregate mean is $m'm''$ and the variance is $m'm''(1 + m'')$. The variance is thus larger than the mean, as required by the observations summarised in Section 3.4.

Increasing the variance of the expected distribution of amounts of use should also improve the stability over time of the shape parameter, k. Data for the use of the short-loan collection illustrate the failure of the negative binomial distribution in this respect. In Table 7.4, a negative binomial distribution (Distribution a) is fitted to the distribution of use for the ten-week period of observation. But the value of the shape parameter used is that estimated for the three-week period. Alongside (Distribution b) is shown the negative binomial distribution fitted with both parameters freely estimated. In each case, the maximum likelihood equation was used to estimate the shape parameter, k. Distribution a) results when the value of k is kept constant over time as required by the model. It has greater variance than Distribution b) and clearly has a greater proportion of zeros and a longer tail. It seems reasonable to conclude that the negative binomial distribution fitted to the three-week data incorporates an adjustment to its gamma parameter to accommodate the excess variance in observed weekly amounts of use noted in Section 3.4. When the distribution is extrapolated to ten weeks, this parameter produces a distribution whose variance exceeds the required ten-week value because it governs the larger proportion of the total variance of the distribution. A more accurate model, therefore, would

TABLE 7.4

Negative binomial distributions fitted to the frequency distribution of Table 2.3 (ten weeks), a) using the shape parameter, k, fitted to the three-week distribution, and b) using the maximum likelihood equation to fit both parameters freely.

Number of uses	Expected numbers of users a)	b)
0	650.9	479.5
1	185.3	210.2
2	113.8	142.0
3	82.4	107.2
4	64.2	85.2
5	52.1	69.7
6	43.4	58.1
7	36.9	49.1
8	31.8	42.0
9	27.7	36.1
10	24.3	31.3
11	21.5	27.2
12	19.1	23.8
13	17.0	20.8
14	15.3	18.3
15	13.7	16.1
16	12.4	14.3
17	11.2	12.6
18	10.2	11.2
19	9.2	9.9
20	8.4	8.8
21	7.7	7.9
22	7.0	7.0
23	6.4	6.3
24	5.9	5.6
25+	72.2	49.8

TABLE 7.4 (continued)

Parameters for negative binomial distributions fitted to the ten-week frequency distribution of recorded use for 1550 users (Table 2.3) and statistics estimated from these parameters.

	a)	b)
Shape parameter, $\underset{\sim}{k}$	0.302	0.48
Scale parameter, $\underset{\sim}{p}$	0.0564	0.0868
Estimated mean	5.05	5.05
Estimated variance	90	58

incorporate extra variance into the distribution of amounts of use per time period rather than into the distribution of rates of use over the population. In Figure 7.5, the behaviour of the fitted negative binomial distribution is compared to such a model. It is clear that an extrapolated negative binomial distribution will always overestimate variance if the observed distribution of amounts of use per time period is more variable than in the Poisson law.

7.2 EXTRAPOLATION OF THE MODIFIED NEGATIVE BINOMIAL DISTRIBUTION

The modified negative binomial distribution incorporates a modest increase in variance in the distribution of amounts of use per time period and should therefore perform better when extrapolated.

Table 7.5 shows the results of fitting the modified negative binomial distribution to observed frequency distributions of use for the short-loan collection (Section 2.2.3). Parameters were estimated as described in Section 6.5.2.1 for three and ten-week periods and fit was then tested. As with the negative binomial distributions, the fit for ten weeks was poor. In fact it was noticeably poorer, although in this case, some part of the cause may be in the method of fitting. With three parameters, estimated by minimizing the total chi-squared statistic for the first 15 terms only, there was room for some miscalculation with respect to the tail of the distribution.

It is clear, however, that the estimated values of the shape parameter, k, differ substantially, and it would be difficult to attribute this difference to the method of fitting. For, although the complementary effect of the two scale parameters, p and a, may confuse their estimation (Section 6.5.4.4), the value of k required to approximate an observed distribution is less equivocal. When estimated parameter values for the six-week period are set between the values for three and ten weeks (Table 7.6), it seems clear that, despite the modification to the model, a progressive change with time in k still occurs, albeit to a smaller degree. The value of the third parameter, a, could also be expected to remain roughly constant with time (if it represented mean use per visit). It not only changes, however, but also resides within a lower range than expected (Section 6.5.4.4).

Distributions for a short period of observation can be fitted with a greater range of parameter values than distributions for longer periods. Nonetheless, parameter values estimated for the ten-week period were not

FIGURE 7.5

A model for the distribution of recorded library use with an increased variance in the distribution of amounts of use per time period.

The gamma distribution with shape parameter, k, and scale parameter, b, has a mean, k/b, and variance, k/b^2. The negative binomial distribution with similar parameters has a mean k/b and variance $\frac{k}{b^2} + \frac{k}{b} = \frac{k}{b^2}(1+b)$.

The variance due to the gamma distribution is thus larger than that due to the Poisson distribution.

In fitting a given observed mean and variance, we substitute an alternative distribution for the negative binomial distribution. It is a gamma mixture of distributions whose variances are greater than their means by a constant ratio, c. This distribution has a mean

$$\frac{uk}{ub}$$

and variance

$$\frac{uk}{(ub)^2} + \frac{cuk}{ub} = \frac{uk}{(ub)^2}(1 + ubc),$$

where $u = 1/(1 + b - bc)$.

To extrapolate the negative binomial and alternative distributions, the scale parameters, b and ub, are divided by t, the ratio of the required time period to the original time period for which the parameters were fitted. The variance of the negative binomial distribution then becomes

$$\frac{k}{\left(\frac{b}{t}\right)^2}\left(1+\frac{b}{t}\right)$$

and the variance of the alternative distribution becomes

$$\frac{uk}{\left(\frac{ub}{t}\right)^2}\left(1 + \frac{ubc}{t}\right).$$

Hence, the ratio of the variance of the negative binomial distribution to the variance of the alternative distribution is

$$\frac{\dfrac{k}{\left(\frac{b}{t}\right)^2}\left(1+\dfrac{b}{t}\right)}{\dfrac{uk}{\left(\frac{ub}{t}\right)^2}\left(1+\dfrac{ubc}{t}\right)} = \frac{b+t}{t+b(c+t-ct)}$$

By definition, c is greater than unity. If t is greater than unity, then the ratio is also greater than unity. Thus, the variance of the extrapolated negative binomial distribution will always exceed that of the extrapolated alternative distribution.

Negative binomial parameter values which are extrapolated from parameter values estimated for a shorter time period will be smaller than parameter values freely estimated. For, if the ratio of the variance to the mean becomes excessive with extrapolation, then the correction of this ratio requires an increase in the value of b. To preserve the value of the mean k must then also be increased.

Let the ratio

$$\frac{\dfrac{k}{\left(\frac{b}{t}\right)^2}+\dfrac{k}{\left(\frac{b}{t}\right)}}{\dfrac{k}{\left(\frac{b}{t}\right)}} = \frac{t}{b}+1$$

be reduced by increasing b to, say, xb, where $x > 1$. To preserve the value of the mean, k must also be increased, to xk. The mean then remains at

$$\frac{xk}{\left(\frac{xb}{t}\right)} = \frac{k}{\left(\frac{b}{t}\right)} = t\left(\frac{k}{b}\right).$$

FIGURE 7.5 (continued)

Thus, in Table 7.4, the freely estimated and extrapolated values of the parameters k and b are related by the same coefficient, 1.59; that is, x = 1.59.

TABLE 7.5

Frequency distributions and statistics for recorded use of the short-loan collection for three and ten weeks with expected frequencies, parameters (k, p and a) and chi-squared test statistics for fitted modified negative binomial distributions.

Number of recorded uses	Observed (Obs) and expected (Exp) numbers of users			
	Three weeks		Ten weeks	
	Obs.	Exp.	Obs.	Exp.
0	955	950.7	470	461.4
1	204	205.6	204	196.5
2	137	130.3	165	144.8
3	86	82.3	118	112.2
4	44	54.2	89	90.3
5-9	96	102.2	260	272.6
10-14	18	19.4	110	127.9
15-24	10	5.3	74	100.6
25+	0	0	60	43.7
Mean use	1.223		5.048	
Variance	6.08		69.38	
k		0.4		0.552
p		0.143		0.0292
a		0.51		0.275
Chi-squared		10.38		51.80
No.of cells		14		31
P		0.4		0.003

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 7.6

Estimates of the parameters (k, p and a) of the modified negative binomial distributions fitted to frequency distributions of recorded use by users of the short-loan collection over three, six and ten weeks.

	Number of weeks observed		
	Three	Six	Ten
k	0.4	0.5	0.552
p	0.143	0.0624	0.0292
a	0.51	0.4	0.275

applicable to the three-week period. Table 7.7 shows the statistics of goodness of fit for negative binomial and modified negative binomial distributions fitted to the observed three-week distribution for short-loan collection users. In each case, the shape parameter, *k*, is that estimated for the ten-week distribution. It does not seem possible, therefore, that a single value of *k* could be found to give reasonable fits for all time periods.

On the evidence of this one set of data, observed distributions would not be well extrapolated using the modified model. For the sample of 309 short-loan collection users (Section 3.1), however, the change in the estimated value of *k* is less consistent than in Table 7.6. At the end of a sixteen week time period, the value falls between the three and the ten-week values. In Chapter 8, these data are used to illustrate the extrapolation of the modified negative binomial distribution.

7.3 UTILITY OF MODELS OF OBSERVED FREQUENCY DISTRIBUTIONS.

The modified negative binomial distribution provided adequate fit to most of the observed distributions of library use and remedied some of the shortcomings of the negative binomial distribution. It was not necessarily better in modelling the extrapolation of these distributions, but it is possible that further modification would remedy this deficiency. Replacing the Poisson distribution of amount of use per spell with a geometric distribution did not appear to improve fit, or the stability of the shape parameter, *k*, over time, however. In a further modification, therefore, it may be necessary to incorporate mean rates of recourse which varied not only between individuals but also over time for each individual. As Cocks and Brookes (96:47) observe, applying 'time parameters to social phenomena assumes that the social behaviour observed is strictly regular over the extended period. This is a bold assumption; it may be found that <u>time</u>-dependent parameters are less reliable than <u>sample-size</u> parameters'.

Before advocating a more complex model, however, it is pertinent to investigate the potential usefulness of the model at its present stage of modification. To this end, two similar stochastic models will be reviewed to determine their usefulness in applications involving decision -making in libraries. They are two of many examples wherein operational research techniques have been consciously applied to library problems. Two useful introductory reviews by Kantor (97) and Rouse (98) describe

TABLE 7.7

Chi-squared statistics for the fit of the negative binomial distribution (k = 0.48) and modified negative binomial distributions (k = 0.552; a = 0.275 and a = 0.51) to the three-week frequency distribution for users of the short-loan collection.

	Parameter values supplied		
	k = 0.48	k = 0.552 a = 0.275	k = 0.552 a = 0.51
Chi-squared	61.53	50.15	30.37
Degrees of freedom	11	10	11
P	$\ll$0.001	$\ll$0.001	0.001

Only one parameter was estimated in each case, therefore degrees of freedom is the number of cells less two. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

some of these problems and the methods used to investigate them. The elegant work of the late Philip Morse (10), a teacher and practitioner of operational research, has been influential and much quoted.

The purpose of fitting a probability distribution to an observed frequency distribution of events is amply explained by Morse (10:18-35). Where the observed events are generated by large numbers of people and governed by numerous but individually modest influences, then the resulting frequency distribution of the variate may appear similar to some simple distribution of probabilities associated with a random variable. The discrepancies between the theoretical model and the sampled data may be no greater than those between the data from two different samples. Once verified against the original data, the model can subsequently be fitted from many fewer data, but may allow insights or predictions not obtainable from the original data because the mathematical function, or its variables, can be more easily manipulated.

7.3.1 Distribution of the numbers of tasks performed per visit.

Morse (10) fits a geometric distribution to data on the numbers of tasks performed per visit by users of the MIT Science Library. Six of the tasks represent the consultation or borrowing of library material; the seventh task being use of the catalogue. The model can be fitted very simply; only a mean is required in order to graduate the whole distribution, and this, Morse suggests, can be estimated from a sample of less than 100 visitors for a homogeneous population.

The practical usefulness of the model is not discussed, although Morse demonstrates its capabilities by predicting the relative hardship which would be suffered by different groups of users if borrowing were restricted to two items per visit. He admits that the rule is 'improbable'. (Indeed, its effect would almost certainly be to nullify his predictions by forcing users to alter their behaviour in compensating for the new constraint.) No other writer seems to have applied the model. It seems that, although the model represents a concise and economical tool for describing and differentiating the behaviour of groups of users, it yields no especially useful information for decision making, even for the numerate library manager.

7.3.2 Distribution of the recorded use of library books.

A most developed essay in the modelling of distributions of recorded use over library collections comes from Burrell and Cane (99). If the stock of a library is partitioned according to the amount of use each item appears to receive over a given period of time, the groups of items so formed, freed of their individuality, can be managed according to their apparent popularity. If necessary, measures (such as duplication or relegation) can be taken to adjust the availability of particular groups to their popularity. Burrell and Cane seek a method of deciding the implications of such measures. They fit both geometric (100) and negative binomial distributions. The geometric distribution is the most manageable because there is a simple relationship between frequency and cumulative probability. Proportions of stock can be related, therefore, to an expected productivity. In both cases, of course, prediction depends upon future use being similar to, or a known function of, past use, at least for large groups of items. A certain degree of complication arises from the inclusion of such a function.

Commentators on Burrell's model (99:463-469,101,102) agree on the need to compromise between building a simple and manageable model and including parameters related to underlying influences; but they disagree, of course, on the degree of simplification required, and doubts are raised not only on the influence upon predictions of factors such as the lengths of loans, the decline in the usefulness of the material with age and the duplication of material, but also on whether purchasing policy, or expenditure, or quality or extent of use, or differences in patterns of use should not rather be studied. A librarian (99:465), while hoping for answers to more of his questions once the description of the use of his collection is encapsulated within a few parameters, concedes that the data may still be insufficiently precise to support these answers. It is also unlikely, he feels, that librarians will readily adopt management methods which involve statistical generalisations.

Nonetheless, it is possible to envisage, as Burrell suggests, that the model could be used to investigate the implications of a policy of relegation and the effects of relegating different proportions of stock (103). Equally, however, it seems likely that, although the model yields 'reasonably useful results', financial and organisational constraints as well as experience and common sense might in many cases decide policy without the policy-maker needing to estimate and minimise the residual demand attaching to the categories of books which are chosen for relegation.

7.4 POTENTIAL UTILITY OF A MODEL OF THE DISTRIBUTION OF RECORDED LIBRARY USE AMONG LIBRARY USERS.

It appears, therefore, that library managers might find little potential usefulness in a model of the distribution of recorded use over library users. If the phenomena observed were physical phenomena, obeying deterministic laws, then predictive models might be more employed by librarians. The number of factors affecting the use of the library is large, however, and the environment is often changing. Librarians prepared to examine quantitative data would, nonetheless, undoubtedly supplement their cumulated experience; and by investigating a fitted model could gain an insight into patterns of use. Rouse (98) stresses the benefits of the heuristic aspects of developing mathematical models of library systems. This insight might diffuse to a wider audience if the patterns recurred regularly or could be encapsulated in a simple rule of thumb such as Trueswell's 80/20 rule (104).

Other positively-skewed distributions of frequency are observed in library work, of course. The frequencies with which the books in a library collection are used or borrowed will often have a very skew distribution (Section 7.3.1). In a given subject field, the frequencies with which articles are distributed over the journal titles which carry them, or over the authors who wrote them, can show a similar variation (15:14-15). Positively-skewed distributions with less variance characterise the variation in performance among individual library users on single occasions, whether it be with respect to length of stay or number of tasks performed (10), or number of uses recorded (Table 6.18).

Although the shapes of these distributions can certainly be similar, they clearly cannot arise from the same causes. In particular, the populations will differ in the extent to which differences between individuals remain fixed over time, change over time or actually evolve over time (as a result of some success-breeds-success mechanism, for example). Thus, although the same statistical model approximates the variation observed both among library books (Section 7.3.1) and among library users (Section 2.2.3), it is clear that not only must different processes generate the variations, but also that these differences are probably discernable even in the fitted models, either from the changes over time in the shapes of the fitted mixing distributions or from the differences in the mixed distributions which give best fit in each case. (Gelman and Sichel (115) suggest the use of binomial mixed distributions for loaned books; but from Figure 2.1 and Section 3.4 it seems that a distribution with a variance greater than the Poisson is required for recorded use by users.)

Two applications of the modified negative binomial distribution model are discussed in the following chapters. A simple method for the extrapolation over time of numbers of non-users is first sought. Then, the effect of patterns of activity among users upon the uptake of material from the collection is investigated.

7.5 SUMMARY

Negative binomial distributions failed to give good fit to frequency distributions of recorded use taken from the literature or to provide a basis for extrapolating the distributions constructed from the author's data. It did not seem that the poor fit could be reliably attributed to heterogeneity in the populations or to the effects of competition among the users, although both of these factors were thought to influence the observed distributions. Increasing the variance of the distribution of amounts of use per time period in the model improved fit, but did not improve extrapolation as much as expected. To secure a further improvement, it would probably be necessary to incorporate variable rather than constant rates of recourse into the model.

It is unlikely that models of frequency distributions of recorded use would find direct applications in management information systems or in extemporaneous policy making unless they yielded simple and widely applicable rules of thumb which served to extend the significance of the basic data. Nonetheless, the use of mathematical models to investigate patterns of library use could, by virtue of the rigour of the method, be expected to result in an enhanced description of user behaviour which could supplement, and provide a quantitative framework for, other sources of information.

CHAPTER 8

FORECASTING NUMBERS OF NON-USERS

8.1 NEGATIVE BINOMIAL MODELS

Under certain conditions, the fitted negative binomial or modified negative binomial models may be used to predict distributions of use beyond the time periods for which data have been collected. The fit of the model to the data must be good to begin with, of course, and must be likely to remain so. The population of potential users and the circumstances of their use must therefore be expected to remain unchanged. In particular, individual mean rates of use and, for the modified negative binomial model, mean amounts of use per spell or visit, must remain stable. For, prediction will be successful only if the inverse scale parameter, $b = p/(1 - p)$, can be extrapolated linearly with time while the other parameters are held constant. Also, behaviour in one time period must be independent of behaviour in another, since in the model it is assumed that each use occurs independently of any other and at random in time for each user. The length of time periods should not be too short therefore since actual library uses will tend to be clustered in time.

These conditions appeared to be largely met in the case of the modified negative binomial distributions fitted to the frequency distributions of use for the sample of 309 short-loan collection users described in Section 3.1. The distributions are shown in Table 8.1. The mean monthly rate of use changes little up to 16 weeks, and the parameters k and a are reasonably stable, although a shows some change in the 16-week distribution. The probability of use being recorded in any weekly time period appears from Table 3.8 (Section 3.3.2) to have been largely independent of use occurring in any other time period. Because of this stable pattern of activity, it seems reasonable to suppose that these distributions could be successfully extrapolated in multiples of one week.

How long the stable pattern of activity would continue is not clear. During the UK academic year, major cycles of activity result in peaks of library use when assignments are due or when examinations are at hand and lulls in library use in the vacations. Patterns of use would not remain stable for more than a few months, therefore, and a population

TABLE 8.1

Frequency distributions of the recorded use of a short-loan textbook collection by 309 students over four and eight weeks with expected frequencies, parameters (*k*, *p* and *a*) and chi-squared statistics for fitted modified negative binomial distributions.

Number of recorded uses	Observed (Obs) and expected (Exp) numbers of users			
	Four weeks		Eight weeks	
	Obs.	Exp.	Obs.	Exp.
0	161	159.1	108	103.7
1	40	40.4	37	36.2
2	32	30.4	29	30.9
3	21	21.3	30	24.5
4	17	15.1	26	19.6
5-9	27	32.5	50	55.7
10-14	8	8.2	9	22.1
15-24	2	2.0	14	13.3
25+	1	0	6	3.0
Mean use	1.9		3.9	
Variance	12		47	
k		0.56		0.7
p		0.195		0.133
a		0.8		0.86
Chi-squared		3.1		18.9
No.of cells		11		16
P		0.9		0.1

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. *P* is the approximate probability of the observed chi-squared value being exceeded in random sampling.

TABLE 8.1 (continued)

Frequency distribution of the recorded use of a short-loan textbook collection by 309 students over sixteen weeks with expected frequencies, parameters (k, p and a) and chi-squared test statistics for a fitted modified negative binomial distribution.

Number of recorded uses	Numbers of users Observed	Expected
0	70	68.2
1	30	29.0
2	24	24.4
3	26	20.3
4	24	17.2
5-9	53	58.8
10-14	32	33.5
15-24	27	33.4
25+	23	24.2
Mean use	8.1	
Variance	192	
k		0.64
p		0.0444
a		0.59
Chi-squared		15.8
No.of cells		23
P		0.7

Chi-squared test: expected frequencies were pooled where necessary to give a minimum cell value of 5.0. P is the approximate probability of the observed chi-squared value being exceeded in random sampling.

would not remain unchanged or in the same environment for more than one year. For the short-loan collection users, 24 weeks was considered the longest time period over which patterns of library use might remain unchanged.

To examine the success of extrapolations of these distributions of use, the scale parameter fitted to the 8-week distribution, $\bar{p} = 0.133$, was used as a base. Extrapolations for 4 and 16 weeks as well as for 24 weeks were then calculated. For simplicity, only extrapolations of the zero term are discussed here. Whole distributions could easily be graduated if required.

In Figure 8.1, the zero term of the modified negative binomial distribution is expressed in terms of the parameters. Table 8.2 shows numbers of potential users not expected to have recorded use during 4, 8, 16 and 24 weeks as estimated from this expression. The estimates are correct to within ±10% of the observed numbers of non-users.

8.2 ESTIMATES BASED ON MEAN AND PROPORTION OF ZEROS

Two statistics, the mean and proportion of zeros, are adequate for estimating the parameters of reversed J-shaped negative binomial distributions. Data collection would be much simplified if these statistics alone were employed in forecasting numbers of non-users. In Figure 8.2, k and b are expressed in terms of the observed mean and proportion of zeros, and in Figure 8.3 the expressions required for extrapolation are shown.

This method of extrapolation would not perform any better than the negative binomial model itself, but the modified negative binomial model could be employed instead, provided that the value of the third parameter, a, was assumed *a priori*. A value of $a = 1$, giving a variance double that of the Poisson distribution in the negative binomial model, seems a reasonable first approximation from Table 3.5. Clearly, it would need to be tested against other sets of data. Figure 8.4 shows the proportion of zeros for such a model and in Figure 8.5, the parameters k and b are expressed in terms of the observed mean and proportion of zeros. In Table 8.3, the value of the extrapolated proportion of zeros, $p_t(0)$ is tabulated for various values of: $p(0)$, the proportion of zeros after unit time; $\bar{x}$, the observed mean use after unit time; and t, the number of time periods required. In Table 8.4 the ratio of $p_t(0)$ to $p(0)$ is shown.

FIGURE 8.1

Zero term of the modified negative binomial distribution

From Figure 6.5, we have

$$p(r) = \frac{p^k}{r!} \sum_s \binom{k+s-1}{k-1} \varepsilon^{-as} (as)^r q^s \quad (s = 0,1,2,\ldots)$$

where $q = 1 - p$.

Hence, the proportion of zeros is

$$p(0) = p^k \sum \binom{k+s-1}{k-1} \varepsilon^{-as} q^s \quad (s = 0,1,2,\ldots)$$

$$= p^k \left(1 + \varepsilon^{-a} kq + \varepsilon^{-2a} k(k+1) q^2/2! + \ldots\right)$$

$$= \left[p / (1 - \varepsilon^{-a} q)\right]^k$$

since the series can be summed* using the binomial expansion,

$(1 - x)^{-n} = 1 + nx + n(n + 1)x^2/2! + \ldots$

The expected frequency of zeros in a population of N potential users is then

$$f(0) = N\left[p/(1 - \varepsilon^{-a} q)\right]^k.$$

If a and k are held constant, then the scale parameter, $b = p/q$, scales linearly with time. For a time period t times that for which b was fitted, b becomes b/t and therefore

$$f(0) = N\left[\frac{\frac{b}{t}}{\frac{b}{t}+1} / \left(1 - \varepsilon^{-a} \frac{1}{\frac{b}{t}+1}\right)\right]^k$$

---

*The author is grateful to Mr. D.M. Ellis for pointing out this simplification.

FIGURE 8.1 (continued)

If, for example, the value of *b* was fitted to data for a period of observation of 8 weeks, then the value for 24 weeks will be

$$\frac{b}{\left(\frac{24}{8}\right)} = \frac{b}{3}\,.$$

TABLE 8.2

Observed and expected numbers of potential users not recording use of the short-loan textbook collection.

Time period	Observed number of non-users	Expected number of non-users	Scale parameter, p
Weeks 1-4	161	147	0.235
Weeks 1-8	108	104	0.133
Weeks 1-16	70	69	0.0712
Weeks 1-24	-	53	0.04865

Extrapolated from parameters fitted to data for Weeks 1-8 (Table 8.1):

$k = 0.7$; $a = 0.86$; $p = \dfrac{\frac{b}{t}}{\left(\frac{b}{t}+1\right)}$ where t = (Time period in weeks)/8.

FIGURE 8.2

Estimating the parameters of the negative binomial distribution from the observed mean and proportion of zeros

---

The negative binomial distribution with shape parameter, k, and scale parameter, b, (Figure 2.2) has a proportion of zeros

$$p(0) = \left(\frac{b}{b+1}\right)^k$$

and a mean (Figure 2.3)

$$m = \frac{k}{b}.$$

The parameters may be estimated by equating these expressions to the observed proportions of non-users, $g(0)$, and the observed mean use, $\bar{x}$, respectively. Then, if $p(0) \equiv g(0)$ and $m \equiv \bar{x}$,

$$k = \frac{\ln g(0)}{\ln\left(\frac{b}{b+1}\right)}$$

and

$$k = b\bar{x}.$$

Hence

$$b \ln \frac{b}{b+1} = \frac{\ln g(0)}{\bar{x}}.$$

For convenience, approximations could be found for the expression on the left-hand side. In general,

$$b \ln \frac{b}{b+1} \approx -b^n,$$

FIGURE 8.2 (continued)

---

where the exponent, n, has a value between 0.6 and unity depending on the value of b. Alternatively

$$b \ln\left(\frac{b}{b+1}\right) \approx -\sqrt{b} + 0.1, \qquad 0.1 < b < 0.4$$

Using the latter approximation,

$$b \approx \left(\frac{\ln g(0)}{\bar{x}} - 0.1\right)^2,$$

and

$$k \approx \bar{x}\left(\frac{\ln g(0)}{\bar{x}} - 0.1\right)^2.$$

These or similar expressions may then be substituted for b and k in Figure 8.1.

FIGURE 8.3

Extrapolation of the negative binomial proportion of zeros

From Figure 2.3, the proportion of zeros of the negative binomial distribution for a single unit of time is

$$p(0) = \left(\frac{b}{b+1}\right)^k .$$

If k remains constant so that b scales linearly with time, then, for t units of time,

$$p_t(0) = \left[\frac{\frac{b}{t}}{\frac{b}{t}+1}\right]^k = \left(\frac{b}{b+t}\right)^k$$

and the ratio $p_t(0)/p(0)$ is

$$\frac{p_t(0)}{p(0)} = \frac{\left[\frac{\frac{b}{t}}{\frac{b}{t}+1}\right]^k}{\left(\frac{b}{b+1}\right)^k} = \left(\frac{b+1}{b+t}\right)^k$$

Alternatively, this ratio could be equated to a particular fraction, say $1/d$, in order to calculate the number of time units, t, which would be required to elapse before the proportion of zeros would fall to this fraction of its initial size. If

$$\left(\frac{b+1}{b+t}\right)^k = \frac{1}{d}$$

FIGURE 8.3 (continued)

---

then

$$t = (b+1)d^{1/k} - b.$$

In each case, k and b could be expressed in terms of the observed mean and proportion of zeros, as shown in Figure 8.2.

FIGURE 8.4

Extrapolation of the proportion of zeros for the modified negative binomial distribution

---

From Figure 8.1, we have

$$p(0) = \left(\frac{p}{1 - q e^{-a}}\right)^k = \left(\frac{b}{b + 1 - e^{-a}}\right)^k$$

If the parameter a is held constant at unity, then the variance will be

$$\frac{k}{b^2}(1 + 2b)$$

compared to the variance of the negative binomial distribution,

$$\frac{k}{b^2}(1 + b)$$

Putting $c = (1 - e^{-a})$ and following Figure 8.3, we have

$$p(0) = \left(\frac{b}{b + c}\right)^k = \left(\frac{b}{b + 0.632}\right)^k$$

and

$$p_t(0) = \left(\frac{b}{b + ct}\right)^k$$

Thus

$$\frac{p_t(0)}{p(0)} = \left(\frac{b + c}{b + ct}\right)^k$$

and

$$t = \frac{(b+c)\,\alpha^{1/k} - b}{c}.$$

FIGURE 8.5

Estimating the parameters of the modified negative binomial distribution from the observed mean and proportion of zeros

---

From Figure 8.4 and following Figure 8.2, we have

$$g(0) \equiv \left(\frac{b}{b + 0{\cdot}632}\right)^{k} \text{ and } \qquad \bar{x} \equiv \frac{ak}{b} \qquad \text{(from Figure 6.5).}$$

If $a = 1$, then

$$b \ \ln\left(\frac{b}{b + 0{\cdot}632}\right) = \frac{\ln[g(0)]}{\bar{x}}$$

For the expression on the left-hand side of the equation we substitute the simpler parabolic approximation $-0.6\sqrt{b}$. Then

$$b \approx \left(\frac{\ln g(0)}{0{\cdot}6\,\bar{x}}\right)^{2}$$

and

$$k \approx \bar{x}\left(\frac{\ln g(0)}{0{\cdot}6\bar{x}}\right)^{2} = \frac{[\ln g(0)]^{2}}{0{\cdot}36\bar{x}} .$$

whence

$$\frac{p_t(0)}{p(0)} = \left[\frac{\left(\frac{\ln g(0)}{0{\cdot}6\bar{x}}\right)^{2} + 0{\cdot}632}{\left(\frac{\ln g(0)}{0{\cdot}6\bar{x}}\right)^{2} + 0{\cdot}632t}\right]^{\frac{[\ln g(0)]^{2}}{0{\cdot}36\bar{x}}}$$

and

$$p_t(0) = g(0)\,\frac{p_t(0)}{p(0)} .$$

TABLE 8.3

Values of the extrapolated proportion of zeros, $p_t(0)$, estimated from the approximation shown in Figure 8.4 for values of $p(0)$ from 0.2 to 0.7, of $\bar{x}$ from 3.0 to 7.0 and of t from 2 to 4.

$p(0)$	$\bar{x}$ 3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
	t = 2								
0.2	.08	.08	.09	.09	.09	.1	.1	.11	.11
0.25	.12	.12	.13	.13	.14	.14	.15	.15	.16
0.3	.16	.17	.17	.18	.19	.19	.2	.2	.21
0.35	.21	.22	.23	.23	.24	.25	.25	.26	.26
0.4	.26	.27	.28	.29	.3	.31	.31	.32	.32
0.45	.32	.33	.34	.35	.36	.37	.37	.38	.38
0.5	.38	.4	.4	.41	.42	.43	.43	.44	.44
0.55	.45	.46	.47	.48	.48	.49	.49	.5	.5
0.6	.51	.52	.53	.54	.54	.55	.55	.56	.56
0.65	.58	.59	.6	.6	.61	.61	.61	.62	.62
0.7	.65	.65	.66	.66	.67	.67	.67	.67	.68
	t = 3								
0.2	.04	.05	.05	.05	.06	.06	.07	.07	.07
0.25	.07	.08	.08	.09	.09	.1	.11	.11	.12
0.3	.11	.11	.12	.13	.14	.15	.15	.16	.17
0.35	.15	.16	.17	.18	.19	.2	.21	.22	.22
0.4	.2	.22	.23	.24	.25	.26	.27	.28	.28
0.45	.26	.28	.29	.3	.31	.32	.33	.34	.34
0.5	.33	.34	.36	.37	.38	.39	.39	.4	.41
0.55	.4	.41	.43	.44	.45	.45	.46	.47	.47
0.6	.47	.48	.5	.51	.51	.52	.53	.53	.54
0.65	.54	.56	.57	.57	.58	.59	.59	.6	.6
0.7	.62	.63	.64	.64	.65	.65	.66	.66	.66
	t = 4								
0.2	.03	.03	.03	.04	.04	.04	.05	.05	.06
0.25	.05	.05	.06	.06	.07	.08	.08	.09	.09
0.3	.08	.09	.09	.1	.11	.12	.13	.14	.14
0.35	.12	.13	.14	.15	.16	.17	.18	.19	.2
0.4	.16	.18	.19	.21	.22	.23	.24	.25	.26
0.45	.22	.24	.26	.27	.28	.29	.3	.31	.32
0.5	.29	.31	.32	.34	.35	.36	.37	.38	.39
0.55	.36	.38	.4	.41	.42	.43	.44	.45	.45
0.6	.44	.46	.47	.48	.49	.5	.51	.52	.52
0.65	.52	.53	.55	.56	.56	.57	.58	.58	.59
0.7	.6	.61	.62	.63	.64	.64	.65	.65	.65

TABLE 8.4

Values of the ratio of the extrapolated and unit proportions of zeros estimated from the approximation shown in Figure 8.4 for values of p(0) from 0.2 to 0.7, of $\bar{x}$ from 3.0 to 7.0 and of t from 2 to 4.

p(0)	$\bar{x}$ 3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
	t = 2								
0.2	.42	.42	.44	.45	.47	.49	.51	.53	.54
0.25	.48	.49	.51	.53	.55	.57	.59	.61	.62
0.3	.54	.56	.58	.6	.62	.64	.66	.68	.69
0.35	.6	.62	.65	.67	.69	.71	.72	.74	.75
0.4	.66	.68	.71	.73	.75	.76	.78	.79	.8
0.45	.72	.74	.76	.78	.8	.81	.82	.84	.84
0.5	.77	.79	.81	.83	.84	.85	.86	.87	.88
0.55	.82	.84	.85	.87	.88	.89	.89	.9	.91
0.6	.86	.87	.89	.9	.91	.91	.92	.93	.93
0.65	.89	.91	.92	.93	.93	.94	.94	.95	.95
0.7	.92	.93	.94	.95	.95	.96	.96	.96	.97
	t = 3								
0.2	.22	.23	.25	.27	.29	.31	.33	.35	.37
0.25	.28	.3	.33	.35	.38	.4	.42	.45	.47
0.3	.35	.38	.41	.44	.46	.49	.51	.54	.56
0.35	.43	.46	.49	.52	.55	.57	.6	.62	.64
0.4	.5	.54	.57	.6	.62	.65	.67	.69	.7
0.45	.58	.61	.64	.67	.69	.72	.73	.75	.76
0.5	.65	.68	.71	.74	.76	.77	.79	.8	.82
0.55	.72	.75	.77	.79	.81	.83	.84	.85	.86
0.6	.78	.81	.83	.84	.86	.87	.88	.89	.89
0.65	.84	.86	.87	.88	.89	.9	.91	.92	.92
0.7	.88	.9	.91	.92	.93	.93	.94	.94	.95
	t = 4								
0.2	.13	.15	.16	.18	.2	.22	.24	.26	.28
0.25	.19	.21	.23	.26	.28	.31	.33	.36	.38
0.3	.26	.29	.32	.35	.37	.4	.43	.45	.47
0.35	.33	.37	.4	.43	.46	.49	.52	.54	.56
0.4	.41	.45	.49	.52	.55	.58	.6	.62	.64
0.45	.5	.54	.57	.6	.63	.65	.68	.69	.71
0.5	.58	.62	.65	.68	.7	.72	.74	.76	.77
0.55	.66	.69	.72	.75	.77	.78	.8	.81	.82
0.6	.73	.76	.79	.81	.82	.84	.85	.86	.87
0.65	.8	.82	.84	.86	.87	.88	.89	.9	.9
0.7	.85	.87	.89	.9	.91	.92	.92	.93	.93

The expressions of Figure 8.5, though cumbersome, are based on simple statistics and are easy to evaluate. The results could be accurate enough to indicate at an early stage whether broad policy objectives were likely to be met. As an example, assume that the proportion of non-users after one month of data collection (preferably not right at the beginning of the session) was 0.4 and that mean recorded use was 3. After three further months (i.e. four months overall), the proportion of zeros could be expected to be around 0.16, perhaps a larger figure than would reflect entirely creditably upon the library. Accordingly, assume that a proportion no larger than 0.05 is the objective. By trial and error, or by calculating the approximate values of k and b, it can be shown that more than 20 months would be required to elapse before the proportion would fall to the required level. If action were taken to increase the rate of use for all potential users equally (that is, if k remained constant and the distributions of propensity within the potential user population remained unchanged), then clearly the monthly mean would have to be raised at least 20/4 = 5 times in order to achieve the required result. If extra books were provided generally or if instruction were given generally, then this would have to be the target. If, however, selective action were taken to alter the propensity of infrequent users (by, for example, instruction or encouragement, or by improving the accessibility or availability of the most popular or perceptibly useful titles in the collection), then mean recorded use may not have to be so massively increased in order to achieve the desired result. In this case, k has to be raised only to 1.2 from 0.7, raising the monthly mean to 5.2. Table 8.5 shows a summary of the distributions for each of these outcomes assuming a population of 100 users and indicates the small loss in accuracy in using the approximated values of b and k.

8.3 SUMMARY

Provided that the propensity of the users and the nature of the environment remain unchanged, distributions of recorded use may be predicted well beyond a period of observation by fitting and extrapolating the modified negative binomial distribution. The probability of zero uses is particularly easy to calculate because the series in the probability function can be summed.

An even simpler method can be employed. Assigning an arbitrary value to the parameter, a, approximations to the parameters b and k can be

TABLE 8.5

Summary of the expected frequencies of use after four months for a population of 100 potential users: a) with monthly mean, 3, and proportion of zeros, 0.4; b) with increased monthly mean; c) with shape parameter, k, increased.

Number of uses	Numbers of users a)	b)	c)
0	17.7 (15.8)	4.9	5.0
1-4	23.3	7.5	13.7
5-9	18.2	7.1	16.0
10+	40.8	80.5	65.3
Mean use	12.0	81.0	20.8
p	0.055 (0.0608)	0.008556	0.055
k	0.699 (0.777)	0.699	1.21
a	1.0	1.0	1.0

Figures in brackets are calculated from the approximation described in Section 8.2

expressed in terms of the observed mean and proportion of zeros. The need to fit the distribution is thus eliminated. For distributions convex to the origin where proportions of zeros will be large enough to be of interest, the loss of accuracy in using the approximations will not be great compared to probable instabilities in the environment or population. The results could be adequate, therefore, for testing performance against broadly defined policy targets.

CHAPTER 9

UPTAKE FROM THE COLLECTION

9.1 RELATIONSHIP BETWEEN DISTRIBUTIONS OF USER ACTIVITY AND DISTRIBUTIONS OF BOOK USE.

Unless users choose library material entirely at haphazard, a relationship will exist between their distribution of recorded library use and the pattern with which material is taken up from the collection. The strength of the relationship will depend upon the extent to which the users share the same preferences and only a detailed record of the interactions of users with items from the collection would make this plain. For more than a small number of users and items, however, the pattern may prove too complex to be profitably analysed.

Of course, if the users share exactly the same preferences and use with absolute independence, then the distribution of recorded uses will define precisely the pattern of uptake from the collection. Heavy users will work through a large range of material from the most popular or preferred items to those of lesser popularity. Light users will follow the same route until their energies or responses to stimuli are exhausted. The effect on uptake from the collection is shown schematically in Figure 9.1. A hypothetical collection of 25 books, ranked from 1 (indicating the item with the highest priority for use by the users) to 25, is associated with a hypothetical group of 20 potential users (named A to T) whose recorded use is distributed similarly to the expected distribution for Class B in Table 6.14. Clearly, 90% of the potential users could be satisfied from a little over half of the collection.

For other groups of users, like the economists in Table 1.2, the distribution of uses will not be so skewed as in Figure 9.1. At the extreme, when the reading is regimented (as in the case (4) cited by Jahoda), most users will use the full range of material and a filled rectangular matrix of interactions between users and items would result.

If the users were to exhibit no preferences in their use of items from the collection, that is, if they used material at random, it would of course be impossible to detect a pattern in the use of the collection. In any period of observation, some books would receive more use than others, but these differences would not be repeated in subsequent periods. Figure 9.2 shows the pattern of use for the hypothetical collection of 25 books

FIGURE 9.1

Pattern of use when users share the same priorities in using library material. Diagram showing the use of a hypothetical collection of 25 books (ranked 1 to 25 according to the amount of use made of them) by 20 potential users (denoted A to T in order of decreasing activity) with amounts of use distributed similarly to Class B in Table 6.14. The books used by each user are shown stacked vertically above their identifying letter.

25																			
24																			
23																			
22																			
21																			
20																			
19																			
18																			
17																			
16	16																		
15	15																		
14	14																		
13	13	13																	
12	12	12																	
11	11	11																	
10	10	10	10																
9	9	9	9																
8	8	8	8	8															
7	7	7	7	7															
6	6	6	6	6	6														
5	5	5	5	5	5	5													
4	4	4	4	4	4	4	4												
3	3	3	3	3	3	3	3	3	3										**Books**
2	2	2	2	2	2	2	2	2	2	2	2								
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1					
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T

Potential Users

FIGURE 9.2

Example of pattern of use when users use library material at random. Diagram showing the use of a hypothetical collection of 25 books (ranked 1 to 25 according to the amount of use made of them) by 20 potential users (denoted A to T in order of decreasing activity) with amounts of use distributed similarly to Class B in Table 6.14. The books used by each user are shown vertically above their identifying letter.

25				25																
24		24																		
23			23																	
22			22																	
21	21																			
20			20				20													
19			19							19										
18		18	18																	
17	17			17																
16	16			16																
15	15	15		15																
14			14					14	14											
13	13	13		13																
12	12	12				12														
11	11	11							11											
10			10		10	10					10									
9	9	9			9		9													
8	8	8							8					8						
7	7	7		7							7									
6	6		6		6	6														
5	5				5		5						5							
4	4	4		4			4	4												
3	3	3	3		3					3										Books
2	2	2	2		2	2														
1	1	1		1		1		1				1								
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	

Potential Users

when used at random by 20 users whose distribution of activity is as shown in Figure 9.1. (Random numbers were drawn with the routine tested (Appendix B) and used in the simulation runs.) Books receive from two to seven uses in a roughly symmetrical distribution. This distribution could be expected to persist if the observations were repeated, although the relative differences between individual books would alter.

Most librarians, however, might expect not only a markedly skewed distribution of use over the books in their collections, but also a marked persistence over time in the relative differences between books (102,105). It seems, therefore, that the preferences of library users (and especially of students sharing the same curriculum) will appear to converge on particular items in library collections, but it is unlikely nonetheless that this convergence would be complete. Thus, even if the distribution of use over the collection is far removed from the symmetrical form associated with random use, students may diversify their recorded use sufficiently to preclude the connection of distributions of activity with patterns of uptake from the collection.

There may be more than one reason for this diversification. It may, of course, be deliberately fostered to promote independence among students: they may be given long reading lists, or no reading lists at all. Even if diversification is not encouraged, competition among users may hinder the exercise of common preferences. If choice is constrained by a lack of availability, the least-used books might be expected to be associated with the last-comers or least-active users, not the heavier users as suggested in Figure 9.1. Rao (106) has noticed a tendency for single-use books to be issued to single-use users in large general academic collections, and it is well-known that a minority of heavy borrowers in academic libraries can be successful in obtaining and retaining popular books to the detriment of their colleagues (107). If groups of users cooperate by exchanging library material which has already been issued (66,108), common preferences may only be partially recorded and an apparent diversification may result. In the extreme case, those that record much use of the library collection would then merely be the students who work alone or those who most frequently serve their colleagues.

It is perhaps for these reasons that the record of library use considered below gives few indications of common preferences among the users. The data consist of recorded library uses by a sample of 20 undergraduate economists using an open-access short-loan collection. As with the collection described in Section 2.2.3, almost any use of library material (with the exception of informal exchange) would have resulted in a transaction record. In this case, however, only numbers of titles used are considered (as proposed in Section 1.4), and the effect of this change in unit on statistics and distributions of use is first noted.

9.2 TITLE USE

For individuals using particular library collections, an association may be observed between the amount of recorded use and the number of titles recorded. In Figure 1.4, numbers of titles used are shown plotted against numbers of recorded uses for the sample of economists. A regular, although perhaps not linear, trend is apparent. In Table 9.1, data for groups of individuals from the sample of 309 short-loan collection users (Section 3.1) are shown. Mean numbers of titles used are calculated for groups of users with similar rates of recorded use. Individual amounts of title use varied up to ±45% about title-use means for the higher numbers of recorded uses, but varied less for those users recording 6, 10 and 11 uses. One individual recording three uses used only one title. The relationship between recorded use and recorded title use was assumed to be of a higher order in the form: (Number of titles used) = (Number of uses)^{a}, where a is constant. This gives the necessary agreements between use and title use at zero and one uses. From inspection, a value of a = 0.86 was adopted.

Similar exponents were fitted by least-squares estimation to data for a sample of humanities and social sciences undergraduates recording the use of short-loan and reserve material in a UK university library. Values of 0.81 and 0.78 gave best fit. Figure 9.3 shows the plots of these data using logarithmic scales on both the 'use' and 'title use' axes.

For each of these three sets of data, the individual variation about the estimated number of titles used was only slightly greater than could have been expected for a sequence of Bernouilli trials (Figure 9.4). The greatest variation occurred in the data of Table 9.1, perhaps because the material was issued for two different loan periods. 13%, 7% and 4% respectively of the values in each set of data exceeded 95% confidence limits.

In Table 9.2, title use by the undergraduate economists is shown. As in Table 9.1, items were available for different lengths of time: one

TABLE 9.1

Numbers of recorded uses, mean numbers of titles used and expected numbers of titles used by a sample of 101 users of a short-loan textbook collection over sixteen weeks.

Number of recorded uses	Number of users in sample	Mean number of titles used	Expected number of titles used
1	30	1.0	1.0
3	26	2.6	2.6
6	11	4.6	4.7
10	9	7.7	7.2
11	5	8.4	7.9
19-21	9	14.9	13.2
29-31	5	18.2	18.6
56-60	4	25.5	32.9
88	1	42	47.0
151	1	85	74.8

$$(\text{Expected number of titles used}) = (\text{Number of recorded uses})^{0.86}$$

FIGURE 9.3

Log-log plot of numbers of recorded uses and numbers of titles recorded for a sample of second-year undergraduates using social science or humanities material from reserve and short-loan collections in a UK academic library over a period of 30 weeks. Fitted lines show the estimated number of titles recorded using the approximations: (Number of recorded uses)$^{0.78}$ for reserve books and (Number of recorded uses)$^{0.81}$ for short-loan books. Exponents were fitted by least squares.

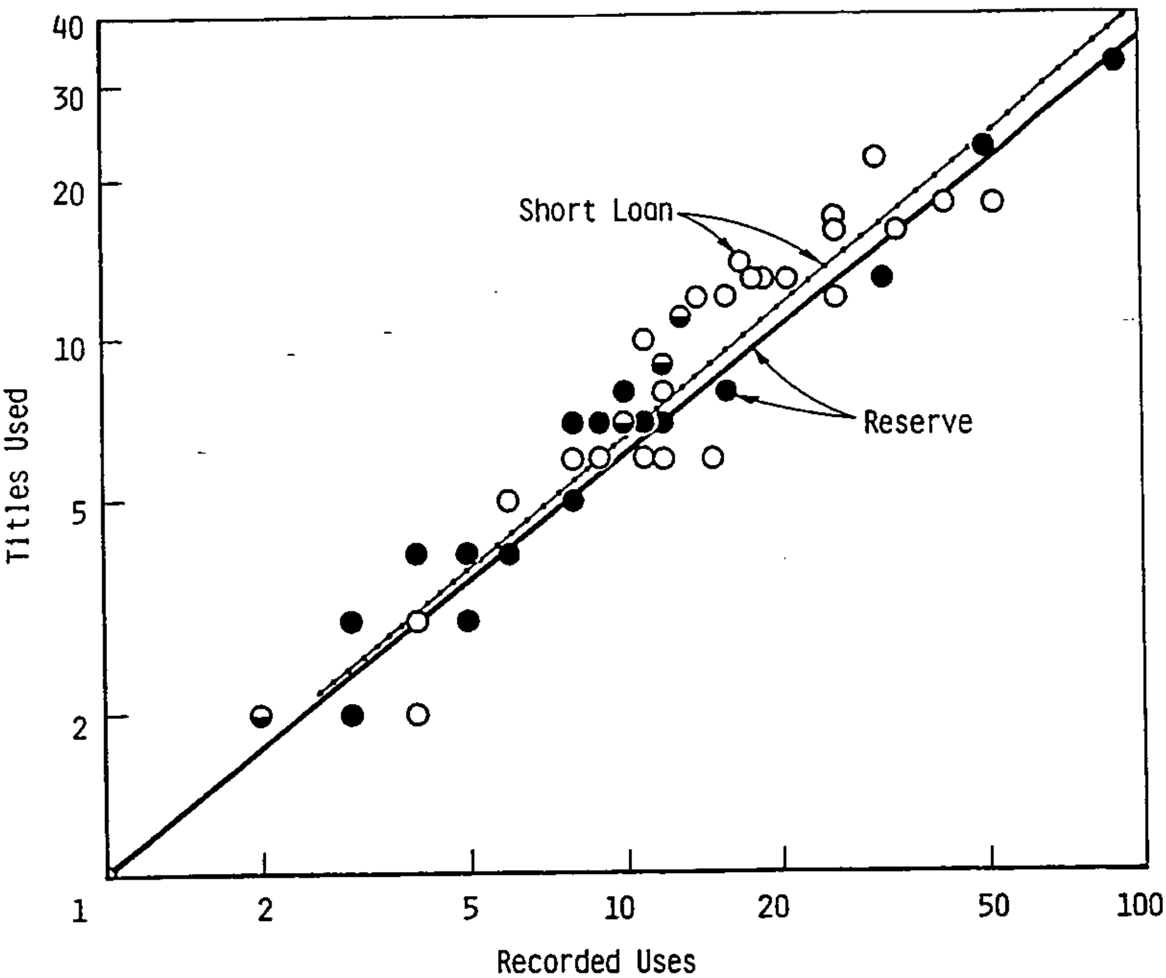


FIGURE 9.4

Confidence limits for estimated numbers of titles used.

Each use is regarded as a Bernouilli trial with the alternative outcomes: (Use of a new title; Use of a title already used). There is one less trial than the observed number of uses, u, because the first trial inevitably results in the use of a previously unused title. The probability at each trial of using a new title is $p = (u^a - 1)/(u - 1)$, where a is the fitted exponent. The expected number of titles is $1 + p(u - 1) = u^a$, as required. The standard deviation of this estimate is $\sigma = \sqrt{p(u - 1)(1 - p)}$, and the 95% confidence limits lie at about $u^a \pm 2\sigma$.

TABLE 9.2

Numbers of recorded uses, numbers of titles used and expected numbers of titles used by a sample of 20 undergraduate economists using a short-loan collection over three terms.

User number	Number of recorded uses	Number of titles used	Expected number of titles used
3	8	8	6.2
15	26	24**	17.4
8	29	19	19.1
5	32	20	20.8
18	34	22	22.0
19	36	29*	23.1
14	46	26	28.6
20	52	31	31.9
1	61	40	36.6
17	62	38	37.2
12	63	38	37.7
9	63	41	37.7
4	70	44	41.3
10	70	49	41.3
13	74	38	43.4
11	74	46	43.4
7	79	44	46.0
6	88	41*	50.5
2	95	53	54.0
16	138	49**	74.9
Total	1200	700	713.1
Mean	60	35	
Variance	852	145	

(Expected number of titles used) = $(\text{Number of recorded uses})^{0.876}$

*Lies less than 0.5 outside 95% confidence limits for the estimate of title use. **Lies more than 1.5 outside 95% confidence limits for the estimate of title use.

week, one day, or within library hours only. Four (20%) of the observed numbers of title uses fell outside 95% confidence limits for the estimate.

It seems possible that the ratio of titles to uses reduces as the period available for use is shortened; that is, where access is restricted users may need to record more uses in order to achieve their objective. Thus in Figure 9.3, the estimated number of reserve titles used per 100 recorded uses is 36, whereas the estimated number of short-loan (three-day loan) titles is 42. A greater difference is observed between loan material and reserve material in the US college library surveyed by Knapp (3). In 4185 reserve collection uses, 1961 titles were used, but in 2547 loans from the main collection, 2444 titles were used. On the other hand, the observed differences may merely reflect the more regular re-use of a core of permanently relevant material held on reserve.

If each of the terms of a frequency distribution of recorded use are transformed with an exponent less than unity, then clearly each term (except zero and one) will be reduced. The mean and variance will therefore also be reduced and the distribution will become less positively skewed. For negative binomial distributions, an increase in both k and p will result. Because this transformation of uses into title uses is performed with a non-linear function, it is not possible to calculate the mean of the transformed values accurately by transforming the mean of the original values. For the positively-skewed distributions encountered so far, an overestimate of more than 10% would result, assuming values of the exponent similar to those given above. But for the more symmetrical distribution associated with the undergraduate economists (Table 9.2), the error is less serious. The transformed mean is $(60)^{0.876} = 36.1$ compared to the mean of the individually transformed values, 35.7.

The approximate relationship demonstrated between use and title-use suggests that users become heavier users not only by using more titles, but also by making rather more use of each title. For an exponent, a, a doubling of total use is accomplished from 2^a times the number of titles and so the increase in the rate of use per title is the proportion $(2^{(1-a)} - 1)$ of the original rate. For the values of the exponent given above, however, this increase is clearly not great, say between 9% and 16%. The range of titles used increases with use, therefore, but at a gradually declining rate. This decline perhaps reflects a change in the marginal utility of consulting each extra title.

9.3 LEVELS OF USE NOT EXCEEDED BY 90% AND 95% OF POTENTIAL USERS

In passing, it is interesting to note that by using the upper values of fitted geometric distributions to approximate the upper values of observed or fitted distributions of recorded use, levels of use not exceeded by the large majority of potential users may be roughly estimated. With the exception of the distribution shown in Table 9.2, the distributions of recorded use examined so far have been very skewed: a proportion of potential users remain unrecorded while a minority are extraordinarily active with rates of recorded use several times the mean. The tails of these distributions may be approximated merely by using the sample mean to estimate the single parameter of a geometric distribution. In Tables 9.3 and 9.4, the correspondence between the upper parts of observed and fitted distributions is shown for two sets of data for large populations. This correspondence will only hold for distributions which are positively skewed. For more symmetrical distributions, such as that of Table 9.2, a different approximation would be necessary and both central tendency and range would probably need to be estimated.

In Figure 9.5, the method of estimating the minimum level of use of the most active 5% or 10% of potential users is described. The results for the sets of data already described are shown in Table 9.5. The actual percentages of users who failed to reach the minimum level of use compare well with the expected percentages of 90% and 95%. More simply, a minimum level of use of three times the mean can be expected to divide off between 5% and 10% of potential users. Column 4 of Table 9.5 shows the actual percentages who failed to reach this level of use. It is clear that most of the percentages are within the predicted range. Three times mean use thus seems a useful and easily calculated statistic to indicate, for positively-skewed distributions, a level of activity not exceeded by at least 90% of all potential users. The other 10% are the heaviest users and they may account, if there are many non-users, for as much as one half of all recorded uses (Table 1.1).

It is possible that this relationship between a level of use and a proportion of potential users could help in defining a minimum collection to satisfy most economically all the recorded uses of a given number of potential users. As noted above, however, it would be necessary for the users to behave as shown in Figure 9.1, with common preferences for the titles they used. The extent of this community of interest is now investigated for the sample of 20 undergraduate economists.

TABLE 9.3

Proportions of population falling within various ranges of amounts of use for the 1550 potential users of a short-loan collection (Table 2.3, Column 6) and a fitted geometric distribution.

Range of amounts of use	Observed proportion of population	Expected proportion of population
0-7	0.794	0.764
0-12	0.890	0.904
0-13	0.901	0.920
0-16	0.927	0.954
0-19	0.941	0.978
0-21	0.951	0.981
0-25	0.964	0.991

Parameter, q, of geometric distribution estimated from observed mean, m, using $q = m/(1 + m)$.

TABLE 9.4

Proportions of population falling within various ranges of amounts of use for the distribution of recorded use reported by Schnaitter (Table 5.3, Column 4) and for a fitted geometric distribution.

Range of amounts of use	Observed proportion of population	Expected proportion of population
0-10	0.769	0.744
0-18	0.880	0.905
0-20	0.896	0.926
0-23	0.917	0.949
0-28	0.940	0.972
0-31	0.951	0.981
0-37	0.963	0.991

Parameter, q, of the geometric distribution estimated from the observed mean, m, using $q = m/(1 + m)$.

FIGURE 9.5

Using the geometric distribution to estimate levels of use not exceeded by a majority of potential users.

The geometric distribution function

$$p(r) = pq^r \quad (r = 0,1,2,\ldots),$$

where $q = 1 - p$, is a special case of the negative binomial distribution function (Figure 2.2)

$$p(r) = \binom{k+r-1}{k-1} p^k q^r \qquad (r = 0,1,2,\ldots)$$

which, in simple or modified form, has been used in the foregoing chapters to model distributions of recorded use by users. If the Poisson component or components of these distributions is ignored and the shape parameter, k, is set to unity, then a geometric distribution with mean, $m = q/(1 - q)$, will result. The single parameter, q, can be estimated by using $q = m/(1 + m)$.

The probability of observing s <u>or more</u> geometric-distributed events is

$$P(s) = p(q^s + q^{s+1} + q^{s+2} + \ldots)$$

$$= q^s.$$

In the library application, the proportion of potential users, a, who record s or more uses can be estimated by setting a equal to q^s and solving for s or a.

If the value of a is given, then we have

$$s = \log(a)/\log(q)$$

$$= \log(a)/\log[m/(1 + m)].$$

If s is given and set to three times the observed mean (i.e. $s = 3m$), then

FIGURE 9.5 (continued)

$$\log(a) = (3m)\log[m/(1 + m)].$$

For values of m between 1.6 and 350, a lies between 0.1 and 0.05. In general, therefore, only a minority of potential users (between 5% and 10%) would be expected to record three or more times mean use.

TABLE 9.5

Observed percentages of potential users recording less than the estimated level of use equalled or exceeded by 10% (90% level) and 5% (95% level) of potential users and observed percentages recording less than three times mean use for various data sets.

Data set		90% level	95% level	3(Mean)
		%	%	%
Table 6.10	Ritter	86.5	90.6	89.7
	Maxted	88.9	92.4	90.4
Table 6.11	Knapp	91.0	93.6	93.0
Table 6.12	Clayton	90.9	95.5	95.6
Table 6.13	Schnaitter W	88.5	92.3	91.5
	Schnaitter M	87.5	91.7	90.8
Table 3.2	Weeks 1-4	91.3	92.9	91.3
	Weeks 1-8	91.9	93.5	93.2
	Weeks 1-16	90.0	93.5	92.6
Table 6.14	Class A	92.2	93.5	93.5
	Class B	93.2	93.2	93.2
Table 6.15	Class C	89.9	94.9	94.9

90% and 95% levels of use estimated from geometric distributions fitted as shown in Figure 9.5.

9.4 TITLE USE BY A SAMPLE OF UNDERGRADUATE ECONOMISTS

Recorded uses by a sample of 20 UK second-year undergraduate economists during one academic session from 26 September to 14 June were examined to determine the extent to which users showed similar preferences in the titles they used from a short-loan collection. Potential users were sampled by taking the first 20 borrower numbers in a class of 86 students. These borrower numbers are represented below by the numbers 1 to 20. Altogether, 1200 uses and 700 title uses were recorded. The distributions of numbers of uses and numbers of title uses are shown in Table 9.2. Negative binomial distributions were fitted to both distributions using the maximum likelihood equation. Goodness of fit was tested with the chi-squared test and appeared adequate ($P \simeq 0.1$, uses; $P \simeq 0.3$, title uses). But, with only four cells and one degree of freedom, the test was poorly founded. The estimated negative binomial parameters were $k = 4.1$ and $p = 0.064$ for uses, and $k = 9.0$ and $p = 0.2045$ for title uses.

Amounts of use appeared to cumulate evenly through the session. The total number of uses in the first third of the session was almost exactly one third of the final total and the individual numbers of uses in the first third could have been random samples from the final individual totals (chi-squared test: $P = 0.67$). Negative binomial distributions were fitted to the distributions of use for one third and two thirds of the session as well as for the whole session. The estimated value of k appeared to change progressively, from 3.3 to 3.8 to 4.1, as the period of observation lengthened.

A total of 280 different titles were used by the 20 students during the session. (Different editions of the same title were not distinguished.) Table 9.6 shows the distribution of titles according to their numbers of users. Almost one half of the titles had only one user. On average, titles with greater numbers of users also received more uses by each user (Table 9.6, Column 3), but they were probably available for shorter periods of loan.

Clearly, no title was used by all the students, and only four titles were used by a majority of users. There is thus little evidence that users shared the same priorities in their use of material from the collection. It is possible, however, that some titles were equivalent, although this seems unlikely. Eighty-five different Dewey class numbers were represented among the 280 titles (class numbers differing by

TABLE 9.6

Distribution of numbers of users for titles in a short-loan collection used by a sample of 20 second-year undergraduate economists and mean number of uses per title per user.

Number of users	Number of titles	Mean use per user per title
1	131	1.2
2	55	1.4
3	28	1.2
4	23	1.5
5	19	2.0
6	8	1.9
7	4	2.8
8	6	2.6
9	2	2.2
11	2	2.9
12	2	2.2
Total	280	

Aggregate mean number of uses per title per user: 1.7.

geographical or '01' suffixes were not distinguished), but no class number was used by all the students, and only four classes were used by a majority of them. The most popular of these was used by sixteen students. At best, class numbers would only be a rough guide to the usefulness of particular titles for particular purposes, but the pattern of their use provides no evidence for shared preferences in the subjects for which students used short-loan books. (Of course, subject needs would also have been met from the main collection.)

Furthermore, it appears that the single-user titles were used not only by the heavy users (as would be expected if the students adopted similar priorities in their use of material), but by all users. The proportion of single-user titles used by the eight less-active users was not significantly different to that for the whole population (Figure 9.6).

The extremes of similarity and dissimilarity between pairs of users in their use of titles were more marked than would have occurred by chance. Table 9.7 shows a coefficient of similarity (called the Czekanowski coefficient by Clifford and Stephenson, 109:55) calculated for each pair of users. The user numbers in each pair are shown separated by a solidus. The coefficient represents the quotient obtained by dividing twice the number of titles which were common to both users in the pair by the sum of the title uses recorded by each of the users. The coefficient therefore takes values between 1.0 (all recorded titles used by both users) and 0 (no titles common to both users). Lists of the title numbers used by each user are given in Appendix D. Table 9.8 shows the relative frequency with which values of the coefficient occurred, and Table 9.9 shows the relative frequency with which values of the coefficient occur when titles are allocated at random to users (assuming the observed rates of use for both users and titles). Clearly the observed range of coefficients is wider than would occur by chance.

Twenty-one pairs of users had no titles in common and used completely dissimilar material. Each of these pairs included one of the eight less-active users (who recorded 31 or less title uses), but only one pair included two such users. Table 9.10 shows how these pairings are distributed over the 20 users. Table 9.11 lists the user numbers involved in the 21 dissimilar pairs. Users 3 or 20 are present in 16 out of the 21 pairs and in 12 cases pair with the same six users. These two users had some similarities; they used three titles (two class numbers) in common out of a possible total of eight titles. Two of the titles and both class numbers were well used, the titles being used by four and

FIGURE 9.6

Contingency table for the comparison of the numbers of title-uses recorded for single-user titles and the numbers of title-uses recorded for all other titles by the eight less-active and the twelve more-active users from the sample of 20 undergraduate economists.

	TITLE-USES		
	Single-user titles	Other titles	Totals
Less-active users*	39 (33.5)	140 (145.5)	179
More-active users	92 (97.5)	429 (423.5)	521
Totals	131	569	700

Expected numbers of title-uses are shown in brackets. Chi-squared statistic (with Yates' continuity correction†) = 1.23; $P \approx 0.25$. The null hypothesis that both groups of users record the use of similar proportions of single-user titles cannot be rejected.

---

*The less-active users recorded the use of between 8 and 31 titles, the more-active users between 38 and 53 titles (Table 9.2).

†The difference between observed and expected numbers was decreased in absolute value by ½ before squaring in order to improve the approximation of the sampling distribution of the calculated statistic to the continuous chi-squared distribution.

TABLE 9.7

Czekanowski coefficients of similarity (C) for recorded title uses by pairs of undergraduate economists (U) with zeros left blank.

U	C	U	C	U	C	U	C
1/2	0.24	1/3		1/4	0.19	1/5	0.07
1/6	0.10	1/7	0.31	1/8	0.03	1/9	0.37
1/10	0.36	1/11	0.49	1/12	0.23	1/13	0.18
1/14	0.09	1/15	0.13	1/16	0.47	1/17	0.23
1/18	0.26	1/19	0.09	1/20		2/3	
2/4	0.06	2/5		2/6	0.17	2/7	0.23
2/8		2/9	0.30	2/10	0.25	2/11	0.18
2/12	0.09	2/13	0.04	2/14	0.15	2/15	
2/16	0.29	2/17	0.18	2/18	0.08	2/19	0.12
2/20	0.12	3/4		3/5	0.21	3/6	0.08
3/7		3/8	0.44	3/9		3/10	0.04
3/11		3/12		3/13	0.13	3/14	0.12
3/15	0.19	3/16		3/17	0.04	3/18	0.07
3/19	0.16	3/20	0.15	4/5	0.03	4/6	0.07
4/7	0.14	4/8	0.13	4/9	0.14	4/10	0.09
4/11	0.13	4/12	0.54	4/13	0.34	4/14	0.03
4/15	0.32	4/16	0.17	4/17	0.12	4/18	0.30
4/19	0.08	4/20		5/6	0.30	5/7	0.03
5/8	0.31	5/9	0.07	5/10	0.23	5/11	0.09
5/12	0.03	5/13	0.28	5/14	0.26	5/15	0.14
5/16	0.06	5/17	0.24	5/18	0.05	5/19	-0.29
5/20	0.20	6/7	0.07	6/8	0.20	6/9	0.05
6/10	0.29	6/11	0.11	6/12	0.05	6/13	0.33
6/14	0.30	6/15	0.18	6/16	0.07	6/17	0.33
6/18	0.06	6/19	0.26	6/20	0.17	7/8	0.10
7/9	0.45	7/10	0.30	7/11	0.33	7/12	0.15
7/13	0.15	7/14	0.03	7/15	0.03	7/16	0.45
7/17	0.29	7/18	0.12	7/19		7/20	
8/9	0.03	8/10	0.12	8/11	0.03	8/12	0.11
8/13	0.35	8/14	0.18	8/15	0.28	8/16	0.06
8/17	0.18	8/18	0.15	8/19	0.17	8/20	0.12
9/10	0.58	9/11	0.41	9/12	0.23	9/13	0.08
9/14	0.03	9/15	0.12	9/16	0.60	9/17	0.33
9/18	0.25	9/19		9/20		10/11	0.36
10/12	0.16	10/13	0.25	10/14	0.11	10/15	0.16
10/16	0.39	10/17	0.44	10/18	0.20	10/19	0.10
10/20	0.03	11/12	0.17	11/13	0.12	11/14	0.08
11/15	0.06	11/16	0.46	11/17	0.31	11/18	0.24
11/19	0.05	11/20		12/13	0.29	12/14	0.03
12/15	0.23	12/16	0.21	12/17	0.18	12/18	0.37
12/19	0.09	12/20	0.06	13/14	0.16	13/15	0.32
13/16	0.11	13/17	0.29	13/18	0.20	13/19	0.21
13/20	0.03	14/15	0.08	14/16	0.05	14/17	0.09
14/18	0.04	14/19	0.40	14/20	0.39	15/16	0.08
15/17	0.13	15/18	0.17	15/19	0.08	15/20	0.04
16/17	0.30	16/18	0.23	16/19	0.03	16/20	
17/18	0.13	17/19	0.15	17/20		18/19	0.08
18/20		19/20	0.37				

TABLE 9.8

Relative frequency of values of coefficient of similarity for recorded title use by pairs of undergraduate economists.

Values of coefficient	Number of economist pairs
0	21
0.01 - 0.04	19
0.05 - 0.09	33
0.10 - 0.14	25
0.15 - 0.19	25
0.20 - 0.24	17
0.25 - 0.29	14
0.30 - 0.34	16
0.35 - 0.39	8
0.40 - 0.44	4
0.45 - 0.49	5
0.50 - 0.54	1
0.55 - 0.59	1
0.60+	1
Total pairs	190

TABLE 9.9

Relative frequency of values of coefficient of similarity for 280 titles allocated at random to undergraduate economists.

Values of coefficient	Number of economist pairs I	II	III	IV	V
0	6	2	1	2	6
0.01 - 0.04	5	9	7	7	4
0.05 - 0.09	33	32	29	25	22
0.10 - 0.14	37	46	44	45	51
0.15 - 0.19	47	33	54	53	44
0.20 - 0.24	30	36	31	35	36
0.25 - 0.29	23	26	18	16	23
0.30 - 0.34	9	5	4	7	3
0.35 - 0.39	0	1	2	0	1
Total pairs	190	190	190	190	190

Titles were assigned to user numbers by calling random numbers (Section 4.2.1.3). Five examples (I to V) are shown.

TABLE 9.10

Number of titles used by undergraduate economists and number of other economists who used no titles in common.

User number	Number of titles used	Number of dissimilar users
3	8	8
8	19	1
5	20	1
18	22	1
15	24	1
14	26	0
19	29	2
20	31	8
12	38	1
13	38	0
17	38	1
1	40	2
6	41	0
9	41	3
4	44	2
7	44	3
11	46	2
10	49	0
16	49	2
2	53	4

There were 21 pairs of users with no titles in common. Each is counted twice in column 3, which therefore sums to 42.

TABLE 9.11

Pairs of undergraduate economists who used no titles in common.

User numbers
3/1
3/2
3/4
3/7
3/9
3/11
3/12
3/16
5/2
8/2
15/2
18/20
19/7
19/9
20/1
20/4
20/7
20/9
20/11
20/16
20/17

seven other users respectively. These two users do not seem to have chosen unusual material, therefore.

The eight pairs of users with the highest coefficients of similarity were drawn only from the more active 12 users. (Of course, it is slightly easier for higher values of the coefficient to occur when users record similar numbers of title uses; certainly the range of numbers of title uses was smaller for the more active users than for the less active users.) Out of the next 12 highest coefficients, six pairs of users were drawn exclusively from the more active users and four pairs from the less active users. In all, seven pairs of users had closely similar numbers of title uses. In this group also, with the ninth highest coefficient, was a pair comprising the two least active users. They shared six out of a possible eight title uses and five out of a possible six classes. (Clearly, had their individual totals of title uses been more similar, the value of the coefficient might have been substantially higher.) Nonetheless, title use by other less active users did not correspond as closely and again there seems no evidence that the economists worked through material from the collection according to a common set of priorities: as Table 9.12 shows, the less-active users did not restrict themselves to the most popular titles; they also used numbers of less-popular (single-user) titles.

The pattern of use for the economists was certainly not similar to that shown in Figure 9.1. But neither, on the evidence of the distribution of the values of the coefficient of similarity, was it random like that shown in Figure 9.2. It is possible that there was widespread collaboration among users (both within and outside the sample) in their use of titles. Perhaps some of the more and some of the less active users (in terms of recorded use) differed in their title use because they formed the most permanent collaborative teams. Of course, the purchase of books, and the sharing of purchased books, might also serve to confuse the pattern of recorded library use.

These possibilities cannot be tested in the record of use. It is interesting to speculate, however, upon the effect which widespread collaboration among students would have on the apparent popularity of books. In Meier's survey (66), almost one half of all the reported use of library material was estimated to have resulted from the unrecorded exchange of highly-sought material between users. In such a situation, misleading gaps in the record of use would almost certainly occur. To

TABLE 9.12

Number of titles used, number of single-user titles used and number of the sixteen most-used titles used by the undergraduate economists.

User number	Number of titles used	Number of single-user titles used	Number of most-used titles used
3	8	1	2
8	19	3	4
5	20	5	5
18	22	5	6
15	24	4	3
14	26	5	4
19	29	4	4
20	31	12	1
17	38	4	11
13	38	7	8
12	38	8	7
1	40	6	12
9	41	2	11
6	41	8	6
4	44	12	5
7	44	12	8
11	46	7	11
10	49	5	14
16	49	5	12
2	53	16	6
Total	700	131	140

assess the effect of such a level of collaboration upon the recorded use of individual titles, a modest simulation was conducted.

9.5 EFFECT OF UNRECORDED USE UPON THE RECORD OF USE

The simple computer simulation program reproduced in Appendix E represents the hypothetical situation described in Section 9.5.1 below. The parameters of the program were arbitrarily, but not unrealistically set to yield roughly equal amounts of recorded and unrecorded (collaborative) use. The titles for which use was simulated differed in popularity but were ranked consistently for preference by all the users. If the simulation confirmed what Meier implied, that individual titles in a library collection might receive different proportions of recorded and unrecorded use depending upon their popularity, then a possible explanation of an apparent lack of common preferences among users might be advanced.

9.5.1 Description of the simulation of recorded and unrecorded library use

A group of ten students was assumed to have been set an assignment which required the use of up to ten titles from a library collection over a period of ten days. Each user tried to use the titles in the same order of preference, but having used each of the four most popular titles, each user would, on subsequent visits to the library, try to re-use these popular titles as well as titles further down the list of preference. The most popular title could be re-used profitably up to twice by any user, and the next three popular titles could be re-used once before losing their attraction. No attempt would be made to re-use any of the other titles. Three copies of the most popular title were available and two copies of the next two most popular titles. All other titles were available only in one copy.

The students were assumed to vary both with respect to their rate of visiting the collection and also in their inclination to collaborate. Three students were expected to visit on two occasions during the ten-day period, four students on six occasions and three students on ten occasions. A visit could take place on up to three occasions each day. Three students always sought at least one title from their colleagues before a visit to the library, and these, and four other students, tended to seek out a title from colleagues if they failed to obtain it first-hand from the collection. Their probability of attempting to collaborate rose linearly during the time period to near certainty on the final day. If collaboration was attempted, the probability of an unrecorded use was 0.5. Three students never attempted to collaborate but collaborative use of a title which they held would occur with a probability of 0.15.

During any visit and subsequent attempts at collaboration and during any pre-visit attempt at collaboration, an average of rather more than three titles would be sought; the probability of each subsequent attempt to find a title after the first being 0.71. When titles were borrowed from the collection they were retained on loan for an average of $4^2/_3$ days. The distribution of retention times was roughly normal with a range from $^1/_3$ day to 7 days.

9.5.2 Result of the simulation

In order to decide an event during the running of the simulation program (for example, whether a visit or a collaboration occurred, or how long a loan lasted), a new pseudo-random number between zero and one was called and compared to the probability assigned to the event. The distribution of these pseudo-random numbers was examined and found to be acceptably uniform (Appendix E). In the initial runs of the simulation program, the parameters were adjusted until roughly equal proportions of recorded and unrecorded use resulted. There was no other evidence save that of Meier to suggest this ratio, but it was not felt unlikely that the economists could have collaborated to this extent.

The simulation program was then run three times with the results shown in Table 9.13, where the numbers of recorded uses and the total number of uses for each of the ten titles is listed, together with the aggregates for all three runs. While recorded use accounts for about half of all use for most titles, it is clear that availability in the face of demand appears to modify this proportion in two cases. Two thirds of the aggregate total use of Title 4, which is popular but represented in only one copy, is unrecorded use. On the other hand, only two fifths of the aggregate total use of Title 1 is unrecorded, apparently because, although the most popular title, extra copies of it were available and would have been found on the shelf more often. It is noticeable that the difference between the aggregate total use of these two titles is exactly the difference between the amounts of recorded use.

TABLE 9.13

Recorded use (R) and total use (T) of ten titles in three runs of the simulation of library use and aggregates for all three runs.

Title*	R	T	R	T	R	T	Aggregate R	Aggregate T
1 (3)	10	19	13	19	10	17	33	55
2 (2)	6	13	6	13	8	13	20	39
3 (2)	7	14	6	13	5	12	18	39
4 (1)	3	13	3	10	4	9	10	32
5 (1)	3	6	4	7	2	6	9	19
6 (1)	4	7	2	5	3	6	9	18
7 (1)	3	7	3	5	4	6	10	18
8 (1)	3	5	2	6	2	4	7	15
9 (1)	3	6	2	6	3	4	8	16
10 (1)	3	4	2	6	3	4	8	14

*Number of copies shown in brackets.

Figure 9.7 shows a typical plot of the recorded title uses resulting from a simulation run. There is a similarity to the plot in Figure 9.2, even though the gaps in Figure 9.7 are caused not by chance non-use, but only by non-availability at the shelf.

9.6 UNRECORDED USE

If, as seems likely, a proportion of the economists' use of library material went unrecorded and if, as also seems likely, this proportion would have been largest for titles where demand most exceeded supply, then a possible explanation for at least part of the incoherence of the economists' record of use is suggested. The extent to which the economists were also genuine diversifiers in their use of library material remains unknown, however: it is impossible to decide from the record of use. Once again the limits of the data have been reached; much supplementary data would be required to pursue the investigation further.

If the users had been engineers sharing a well-defined syllabus, then their pattern of use may have been more like that in Figure 9.1, but the problem of unrecorded use would still remain. Within a group of users sharing a common task, the unrecorded exchange of books seems inevitable. The poorer the availability, or the greater the incentive to use only particular titles, then the more important will collaborative exchange become. Indeed, for some users it may possibly be a more direct and congenial alternative to formal library circulation whatever the circumstances. The importance of unrecorded exchange, noted by Meier, seems confirmed by student responses in the surveys reported by COPOL (79). Unrecorded exchange is not, however, included by Warwick (110) in the model of user behaviour which he uses to suggest a policy for the economic duplication of recommended texts; but it is not clear whether this was a deliberate omission or not.

9.7 SUMMARY

Two hypothetical patterns of user behaviour would produce discernable patterns of use in the collection.

If the users selected titles entirely at haphazard, then the form of the distribution of use over the titles in the collection would be demonstrably random and could be expected to persist over time, albeit with never the same individual titles consistently receiving the same

FIGURE 9.7

Pattern of use when users share the same priorities in using library material but compete for material and exchange material already on loan. Ten books (ranked 1 to 10 according to the amount of use made of them) are shown stacked vertically above their users (denoted A to J in order of decreasing activity). Use was simulated as described in Section 9.6.

10	10								
9	9					9			
8		8	8						
7	7	7							
6	6	6							
5	5	5	5						
4	4	4						4	
3	3	3	3	3					
2			2	2	2		2		
1		1	1	1	1	1	1		
A	B	C	D	E	F	G	H	I	J

Books

Potential Users

amounts of use. If, on the other hand, all the users shared the same preferences for library material and were able to record the use of any title they sought, then the relative amount of use received by each title would vary little over time and could be directly related to the activity of the users. For positively-skewed distributions of user activity, a level of title provision expressed in multiples of mean title use could then be related to a proportion of potential users completely satisfied. This relationship depends, however, not only on the validity of the assumption that users are like-minded in their choice of material from the collection, but also upon the validity of the assumptions which are necessarily adopted in utilizing readily-available data. For example, the use of recorded transaction data involves the assumptions that recorded use is proportional to total use and that total use is proportional to demand. In the case investigated above, none of these assumptions could be verified. It may be that for a well-defined syllabus, in science or engineering perhaps, the first assumption will be justified. But in all cases, it would seem that supplementary data on unrecorded use would need to be available, first to verify the first assumption and then, if it were valid, to make use of the resulting relationship between user activity and title use.

CHAPTER 10

CONCLUSION

The aim of the investigation reported above was to describe and analyse distributions of recorded use for students using academic library collections. It was acknowledged that only counts of issue (circulation) transactions would yield sufficient data for the statistical analyses which were proposed. The shortcomings of such data were clearly recognized, and indeed proved to limit the extent to which productive analyses could be performed. (A count of recorded uses is at best an imperfect index of the relative differences between individual users in their total amounts of library use or in their exposure to library materials.)

In the eyes of many librarians quantitative or statistical studies such as this have little value. Of course, it is never easy to apply information on past use predictively, and to apply it with enough detail and confidence to dispense with human judgement is clearly impossible, even in a situation where new initiatives are excluded. What was sought here, however, was not bibliothecal determinism, but merely a quantitative background against which to set intuition and experience, or merely a rough model with which to test hypotheses. Neither objective implies a purely reactive or mechanistic approach to management problems, therefore.

It seems reasonable to assume that individual students will seek out different amounts of information in performing their assigned tasks, and will adopt different strategies in doing so. Persistent differences among multidisciplinary groups of students, because of the different nature of their courses or the different expectations of their teachers, are understandable, of course. But persistent differences among peers are also understandable if information-seeking behaviour is motivated for the satisfaction of individual psychological as well as cognitive needs. Information seeking in libraries is a particularly laborious business and requires an industriousness which is perhaps not widely prized, at least among adolescents (111). Consequently, although librarians in academic libraries may wish to see their users adopting efficient, energetic, critical and, above all, library-centred methods of information gathering, they will often it seems find that, except in highly regimented situations involving great motivation or stimulation, the response of the students is varied - just as it is in other tests of scholarship, aptitude or enthusiasm. The aim in the preceding chapters has merely been to investigate, using such data as it was possible to procure, the form of the distribution of the resulting 'scores'.

Observed distributions of recorded use were approximated using a fairly simple mathematical function relating numbers of users to given amounts of use. These two distributions were not generated by the same mechanism, of course, but in some respects the superficial pattern appeared similar. No other factor save individual propensity for library use among the users seemed likely to have been predominantly responsible for the form of the observed distributions; and no other factor governed the form of the approximating distributions. No doubt, local conditions in particular libraries favoured the predilections of different users, but the observed distributions of use did not appear to have been the artefact of particular conditions of supply and demand, or of particular methods of provision.

To the extent that numbers of recorded uses tend to increase as numbers of attempts at use increase, such distributions of recorded use could also be expected to reflect differences in rates of recourse to the library. Substantial individual variations in the relationship between total activity and recorded use are likely, of course, although no factor which would yield a large systematic variation was identified in the work reported here.

Ample scope remains for extending the work performed here and for the testing of alternative types of probability distribution against observed distributions of recorded use. Whether regarded as data-fitting or modelling, such analyses can be performed perfectly justifiably for their own sake and in their own right. Useful practical applications for the information derived from the analyses are less arguable and certainly were not obvious here. The parameters of the fitted distribution of use were too numerous and depended too greatly upon local factors for universal rules of thumb to be distilled out. Nonetheless, for positively-skewed distributions, simplifications of the distribution allowed some interesting speculations: on methods of predicting proportions of non-users; on the connection between mean use and the rate of use observed for the most active users; and therefore on the number of titles required to satisfy given proportions of users. In pursuing these speculations, some limitations in the data and in the fitted distribution were encountered. Accurate extrapolation was hindered by the discrepancy between the observed and expected variances of the rate of use by users; but the severity of the assumption that the environment surrounding the features encompassed by the model remained constant would also be important and was thought to limit the extent to which it was worth refining the fit of any model. Similarly, the incompleteness of the record of use may have prevented the connection of the pattern of user activity to the pattern of uptake from the collection, but it is likely that the users in the sample studied were naturally diverse enough in their choice of library material to prevent the connection anyway.

Consequently, although a further modification to the already modified model was noted, it was felt unprofitable to seek a greater sophistication in the model than the data or assumptions permitted. (In itself, nonetheless, the exercise would be of interest since it would entail further work in modelling fundamental phenomena, especially the incidence of use over time for individual users.) Simplification of the model yielded a method of extrapolation and a method of estimating maximum majority use. For extemporized work, in the absence of better information, these may prove useful. At present, however, it seems most likely that the main value of quantitative (and qualitative) studies of user behaviour will lie not in the incorporation of their techniques or findings into management procedures, but in their contribution to a gradually deepening awareness among library managers of the subtly complex nature of library use.

APPENDIX A

Weekly amounts of recorded use by 241 students using a short-loan collection over a period of 17 'weeks' (see Section 3.1)

Weekly numbers of recorded uses are listed for all 241 students on the following five pairs of sheets. Students are listed 60 at a time. On the first (summary) sheet are shown:

Column 1: Running number, 1 to 60;
Column 2: Total number of recorded uses in the 17 'weeks';
Column 3: Number of recorded uses in Weeks 1 to 9 and 11 to 16.

On the second sheet is shown the array of weekly numbers of recorded uses:

Column 1: Running number, 1 to 60;
Columns 2 to 18; Numbers of uses recorded in each of the Weeks 1 to 17.

1	15	12
2	9	9
3	8	6
4	1	1
5	18	16
6	7	7
7	4	4
8	6	2
9	4	4
10	12	11
11	156	139
12	6	6
13	13	12
14	4	3
15	7	7
16	13	13
17	6	6
18	1	1
19	9	8
20	2	2
21	2	2
22	7	7
23	4	3
24	8	8
25	3	3
26	8	5
27	12	12
28	12	12
29	12	11
30	31	27
31	35	29
32	9	3
33	1	1
34	14	13
35	14	13
36	1	1
37	12	10
38	2	1
39	21	18
40	10	10
41	2	2
42	1	1
43	15	13
44	3	3
45	4	4
46	17	12
47	2	2
48	1	1
49	3	3
50	9	9
51	8	8
52	1	1
53	19	17
54	13	13
55	1	1
56	4	4
57	4	4
58	24	19
59	3	3
60	19	19

1	1	0	1	0	2	0	1	1	0	3	0	1	2	1	1	1	0
2	2	2	0	1	1	1	0	0	0	0	0	1	1	0	0	0	0
3	0	2	0	0	0	3	1	0	0	2	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
5	0	0	0	0	0	1	2	1	1	0	0	1	4	5	0	1	2
6	0	1	0	0	0	0	0	0	0	0	3	1	1	0	1	0	0
7	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	3	0	0	0	0	0	1	1
9	0	0	0	0	1	1	0	2	0	0	0	0	0	0	0	0	0
10	0	0	0	2	0	0	3	2	0	1	2	1	0	1	0	0	0
11	6	10	4	8	8	15	15	5	11	12	5	5	8	6	22	11	5
12	0	0	0	0	0	2	1	0	0	0	1	0	0	1	0	1	0
13	1	2	0	0	0	2	0	3	0	1	0	0	2	2	0	0	0
14	0	1	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	4	1	0	0	0	0	1	1	0	0
16	0	0	2	0	0	1	3	0	0	0	0	2	1	3	0	1	0
17	0	0	0	0	0	1	0	0	0	0	0	1	0	0	3	1	0
18	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	1	0	1	1	0	3	0	1	1	1	0
20	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
22	0	1	0	3	0	1	0	0	0	0	1	1	0	0	0	0	0
23	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1
24	0	0	0	0	0	1	0	1	5	0	0	0	0	0	0	1	0
25	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
26	0	0	0	0	0	2	1	0	0	3	2	0	0	0	0	0	0
27	1	1	1	1	1	2	0	0	1	0	0	0	0	2	2	0	0
28	2	1	1	1	1	1	3	0	0	0	1	0	0	0	1	0	0
29	0	0	1	1	1	1	1	0	0	1	2	2	0	1	1	0	0
30	1	1	2	0	8	6	3	2	4	4	0	0	0	0	0	0	0
31	1	3	4	6	0	2	0	3	2	6	1	1	2	2	2	0	0
32	0	0	0	1	0	0	0	0	0	2	2	0	0	0	0	0	4
33	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
34	2	0	0	0	3	0	0	4	0	0	1	1	1	0	0	1	1
35	0	2	0	1	1	0	1	1	0	1	4	1	1	1	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
37	0	0	4	1	0	0	0	0	0	2	0	2	0	2	1	0	0
38	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
39	0	0	1	1	0	2	3	3	0	3	3	2	0	2	0	1	0
40	0	0	0	1	1	0	0	2	1	0	1	0	1	2	0	1	0
41	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
42	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	5	0	2	0	2	0	0	1	2	2	1	0	0	0	0	0	0
44	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0
45	0	0	1	0	0	2	1	0	0	0	0	0	0	0	0	0	0
46	0	0	0	5	1	2	0	0	1	3	1	0	1	0	0	1	2
47	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0
50	0	1	0	3	1	0	0	1	0	0	0	0	0	1	2	0	0
51	0	0	0	0	0	0	2	0	0	0	0	3	3	0	0	0	0
52	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
53	0	4	0	3	0	2	0	0	0	2	2	0	4	0	2	0	0
54	0	0	1	1	0	2	0	0	2	0	2	4	0	0	0	1	0
55	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
56	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0
57	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
58	0	1	1	4	0	2	0	0	0	5	4	1	2	1	3	0	0
59	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0
60	4	0	4	2	3	1	2	0	0	0	0	0	0	1	2	0	0

1	3	3
2	1	0
3	11	10
4	3	0
5	31	29
6	2	2
7	17	11
8	5	5
9	5	3
10	4	4
11	5	5
12	3	3
13	2	2
14	1	1
15	6	6
16	9	9
17	9	9
18	22	18
19	26	26
20	9	7
21	1	1
22	8	7
23	3	1
24	2	2
25	9	7
26	30	22
27	1	1
28	1	1
29	24	19
30	2	2
31	12	11
32	10	6
33	18	16
34	2	2
35	30	29
36	18	14
37	1	1
38	2	2
39	4	0
40	11	10
41	6	6
42	1	1
43	4	3
44	4	2
45	3	3
46	7	7
47	30	30
48	12	8
49	8	8
50	9	8
51	4	4
52	2	2
53	4	4
54	1	1
55	4	4
56	3	3
57	20	19
58	10	10
59	13	12
60	19	15

1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3	0	1	3	1	0	0	0	0	2	1	0	1	1	1	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
5	1	5	0	3	1	4	5	3	2	0	1	0	0	2	2	0	2
6	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	8	6	0	0	1	2	0	0	0
8	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	0	0
9	0	0	0	0	1	0	0	0	0	2	1	0	0	0	0	1	0
10	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
11	0	1	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0
12	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0
13	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
15	0	0	0	2	0	0	0	3	0	0	0	0	1	0	0	0	0
16	3	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	1	1	3	2	1	0	0	0	0	1	0	0	0	0
18	2	1	2	2	0	1	0	1	0	0	1	1	1	2	1	3	4
19	5	0	2	4	3	2	3	0	1	0	0	1	0	1	0	4	0
20	0	0	0	1	1	0	0	0	1	2	3	0	0	0	1	0	0
21	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
22	0	0	0	1	1	0	1	0	2	1	0	2	0	0	0	0	0
23	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
25	0	3	0	0	0	0	1	2	1	2	0	0	0	0	0	0	0
26	2	3	0	0	1	1	0	1	0	3	1	4	0	2	3	4	5
27	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
28	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
29	1	0	0	0	2	3	2	2	2	5	1	1	3	1	0	1	0
30	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
31	0	0	1	1	0	0	1	0	0	1	0	2	1	2	3	0	0
32	0	0	0	0	1	5	0	0	0	4	0	0	0	0	0	0	0
33	0	0	1	0	1	2	1	1	1	2	1	1	2	0	4	1	0
34	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
35	2	0	0	3	3	0	1	2	0	0	1	6	6	0	3	2	1
36	0	0	4	0	0	0	0	0	1	4	0	4	3	0	2	0	0
37	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
38	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
40	0	0	1	0	2	0	0	1	0	1	1	3	1	1	0	0	0
41	0	0	0	0	0	1	0	0	1	0	0	1	1	1	1	0	0
42	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
43	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0
44	0	0	0	0	0	1	0	0	0	2	1	0	0	0	0	0	0
45	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	2	1	0	0	0	2	0	1	0	0	1	0
47	2	3	2	0	1	6	2	2	5	0	0	2	0	0	3	2	0
48	0	1	3	0	0	0	0	2	0	4	0	0	0	2	0	0	0
49	0	3	0	1	1	0	0	2	1	0	0	0	0	0	0	0	0
50	2	1	1	0	0	0	0	0	4	1	0	0	0	0	0	0	0
51	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0
52	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
53	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0
54	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
55	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0
56	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
57	2	0	0	1	1	0	1	4	3	1	0	3	1	3	0	0	0
58	0	0	2	0	0	1	0	0	0	0	1	1	1	2	1	1	0
59	0	0	0	0	3	1	2	1	2	0	0	0	1	0	1	1	1
60	3	1	0	0	0	0	0	0	2	4	0	1	2	3	2	1	0

1	7	7
2	2	2
3	12	11
4	25	21
5	2	2
6	31	26
7	57	47
8	29	27
9	44	42
10	8	8
11	10	10
12	5	4
13	11	6
14	4	3
15	8	8
16	3	3
17	1	1
18	10	9
19	3	3
20	10	6
21	7	7
22	3	3
23	11	7
24	10	9
25	2	2
26	8	8
27	2	2
28	61	57
29	4	4
30	5	5
31	2	2
32	6	5
33	2	2
34	6	6
35	18	17
36	13	9
37	8	8
38	1	1
39	3	3
40	3	3
41	4	4
42	15	8
43	1	0
44	1	1
45	12	11
46	3	3
47	5	5
48	6	5
49	17	17
50	10	9
51	9	6
52	6	6
53	88	80
54	1	1
55	4	4
56	1	1
57	3	2
58	7	7
59	2	2
60	56	52

1	0	0	0	1	2	0	0	1	0	0	2	0	0	0	0	1	0
2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	3	1	0	0	1	2	1	1	1	0	1	0	0	0	0
4	2	0	0	2	0	2	1	2	1	4	3	1	2	2	1	2	0
5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
6	0	2	1	0	1	0	0	0	3	4	2	6	0	5	3	3	1
7	0	5	4	4	0	7	8	0	6	10	0	4	0	6	1	2	0
8	0	1	1	1	1	0	0	0	2	0	2	2	7	3	4	3	2
9	0	0	0	2	2	4	6	3	4	2	1	6	5	1	3	5	0
10	0	0	1	1	1	1	0	0	0	0	0	1	3	0	0	0	0
11	2	0	1	0	0	0	0	0	1	0	2	1	1	1	1	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	1
13	0	1	0	1	0	0	2	0	0	5	0	0	2	0	0	0	0
14	0	0	0	0	0	0	0	0	0	1	1	0	2	0	0	0	0
15	0	0	0	0	0	0	1	1	3	0	1	0	1	0	1	0	0
16	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
18	1	1	0	0	0	0	0	1	1	1	1	0	2	1	1	0	0
19	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	1	0	3	0	4	0	2	0	0	0	0	0
21	1	3	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0
23	0	0	0	0	0	0	2	2	1	4	0	0	1	1	0	0	0
24	0	1	0	0	0	2	0	0	1	1	0	1	1	1	2	0	0
25	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
26	0	0	0	0	2	0	0	0	3	0	3	0	0	0	0	0	0
27	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	4	4	3	9	2	2	2	2	6	2	6	6	4	3	3	1	2
29	0	1	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	0
31	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
32	0	1	0	0	2	0	2	0	0	1	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
34	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
35	1	0	1	2	1	0	0	2	1	0	0	1	2	2	4	0	1
36	0	1	1	1	0	0	1	1	1	4	0	0	1	0	1	1	0
37	0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	3	0
38	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0
40	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
41	0	0	0	0	1	1	0	0	2	0	0	0	0	0	0	0	0
42	0	0	0	0	0	2	0	4	0	7	0	0	0	0	2	0	0
43	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
45	0	0	0	0	0	0	0	0	0	1	0	5	0	1	5	0	0
46	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0
47	1	1	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0
48	1	0	0	1	2	0	0	1	0	1	0	0	0	0	0	0	0
49	0	0	0	2	2	3	1	4	0	0	2	1	0	0	1	1	0
50	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	3	1
51	0	3	0	1	1	0	0	1	0	3	0	0	0	0	0	0	0
52	0	0	0	1	0	0	0	0	0	0	0	2	2	0	0	1	0
53	6	4	5	9	7	1	5	6	3	8	5	8	7	6	5	3	0
54	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
55	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
56	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
57	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
58	1	0	0	2	0	0	0	1	1	0	0	0	1	0	1	0	0
59	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	6	2	3	2	3	0	2	2	0	4	7	0	8	8	6	3	0

1	9	7
2	47	42
3	11	11
4	2	2
5	26	24
6	60	57
7	19	15
8	7	5
9	16	16
10	4	4
11	5	5
12	43	38
13	6	4
14	4	3
15	1	1
16	21	18
17	7	6
18	3	3
19	6	5
20	10	10
21	4	3
22	2	2
23	16	11
24	4	4
25	2	2
26	3	3
27	34	34
28	17	15
29	1	1
30	3	3
31	1	1
32	3	3
33	3	3
34	23	23
35	3	3
36	1	1
37	1	0
38	14	12
39	4	4
40	1	1
41	21	21
42	7	7
43	21	21
44	3	2
45	2	2
46	2	2
47	1	1
48	4	3
49	18	14
50	6	6
51	8	8
52	40	32
53	1	1
54	28	28
55	1	1
56	3	0
57	14	13
58	5	5
59	4	3
60	11	10

1	2	1	0	0	0	1	1	0	1	1	0	0	0	0	1	0	1
2	1	2	1	2	2	7	2	3	0	4	1	4	8	4	2	3	1
3	0	0	0	4	1	1	0	3	0	0	0	0	2	0	0	0	0
4	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5	1	0	2	5	2	2	1	2	1	1	3	1	0	0	2	2	1
6	3	3	1	3	3	4	8	2	3	3	6	4	4	3	4	6	0
7	1	4	1	2	1	1	0	1	0	4	3	0	0	0	1	0	0
8	3	0	0	0	0	0	0	1	0	2	0	0	1	0	0	0	0
9	0	2	3	1	3	0	1	1	1	0	0	0	0	2	2	0	0
10	0	0	0	0	0	0	2	0	1	0	0	0	1	0	0	0	0
11	0	1	1	0	0	0	0	0	0	0	2	0	0	0	0	1	0
12	0	0	0	8	1	2	11	6	0	5	1	1	2	2	1	3	0
13	0	2	1	0	0	0	0	0	1	2	0	0	0	0	0	0	0
14	0	2	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
16	2	3	0	0	2	0	0	0	1	3	5	3	2	0	0	0	0
17	0	0	0	0	0	3	1	1	0	0	0	0	0	1	0	0	1
18	0	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
19	0	0	0	0	0	0	1	2	0	0	0	0	2	0	0	0	1
20	0	0	0	1	1	1	0	0	0	0	0	2	1	1	1	2	0
21	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
23	0	1	0	0	0	1	0	2	0	4	2	1	1	2	0	1	1
24	0	0	0	2	0	0	1	1	0	0	0	0	0	0	0	0	0
25	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
26	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	1	6	2	1	4	5	2	1	3	0	2	0	3	0	3	1	0
28	4	3	1	1	0	0	0	0	0	2	1	0	0	0	2	3	0
29	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
30	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
31	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
33	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	3	1	0	0	2	0	0	0	2	4	5	2	3	1	0
35	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0
36	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
38	0	0	1	1	4	0	1	0	3	2	0	1	0	1	0	0	0
39	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
41	1	0	1	1	1	5	1	1	3	0	0	2	1	2	1	1	0
42	2	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
43	0	3	3	5	1	3	0	0	0	0	0	0	3	0	0	3	0
44	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0
45	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
46	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
47	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	0	0	0	0	0	0	0	0	2	1	0	0	0	1	0	0	0
49	1	0	2	2	1	0	0	2	0	4	2	0	2	1	1	0	0
50	0	0	0	0	1	0	0	2	1	0	0	0	0	0	0	2	0
51	0	0	1	1	0	0	2	0	0	0	0	0	1	1	1	1	0
52	3	2	3	1	3	3	2	0	0	7	4	3	1	2	3	2	1
53	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
54	0	0	2	3	1	5	5	3	1	0	0	2	5	0	0	1	0
55	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
56	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
57	1	0	3	0	1	1	3	0	4	1	0	0	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0	0	3	0	0	2	0	0
59	1	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0
60	0	0	0	0	3	2	0	0	3	0	0	0	0	0	0	2	1

1 5 4

1	0	0	0	0	0	3	0	0	0	1	0	0	1	0	0	0	0

APPENDIX B

Computer program to simulate the results of competition among 100 students attempting to use potentially useful library material over a period of 50 days

CONTENTS

B.1 Description of simulation program

The program for a single run of the simulation is shown overleaf. It is written in Commodore Microsoft BASIC and was run on a Commodore 3032 PET microcomputer. The identities of the variables and constants used in the program are listed on the page following.

B.1.1 Program summary

Lines 10 to 16 initialize the variables and set the parameters.

Lines 30 to 97 are subroutines for: deciding the loan period for an issued book (30-32); totalling the numbers of visits and revisits (40-42); recording the number of uses per visit (50); adjusting the spacing in printing (60-75); setting the tabs on the printer (80-88); graduating the distribution of probability of visit (90-97).

From line 100 to 199, the program loops through 50 'days'. In lines 100 to 105, the numbers of useful books and useful books available are updated and recorded, and the numbers of shorter-loan books are updated. From line 110 to 199, the program loops through the 100 users, deciding whether a visit occurs (110-115); whether a use occurs (120) and if so, with what length of loan (125-130); whether another attempt to use is made (140-150); whether a substitute book is found (160-170); whether a second visit occurs in the same day [for the most frequent visitors under the geometric distribution of rates of visit] (180).

In lines 200 to 499, the summary tables, 'COLLECTION', 'USAGE' and 'DISTRIBUTION', are printed out. At line 499 the program terminates.

In lines 530 to 930, the values of various parameters are requested. The distribution of rate of visit is graduated starting from lines 550, 570, 590 and 610 depending upon the type of distribution stipulated in lines 510-520. From line 650 to 845, the first summary table, 'USERS', is printed if required. In lines 700 and 705, the probabilities of further attempt are allocated across the user numbers, and in lines 720 to 760 users are assigned their types. In lines 850 to 860, daily probabilities of visit are calculated for each user.

In line 990, the proportions of shorter-loan books are converted into numbers (which become integers in the routine at line 105).

Lines 992 to 998 provide a summary of the run on the screen if the option to print out a summary is not taken.

In several lines, displays are put up on the screen (i.e. using the 'print' rather than the 'print£2,' command) in order to confirm values that have been input, monitor progress or to aid calibration.

B.1.2 Simulation program

```
10 i=0:a=0:c=0:d=0:j=0:b=0:z=0:q=50:h=100:p=1:e=0:f=0:g=0:k=0:m=0:n=0:o=0:l=0
15 dimv(100):dimx(100):dimy(55):dimr(100):dims(100):dimu(100):dimw(100)
16 dimt(100):a=rnd(i):goto500
30 z=rnd(p):ifz<f/atheno=o+p:f=f-p:x(71)=x(71)+p:return
31 ifz<(g+f)/atheny(j+5)=y(j+5)+p:g=g-p:x(72)=x(72)+p:return
32 x(j+20)=x(j+20)+p:x(73)=x(73)+p:return
40 forz=ptok
41 y(0)=y(0)+p:r(i)=r(i)+p:v(0)=v(0)+p:ifd*a/c>rnd(p)thenx(91)=x(91)+p:goto130
42 x(90)=x(90)+p:next:goto140
50 x(80+1)=x(80+1)+p:return
60 ifz=0thenprint£2,chr$(32)chr$(9);:return
61 iflen(str$(z))=5thenprint£2,chr$(8)chr$(8)zchr$(9);:return
70 on(len(str$(z))-p)goto71,72,73,73,73
71 print£2," "zchr$(9);:return
72 print£2,zchr$(9);:return
73 print£2,chr$(8)zchr$(9);:return
75 return
80 print"Turn on printer:insert paper:then 5"
81 getz:ifz=5thenprint"OK":goto83
82 goto81
83 open2,4:print£2,"                          "chr$(17);
84 fori=pto3:print£2,"           "chr$(18);:next:print£2,"                 "chr$(18);
85 fori=pto3:print£2,"           "chr$(18);:next:print£2,chr$(13)
86 print"5 to proceed"
87 getz:ifz=5thenprint"OK":return
88 goto87
92 g=f+e:ifint(g)>othen94
93 e=g-int(g):return
94 forj=z+ptoz+g:v(h-j+p)=i:a=a+i:ifj=h-pthen97
96 next:z=z+int(g):goto93
97 j=int(m*h-a+.5):v(p)=j:e=10-b-c-d:goto650
100 forj=ptoq:printj;:z=p+p:o=o/5:y(j)=y(j)/(z+p):x(j)=x(j)/(z+p)
103 a=int(a+b+m+n+(z+z)*o+z*y(j)+z*x(j)+.5):c=int(c+b+m+n-o-y(j)-x(j)+.5)
105 f=int(f+b+o*(z+z)+.5):g=int(g+m+y(j)*z+.5):x(j)=0:y(j)=a:w(j+q)=c:o=0
110 fori=ptoh:ifv(i) rnd(p)then199
115 r(i)=r(i)+p:l=0
120 y(0)=y(0)+p:ifd*a/c<rnd(p)thenonu(i)goto180,40,160,140
125 l=l+p
130 gosub30:s(i)=s(i)+p:a=a-p:x(j)=x(j)+p
140 z=rnd(p):ifz<t(i)then120
145 iflthenifz<(t(i)+t(i)/(l*l))then120
150 goto180
160 y(0)=y(0)+p:ife<rnd(p)then140
170 l=l+p:s(i)=s(i)+p:x(j)=x(j)+p:x(74)=x(74)+p:goto140
180 ifi<q+pthenifw(i)>rnd(p)thengosub50:goto115
190 gosub50
199 nexti:x(0)=x(0)+x(j):next:printchr$(13)
```

```
200 gosub992:gosub80:print£2,"COLLECTION"chr$(13)
210 print£2,"Initial no. useful books ="w(0)chr$(9)"Daily addition ="b+m+n
220 print£2,"Prop'n day = 0"chr$(8)t(0)", add'n ="b", uses ="x(71)
225 print£2,"Prop'n week = 0"chr$(8)u(0)", add'n ="m", uses ="x(72)
230 print£2,"Prop'n month = 0"chr$(8)int(h*(p-t(0)-u(0))+.5)/h", add'n ="n;
235 print£2,", uses ="x(73)
240 print£2,"Max success base rate = 0"chr$(8)dchr$(8);
245 print£2,", prob of substitute = 0"chr$(8)e
246 print£2,"Total of daily uses ="x(0)chr$(8)", substitute uses ="x(74)
250 print£2,chr$(13)"Day"chr$(9)"Bks"chr$(9)"Avb"chr$(9)"Uses"chr$(9);
255 print£2,"Uses"chr$(9)"Vsts"chr$(9)"Rvts"chr$(13)
260 fori=ptoq
270 z=i:gosub70:z=w(i+q):gosub70:z=y(i):gosub70:z=x(i):gosub70
280 ifi<llthenz=i-p:gosub70:z=x(79+i):gosub60:z=x(89+i):gosub60
285 print£2,chr$(32):next
290 fori=ptoh:r(0)=r(0)+r(i):s(0)=s(0)+s(i):next
300 print"Next sheet: ";:gosub86:print£2,"USAGE"chr$(13)
310 print£2,"Total visits ="r(0)chr$(8)", including"v(0)"revisits"
315 print£2,"Total uses ="s(0)chr$(13)
320 print£2,"No."chr$(9)"Vsts"chr$(9)"Uses"chr$(9)"A+Ty"chr$(9)"No."chr$(9);
330 print£2,"Vsts"chr$(9)"Uses"chr$(9)"A+Ty"chr$(13)
340 fori=ptoq
350 z=i:gosub70:z=r(i):gosub70:z=s(i):gosub70:z=t(i)+u(i):gosub70:z=i+q
355 gosub70:z=r(i+q):gosub70:z=s(i+q):gosub70:z=t(i+q)+u(i+q):print£2,chr$(8)z
360 next
400 z=0:fori=ztoh:v(i)=z:next:a=p:h=h-p
405 fori=ptoh+p:ifs(i)>hthenx(a)=s(i):a=a+p:next:goto420
410 v(s(i))=v(s(i))+p:next
420 d=z:z=h*q:c=d
430 fori=ptoa-p:ifx(i)<zthenz=x(i):j=i:next
440 r(c)=z:x(j)=q*h+p:z=q*h:c=c+p:ifc a-pthen430
450 print"Next sheet: ";:gosub86:print£2,"DISTRIBUTION"chr$(13)
453 print£2,"Total attempts ="y(0)chr$(9);
455 print£2,"Total uses ="s(0)chr$(13)"Mean use ="s(0)/(h+p);chr$(13)
460 print£2,"Uses"chr$(9)"Usrs"chr$(9)chr$(9)chr$(9);
465 print£2,"Uses"chr$(9)"Usrs"chr$(13)
470 fori=0toq-p
480 z=i:gosub70:z=v(i):gosub60:print£2,chr$(9)chr$(9);
485 z=i+q:gosub70:z=v(i+q):gosub60:print£2,chr$(9);
490 ifi<a-pthenprint£2,r(i):next:goto499
495 print£2,chr$(32):next
499 close2:print"Turn off printer":end
```

```
500 printchr$(147)"USERS":input"Mean exp visits";m
510 print"Pois,1: Unif,2: Geom,3: Undist,4"
520 input"Dist'n visits & run: 1, 2, 3, 4";q
525 input"Attempts dist/not dist, 0/1";n
530 input"Max no of revisits";k
540 input"Tenths who Ren, Rev, Sub";b,c,d:i=p:onqgoto550,570,590,610
550 f=h*m/exp(m)
560 gosub92:i=i+p:f=f*m/i:goto560
570 f=h/(m+m-p)
580 gosub92:i=i+p:goto580
590 f=h/m
600 gosub92:i=i+p:f=f-f/m:goto600
610 fori=ptoh:v(i)=m:next:e=10-b-c-d
650 input"5 to print; 9 to skip";l:ifl>5then700
660 gosub80:print£2,"USERS"chr$(13)chr$(13)"Dist'n of visits and run no. ="q
665 print£2,"Mean no. of visits ="mchr$(8)", attempts per visit = 1.4";
666 ifnthenprint£2,chr$(32):goto670
667 print£2," distributed"
670 print£2,"Prop'n who renege, [type 1] = 0."chr$(8)bchr$(8);
673 print£2,", who revisit, [type 2] = 0."chr$(8)c
675 print£2,"Prop'n who substitute, [type 3] = 0."chr$(8)dchr$(8);
678 print£2,", balance, [type 4] = 0."chr$(8)e
680 print£2,"Max no. of revisits ="kchr$(13)
700 ifnthena=.31:forj=ptoh:t(j)=a:next:goto710
705 forj=ptoh-pstep3:t(j)=.25:t(j+p)=.45:t(j+p+p)=.1:next:t(h)=.31
710 q=10:a=0
720 ifethenforj=ptoe:u(j+a)=p+p+p+p:next
730 ifcthenforj=p+etoc+e:u(j+a)=p+p:next
740 ifbthenforj=p+e+ctob+e+c:u(j+a)=p:next
750 ifdthenforj=p+e+c+btod+e+c+b:u(j+a)=p+p+p:next
760 a=a+q:ifa<hthen720
800 q=50:ifl>5then850
805 print£2,"No."chr$(9)"EVt"chr$(9)"EPrA"chr$(9)"Type"chr$(9);
810 print£2,"No."chr$(9)"EVt"chr$(9)"EPrA"chr$(9)"Type"chr$(13)
820 fori=ptoq
830 z=i:gosub70:z=v(i):gosub70:print£2,t(i)chr$(9)u(i)chr$(9);
840 z=i+q:gosub70:z=v(i+q):gosub70:print£2,t(i+q)chr$(9)u(i+q)
845 next:print"Turn off printer for now":close2
850 fori=ptoh:v(i)=v(i)/q:ifv(i)>pthenw(i)=v(i)-p:v(i)=p
860 next
870 fori=ptoq:printw(i);:next
900 print" ":print"COLLECTION":input"Initial coll'n available";c:w(0)=c
910 input"Daily addition: d,w,m";b,m,n:input"Probability of substitute";e
920 input"Initial prop'n: d,w";f,g:t(0)=f:u(0)=g
930 input"Max success base rate";d
990 print"Day":g=g*c:f=f*c:a=c:goto100
992 input"5 for printout";z
993 ifz=5thenreturn
994 fori=ptoh:r(0)=r(0)+r(i):s(0)=s(0)+s(i):next
995 print"Visits"r(0)"incl revisits"v(0)chr$(13)
996 print"Attempts"y(0)chr$(13)
997 print"Daily uses"x(0)"User total"s(0)chr$(13)
998 print"Substitute uses"x(74):end
```

B.1.3 Variables

a Miscellaneous variable, especially number of useful books on shelf.

b Miscellaneous variable, especially proportion of users who renege and daily addition of new one day books.

c Miscellaneous variable, especially current number of useful books and proportion of users who revisit.

d Miscellaneous variable, especially proportion of users who use substitutes, and constant.

e Miscellaneous variable, especially proportion of uncharacterised borrowers, and constant.

f Denotes distribution of visits or current number of day books.

g Denotes distribution of visits or current number of week books.

i Miscellaneous variable and counter, especially for number of users.

j Counter, especially for number of days.

l Set for skip print; number of uses in current visit.

m Daily addition of week books and constant.

n Set for attempt rate distributed; daily addition of week books.

o Number of day books to be returned next day.

q Miscellaneous variable.

z Miscellaneous variable.

B.1.4 Constants

d = 0.8; success rate in finding correct place on shelf.

h = 100 or 99

k = 2; maximum number of revisits.

m = 10; mean number of visits.

o = 0

p = 1

q = 50

B.1.5 Arrays

r(0)	Total visits.
r(1-100)	Numbers of visits; subtotal of visits per user; high values for distribution of visits.
s(0)	Total uses.
s(1-100)	Subtotal of uses per user.
t(0)	Set for attempts per visit distributed; initial proportion of day books.
t(1-100)	Probability of another attempt.
u(0)	Initial proportion of week books.
u(1-100)	User type.
v(0)	Number of revisits.
v(1-100)	Probability of visit; numbers of users in distribution of use.
w(0)	Initial number of useful books.
w(1-50)	Probability of second visit in day.
w(51-100)	Number of books in useful collection at start of day.
x(0)	Total uses (sum of daily totals).
x(1-70)	Number of month books to be returned on given day.
x(71-73)	Subtotals of uses for each book type.
x(80-89)	Subtotals for distribution of numbers of uses per visit.
x(90-91)	Numbers of revisits with 0,1 use.
y(0)	Number of attempts at use.
y(1-55)	Number of week books to be returned on given day; books available at start of each day.

B.2 Summary sheets for the results of three-runs for each permutation of two collection sizes and two distributions of rate of visit

Summary sheets follow for the results of simulated use by users with the Poisson distribution of visits ('USERS: Dist'n of visits = 1') and the geometric distribution of visits ('USERS: Dist'n of visits = 3') using the initial collection ('COLLECTION: Daily addition = 8') and the reduced collection ('COLLECTION: Daily addition = 5'). The USAGE sheet shows the characteristics for each numbered user together with numbers of uses and numbers of visits. The three DISTRIBUTION sheets show the distributions resulting from three separate runs, the first of which is that represented on the COLLECTION and USAGE sheets.

On the USERS sheets are listed the expected number of visits, the expected probability of another attempt on any visit and the user type number (explained in the heading) for each of the numbered users. The COLLECTION sheet shows the state of the collection on each day of the simulated period of observation and the number of uses generated. The numbers of visits and revisits which resulted in 0,1,2... uses is also shown. The USAGE sheet shows the numbers of visits and uses made by each user and, for convenience of comparison, their type number and probability of another attempt. Finally, on the DISTRIBUTION sheets, the distribution of numbers of users over numbers of uses is shown.

Abbreviations used at the heads of columns are as follows:

EVt	Expected number of visits.
EPrA	Expected probability of another attempt.
Bks	Number of useful books in the collection.
Avb	Number of useful books available for use at the start of the day.
Vsts	Visits.
Rvts	Revisits.
A+Ty	User type number and probability of another attempt.
Usrs	Users.

USERS

Dist'n of visits = 1
Mean no. of visits = 10, attempts per visit = 1.45distributed
Prop'n who renege, [type 1] = 0.1, who revisit, [type 2] = 0.3
Prop'n who substitute, [type 3] = 0.1, balance, [type 4] = 0.5
Max no. of revisits = 2

No.	EVt	EPrA	Type	No.	EVt	EPrA	Type
1	12	.25	4	51	10	.1	4
2	18	.45	4	52	10	.25	4
3	17	.1	4	53	10	.45	4
4	16	.25	4	54	10	.1	4
5	16	.45	4	55	10	.25	4
6	15	.1	2	56	9	.45	2
7	15	.25	2	57	9	.1	2
8	15	.45	2	58	9	.25	2
9	15	.1	1	59	9	.45	1
10	14	.25	3	60	9	.1	3
11	14	.45	4	61	9	.25	4
12	14	.1	4	62	9	.45	4
13	14	.25	4	63	9	.1	4
14	14	.45	4	64	9	.25	4
15	13	.1	4	65	9	.45	4
16	13	.25	2	66	9	.1	2
17	13	.45	2	67	9	.25	2
18	13	.1	2	68	8	.45	2
19	13	.25	1	69	8	.1	1
20	13	.45	3	70	8	.25	3
21	13	.1	4	71	8	.45	4
22	12	.25	4	72	8	.1	4
23	12	.45	4	73	8	.25	4
24	12	.1	4	74	8	.45	4
25	12	.25	4	75	8	.1	4
26	12	.45	2	76	8	.25	2
27	12	.1	2	77	8	.45	2
28	12	.25	2	78	8	.1	2
29	12	.45	1	79	7	.25	1
30	12	.1	3	80	7	.45	3
31	12	.25	4	81	7	.1	4
32	11	.45	4	82	7	.25	4
33	11	.1	4	83	7	.45	4
34	11	.25	4	84	7	.1	4
35	11	.45	4	85	7	.25	4
36	11	.1	2	86	7	.45	2
37	11	.25	2	87	7	.1	2
38	11	.45	2	88	6	.25	2
39	11	.1	1	89	6	.45	1
40	11	.25	3	90	6	.1	3
41	11	.45	4	91	6	.25	4
42	11	.1	4	92	6	.45	4
43	10	.25	4	93	6	.1	4
44	10	.45	4	94	6	.25	4
45	10	.1	4	95	5	.45	4
46	10	.25	2	96	5	.1	2
47	10	.45	2	97	5	.25	2
48	10	.1	2	98	5	.45	2
49	10	.25	1	99	4	.1	1
50	10	.45	3	100	3	.31	3

COLLECTION

Initial no. useful books = 0   Daily addition = 8
Prop'n day = 0 , add'n = 2 , uses = 283
Prop'n week = 0 , add'n = 3 , uses = 272
Prop'n month = 0 , add'n = 3 , uses = 182
Max success base rate = 0.8, prob of substitute = 0.5
Total of daily uses = 808, substitute uses = 71

Day	Bks	Avb	Uses
1	8	8	8
2	16	10	8
3	23	13	15
4	30	14	13
5	37	15	9
6	43	20	10
7	50	24	20
8	55	23	14
9	61	24	6
10	68	31	9
11	74	34	20
12	79	33	16
13	84	35	18
14	91	32	18
15	96	35	15
16	101	38	16
17	107	40	16
18	111	45	16
19	117	46	15
20	121	51	20
21	125	50	22
22	129	50	16
23	133	53	13
24	136	61	16
25	139	65	14
26	143	72	19
27	146	74	12
28	151	79	7
29	156	86	23
30	160	84	21
31	162	86	19
32	167	85	28
33	172	78	9
34	176	85	22
35	180	86	16
36	184	87	23
37	187	86	21
38	192	84	13
39	196	88	14
40	198	96	25
41	200	95	16
42	203	100	17
43	208	101	20
44	213	96	15
45	216	104	17
46	221	109	21
47	226	108	14
48	228	116	11
49	232	123	10
50	236	129	32

Uses	Vsts	Rvts
0	608	469
1	252	182
2	119	
3	32	
4	5	
5	4	
6		
7		
8		
9		

USAGE

Total visits = 1671, including 651 revisits
Total uses = 808

No.	Vsts	Uses	A+Ty	No.	Vsts	Uses	A+Ty
1	9	2	4.25	51	10	3	4.1
2	21	22	4.45	52	10	7	4.25
3	15	8	4.1	53	8	3	4.45
4	12	6	4.25	54	8	3	4.1
5	8	9	4.45	55	13	5	4.25
6	23	14	2.1	56	40	8	2.45
7	41	21	2.25	57	22	6	2.1
8	60	40	2.45	58	36	14	2.25
9	18	7	1.1	59	6	3	1.45
10	16	19	3.25	60	10	8	3.1
11	16	15	4.45	61	11	3	4.25
12	18	8	4.1	62	7	5	4.45
13	19	13	4.25	63	10	2	4.1
14	12	11	4.45	64	15	4	4.25
15	11	5	4.1	65	10	1	4.45
16	40	15	2.25	66	7	3	2.1
17	57	32	2.45	67	17	10	2.25
18	31	7	2.1	68	51	14	2.45
19	15	5	1.25	69	4	1	1.1
20	10	19	3.45	70	11	7	3.25
21	9	2	4.1	71	7	12	4.45
22	16	13	4.25	72	10	3	4.1
23	9	11	4.45	73	6	2	4.25
24	10	4	4.1	74	11	9	4.45
25	13	7	4.25	75	7	2	4.1
26	72	36	2.45	76	33	7	2.25
27	30	8	2.1	77	27	9	2.45
28	52	21	2.25	78	29	5	2.1
29	12	5	1.45	79	6	2	1.25
30	14	14	3.1	80	8	17	3.45
31	17	7	4.25	81	4	1	4.1
32	11	9	4.45	82	9	3	4.25
33	14	6	4.1	83	3	3	4.45
34	14	7	4.25	84	8	2	4.1
35	6	2	4.45	85	10	2	4.25
36	35	11	2.1	86	29	7	2.45
37	32	11	2.25	87	14	2	2.1
38	24	9	2.45	88	14	4	2.25
39	10	2	1.1	89	7	3	1.45
40	6	13	3.25	90	7	5	3.1
41	14	6	4.45	91	5	4	4.25
42	16	4	4.1	92	5	0	4.45
43	12	4	4.25	93	10	6	4.1
44	11	8	4.45	94	5	3	4.25
45	13	5	4.1	95	7	9	4.45
46	30	9	2.25	96	8	3	2.1
47	28	9	2.45	97	14	5	2.25
48	41	12	2.1	98	9	4	2.45
49	13	7	1.25	99	4	0	1.1
50	10	23	3.45	100	3	1	3.31

DISTRIBUTION

Total uses = 808
Mean use = 8.08

Uses	Usrs	Uses	Usrs
0	2	50	
1	4	51	
2	11	52	
3	12	53	
4	7	54	
5	9	55	
6	5	56	
7	10	57	
8	6	58	
9	8	59	
10	1	60	
11	4	61	
12	2	62	
13	3	63	
14	4	64	
15	2	65	
16		66	
17	1	67	
18		68	
19	2	69	
20		70	
21	2	71	
22	1	72	
23	1	73	
24		74	
25		75	
26		76	
27		77	
28		78	
29		79	
30		80	
31		81	
32	1	82	
33		83	
34		84	
35		85	
36	1	86	
37		87	
38		88	
39		89	
40	1	90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

DISTRIBUTION

Total attempts = 2532   Total uses = 809
Mean use = 8.09

Uses	Usrs	Uses	Usrs
0	3	50	
1	5	51	
2	9	52	
3	10	53	
4	9	54	
5	10	55	
6	7	56	
7	6	57	
8	5	58	
9	6	59	
10	3	60	
11	5	61	
12	2	62	
13	2	63	
14		64	
15	4	65	
16	4	66	
17	1	67	
18		68	
19	1	69	
20	1	70	
21	1	71	
22		72	
23	1	73	
24		74	
25	2	75	
26	1	76	
27		77	
28	1	78	
29	1	79	
30		80	
31		81	
32		82	
33		83	
34		84	
35		85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

DISTRIBUTION

Total attempts = 2492   Total uses = 812
Mean use = 8.12

Uses	Usrs	Uses	Usrs
0	7	50	
1	9	51	
2	7	52	
3	10	53	
4	7	54	
5	9	55	
6	5	56	
7	6	57	
8	9	58	
9	2	59	
10	4	60	
11	2	61	
12	2	62	
13	5	63	
14	1	64	
15		65	
16	1	66	
17	3	67	
18	1	68	
19		69	
20	1	70	
21		71	
22		72	
23	2	73	
24		74	
25	1	75	
26	1	76	
27	2	77	
28		78	
29	1	79	
30		80	
31		81	
32		82	
33		83	
34	1	84	
35		85	
36		86	
37		87	
38		88	
39		89	
40	1	90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

COLLECTION

Initial no. useful books = 0 Daily addition = 5
Prop'n day = 0 , add'n = 1 , uses = 179
Prop'n week = 0 , add'n = 2 , uses = 202
Prop'n month = 0 , add'n = 2 , uses = 131
Max success base rate = 0.8, prob of substitute = 0.5
Total of daily uses = 593, substitute uses = 81

Day	Bks	Avb	Uses
1	5	5	6
2	10	6	10
3	15	7	8
4	19	7	10
5	23	7	10
6	27	9	5
7	31	13	7
8	35	15	9
9	39	15	15
10	42	11	5
11	46	13	11
12	49	13	15
13	51	14	8
14	54	19	15
15	58	16	10
16	61	17	6
17	64	22	9
18	67	23	9
19	70	22	10
20	73	23	18
21	75	19	6
22	78	24	13
23	79	29	9
24	81	32	12
25	83	33	14
26	86	29	21
27	88	24	7
28	91	26	21
29	92	29	8
30	93	35	16
31	94	39	6
32	97	42	13
33	100	40	19
34	102	33	6
35	104	39	18
36	107	37	11
37	109	38	15
38	109	43	18
39	112	37	17
40	113	36	14
41	116	36	19
42	118	30	9
43	118	39	20
44	119	35	9
45	121	40	13
46	121	44	10
47	124	47	12
48	125	50	18
49	127	46	11
50	129	48	12

Uses	Vsts	Rvts
0	702	719
1	174	148
2	87	
3	21	
4	6	
5	2	
6		
7		
8		
9		

USAGE

Total visits = 1859, including 867 revisits
Total uses = 593

No.	Vsts	Uses	A+Ty	No.	Vsts	Uses	A+Ty
1	9	5	4.25	51	7	1	4.1
2	15	11	4.45	52	9	0	4.25
3	24	5	4.1	53	14	6	4.45
4	18	4	4.25	54	10	3	4.1
5	20	15	4.45	55	3	0	4.25
6	42	10	2.1	56	61	11	2.45
7	49	12	2.25	57	39	3	2.1
8	98	27	2.45	58	28	7	2.25
9	18	6	1.1	59	5	0	1.45
10	18	20	3.25	60	8	6	3.1
11	15	16	4.45	61	7	2	4.25
12	11	5	4.1	62	6	1	4.45
13	13	6	4.25	63	6	1	4.1
14	14	9	4.45	64	7	2	4.25
15	15	3	4.1	65	5	3	4.45
16	41	8	2.25	66	24	4	2.1
17	59	18	2.45	67	37	8	2.25
18	29	9	2.1	68	68	16	2.45
19	15	4	1.25	69	9	0	1.1
20	15	25	3.45	70	10	13	3.25
21	12	4	4.1	71	9	8	4.45
22	5	3	4.25	72	6	0	4.1
23	16	13	4.45	73	9	4	4.25
24	10	1	4.1	74	14	8	4.45
25	14	5	4.25	75	7	2	4.1
26	71	13	2.45	76	27	5	2.25
27	21	9	2.1	77	62	12	2.45
28	39	8	2.25	78	31	7	2.1
29	15	10	1.45	79	6	0	1.25
30	11	5	3.1	80	6	5	3.45
31	13	4	4.25	81	4	0	4.1
32	8	0	4.45	82	9	2	4.25
33	6	2	4.1	83	7	4	4.45
34	11	3	4.25	84	10	2	4.1
35	16	10	4.45	85	7	2	4.25
36	22	6	2.1	86	20	4	2.45
37	25	3	2.25	87	15	1	2.1
38	72	18	2.45	88	13	4	2.25
39	10	2	1.1	89	6	0	1.45
40	14	13	3.25	90	3	4	3.1
41	15	9	4.45	91	5	3	4.25
42	13	2	4.1	92	5	1	4.45
43	9	3	4.25	93	8	0	4.1
44	7	7	4.45	94	9	3	4.25
45	9	1	4.1	95	4	0	4.45
46	50	11	2.25	96	7	2	2.1
47	55	16	2.45	97	15	1	2.25
48	25	2	2.1	98	20	3	2.45
49	3	2	1.25	99	3	0	1.1
50	12	16	3.45	100	2	0	3.31

DISTRIBUTION

Total uses = 593
Mean use = 5.93

Uses	Usrs	Uses	Usrs
0	13	50	
1	8	51	
2	12	52	
3	11	53	
4	10	54	
5	7	55	
6	5	56	
7	3	57	
8	5	58	
9	4	59	
10	3	60	
11	3	61	
12	2	62	
13	4	63	
14		64	
15	1	65	
16	4	66	
17		67	
18	2	68	
19		69	
20	1	70	
21		71	
22		72	
23		73	
24		74	
25	1	75	
26		76	
27	1	77	
28		78	
29		79	
30		80	
31		81	
32		82	
33		83	
34		84	
35		85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

DISTRIBUTION

Total attempts = 2794   Total uses = 588
Mean use = 5.88

Uses	Usrs	Uses	Usrs
0	12	50	
1	12	51	
2	14	52	
3	6	53	
4	6	54	
5	7	55	
6	11	56	
7	7	57	
8	6	58	
9	2	59	
10	1	60	
11		61	
12	3	62	
13	2	63	
14	2	64	
15	1	65	
16		66	
17	1	67	
18	2	68	
19	1	69	
20		70	
21		71	
22	1	72	
23	1	73	
24		74	
25		75	
26		76	
27	1	77	
28		78	
29		79	
30		80	
31	1	81	
32		82	
33		83	
34		84	
35		85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

DISTRIBUTION

Total attempts = 2695 Total uses = 617
Mean use = 6.17

Uses	Usrs	Uses	Usrs
0	4	50	
1	13	51	
2	13	52	
3	10	53	
4	14	54	
5	10	55	
6	3	56	
7	4	57	
8	4	58	
9	3	59	
10		60	
11	3	61	
12	5	62	
13	2	63	
14	1	64	
15	1	65	
16	1	66	
17	4	67	
18	1	68	
19		69	
20	1	70	
21	1	71	
22		72	
23	2	73	
24		74	
25		75	
26		76	
27		77	
28		78	
29		79	
30		80	
31		81	
32		82	
33		83	
34		84	
35		85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

USERS

Dist'n of visits = 3
Mean no. of visits = 10, attempts per visit = 1.45distributed
Prop'n who renege, [type 1] = 0.1, who revisit, [type 2] = 0.3
Prop'n who substitute, [type 3] = 0.1, balance, [type 4] = 0.5
Max no. of revisits = 2

No.	EVt	EPrA	Type	No.	EVt	EPrA	Type
1	33	.25	4	51	7	.1	4
2	44	.45	4	52	7	.25	4
3	38	.1	4	53	7	.45	4
4	34	.25	4	54	7	.1	4
5	31	.45	4	55	6	.25	4
6	29	.1	2	56	6	.45	2
7	27	.25	2	57	6	.1	2
8	26	.45	2	58	6	.25	2
9	24	.1	1	59	6	.45	1
10	23	.25	3	60	6	.1	3
11	22	.45	4	61	5	.25	4
12	21	.1	4	62	5	.45	4
13	21	.25	4	63	5	.1	4
14	20	.45	4	64	5	.25	4
15	19	.1	4	65	5	.45	4
16	19	.25	2	66	5	.1	2
17	18	.45	2	67	4	.25	2
18	17	.1	2	68	4	.45	2
19	17	.25	1	69	4	.1	1
20	16	.45	3	70	4	.25	3
21	16	.1	4	71	4	.45	4
22	15	.25	4	72	4	.1	4
23	15	.45	4	73	4	.25	4
24	14	.1	4	74	3	.45	4
25	14	.25	4	75	3	.1	4
26	14	.45	2	76	3	.25	2
27	13	.1	2	77	3	.45	2
28	13	.25	2	78	3	.1	2
29	13	.45	1	79	3	.25	1
30	12	.1	3	80	3	.45	3
31	12	.25	4	81	3	.1	4
32	12	.45	4	82	2	.25	4
33	11	.1	4	83	2	.45	4
34	11	.25	4	84	2	.1	4
35	11	.45	4	85	2	.25	4
36	10	.1	2	86	2	.45	2
37	10	.25	2	87	2	.1	2
38	10	.45	2	88	2	.25	2
39	10	.1	1	89	2	.45	1
40	9	.25	3	90	2	.1	3
41	9	.45	4	91	1	.25	4
42	9	.1	4	92	1	.45	4
43	9	.25	4	93	1	.1	4
44	9	.45	4	94	1	.25	4
45	8	.1	4	95	1	.45	4
46	8	.25	2	96	1	.1	2
47	8	.45	2	97	1	.25	2
48	8	.1	2	98	1	.45	2
49	7	.25	1	99	1	.1	1
50	7	.45	3	100	1	.31	3

COLLECTION

Initial no. useful books = 0 Daily addition = 8
Prop'n day = 0 , add'n = 2 , uses = 284
Prop'n week = 0 , add'n = 3 , uses = 275
Prop'n month = 0 , add'n = 3 , uses = 181
Max success base rate = 0.8, prob of substitute = 0.5
Total of daily uses = 800, substitute uses = 60

Day	Bks	Avb	Uses
1	8	8	8
2	16	11	13
3	23	11	12
4	30	12	11
5	37	14	8
6	44	19	15
7	50	22	15
8	56	23	5
9	63	30	8
10	70	35	19
11	75	35	17
12	80	35	19
13	86	32	15
14	91	34	9
15	97	40	15
16	103	41	14
17	108	45	11
18	114	49	16
19	119	53	23
20	124	49	14
21	130	50	7
22	135	57	11
23	140	62	10
24	143	72	21
25	146	71	17
26	151	70	20
27	155	71	17
28	161	69	19
29	164	72	14
30	167	78	19
31	169	85	32
32	171	78	17
33	175	80	22
34	180	77	19
35	184	77	12
36	186	87	36
37	189	75	15
38	193	79	23
39	195	81	16
40	199	84	12
41	201	94	27
42	204	91	23
43	206	91	12
44	210	98	14
45	215	106	17
46	219	107	13
47	222	115	17
48	226	118	14
49	230	120	22
50	232	120	15

Uses	Vsts	Rvts
0	627	461
1	255	184
2	125	
3	23	
4	8	
5	2	
6		
7		
8		
9		

USAGE

Total visits = 1685, including 645 revisits
Total uses = 800

No.	Vsts	Uses	A+Ty	No.	Vsts	Uses	A+Ty
1	35	13	4.25	51	7	4	4.1
2	43	55	4.45	52	6	2	4.25
3	39	20	4.1	53	6	4	4.45
4	33	20	4.25	54	7	2	4.1
5	31	21	4.45	55	4	2	4.25
6	56	23	2.1	56	39	11	2.45
7	71	30	2.25	57	13	4	2.1
8	122	73	2.45	58	16	8	2.25
9	22	5	1.1	59	8	3	1.45
10	21	29	3.25	60	7	9	3.1
11	27	23	4.45	61	4	4	4.25
12	24	12	4.1	62	3	0	4.45
13	27	16	4.25	63	8	0	4.1
14	20	12	4.45	64	1	0	4.25
15	18	6	4.1	65	5	2	4.45
16	41	10	2.25	66	6	2	2.1
17	56	31	2.45	67	12	6	2.25
18	67	18	2.1	68	17	4	2.45
19	25	3	1.25	69	6	2	1.1
20	14	27	3.45	70	5	3	3.25
21	12	3	4.1	71	8	6	4.45
22	18	8	4.25	72	6	0	4.1
23	16	14	4.45	73	7	3	4.25
24	16	3	4.1	74	2	4	4.45
25	16	7	4.25	75	1	0	4.1
26	59	25	2.45	76	2	1	2.25
27	40	9	2.1	77	3	0	2.45
28	46	13	2.25	78	13	1	2.1
29	20	8	1.45	79	2	0	1.25
30	11	10	3.1	80	1	5	3.45
31	11	3	4.25	81	2	1	4.1
32	8	2	4.45	82	2	2	4.25
33	14	3	4.1	83	3	0	4.45
34	12	5	4.25	84	1	1	4.1
35	7	3	4.45	85	2	1	4.25
36	32	11	2.1	86	35	16	2.45
37	18	7	2.25	87	9	0	2.1
38	47	21	2.45	88	3	1	2.25
39	8	2	1.1	89	3	0	1.45
40	8	7	3.25	90	1	1	3.1
41	12	11	4.45	91	3	0	4.25
42	13	2	4.1	92	0	0	4.45
43	9	3	4.25	93	1	0	4.1
44	13	12	4.45	94	3	2	4.25
45	12	3	4.1	95	2	1	4.45
46	25	5	2.25	96	1	1	2.1
47	56	17	2.45	97	3	0	2.25
48	23	6	2.1	98	3	1	2.45
49	3	3	1.25	99	1	0	1.1
50	4	6	3.45	100	1	1	3.31

DISTRIBUTION

Total uses = 800
Mean use = 8

Uses	Usrs	Uses	Usrs
0	15	50	
1	11	51	
2	11	52	
3	12	53	
4	6	54	
5	4	55	1
6	5	56	
7	3	57	
8	3	58	
9	2	59	
10	2	60	
11	3	61	
12	3	62	
13	2	63	
14	1	64	
15		65	
16	2	66	
17	1	67	
18	1	68	
19		69	
20	2	70	
21	2	71	
22		72	
23	2	73	1
24		74	
25	1	75	
26		76	
27	1	77	
28		78	
29	1	79	
30	1	80	
31	1	81	
32		82	
33		83	
34		84	
35		85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

DISTRIBUTION

Total attempts = 2666 Total uses = 822
Mean use = 8.22

Uses	Usrs	Uses	Usrs
0	15	50	
1	8	51	
2	12	52	
3	8	53	
4	6	54	
5	10	55	
6	5	56	
7	3	57	
8	5	58	1
9	3	59	
10	5	60	
11		61	
12	1	62	
13		63	
14	1	64	
15		65	
16	1	66	
17	2	67	
18	1	68	
19	3	69	
20		70	
21	2	71	
22		72	
23		73	
24	1	74	
25		75	
26		76	
27		77	
28	2	78	
29		79	
30		80	
31		81	
32		82	
33		83	
34		84	
35	1	85	
36	1	86	
37		87	
38	2	88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46	1	96	
47		97	
48		98	
49		99	

DISTRIBUTION

Total attempts = 2543   Total uses = 818
Mean use = 8.18

Uses	Usrs	Uses	Usrs
0	15	50	
1	14	51	1
2	10	52	
3	3	53	
4	4	54	
5	7	55	
6	4	56	
7	4	57	
8	6	58	
9	4	59	
10	8	60	
11		61	
12	1	62	
13	1	63	
14		64	
15	1	65	
16	1	66	
17	2	67	
18	2	68	
19		69	
20		70	
21	1	71	
22		72	
23	2	73	
24	3	74	
25	1	75	
26		76	
27	1	77	
28	1	78	
29		79	
30	1	80	
31		81	
32		82	
33		83	
34		84	
35	1	85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47	1	97	
48		98	
49		99	

COLLECTION

Initial no. useful books = 0   Daily addition = 5
Prop'n day = 0 , add'n = 1 , uses = 170
Prop'n week = 0 , add'n = 2 , uses = 198
Prop'n month = 0 , add'n = 2 , uses = 130
Max success base rate = 0.8, prob of substitute = 0.5
Total of daily uses = 566, substitute uses = 68

Day	Bks	Avb	Uses
1	5	5	6
2	10	6	11
3	15	7	10
4	19	7	9
5	23	7	11
6	27	9	8
7	31	10	11
8	35	11	5
9	39	13	13
10	42	13	8
11	46	14	10
12	49	16	10
13	53	16	10
14	56	17	11
15	59	19	12
16	62	17	9
17	66	16	6
18	69	19	9
19	72	20	10
20	75	22	15
21	77	23	14
22	79	21	7
23	82	24	14
24	84	23	8
25	86	26	14
26	87	27	5
27	90	32	21
28	91	30	12
29	94	32	14
30	95	34	10
31	98	33	14
32	99	34	13
33	102	32	18
34	104	30	15
35	105	32	14
36	105	36	8
37	107	42	15
38	109	41	8
39	112	43	13
40	114	44	12
41	117	44	17
42	119	43	15
43	121	42	11
44	123	44	17
45	124	43	8
46	127	45	9
47	128	51	8
48	130	55	16
49	131	55	10
50	133	58	12

Uses	Vsts	Rvts
0	741	663
1	168	139
2	95	
3	17	
4	2	
5	2	
6		
7		
8		
9		

USAGE

Total visits = 1827, including 802 revisits
Total uses = 566

No.	Vsts	Uses	A+Ty	No.	Vsts	Uses	A+Ty
1	28	14	4.25	51	6	0	4.1
2	43	43	4.45	52	10	4	4.25
3	44	9	4.1	53	4	1	4.45
4	36	16	4.25	54	5	0	4.1
5	31	26	4.45	55	8	5	4.25
6	62	25	2.1	56	19	2	2.45
7	81	16	2.25	57	26	4	2.1
8	144	35	2.45	58	28	6	2.25
9	20	4	1.1	59	6	0	1.45
10	28	33	3.25	60	7	5	3.1
11	25	15	4.45	61	3	3	4.25
12	19	6	4.1	62	5	2	4.45
13	22	2	4.25	63	5	1	4.1
14	15	3	4.45	64	5	0	4.25
15	16	6	4.1	65	11	9	4.45
16	55	15	2.25	66	21	3	2.1
17	157	16	2.45	67	6	1	2.25
18	51	9	2.1	68	27	3	2.45
19	20	6	1.25	69	5	0	1.1
20	18	35	3.45	70	5	7	3.25
21	23	3	4.1	71	6	1	4.45
22	18	4	4.25	72	4	0	4.1
23	11	7	4.45	73	4	0	4.25
24	12	1	4.1	74	3	0	4.45
25	18	3	4.25	75	3	0	4.1
26	87	18	2.45	76	8	2	2.25
27	28	6	2.1	77	4	1	2.45
28	49	8	2.25	78	12	2	2.1
29	15	6	1.45	79	3	0	1.25
30	9	5	3.1	80	4	3	3.45
31	16	4	4.25	81	2	0	4.1
32	13	9	4.45	82	2	0	4.25
33	9	3	4.1	83	1	3	4.45
34	13	2	4.25	84	0	0	4.1
35	9	4	4.45	85	1	0	4.25
36	36	5	2.1	86	5	0	2.45
37	42	8	2.25	87	3	0	2.1
38	39	13	2.45	88	8	2	2.25
39	11	1	1.1	89	5	0	1.45
40	7	10	3.25	90	1	2	3.1
41	8	3	4.45	91	0	0	4.25
42	8	1	4.1	92	1	0	4.45
43	7	4	4.25	93	1	0	4.1
44	11	0	4.45	94	2	0	4.25
45	5	1	4.1	95	1	1	4.45
46	23	6	2.25	96	8	2	2.1
47	34	7	2.45	97	3	0	2.25
48	28	6	2.1	98	3	0	2.45
49	3	0	1.25	99	1	0	1.1
50	6	5	3.45	100	3	4	3.31

DISTRIBUTION

Total uses = 566
Mean use = 5.66

Uses	Usrs	Uses	Usrs
0	26	50	
1	10	51	
2	9	52	
3	10	53	
4	8	54	
5	5	55	
6	8	56	
7	3	57	
8	2	58	
9	4	59	
10	1	60	
11		61	
12		62	
13	1	63	
14	1	64	
15	2	65	
16	3	66	
17		67	
18	1	68	
19		69	
20		70	
21		71	
22		72	
23		73	
24		74	
25	1	75	
26	1	76	
27		77	
28		78	
29		79	
30		80	
31		81	
32		82	
33	1	83	
34		84	
35	2	85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43	1	93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

DISTRIBUTION

Total attempts = 2845   Total uses = 589
Mean use = 5.89

Uses	Usrs	Uses	Usrs
0	22	50	
1	16	51	
2	12	52	
3	5	53	
4	4	54	
5	8	55	
6	1	56	
7	5	57	
8	2	58	
9	3	59	
10	2	60	
11	4	61	
12	1	62	
13		63	
14		64	
15	1	65	
16	2	66	
17	2	67	
18		68	
19	2	69	
20	2	70	
21	1	71	
22		72	
23		73	
24	1	74	
25		75	
26	2	76	
27	1	77	
28		78	
29		79	
30		80	
31		81	
32		82	
33		83	
34		84	
35	1	85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

DISTRIBUTION

Total attempts = 2500   Total uses = 558
Mean use = 5.58

Uses	Usrs	Uses	Usrs
0	20	50	
1	19	51	
2	7	52	
3	5	53	
4	3	54	
5	9	55	
6	6	56	
7	4	57	
8	4	58	
9	4	59	
10	2	60	
11	3	61	
12	2	62	
13		63	
14	4	64	
15		65	
16		66	
17	1	67	
18	1	68	
19	1	69	
20	2	70	
21		71	
22		72	
23		73	
24		74	
25		75	
26	1	76	
27		77	
28		78	
29		79	
30		80	
31		81	
32		82	
33	1	83	
34		84	
35	1	85	
36		86	
37		87	
38		88	
39		89	
40		90	
41		91	
42		92	
43		93	
44		94	
45		95	
46		96	
47		97	
48		98	
49		99	

B.3 Methods of estimating the collection size and the numbers of books available

A mean of eight recorded uses per user during the simulation period of ten weeks was thought reasonable for course-related use by science and technology students (Section 4.2.1). Rates for social science or humanities students might be two or three times higher than this, so that averages for UK universities are usually also higher, especially in the 1980's (Table 1.6). On average, about five different titles could be expected to be used during the simulation period (Section 9.2 and Table 9.1). A collection adequate for the requirements of the large majority of the users might afford up to three times the average number of recorded uses (Section 9.3) and therefore very roughly three times the number of title uses per user, say 15 titles ($24^{0.86} = 15.4$). For a class of 100 students, it might be reasonable to assume that, at the minimum, one copy of each title or one alternative title was available for every ten students. About 150 to 200 useful items could be expected to be provided for the simulated students according to this method of estimation, therefore.

Another estimate was based upon a collection known to the author and serving over 1000 science and technology students. In this collection, the average recorded usage per book (including both short-loan and main-collection books) was rather less than two, indicating that the simulated collection should comprise around 400 items. Even this would not be judged large by university standards, since the collection in question was regularly weeded.

At the start of calibration, a generous collection of 500 items was therefore assumed, with 100 items initially useful and a further eight new items becoming useful each day. Usefulness was assumed to result from recommendation by a lecturer or from the setting of an assignment. The average number of visits during the period was initially set at 15 and the probability of a further attempt at 0.31, increased by 0.3 after a success.

During calibration the size of collection was reduced to 400 with no initially useful items. One hundred items out of the 400 in the collection were assumed to be retained for one day only, and using data from the Centre for Interfirm Comparison (68) were expected to yield about 300 issues out of the required total of about 730 (about 70 issues came from substituted items outside the collection).

B.3.1 Retention periods

The majority of books issued in academic libraries appear to be returned close to the date due, even though a substantial and predictable minority may return early and late (15:79,42:Appendix 12,68,112,115). For simplicity, retention times rather than loan periods were represented in the simulation and only three lengths of retention were included: one, five and twenty days. Unavailability due to differences between a stipulated loan period and an actual retention period, and especially to the late return of popular material (15:87), was thought to be too sophisticated a factor to incorporate simply. Because availability, measured by the success rate, was being used as the independent variable it was thought important that it should be adjusted by only one factor; namely, the size of the collection.

B.4 Method of estimating numbers of attempts per visit

The numbers of times during each visit that the simulated users were to attempt to find useful material was estimated from data on observed numbers of recorded uses per visit or spell (Table 6.18) by making two simplifying assumptions. It was assumed that the success rate in attempting to use material and the probability of making a further attempt were common to all users and constant over time. Expressions for 0,1,2 ... successes were then written out in terms of an attempt rate, a, and a success rate, s. Thus, the probability of observing, for example, two recorded uses (i.e. two successes) is:

$$
\begin{aligned}
P(2) &= P(2|2) + P(2|3) + \dots \\
&= s^2ab + 3s^2a^2fb + 6s^2a^3f^2b + \dots \\
&= s^2ab(1 + 3af + 6a^2f^2 + 10a^3f^3 + \dots) \\
&= s^2ab/(1 - fa)^3.
\end{aligned}
$$

where $P(m|n)$ is the probability of observing m successes in n attempts and $b = 1 - a$ and $f = 1 - s$. The sum of all the probabilities, $P(0)$, $P(1)$, $P(2)$, ..., is unity as required.

By trial and error, values of a and s were found which yielded distributions similar to those shown in Table 6.18. For $s \simeq 0.5$, a was found to be between 0.45 and 0.6 with $P(0)$ between 0.25 and 0.35.

The distributions of individual mean rates of recorded use per visit (summarised in Table 6.17) gave some indication of the range of values which should be represented in the simulated population. Individuals were assigned one of three rates of attempt per visit. Random variation about these rates ensured a broad distribution of actual individual rates.

During the calibration runs, it was found that attempt rates lower than those suggested above, but enhanced after a success, were necessary to produce a distribution of recorded uses per visit similar to those observed.

B.5 Failure rates in academic libraries

Typical rates of failure for users seeking known items in academic libraries were required not only (i) to estimate a maximum expected success rate for users in the simulation (variable *d*, program line 930) but also (ii) to estimate the minimum expected availability or success rate which might be observed in actual libraries.

It was assumed that, for a given capability on the part of the library in meeting the users' demands, a negative-feedback loop connected the level of expectation among the users with the performance of the library: the higher the level of expectation among the users, the lower the performance level. The intermediate causal relationships include the level of utilization of library material, which affects the availability of library material and therefore the success rate of potential users seeking material (6,15:130,113). An observed success rate or level of availability could be expected to vary only modestly and briefly, therefore, if the balance between the capacity and the utilization of the library was disturbed.

Excellent reviews of previous work on availability and rates of failure in academic libraries have been contributed by Mansbridge (80) and Revill (94). Both stress the advantages of viewing the attempt to use the library as depending upon a sequence of conditions: the sought item must be held by the library; the record of its location and the location itself must be correct, helpfully communicated and correctly found by the prospective user; the item must be present in the library and correctly located. Failure at each step can be quantified and the cause attributed. The negative-feedback loop suggested above could be expected to maintain an equilibrium in the users' overall rate of failure. By continually minimising the dissonance between the expectations of the users and the performance of the library, the relationships within the sytem should maintain the rate within predictable bounds (say, between 20% and 50%).

B.5.1 Maximum success rate

It was assumed that most of the items sought by the simulated users were recommended in some way and were therefore in the stock of the library. Whether seeking known and stocked items or information from unknown stocked items, however, the users were expected to err or be misled. They could, for example, fail to find the correct place on the shelves or could overlook the item, especially if it was misplaced. Their reference might be at fault, the cataloguing rules might defeat them or they might ignore the catalogue altogether. They might transcribe the call number incorrectly or misinterpret the location (or temporary location if it was a library which operated a temporary reserve). They might just guess this information. The item might be awaiting shelving or missing. The library might be poorly laid out, poorly guided or poorly administered.

Various rates of failure were found in the literature for users seeking items which were potentially available. For example, 12% (92), about 20% (91), more than 21% (79), 22% to 25% (90), up to 25% (114), about $^1/_3$ (93). A 20% minimum error rate for user performance is quoted as a typical value by Saracevic (90) and seems reasonable in the light of these figures. The maximum possible success rate for the simulated users was therefore set at 80%.

B.5.2 Minimum observed success rates

In the references and reviews cited above, average rates of failure due to competition by other users for library material are reported to be between 20% and 40%. The combined rate of failure caused either by user error or by circulation interference appears unlikely to be greater than 60% therefore. (Such a rate would not necessarily deter users. If visits to the library usually entailed searches for more than one item, the chances of coming away with something of use would be greater than 40%.) In fact, the authors cited above report success rates for single attempts varying from 50% to over 80%.

For most of the simulated period of library use, the probability of success for a single attempt fell below 40% for the initial collection and well below 40% for the reduced collection. These levels are below those likely to be observed in reality.

B.6 Test of the uniformity of the distribution of pseudo-random numbers

Ten thousand random numbers between zero and one were called using the random number routine in the simulation program. To test the uniformity of the distribution of numbers, the range was divided into 100 equal divisions and the frequencies with which numbers fell into these divisions were compared using the chi-squared test. Each random number was multiplied by 100 and truncated after the units digit. The frequency of occurrence of the resulting numbers (0 to 99) was tabulated. The distribution of frequencies was tested with the chi-squared test to determine the probability with which it could have occurred in random sampling from a population in which the distribution was uniform.

The test was conducted three times, yielding chi-squared statistics of 97.1, 89.9 and 101.6. These values indicated that the distributions of frequencies could have been expected to occur by chance with probabilities of between 0.6 and 0.8 if sampling had been done from a uniform population. The distribution of pseudo-random numbers was judged to have been uniform, therefore.

APPENDIX C

Private communication from Mr. D.M. Ellis showing a derivation of the mean and variance for the modified negative binomial distribution (Figure 6.5)

Your distribution for the use of books does not seem to lead to an easily defined general formula. However, I have derived the Probability Generating Function, and this gives a recurrence relation for the probabilities, and also a means of deriving the cumulants (and, in particular, from them, the mean and the variance).

The distribution is

$$f(s) = \sum_{r=0}^{\infty} \frac{\Gamma(\kappa+r)}{\Gamma(\kappa)\, r!} p^{\kappa} q^{r} \frac{e^{-r\lambda}(\lambda r)^{s}}{s!}$$

The probability generating function is

$$G(\theta) = \sum_{s=0}^{\infty} f(s)\, \theta^{s}$$

$$= \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \frac{\Gamma(\kappa+r)}{\Gamma(\kappa)\, r!} p^{\kappa} q^{r} \frac{e^{-r\lambda}(\lambda r)^{s}\theta^{s}}{s!}$$

$$= \sum_{r=0}^{\infty} \frac{\Gamma(\kappa+r)}{\Gamma(\kappa)\, r!} p^{\kappa} q^{r} e^{-r\lambda} e^{\lambda r\theta} \quad \text{(using the well-known expansion } e^{x} = \sum_{s=0}^{\infty} \frac{x^{s}}{s!}\text{)}$$

$$= p^{\kappa} \sum_{r=0}^{\infty} \frac{\Gamma(\kappa+r)}{\Gamma(\kappa)\, r!} \left(q e^{\lambda(\theta-1)}\right)^{r}$$

$$= p^{\kappa}\left(1 - q e^{\lambda(\theta-1)}\right)^{-\kappa} \quad \text{(using the Binomial theorem)}$$

So

PROBABILITY GENERATING FUNCTION

$$G(\theta) = p^{\kappa}\left(1 - q e^{\lambda(\theta-1)}\right)^{-\kappa}$$

Now $s!\, f(s) = \left.\frac{d^{s}}{d\theta^{s}} G(\theta)\right|_{\theta=0}$ (or $G^{(s)}(0)$)

The r^{th} cumulant $\kappa_r = \mathcal{E}(s(s-1)\ldots(s-r+1)) = G^{(r)}(1)$

Since $\left(1 - q e^{\lambda(\theta-1)}\right)^{\kappa} G(\theta) = p^{\kappa}$, it follows on differentiation with respect to θ that

$$\kappa\left(1 - q e^{\lambda(\theta-1)}\right)^{\kappa-1}\left(-\lambda q e^{\lambda(\theta-1)}\right) G(\theta) + \left(1 - q e^{\lambda(\theta-1)}\right)^{\kappa} G'(\theta) = 0$$

i.e. $\left(1 - q e^{\lambda(\theta-1)}\right) G'(\theta) = \lambda \kappa q\, e^{\lambda(\theta-1)} G(\theta)$

and $e^{-\lambda\theta} G'(\theta) = q e^{-\lambda}\left(\lambda\kappa G(\theta) + G'(\theta)\right)$

If we differentiate this n times, by using Leibniz' Theorem

$$q e^{-\lambda}\left(\lambda\kappa G^{(n)}(\theta) + G^{(n+1)}(\theta)\right) = \sum_{r=0}^{n} {}^{n}C_r\, G^{(r+1)}(\theta)(-\lambda)^{n-r} e^{-\lambda\theta}$$

$$= \sum_{r=0}^{n-2} {}^{n}C_r\, G^{(r+1)}(\theta)(-\lambda)^{n-r} e^{-\lambda\theta} + {}^{n}C_{n-1}\, G^{(n)}(\theta)(-\lambda) e^{-\lambda\theta} + {}^{n}C_n\, G^{(n+1)}(\theta) e^{-\lambda\theta}$$

Whence

$$G^{(n+1)}(\theta)\left(e^{-\lambda\theta} - q e^{-\lambda}\right) = \lambda\left(n e^{-\lambda\theta} + \kappa q e^{-\lambda}\right) G^{(n)}(\theta) - \sum_{r=0}^{n-2} {}^{n}C_r\, G^{(r+1)}(\theta)(-\lambda)^{n-r} e^{-\lambda\theta}$$

Since $s!\, f(s) = G^{(s)}(0)$ and $\kappa_r = G^{(r)}(1)$

Hence

$$(n+1)!\, f(n+1)(1-qe^{-\lambda}) = \lambda n!\,(n+\kappa q e^{-\lambda}) f(n) - \sum_{r=0}^{n-1} \frac{n!\,(r+1)!}{r!\,(n-r)!} f(r+1)(-\lambda)^{n-r}$$

i.e. $$\underline{(n+1)(1-qe^{-\lambda}) f(n+1) = \lambda(n+\kappa q e^{-\lambda}) f(n) - \sum_{u=1}^{n-1} \frac{u\, f(u)\,(-\lambda)^{n-u+1}}{(n-u+1)!}}$$

$$f(0) = G(0) = p^{\kappa}(1-qe^{-\lambda})^{-\kappa}$$
$$f(1) = G'(0) = \lambda \kappa q e^{-\lambda} p^{\kappa}(1-qe^{-\lambda})^{-(\kappa+1)}$$

Then $$2(1-qe^{-\lambda}) f(2) = \lambda(1+\kappa q e^{-\lambda}) f(1)$$

So $$f(2) = \frac{\lambda^2}{2} \kappa q e^{-\lambda} p^{\kappa}(1+\kappa q e^{-\lambda})(1-qe^{-\lambda})^{-(\kappa+2)}$$

$$3(1-qe^{-\lambda}) f(3) = \lambda(2+\kappa q e^{-\lambda}) f(2) - \frac{f(1)(-\lambda)^2}{2!}$$

from which, after some algebra, we find

$$f(3) = \frac{\lambda^3}{3!} \kappa q e^{-\lambda} p^{\kappa}\left\{1+(3\kappa+1)qe^{-\lambda}+\kappa^2(qe^{-\lambda})^2\right\}(1-qe^{-\lambda})^{-(\kappa+3)}$$

Similarly

$$f(4) = \frac{\lambda^4}{4!} \kappa q e^{-\lambda} p^{\kappa}\left\{1+(7\kappa+4)qe^{-\lambda}+(6\kappa^2+4\kappa+1)(qe^{-\lambda})^2+\kappa^3(qe^{-\lambda})^3\right\}(1-qe^{-\lambda})^{-(\kappa+4)}$$

$$f(5) = \frac{\lambda^5}{5!} \kappa q e^{-\lambda} p^{\kappa}\frac{\left\{1+(15\kappa+11)qe^{-\lambda}+(25\kappa^2+30\kappa+11)(qe^{-\lambda})^2+(10\kappa^3+10\kappa^2+5\kappa+1)(qe^{-\lambda})^3+\kappa^4(qe^{-\lambda})^4\right\}}{(1-qe^{-\lambda})^{\kappa+5}}$$

Some features of the pattern are clear, but the development of the term in curly brackets is not.

For the cumulants

$$e^{-\lambda}(1-q)\kappa_{n+1} = \lambda e^{-\lambda}(n+\kappa q)\kappa_n - \sum_{u=1}^{n-1} {}^{n}C_{u-1}\, \kappa_u (-\lambda)^{n-u+1} e^{-\lambda}$$

or $$(1-q)\kappa_{n+1} = \lambda\left\{(n+\kappa q)\kappa_n + \sum_{u=1}^{n-1} {}^{n}C_{u-1}\, \kappa_u (-\lambda)^{n-u}\right\}$$

$$\mathcal{E}(s) = \kappa_1 = G'(1) = \frac{\lambda\kappa q\, G(1)}{(1-q)} = \lambda\kappa\frac{q}{1-q} \quad (\text{since } G(1)=1)$$

$$\mathcal{E}(s(s-1)) = \kappa_2, \quad \text{but } (1-q)\kappa_2 = \lambda(1+\kappa q)\kappa_1 = \frac{\lambda^2\kappa q(1+\kappa q)}{(1-q)}$$

so $$\kappa_2 = \frac{\lambda^2\kappa q(1+\kappa q)}{(1-q)^2}$$

Now $$Var(s) = \mathcal{E}(s^2) - \mathcal{E}^2(s) = \mathcal{E}(s(s-1)) + \mathcal{E}(s) - \mathcal{E}^2(s)$$

$$= \frac{\lambda^2\kappa q(1+\kappa q)}{(1-q)^2} + \frac{\lambda\kappa q}{(1-q)} - \frac{\lambda^2\kappa^2 q^2}{(1-q)^2}$$

which reduces to $$\frac{\lambda\kappa q(\lambda+1-q)}{(1-q)^2}$$

That is

$\underline{\text{MEAN of } s = \lambda\kappa q/(1-q)}$

$\underline{\text{VARIANCE of } s = \lambda\kappa q(\lambda+1-q)/(1-q)^2}$

APPENDIX D

Titles used by 20 undergraduate economists recording the use of a short-loan collection during one academic session (see Section 9.4)

Borrowers are numbered from 1 to 20 under the heading 'User:' on the following four sheets. Below each borrower number are listed the numbers of the titles used (left-hand column) and the numbers of the class-marks borne by these titles (right-hand column). In all, 280 separate titles were used: they are numbered between 1 and 285 (obviously five numbers are not used). Eighty-five different class-marks were represented: they are numbered between 1 and 86 (one number is not used).

User:

1

21 6
22 6
27 6
78 22
101 33
102 33
109 37
110 37
113 37
115 39
117 40
124 41
131 45
133 45
134 46
135 46
137 48
165 60
166 60
172 60
177 62
183 62
193 63
206 68
209 68
215 68
223 69
227 69
233 70
236 71
237 72
238 72
244 72
245 73
247 74
268 84
272 85
279 85
280 85
282 86

2

11 4
80 24
81 25
82 25
83 25
84 25
86 25
87 25
88 26
89 26
90 27
92 29
93 29
94 29
96 29
98 31
99 31
104 34
130 44
133 45
134 46
135 46
137 48
140 49
167 60
169 60
170 60
171 60
173 60
174 60
178 62
184 62
186 63
187 63
188 63
189 63
190 63
193 63
197 64
200 65
209 68
210 68
216 68
217 68
223 69
233 70
237 72
238 72
240 72
247 74
248 75
251 77
255 78

3

35 10
37 10
39 11
58 14
105 35
149 53
151 53
283 86

4

1 1
3 3
4 4
6 4
7 4
8 4
11 4
15 6
17 6
19 6
21 6
22 6
27 6
30 6
36 10
53 14
54 14
55 14
63 14
65 15
66 16
70 17
75 21
76 21
167 60
203 67
220 69
222 69
223 69
224 69
226 69
230 69
232 69
245 73
250 76
253 78
258 78
261 79
269 85
270 85
273 85
279 85
280 85
282 86

5

18 6
27 6
29 6
31 6
32 7
34 9
35 10
38 10
43 13
48 13
105 35
120 41
123 41
124 41
126 41
127 41
128 42
150 53
151 53
158 58

User:

6		7		8		9		10	
15	6	10	4	21	6	20	6	22	6
21	6	21	6	25	6	22	6	24	6
23	6	38	10	32	7	25	6	25	6
27	6	54	14	35	10	27	6	27	6
33	8	57	14	38	10	56	33	78	22
86	25	60	14	39	11	101	33	105	35
87	25	61	14	40	12	103	33	109	37
88	26	63	14	50	13	109	37	118	41
92	29	71	18	55	14	130	44	120	41
94	29	101	33	58	14	131	45	123	41
95	29	102	33	63	14	132	45	124	41
98	31	103	33	66	16	133	45	126	41
104	34	109	37	68	16	137	48	127	41
105	35	130	44	105	35	158	58	131	45
106	35	131	45	127	41	166	60	132	45
107	36	132	45	149	53	168	60	133	45
108	37	133	45	151	53	170	60	137	48
111	37	136	47	207	68	173	60	158	58
112	37	161	58	212	68	174	60	161	58
113	37	175	61			175	61	164	60
118	41	176	61			176	61	170	60
120	41	177	62			177	62	174	60
121	41	181	62			179	62	177	62
123	41	182	62			187	63	179	62
124	41	185	63			189	63	180	62
125	41	187	63			193	63	182	62
126	41	189	63			195	64	187	63
127	41	196	64			197	64	193	63
147	53	197	64			206	68	195	64
151	53	199	64			209	68	197	64
158	58	202	67			214	68	200	65
159	58	205	68			221	69	206	68
161	58	206	68			223	69	209	68
185	63	209	68			225	69	210	68
207	68	223	69			226	69	211	68
210	68	228	69			230	69	212	68
212	68	230	69			235	71	213	68
213	68	232	69			237	72	221	69
263	80	233	70			243	72	223	69
264	80	234	71			247	74	225	69
265	81	237	72			261	79	227	69
		239	72					229	69
		240	72					230	69
		247	74					231	69
								235	71
								238	72
								243	72
								247	74
								276	85

User:

11		12		13		14		15	
20	6	1	1	21	6	16	6	22	6
21	6	3	3	22	6	22	6	27	6
22	6	5	4	23	6	32	7	55	14
23	6	6	4	26	6	35	10	58	14
27	6	8	4	38	10	47	13	59	14
28	6	9	4	43	13	48	13	65	15
29	6	11	4	54	14	79	23	66	16
78	22	20	6	55	14	87	25	67	16
101	33	21	6	65	15	88	26	70	17
102	33	22	6	66	16	91	29	76	21
103	33	26	6	70	17	92	29	149	53
115	39	27	6	72	18	96	29	151	53
116	39	52	14	75	21	98	31	158	58
124	41	55	14	105	35	99	31	159	58
130	44	62	14	114	38	102	33	161	58
131	45	63	14	120	41	118	41	162	59
132	45	64	15	121	41	121	41	212	68
133	45	69	17	122	41	124	41	220	69
134	46	70	17	123	41	127	41	226	69
137	48	73	19	124	41	129	43	227	69
138	48	75	21	125	41	147	53	246	73
139	48	76	21	127	41	148	53	269	85
163	60	100	32	149	53	151	53	274	85
172	60	141	50	151	53	152	53	279	85
177	62	142	50	160	58	153	54		
182	62	167	60	207	68	154	55		
185	63	183	62	208	68				
191	63	193	63	210	68				
193	63	206	68	212	68				
198	64	221	69	219	69				
201	66	223	69	220	69				
206	68	226	69	223	69				
211	68	230	69	227	69				
223	69	232	69	230	69				
235	71	269	85	232	69				
237	72	270	85	262	79				
238	72	280	85	280	85				
242	72	282	86	282	86				
247	74								
256	78								
257	78								
258	78								
259	78								
261	79								
266	82								
272	85								

User:

16		17		18		19		20	
2	2	21	6	8	4	35	10	13	5
17	6	27	6	12	5	41	13	14	5
20	6	33	8	20	6	44	13	32	7
21	6	38	10	21	6	47	13	35	10
22	6	76	21	22	6	48	13	37	10
25	6	105	35	24	6	49	13	41	13
26	6	107	36	25	6	51	13	42	13
27	6	119	41	27	6	67	16	44	13
101	33	120	41	58	14	75	21	45	13
102	33	121	41	75	21	77	22	46	13
103	33	123	41	131	45	81	25	47	13
131	45	124	41	133	45	83	25	48	13
132	45	127	41	167	60	85	25	49	13
133	45	131	45	223	69	88	26	80	24
134	46	137	48	226	69	92	29	87	25
137	48	138	48	249	76	98	31	88	26
155	56	139	48	254	78	105	35	91	29
157	58	156	57	257	78	107	36	92	29
158	58	157	58	277	85	120	41	97	30
163	60	161	58	278	85	124	41	98	31
165	60	185	63	280	85	127	41	126	41
166	60	189	63	282	86	129	43	141	50
175	61	193	63			148	53	142	50
177	62	194	64			151	53	143	51
187	63	195	64			260	78	144	52
189	63	197	64			270	85	145	52
190	63	200	65			271	85	146	52
192	63	206	68			272	85	148	53
193	63	209	68			280	85	151	53
197	64	212	68					284	87
201	66	214	68					285	87
204	67	223	69						
205	68	232	69						
206	68	235	71						
209	68	240	72						
214	68	241	72						
216	68	242	72						
223	69	243	72						
228	69								
230	69								
233	70								
237	72								
238	72								
243	72								
245	73								
247	74								
260	78								
261	79								
267	83								

APPENDIX E

Computer program to simulate recorded and collaborative use of ten titles by ten students

The program for a single run of the simulation is shown overleaf. It is written in Commodore Microsoft BASIC and was run on a Commodore 3032 PET microcomputer. For clarity, routines to print out the results are not shown. The identities of the variables and constants used in the program are listed on the page following.

E.1 Program summary

Lines 10 to 17 initialize the variables and set the parameters.

Lines 20 to 71 are subroutines for: pre-visit search for collaborators (20-25); search for collaborators after failure (30-33); attempt to use next book and, on failure, decision whether to search for collaborator (38-43); use of book and determination of period of retention (60-61); decision whether to attempt further use (70-71).

From lines 100 to 165, the program loops through 30 possible occasions (ten days) on which use may occur. Lines 105 to 115 reset the counter which shows for each user the highest preference title still not used, and make available any copies of titles whose retention period has expired. Lines 120 to 140 decide which user first attempts use on the current occasion and thereafter work backwards through the user numbers. At line 155, pre-visit collaborations occur and at line 160, library visits occur including attempts to re-use Titles 1 to 4 by users who have already used these titles.

E.2 Simulation program

```
10 a=0:h=15:i=0:x=0:k=0:u=1:t=10:s=7:j=0:y=.71:z=.21:g=.5
11 dimq(9,9):dimb(9,9):dimd(9,9):dimm(9):dimn(9):dimv(9):dime(9)
12 dimc(9):v(0)=1/h:v(3)=v(0):v(7)=v(0):v(1)=3/h:v(4)=v(1):v(5)=v(1):v(8)=v(1)
13 v(2)=5/h:v(6)=v(2):v(9)=v(2):printchr$(147):a=rnd(0)
15 dimw(9):w(0)=3:w(1)=2:w(2)=2:w(3)=2:w(4)=u:w(5)=u:w(6)=u:w(7)=u:w(8)=u
16 w(9)=u
17 c(0)=2:c(u)=u:c(2)=u:goto100
20 forh=e(a)to9:goto24
21 forh=0to9
24 ifq(a,h)=w(h)thennexth:return
25 gosub30:gosub70:nexth:return
30 fork=0toc(h):ifb(h,k)<sthenx=y:goto32
31 x=z
32 ifrnd(u)<(x*y)thenm(h)=m(h)+u:q(a,h)=q(a,h)+u:k=t
33 nextk:return
38 forh=e(a)to9:goto40
39 forh=0to9
40 ifq(a,h)=w(h)thennexth:return
41 fork=0toc(h):ifb(h,k)=tthen60
42 nextk:ifint(3*t*rnd(u))<jthenifa<sthengosub30
43 gosub70:nexth:return
60 b(h,k)=a:q(a,h)=q(a,h)+u:n(h)=n(h)+u
61 d(h,k)=j+int(t*rnd(u))+int(t*rnd(u))+u:goto43
70 ifrnd(u)<ythenreturn
71 h=t:return
100 forj=0to29
105 forh=0to9:fora=0to9:ifq(h,a)=0thene(h)=a:a=t
106 nexta:nexth
110 fora=0to9:forh=0toc(a):ifd(a,h)=jthenb(a,h)=t
115 nexth:nexta
120 h=int(t*rnd(u))
130 fori=htoh-9step-u:ifi<0thena=i+t:goto150
140 a=i
150 ifrnd(u)>v(a)thennexti:nextj:goto170
155 ifa<3thengosub20:ife(a)>3thengosub21
160 gosub38:ife(a)>3thengosub39
165 nexti:nextj
169 end
```

E.3 Variables

a	Counter and variable used especially for user number.
$b(9,9)$	User number to which (title,copy) is issued. 10 if not issued.
$c(9)$	Number of extra copies of (title) available.
$d(9,9)$	Occasion when (title,copy) will become available again.
$e(9)$	Highest priority title yet unused by (user).
h	Counter and variable used especially for title number.
i	Counter, especially for user numbers.
j	Counter for occasions.
k	Counter, especially for copy numbers.
$m(9)$	Running total of unrecorded uses by (user).
$n(9)$	Running total of recorded uses by (user).
$q(9,9)$	Number of uses by (user, of title).
$v(9)$	Probability that (user) will visit on any occasion.
$w(9)$	Maximum profitable number of uses of (title).
x	Miscellaneous variable.

E.4 Constants

$g = 0.5$

$s = 7$

$t = 10$

$u = 1$

$y = 0.71$

$z = 0.21$.

REFERENCES

1) HARROP, C. The Information Needs of Undergraduates Project: some preliminary findings. *CRUS News*, 1981, (11), 2-6.

2) FORD. N. Psychological determinants of information needs: a small-scale study of higher education students. *Journal of Librarianship*, 1986, **18**, 47-62.

3) KNAPP, P.B. *College teaching & the college library*, 1959.

4) JAHODA, G., HUBBARD, C.L. & STURSA, M.L. Academic library procedures for providing students with required reading materials. *College & Research Libraries*, 1970, **31**, 103-106.

5) LUBANS, J. Nonuse of an academic library. *College & Research Libraries*, 1971, **32**, 362-367.

6) WILLS, G. & OLDMAN, C. *The beneficial library*, 1977. (BLRDR 5389).

7) RITTER, R.V. An investigation of classroom-library relationships on a college campus as seen in recorded circulation & GPA's. *College & Research Libraries*, 1968, **29**, 30-40.

8) CLAYTON, H. *An investigation of various social & economic factors influencing student use of one college library*, 1965, Ph.D. dissertation, University of Oklahoma.

9) HARROP, C. The Information Needs of Undergraduates Project. In: Fox, D. & Malley, I. (eds). *Third International Conference on Library User Education, Proceedings*, 1983, 27-35.

10) MORSE, P.M. *Library effectiveness: a systems approach*, 1968.

11) LINE, M.B. On the design of information systems for human beings. *Aslib Proceedings*, 1970, **22**, 320-335.

12) LUBANS, J., HARPER, W.A. & ERISMAN, R.E. *A study with computer-based circulation data of the non-use & use of a large academic library*, 1973.

13) HARRIS, C. User needs & user education. In: Fox, P. (ed). *Library user education; are new approaches needed?*, 1980, 12-16. (BLRDR 5503).

14) MAYS, T. Do undergraduates need their libraries? *Australian Academic & Research Libraries*, 1986, **17**, 51-62.

15) BUCKLAND, M.K. *Book availability & the library user*, 1975.

16) BAKER, D. (ed). *Student reading needs & higher education*, 1986.

17) GRAHAM, T. A case of tensions: university libraries, students & reading provision. In: Baker, D. (ed). *Student reading needs & higher education*, 1986, 1-22.

18) HARROP, C. The Information Needs of Undergraduates Project: library use by some first year social science students. *CRUS News*, 1981, (12), 5-10.

19) KNAPP, P.B. The reading of college students. *Library Quarterly*, 1968, **38**, 301-308.

20) WALL, T. Distribution of use among users of an academic library collection. *Library Research*, 1980-81, **2**, 177-180.

21) HOSTROP, R.W. *Teaching & the community college library*, 1968.

22) THOMPSON, R.I. & NICHOLSON, J.B. Significant influences on general circulation in a small college library. *Library Quarterly*, 1941, **11**, 142-185.

23) HARDESTY, L. *Student use of the library at De Pauw University*, 1980. (ED-187335).

24) SCHNAITTER, A.F. *Native & transfer students in one midwestern university: a comparison of their book borrowing & other library use*, 1972, Ph.D. dissertation, Indiana University.

25) MAXTED, I.F. & SMITH, G.C.K. Borrowing from Sorby Hall Library, University of Sheffield 1967-68. *Research in Librarianship*, 1969, **2**, 127-140.

26) TAYLOR, P.H. Survey of library use by second year students at a training college for teachers. *Education Libraries Bulletin*, 1960, (7), 5-16.

27) WEATHERFORD, J. Student library habits. *College & Research Libraries*, 1961, **22**, 369-371.

28) DAVIS, E. The unchanging profile. *Library College Journal*, 1970, **3**, 11-19.

29) KINGSTON, P. Teachers & books in the 1980s. In: Baker, D. (ed). *Student reading needs & higher education*, 1986, 171-179.

30) MANN, P.H. Students & books in the 1980s. In: Baker, D. (ed). *Student reading needs & higher education*, 1986, 180-189.

31) OPPENHEIM, A.N. Reading habits of students: a survey of students at the London School of Economics. *Journal of Documentation*, 1962, **18**, 42-57.

32) McDIARMID, E.W. *Conditions affecting use of the college library*, 1934, Ph.D. dissertation, University of Chicago. (Abstracted in: United States Department of Health, Education & Welfare. Office of Education. *Library science dissertations: 1925-1960*, 1963, 105.)

33) KRAMER, L.A. & KRAMER, M.B. The college library & the drop-out. *College & Research Libraries*, 1968, **29**, 310-312.

34) McDIARMID, E.W. Conditions affecting use of the college library. *Library Quarterly*, 1935, **5**, 59-77.

35) BARKEY, P. Patterns of student use of a college library. *College & Research Libraries*, 1965, **26**, 115-118.

36) STEIG, L. Circulation records & the study of college-library use. *Library Quarterly*, 1942, **12**, 94-108.

37) WILLIAMSON, E. & BRETHERTON, M.H. *Tables of the negative binomial probability distribution*, 1963.

38) CAMMACK, F. & MANN, D. Institutional implications of an automated circulation study. *College & Research Libraries*, 1967, **28**, 129-132.

39) EHRENBERG, A.S.C. *Repeat-buying*, 1972.

40) WAPLES, D. *The evaluation of higher institutions*, 1936. (Quoted in: Broadus, R.N. Analysis of faculty circulation in a university library. *College & Research Libraries*, 1963, **24**, 323-325.)

41) McDOWELL, E. Part-time students at Newcastle Polytechnic: users, marginal users & non-users. *CRUS News*, 1984, (19), 3-5.

42) WALL, T. *Estimating the distribution of use among users of an academic library short loan collection*, 1978, Dissertation, School of Librarianship, The Polytechnic of North London.

43) MUSAVI, N. *Users & nonusers in college libraries*, 1977, Ph.D. dissertation, University of Pittsburgh.

44) MADDEN, M. Marketing survey spinoff: library user/nonuser lifestyles. *American Libraries*, 1979, **10**, 78-81.

45) ROHLF, F.J. & SOKAL, R.R. *Statistical tables*, 2nd ed., 1981.

46) MASSEY, M.E. Market analysis & audience research for libraries. *Library Trends*, 1976, **24**, 473-481.

47) McELROY, R. Courses, collections & colleges: a case study & a model. In: Baker, D. (ed). *Student reading needs & higher education*, 1986, 35-56.

48) WILSON, T.D. On user studies & information needs. *Journal of Documentation*, 1981, **37**, 3-15.

49) DUNN, K. Psychological needs & source linkages in undergraduate information-seeking behavior. *College & Research Libraries*, 1986, **47**, 475-481.

50) AITCHISON, J. & BROWN, J.A.C. *The lognormal distribution with special reference to its uses in economics*, 1957.

51) SLATER, M. Non-users in industry & commerce. *Library Association Record*, 1985, **87**, 302-304.

52) WHITLATCH, J.B. Library use patterns among full- & part-time faculty & students. *College & Research Libraries*, 1983, **44**, 141-152.

53) STRAIN, P.M. Circulation pattern of one technical library. *Special Libraries*, 1965, **56**, 312-317.

54) BLAGDEN, J. Market penetration of library services. *CIIG Review*, 1979, (3), 10.

55) OSEASOHN, R. Borrower use of a modern medical library by practicing physicians. *Bulletin of the Medical Library Association*, 1970, **58**, 58-59.

56) KOVACS, H. Analysis of one year's circulation at the Downstate Medical Center Library. *Bulletin of the Medical Libraries Association*, 1968, 54, 42-47.

57) LAW, D. Doctors & books. In: Baker, D. (ed). *Student reading needs & higher education*, 1986, 88-98.

58) WARD, M.L. *Readers & library users*, 1977.

59) SURRIDGE, R.G. *Management information in the Bromley computerised book charging system*, 1974.

60) GOFFMAN, W. & MORRIS, T.G. Bradford's Law & library acquisition. *Nature*, 1970, 226, 922-923.

61) HARRIS, C. *The subject intensity of library use*, 1975. (BLRDR 5267).

62) CLARK, P.M. New approaches to the measurement of public library use by individual patrons. *Occasional Paper, University of Illinois Graduate School of Library & Information Science*, 1983, (162).

63) SAUNDERS, W.L. et al. Survey of borrowing from the University of Sheffield during one academic year. In: Saunders, W.L. (ed). *The provision & use of library & documentation services*, 1966, 115-143.

64) HARROP, C. Sources of information used by students to carry out work set for tutorials. *CRUS News*, 1982, (14), 3-5.

65) KIDSTON, J.S. The validity of questionnaire responses. *Library Quarterly*, 1985, 55, 133-150.

66) MEIER, R.L. Efficiency criteria for the operation of large libraries. *Library Quarterly*, 1961, 31, 215-234.

67) AMERICAN NATIONAL STANDARDS INSTITUTE. *American national standard for library & information sciences & related publishing practices - library statistics*, 1983. (ANSI Z39.7-1983).

68) CENTRE FOR INTERFIRM COMPARISON. *Inter-library comparisons in academic libraries*, 1984. (BLRDR 5763).

69) FORD, G. A statistical database for British & Irish universities. *Library & Information Research News*, 1985, 7 (29), 10-12.

70) NAYLOR, B. Library output & performance measures. *UC&R Newsletter*, 1986, (18), 5-9.

71) STANDING CONFERENCE OF NATIONAL & UNIVERSITY LIBRARIES. *SCONUL statistical return part two 1985-1986. Notes for guidance to respondents*, 1985. (SCONUL Doc.85/161).

72) HAMBURG, M., RAMIST, L.E. & BOMMER, M.R.W. Library objectives & performance measures & their use in decision making. *Library Quarterly*, 1972, 42, 107-128.

73) TAYLOR, L.W. How complex should a model be? In: *Proceedings of the Joint Automatic Control Conference, Atlanta, Georgia*, 1970, Session paper 18-B, 441.

74) GREENWOOD, M. & YULE, G.U. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 1920, **83**, 255-279.

75) SICHEL, H.S. Repeat-buying & the Generalized Inverse Gaussian-Poisson distribution. *Applied Statistics*, 1982, **31**, 193-204.

76) FROGGATT, P., DUDGEON, M.Y. & MERRETT, J.D. Consultations in general practice, analysis of individual frequencies. *British Journal of Preventive & Social Medicine*, 1969, **23**, 1-11.

77) ELLIOTT, J.M. *Some methods for the statistical analysis of samples of benthic invertebrates*, 2nd ed., 1977.

78) SOKAL, R.R. & ROHLF, F.J. *Biometry*, 1969.

79) COUNCIL OF POLYTECHNIC LIBRARIANS. *Working papers on book availability*, 1983. (BLRDR 5772).

80) MANSBRIDGE, J. Availability studies in libraries. *Library & Information Science Research*, 1986, **8**, 299-314.

81) CHATFIELD, C., EHRENBERG, A.S.C. & GOODHARDT, G.J. Progress on a simplified model of stationary purchasing behaviour. *Journal of the Royal Statistical Society*, 1966, **A129**, 317-367.

82) CASSIE, R.M. Frequency distribution models in the ecology of plankton & other organisms. *Journal of Animal Ecology*, 1962, **31**, 65-92.

83) HINES, W.W. & MONTGOMERY, D.C. Probability & statistics in engineering & management science, 2nd ed., 1980.

84) DAVID, F.N. *Probability theory for statistical methods*, 1951.

85) EVANS, D.A. Experimental evidence concerning contagious distributions in ecology. *Biometrika*, 1953, **40**, 186-211.

86) BROWNSEY, K.W.R. & BURRELL, Q.L. Library circulation distributions: some observations on the PLR sample. *Journal of Documentation*, 1986, **42**, 22-45.

87) SICHEL, H.S. A bibliometric distribution which really works. *Journal of the American Society for Information Science*, 1985, **36**, 314-321.

88) ELLIS, D.M. Private communication. (See Appendix C).

89) NOSIK, B.S. *A stochastic model to predict demand for library services*, 1974, Ph.D dissertation, University of California, Berkeley.

90) SARACEVIC, T. et al. Causes & dynamics of user frustration in an academic library. *College & Research Libraries*, 1977, **38**, 7-18.

91) SMITH, R.H. & GRANADE, V. Use & library failures in an undergraduate library. *College & Research Libraries*, 1978, **39**, 467-473.

92) SCHOFIELD, J.L. et al. Evaluation of an academic library's stock effectiveness. *Journal of Librarianship*, 1975, **7**, 207-227.

93) RADFORD, N.A. Failure in the library - a case study. *Library Quarterly*, 1983, **53**, 328-339.

94) REVILL, D.H. 'Availability' as a performance measure for academic libraries. *Journal of Librarianship*, 1987, **19**, 14-30.

95) SHOOLBRED, M. & ALABASTER, H. A survey of resource centres at City of Birmingham Polytechnic. *Library & Information Research News*, 1985, **8** (30), 10-17.

96) COCKS, T.M. & BROOKES, B.C. Sichel's unification of bibliometric frequency distributions. *Journal of Information Science*, 1986, **12**, 45-51.

97) KANTOR, P.B. A review of library operations research. *Library Research*, 1979, **1**, 295-345.

98) ROUSE, W.B. Tutorial: mathematical modeling of library systems. *Journal of the American Society for Information Science*, 1979, **30**, 181-192.

99) BURRELL, Q.L. & CANE, V.R. The analysis of library data. *Journal of the Royal Statistical Society*, 1982, **A145**, 439-471.

100) BURRELL, Q. A simple stochastic model for library loans. *Journal of Documentation*, 1980, **36**, 115-132.

101) HINDLE, A. & WORTHINGTON, D. Simple stochastic models for library loans. *Journal of Documentation*, 1980, **36**, 209-213.

102) BAGUST, A. A circulation model for busy public libraries. *Journal of Documentation*, 1983, **39**, 24-37.

103) BURRELL, Q.L. A third note on ageing in a library circulation model: applications to future use & relegation. *Journal of Documentation*, 1987, **43**, 24-45.

104) TRUESWELL, R.W. Some behavioral patterns of library users: the 80/20 rule. *Wilson Library Bulletin*, 1969, **43**, 458-461.

105) BURRELL, Q.L. A second note on ageing in a library circulation model: the correlation structure. *Journal of Documentation*, 1986, **42**, 114-128.

106) RAVICHANDRA RAO, I.K. Bivariate distributions of transactions: who borrows what? *Library Science with a Slant to Documentation*, 1981, **18**, 155-173.

107) McDOWELL, E. Book availability survey. In: Council of Polytechnic Librarians. *Working papers on book availability*, 1983. (BLRDR 5772).

108) BLACKIE, E. & SMITH, J.M. Student information needs & library user education. *Education Libraries Bulletin*, 1981, **24** (3), 16-23.

109) CLIFFORD, H.T. & STEPHENSON, W. *An introduction to numerical classification*, 1975.

110) WARWICK, J.P. Duplication of texts in academic libraries: a behavioural model for library management. *Journal of Librarianship*, 1987, **19**, 41-52.

111) RAVIV, A. et al. Students reactions to attributions of ability & effort. *British Journal of Educational Psychology*, 1983, **53**, 1-13.

112) CLINTON, M. Study of the effect of fines on circulation. *Canadian Library Journal*, 1972, **29**, 248-252.

113) ORR, R.H. Measuring the goodness of library services: a general framework for considering quantitative measures. *Journal of Documentation*, 1973, **29**, 315-332.

114) SEYMOUR, C.A. & SCHOFIELD, J.L. Measuring reader failure at the catalog. *Library Resources and Technical Services*, 1973, **17**, 6-24.

115) GELMAN, E. & SICHEL, H.S. Library book circulation and the beta-binomial distribution. *Journal of the American Society for Information Science*, 1987, **38**, 4-12.

ANNEXE

PUBLISHED PAPER

WALL, T. Frequency distributions of recorded use for students using academic library collections. *Collection Management*, 1984, 6, 11-24.

Frequency Distributions of Recorded Use for Students Using Academic Library Collections

T. Wall

ABSTRACT. The distribution of activity among students using academic library collections is discussed. Frequency distributions of recorded use are considered for six libraries. A mixture of Poisson distributions with a negative binomial distribution of means is used to approximate the observed distributions. The extrapolation of distributions of use from this model is described. Expected numbers of non-users in time periods of differing lengths are extrapolated for one set of data. Levels of use not exceeded by 90% and 95% of potential users are estimated from a geometric distribution fitted to the data. The value of three times the sample mean is shown to lie between these levels.

INTRODUCTION

Academic libraries commonly rate their collections by the amount that they are used, often employing use recorded as transaction or circulation data as an indicator of total collection use.[2,9] Potential users do not participate equally, however, in generating collection use, and it is well known that much of the use recorded by academic libraries results from the activities of only a minority of potential users. In this paper, frequency distributions of recorded use by students are presented, modelled and extrapolated in order to il-

T. Wall is with the Bute Library, University of Wales Institute of Science and Technology, Cardiff CF3 9XJ, Wales, United Kingdom.

The author is grateful to Mr. D. M. Ellis, Department of Mathematics, UWIST, for his advice and to Professor P. Havard-Williams, Department of Library and Information Studies, Loughborough University of Technology, for guidance.

Collection Management, Vol. 6(3/4), Fall/Winter 1984

lustrate patterns of uptake by potential users from academic library collections.

Students using a library collection differ in the amount of library material they use and in their frequency of use. Even those with similar tasks to perform can differ widely. A large amount of recorded use is generated by only a small proportion of all potential users: typically 10% to 20% of potential users account for 50% of all use. Wall[15] presents a plot of the relationship between proportion of actual users and proportion of total use. Harrop,[7] Oldman[12] and Whitlatch,[17] among others, illustrate differences in library use among students, and discuss factors that may cause them. Musavi[11:118] has suggested that potential library users may differ fundamentally along a dimension that is reflected in other forms of academic behaviour as well as observed library use. Academic performance itself, however, rarely correlates well with amount of library use.[6,8,13]

Wall[15] used negative binomial distributions to approximate the relative frequencies of recorded use by students using a short-loan textbook collection. This distribution was used by Greenwood and Yule[5] in 1920 to describe the frequency of industrial accidents among workers, and it has found a number of subsequent applications.[4:133,18:14] It can be generated as a mixture of Poisson distributions with a gamma distribution of means. Thus, in the case of library users, potential users use a library collection randomly but with constant mean individual rates of use that are distributed according to a continuous gamma distribution. There are no potential users with zero expectation of use.

This representation is not a simulation model of user behaviour, but an approximation allowing relative frequencies of collection use among users to be summarized and predicted. The gamma distribution is an outcome of the process of collection use; it is not necessarily an inherent characteristic of the user population. A less skew distribution can be shown to generate distributions of use that are similar to those observed and that are approximated by negative binomial distributions. For example, a normally-distributed propensity for library use among users, coupled with probabilistic failure at a constant rate and simple mechanisms for modestly promoting or temporarily discouraging use after success or repeated failure, can, after a few iterations, lead to such outcomes. Underlying the distribution of propensity are likely to be complex diversities in need, role and task among students seeking information.

The shape of the negative binomial distribution is quite flexible though unimodal. There are two parameters. In the notation of Williamson and Bretherton,[18:9] the shape parameter, k, is estimated from the sample mean and the proportion of zeros in the observed distribution, or from the maximum likelihood equation. An inverse scale parameter, p, is estimated from k and the sample mean. The number of users recording r uses, $f(r)$, out of a total of N potential users is seen in Figure 1 where $q = 1 - p$. Each rth term is more easily evaluated as the $(r - 1)$th term multiplied by $q(k + r - 1)/r$. The mean of the distribution is kq/p and the variance kq/p^2.

DATA

The fit of negative binomial distributions was tested for the following sets of data from academic libraries. Ritter[13] gives nine-week circulation totals for 468 students in a liberal arts college. Maxted[10] records borrowing over two terms from an unsupervised library in a hall of residence by 342 students. Knapp[8] records reserve and general collection withdrawals by 738 students from a college library during one semester. Clayton[3] presents similar data for 545 students in another college, and Schnaitter[14] for 3755 junior students (1598 women and 2157 men) at a university.

Most sets of data are grouped in some way. They have been regrouped in Tables 1 and 2 for conciseness, but all analyses were performed on the original data. Authors' estimates of the size of potential user populations were accepted. Ritter discounts absent or uninvolved students, but other authors probably give gross totals derived from enrollment records. It was impossible to estimate sizes of potential user populations for the use of particular parts of collections (e.g., withdrawals from reserve collections), and subdivisions of the data sets were therefore not analysed. It is likely that a proportion of use in Maxted's survey went unrecorded.

Figure 1

$$f(r) = N \binom{k + r - 1}{k - 1} p^k q^r \qquad (r = 0,1,2,\ldots)$$

TABLE 1

Distributions of recorded library use from Ritter (13), Maxted (10) and Knapp (8) and expected frequencies, fitted parameters and observed values of the chi-square statistic for the proposed model.

Number of recorded uses	Observed (Obs.) and expected (Exp.) numbers of users					
	Ritter		Maxted		Knapp	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	150	151.1	149	145.3	111	108.7
1	27	30.5	35	31.8	58	60.0
2	41	33.1	20	28.0	61	55.7
3	23	29.3	25	22.1	54	49.6
4	23	24.9	13	17.5	49	44.3
5-9	84	82.9	58	51.0	192*	186.5
10-14	41	45.4	19	22.6	90	88.7
15-24	50	42.9	18	17.0	62	87.9
25-49	26	25.1	4	6.7	61	56.6
50+	3	2.8	1	0	-	-
Total	468		342		738	
Mean use	6.8		4.0		9.1	
Variance	84		45		114	
Parameters: k		0.6		0.47		0.91
p		0.126		0.113		0.0654
j		1.63		1.08		0.7
Chi-square value		30.4		13.5		10.0
Number of cells		26		16		11
P		0.1		0.3		0.2

Notes.

Parameters: k, negative binomial shape parameter
p, negative binomial scale parameter
j, Poisson parameter.

Chi-square test.
Expected frequencies are pooled to give a minimum cell value of 5.0.
P is the approximate probability of the observed chi-square value being exceeded in random sampling.

* Numbers of recorded uses are grouped: 5-10; 11-15; 16-25; 26+.

TABLE 2

Distributions of recorded library use from Clayton (3) and Schnaitter (14) and expected frequencies, fitted parameters and observed values of the chi-square statistic for the proposed model.

Number of recorded uses	Observed (Obs.) and expected (Exp.) numbers of users: Clayton Obs.	Clayton Exp.	Schnaitter W Obs.	Schnaitter W Exp.	Schnaitter M Obs.	Schnaitter M Exp.
0	48	48.0	392	394.6	916	915.5
1	28	24.6	111	111.4	187	186.6
2	20	26.4	112	102.5	179	157.7
3	23	25.4	92	86.4	116	123.0
4	22	24.2	79	73.7	89	98.2
5-9	102	103.3	242	259.4	284	299.4
10-14	73	77.6	185	161.0	164	150.4
15-24	114	100.1	177	185.8	131	136.4
25-49	94	91.5	147	169.5	79	79.2
50+	21	23.9	61	53.7	12	10.6
Total	545		1598		2157	
Mean use	15.5		11.0		5.0	
Variance	216		-		-	
Parameters: k		1.08		0.53		0.38
p		0.0749		0.047		0.067
j		1.16		1.025*		0.95*
Chi-square value		17.9		35.6		19.8
Number of cells		23		29		28
P		0.5		0.1		0.7

Notes.

Schnaitter W: women; Schnaitter M: men.
* Rounded

Parameters: k, negative binomial shape parameter
p, negative binomial scale parameter
j, Poisson parameter.

Chi-square test.
Expected frequencies are pooled to give a minimum cell value of 5.0.
P is the approximate probability of the observed chi-square value being exceeded in random sampling.

A number of factors makes recorded use an unsatisfactory unit to support consistent analysis. Records will represent a whole range of types and lengths of use, and many valid uses will not be recorded. Differences in academic requirements and in the accessibility and availability of material will vitiate comparisons between all but the most demonstrably similar users in the same library. Crude as the unit is, however, it is the most accessible indicator available. Among users it is likely to be regarded as clearly reflecting purposeful library activity, whether productive or not.[1]

FIT

Negative binomial distributions with parameters estimated from the maximum likelihood equation did not fit the observed distributions of Tables 1 and 2 closely. Observed values of the chi-square test statistic would have been exceeded in random sampling with probabilities ranging from 0.002 to 0.2. The fit to the zero or early terms of some distributions was poor. Nonetheless, the essential shape of the observed distributions was reproduced, so that the negative binomial distribution appeared to be a serviceable base for approximation. It was more successful than lognormal, Neyman Type A and arbitrary distributions. To improve the fit in particular areas of the distribution, an extra parameter, j, was introduced and a new model proposed. This model comprised a mixture of Poisson distributions with individual means distributed according to a negative binomial distribution and multiplied by the constant, j. For each user, the probability of s uses is the negative binomial probability of r activities multiplied by the Poisson probability of s uses given a constant mean (over all users) of j uses per activity. The number of users recording s uses, $f(s)$, out of a total of N potential users is then seen in Figure 2 where $e = 2.718$ and the other notation is as before.

Figure 2

$$f(s) = \frac{N p^k}{s!} \sum_r \binom{k + r - 1}{k - 1} e^{-rj} (rj)^s q^r \qquad (s = 0,1,2,\ldots)$$

It is possible that the two components of the model (negative binomial distribution and Poisson distribution) could represent actual and observable activities among users. Data collected from three UK academic libraries suggested that a Poisson distribution roughly approximated actual distributions of numbers of uses per library visit or attempt at use. Individual rates of visiting and use did not, however, appear to remain constant as assumed in the model. The model does not therefore necessarily simulate user activities in collection use.

Tables 1 and 2 show the fit of this model to the sets of data already described. Parameters were estimated from the sample mean and proportion of zeros. Tables 3 and 4 show the fit to six sets of data derived from a study by Wall.[16] Table 3 shows the use of a short-loan textbook collection by a purposive sample of 309 students over periods of four, eight and sixteen weeks. Table 4 shows eight weeks' use of the same collection by all students in each of the three classes (years) of a science course.

Observed values of the chi-square statistic were used to test goodness of fit. The results are shown in Tables 1 to 4. The null hypothesis could not be rejected at the 5 percent level of significance for any of the sets of data. The probability, P, that chi-square values would be exceeded in random sampling ranged from 0.1 to 0.9 with three values greater than 0.5. The fit of the model was therefore judged to be satisfactory.

EXTRAPOLATION

Under certain conditions, the model may be used to extrapolate distributions of use beyond the time period for which data have been collected. The accuracy of the forecast will depend upon the degree to which both the parameter k and the mean rate of use per time period remain stable. If these conditions are met, then the expression q/p scales linearly with the mean and with time. Time periods should be of such a length that behaviour in one time period appears independent of behaviour in another, since this is a property of the model.

The method is set out by Chatfield, Ehrenberg[4:137] and others who studied brand loyalty among buyers of regularly-purchased consumer goods. They extrapolate negative binomial distributions fitted to frequency distributions of numbers of "purchase occasions" re-

TABLE 3

Distributions of the recorded use of a short-loan textbook collection by 309 students over four, eight and sixteen weeks and expected frequencies, fitted parameters and observed values of the chi-square statistic for the proposed model.

Number of recorded uses	Observed (Obs.) and expected (Exp.) numbers of users					
	Four weeks		Eight weeks		Sixteen weeks	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	161	159.1	108	103.7	70	68.2
1	40	40.4	37	36.2	30	29.0
2	32	30.4	29	30.9	24	24.4
3	21	21.3	30	24.5	26	20.3
4	17	15.1	26	19.6	24	17.2
5-9	27	32.5	50	55.7	53	58.8
10-14	8	8.2	9	22.1	32	33.5
15-24	2	2.0	14	13.3	27	33.4
25+	1	0	6	3.0	23	24.2
Total	309		309		309	
Mean use	1.9		3.9		8.1	
Variance	12		47		192	
Parameters: k		0.56		0.7		0.64
p		0.195		0.133		0.0444
j		0.8		0.86		0.59
Chi-square value		3.1		18.9		15.8
Number of cells		11		16		23
P		0.9		0.1		0.7

Notes.

Parameters: k, negative binomial shape parameter
p, negative binomial scale parameter
j, Poisson parameter.

Chi-square test.
Expected frequencies are pooled to give a minimum cell value of 5.0.
P is the approximate probability of the observed chi-square value being exceeded in random sampling.

TABLE 4

Distributions of the recorded use of a short-loan textbook collection by students in three classes of a science course over eight weeks and expected frequencies, fitted parameters and observed values of the chi-square statistic for the proposed model.

Number of recorded uses	Observed (Obs.) and expected (Exp.) numbers of users					
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	31	30.7	18	18.0	13	12.8
1	8	8.0	9	9.0	9	8.8
2	7	6.6	6	7.4	8	7.4
3	4	5.1	8	6.0	8	6.3
4	6	4.0	3	4.9	4	5.5
5-9	14	11.8	22	15.1	13	18.2
10-14	2	5.3	2	6.7	12	9.3
15-24	3	3.9	1	4.5	11	7.5
25+	2	1.6	4	1.4	1	3.2
Total	77		73		79	
Mean use	4.1		5.0		6.8	
Variance	58		56		47	
Parameters: k		0.48		0.77		0.88
p		0.09		0.07		0.028*
j		0.85		0.49		0.22
Chi-square value		4.7		4.0		9.7
Number of cells		8		8		11
P		0.3		0.4		0.2

Notes.

Parameters: k, negative binomial shape parameter
p, negative binomial scale parameter
j, Poisson parameter.

Chi-square test.
Expected frequencies are pooled to give a minimum cell value of 5.0.
P is the approximate probability of the observed chi-square value being exceeded in random sampling.

* Rounded

ported by consumers and show how numbers of loyal buyers per sales period and sales volume to loyal and to irregular buyers may be predicted for the brand under consideration. Perhaps most useful in the library context will be the prediction of numbers of non-users and maximum levels of use. These aspects are now briefly treated.

Consider the data in Table 3. The parameter k and the mean monthly rate of use are reasonably stable up to 16 weeks. Table 5 shows mean numbers of users recorded in any two weeks separated by varying lengths of time, arbitrarily chosen. Weekly numbers of re-users appear largely independent of the length of time and amount of use intervening. The data of Table 3 appear, therefore, to meet the conditions for extrapolation in multiples of one week.

By extrapolating the scale parameter, p, we can now calculate expected distributions of use for any period of time over which conditions of use would have remained unchanged (in this case, up to 24 weeks). As an example, we use the scale parameter, $p = 0.133$, of the eight-weeks data in Table 3. In calculating numbers of non-users, only the zero term of the distribution is required. Zero terms are estimated from: $f(0) = Np(t)^k\ (1 + e^{-j}kq + e^{-2j}k\ (k + 1)\ q^2/2! + \ldots)$ which simplifies to: $f(0) = Np(t)^k/(1 - qe^{-j})^k$, where $p(t)$ is the scale parameter, p, adjusted for time period t and $q = 1 - p(t)$. Table 6 shows expected numbers of potential users

TABLE 5

Effect of recorded use in any week of a sixteen week period on use in subsequent weeks: mean number of users recorded in the first, second, third, eighth and tenth weeks after previous use, irrespective of intervening use, for 309 potential users of a short-loan textbook collection.

	Week after previous use				
	First	Second	Third	Eighth	Tenth
Mean number of users	39.7	36.1	35.2	32.9	32.3
95% confidence interval for the mean	±2.2	±2.4	±2.7	±2.9	±5.8

TABLE 6

Observed and expected numbers of potential users not recording use of a short-loan textbook collection.

Time period	Observed number of non-users	Expected number of non-users	Scale parameter, p(t)
Weeks 1-4	161	147	0.235
Weeks 1-8	108	104	0.133
Weeks 1-16	70	69	0.0712
Weeks 1-24	-	53	0.04865

Extrapolated from parameters fitted to data for weeks 1-8 (Table 3): k = 0.7; j = 0.86; p(t) as shown.

not recording use in 4, 8, 16 and 24 weeks estimated from this expression and using the scale parameter, $p(t)$, as indicated.

MAXIMUM MAJORITY USE

It is clear from Tables 1 to 4 that distributions of use are very skewed. A proportion of potential users remains unrecorded while a minority is extraordinarily active with transaction rates several times the mean. No single statistic characterises the distribution well, although the sample mean and the proportion of non-users will be important. The upper end of the distribution is not defined, and the range can often be exaggerated by a few high values well separated from the rest of the distribution. In such cases, the median and percentiles could be used to indicate levels of activity not exceeded by given proportions of the population, with calculations made either from data or with extrapolations from the model. More simply, however, a statistic describing a level of activity not exceeded by a majority of the population (say, 90% or 95%) can be estimated directly from the sample mean.

For this purpose, the model is simplified by ignoring the Poisson component and setting the shape parameter, k, to unity. The resulting geometric distribution has values similar to those of the model over the range required. The parameter, p, is estimated from the sample mean. If $q = 1 - p$, then, for mean, m, $q = m/(1 + m)$. The proportion, a, of potential users who each record r or more uses is q^r. Thus $r = \log a / \log [m/(1 + m)]$. Each of the proportion $(1 - a)$ of potential users is expected to record less than r uses. If a is set to 0.1 or 0.05, then r becomes the level of use below which 90% or 95% of the population are expected to record.

In Table 7 actual percentages of potential users recording less than the 90% and 95% levels of use are shown for each set of data from Tables 1 to 4. The 95% level of use tends to be underestimated, but both approximations are close enough to be useful indi-

TABLE 7

Observed percentages of potential users recording less than the 90% and 95% levels of use and three times the sample mean for the data of Tables 1 to 4

Data set		Observed percentages not exceeding: 90% level	95% level	3(mean)
		%	%	%
Table 1:	Ritter	86.5	90.6	89.7
	Maxted	88.9	92.4	90.4
	Knapp	91.0	93.6	93.0
Table 2:	Clayton	90.9	95.5	95.6
	Schnaitter W	88.5	92.3	91.5
	Schnaitter M	87.5	91.7	90.8
Table 3:	Four weeks	91.3	92.9	91.3
	Eight weeks	91.9	93.5	93.2
	Sixteen weeks	90.0	93.5	92.6
Table 4:	Column 1	92.2	93.5	93.5
	Column 2	93.2	93.2	93.2
	Column 3	89.9	94.9	94.9

90% and 95% levels of use are estimated from geometric distributions fitted to the data.

cators of the maximum activity of a majority of potential users. The value of three times the sample mean usually falls between the 90% and 95% levels of use whether calculated from the data or from the geometric distribution, where $[m/(1 + m)]^{3m}$ lies between 0.1 and 0.05 for m between 1.6 and 350.

DISCUSSION

Good approximations to observed distributions of recorded library use from six sources have been obtained using mixtures of Poisson distributions with negative binomial distributions of means. Although further elaboration of this model might improve some of the fits, it would be difficult to demonstrate a general improvement using only data-fitting methods. The model may, in any case, not accommodate all outcomes of collection use. Polymodal distributions, for example, could only be fitted by subdividing populations of users. For the data available, however, the model appears to offer a good compromise between approximation and simplicity.

The transaction data from which the parameters of the model are estimated are precise and easily obtained. Transaction data cannot be assumed to typify the total use of the collections concerned, but the distribution of activity among users appears real and persistent and the model allows these user data to be summarised and, where patterns of use are stable, extrapolated.

Two basic analyses have been illustrated; other applications using the model in combination with local information will be readily suggested. Reference to Ehrenberg[4] will assist the reader in manipulating negative binomial distributions and adapting the model to his or her purposes.

REFERENCES

1. Bookstein, A. "Sources of Error in Library Questionnaires." *Library Research* 4 (1982): 85-94.

2. Burns, R. W. "Library Use As a Performance Measure: Its Background and Rationale." *Journal of Academic Librarianship* 4 (1978): 4-11.

3. Clayton, H. *An Investigation of Various Social and Economic Factors Influencing Student Use of One College Library*. Ph.D. Dissertation. University of Oklahoma. 1965. (University Microfilms No. 65-12993.)

4. Ehrenberg, A. S. C. *Repeat-buying*. Amsterdam: North-Holland. 1972.

5. Greenwood, M., and Yule, G. U. "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence

of Multiple Attacks of Disease or of Repeated Accidents." *Journal of the Royal Statistical Society* 83 (1920): 255-279.

6. Hardesty, L. *Student Use of the Library at De Pauw University.* Greencastle, Ind.: De Pauw University. 1980. (ED 187 335.)

7. Harrop, C. "The Information Needs of Undergraduates Project: Some Preliminary Findings." *CRUS News* No. 11 (1981): 2-6.

8. Knapp, P. B. *College Teaching and the College Library.* Chicago: American Library Association. 1959.

9. Lancaster, F. W. "Evaluating Collections by Their Use." *Collection Management* 4 (1982): 15-43.

10. Maxted, I. F., and Smith, G. C. K. "Borrowing from Sorby Hall Library, University of Sheffield 1967-68." *Research in Librarianship* 2 (1969): 127-140.

11. Musavi, N. *Users and Non-users in College Libraries.* Ph.D. Dissertation. University of Pittsburgh. 1977. (University Microfilms No. 77-23574.)

12. Oldman, C., and Wills, G. *The Beneficial Library.* Bradford: MCB Publications. 1977.

13. Ritter, R. V. "An Investigation of Classroom-Library Relationships on a College Campus as Seen in Recorded Circulation and GPA's." *College and Research Libraries* 29 (1968): 30-40.

14. Schnaitter, A. F. *Native and Transfer Students in One Midwestern University: A Comparison of Their Book Borrowing and Other Library Use.* Ph.D. Dissertation. Graduate Library School, Indiana University. 1972. (University Microfilms No. 73-2751.)

15. Wall, T. "Distribution of Use Among Users of an Academic Library Collection." *Library Research* 2 (1980-81): 177-180.

16. Wall, T. *Estimating the Distribution of Use Among Users of an Academic Library Short Loan Collection.* Unpublished Dissertation. School of Librarianship, The Polytechnic of North London. 1978.

17. Whitlatch, J. B. "Library Use Patterns Among Full- and Part-Time Faculty and Students." *College and Research Libraries* 44 (1983): 141-152.

18. Williamson, E., and Bretherton, M. H. *Tables of the Negative Binomial Probability Distribution.* London: Wiley. 1963.