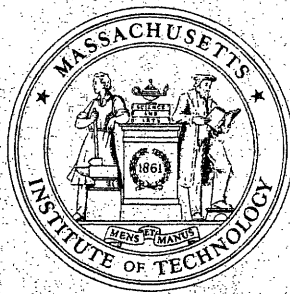


OPERATIONS RESEARCH CENTER

working paper



**MASSACHUSETTS INSTITUTE
OF TECHNOLOGY**

DEMAND FOR LIBRARY MATERIALS
AN EXERCISE IN PROBABILITY ANALYSIS

by

Philip M. Morse

OR 059-76

October 1976

DEMAND FOR LIBRARY MATERIALS.

An Exercise in Probability Analysis.

by Philip M. Morse
Operations Research Center
Mass. Inst. of Technology

Abstract - To evaluate the effect, on the use of library materials, of various possible changes in library policy on circulation rules, for example, or ^{on} the buying of duplicate copies, one must estimate the potential demand for the material, not just the actual use under existing policy. Although the concept of the potential demand, for a book for instance, is a rather vague one, this paper shows how it can be defined and evaluated in terms of the more definite and more easily measurable quantities, yearly circulation rate and mean loan period for borrowed books. The estimates are statistical ones, the average demand per book, the probability that a book that circulates \underline{m} times a year has a demand λ , etc. Graphs and Tables are given that show how these quantities can be evaluated once one knows the mean per book circulation and the mean length of time a book is out of the library per circulation, for a portion of the library that is fairly homogeneous in regard to use (such as all science books, or all biographies). The analysis is then used to show how one can, by the use of the tables and graphs, estimate how much a change in the allowed length of loan period will change the average per book circulation, or what the quantitative effect would be if duplicate copies were bought for all books that circulated more than \underline{m} times, as well as other measures of library utility that depend on demand rather than directly on past circulation.

Rationale - The amount of use of the books in a library is one of the important measures of the library's effectiveness in serving its public. Data on use are helpful in allocating book budgets, they are important in deciding which books to retire from overcrowded open shelves and they are crucial in picking the few books that need to have a duplicate copy bought. As with other measures of library effectiveness, it is important to devise logical means to make the most of the few data that can be collected without encroaching over much on staff time.

At a minimum, mean circulation rate and return rate are needed, plus some idea about the distribution of circulation rates in the collection¹. Circulation rates are the more easily collectible measures of book use and return rates are the only measurable indicator of the discrepancy between potential demand and actual use for the more popular books². Many librarians use the Bradford distribution³ to describe the distribution in use of their collections. But there are several drawbacks⁴ to using the Bradford distribution in a deeper analysis of book use; correspondence with actual data is often imprecise, divisions into "core" and "outer zones" are not "fine-grained" enough on which to base further analysis and, not least, the mathematical form of the Bradford distribution is so involved, that only a closer correspondence with reality than is the case, would justify using it as a basis for deeper study.

As shown elsewhere, if the analysis deals, not with the library as a whole, but with a small number of more homogeneous collections (such as all the books with Dewey classification 600-700, for example, or those with LC classification D) then the modified geometric distribution fits the data at

least as well as the Bradford distribution⁴. Furthermore the modified geometric distribution enables one easily to use the procedures developed for the analysis of queues⁵ to begin to estimate the discrepancies between potential demand and actual use of the collections.

In this paper we shall first define the few quantities, some to be measured, some to be determined indirectly, that are the basis of the logical analysis. Then we shall outline the steps in the analysis, illustrated by graphs and tables, with the mathematical details relegated to an Appendix. The conclusion is that if one measures ^{three} characteristics of a homogeneous collection, then one can derive other measures that will assist in many policy decisions, including those mentioned in the first paragraph.

Definitions - The quantities defined here, some to be measured and some to arise from the analysis, are properties of a homogeneous collection of books or periodicals (which has already been defined). Because library usage is a stochastic phenomenon, all the quantities must be expressed in probabilistic terms; there can be no precise measurement of individual book performance; we must be satisfied with average values and with probabilities. For the determination of general policy, indeed, they are quite sufficient.

First we define the three quantities that must be measured, if any quantitative picture is to be formed of the usage characteristics of the various homogeneous collections that make up the library.

The first is simply N , the number of books (or periodicals) in the collection at a given time; all the books with Dewey number $300_{\wedge}^{\text{to } 400}$, for example, or with LC letter ψC , if one wishes to subdivide in more detail. An average value over the year is sufficiently precise for our purposes.

Integer m , the yearly circulation of a particular book, is the number of times that book circulated during the year in question. Note that a renewal does not count as an additional circulation; each continuous removal of the book from the library by an individual, whether for a long or a short interval, counts as a single circulation. Circulations can be counted by counting the number of withdrawals or else by counting the returned books as they are reshelved; the results do not differ appreciably as long as the library use is fairly constant.

More important is R , the mean yearly circulation rate, per book of the collection. The value of R can be found by dividing the total number of circulations of all the books in the collection, during the year, by N , the total number of books in the collection. If circumstances change markedly during a part of the year, such as during the three summer months, this may be treated separately by counting the circulation during the summer, dividing by N and then multiplying the result by 4 to put it on a per-year basis; the rest of the year has mean circulation rate obtained by dividing total circulation during the other 9 months by N and then multiplying by $4/3$.

The average return rate μ requires a bit more explanation. Its measure² makes possible an estimate of the fraction of time an average book is not available to another potential borrower because it is already borrowed. If book cards are

used for recording circulation, the cards for the books on loan at any time being held at the circulation desk, then μ can be easily evaluated. One counts the number \underline{J} of cards for the given collection, at the circulation desk at any given time. This is the number of books of the collection that are out of the library, unavailable to others, the day the count is made. If the collection is sufficiently large (\underline{N} larger than about 1000) this number \underline{J} will not vary much from day to day if the library use is steady, but it is useful to make this count four or more times during the year (or during the summer months, if the effects of seasonal usage are being investigated) and averaging the \underline{J} 's, to reduce possible fluctuations. Obviously $(\underline{J}/\underline{N})$ is the average fraction of the collection that is out on loan at any time.

But this fraction $(\underline{J}/\underline{N})$ can be expressed in a quite different way⁶; it is the product of the mean circulation rate \underline{R} times the average length of time (the average fraction of the year) the book is out of the library per circulation (which is what we wish to evaluate). Suppose we call this mean loan period (in units of a year) by the symbol $(1/\mu)$; this times the average number of circulations per year for a book of the collection is equal to the fraction of the year the average book is off the shelf, unavailable to another potential user. The fraction of the year an average book is off the shelf must equal the average fraction of the collection that is out on loan at any time; if library usage is statistically steady (if the average book is out half the time, for example, then half the collection will be out at any time, on the average).

Therefore

$$(\underline{R}/\underline{\mu}) = (\underline{J}/\underline{N}) \quad \text{or} \quad (1/\underline{\mu}) = (\underline{J}/\underline{NR}) \quad \text{or} \quad \underline{\mu} = (\underline{NR}/\underline{J}) \quad (1)$$

with the mean loan period $(1/\underline{\mu})$ being given in fractions of a year (or $52/\underline{\mu}$ being given in weeks) if \underline{R} is the mean yearly circulation of a book of the collection. A library with a 2-week loan rule would have $(1/\underline{\mu})$ equal to about $1/25$ year; a few renewals would raise this to $1/20$ or $\underline{\mu} = 20$. University libraries, with extended loan privileges to faculty, turn out to have values of $\underline{\mu}$ of 10 or less (mean loan period a month or more).

The reciprocal of $(1/\underline{\mu})$ is $\underline{\mu}$, called the mean return rate for a book of the collection. If a book happened to be borrowed as soon as it was returned, so it was never on the shelf throughout the year, then its circulation rate would ^{on the average} equal $\underline{\mu}$. This justifies our calling it a rate; its value depends on how fast the average borrower returns books. Naturally a large fraction of renewals will increase the mean loan period $(1/\underline{\mu})$, will diminish the return rate $\underline{\mu}$. Most libraries have return rates between 10 and 20, though some have $\underline{\mu}$'s as small as 5 (mean loan period more than 2 months) and a few have $\underline{\mu}$'s greater than 20.

When these three quantities \underline{N} , \underline{R} and $\underline{\mu}$ are determined for a given year (or other period, if so desired) and for a given homogeneous collection, we can next apply the logic of probability analysis to evaluate indirectly certain other quantities that are more difficult to measure directly but are important in evaluating how effectively the library satisfies prospective borrowers. One of these quantities is an average

value, just as R is; the others are probabilities measuring how usage is spread over the collection, how many books circulate 5 or more times, for example, and what the chance is that, although a book was borrowed 3 times during the year, 2 more persons would have borrowed it if they had arrived when the book was in the library.

An important quantity is λ , the demand parameter for a book, the expected number of persons who come into the library during the year for a given book, who borrow the book if it is on the shelf or who are, to some extent, frustrated if the book is out on loan. It is not the number of persons who come to the library intending ahead of time to borrow that particular book. Even if the book is on the shelf, it is occasionally (or often) borrowed by someone who had not been aware, ahead of time, that this book was the desired one. Thus λ is the expected number per year of persons who would borrow the book if it were on the shelf, whether or not they wanted that book beforehand. The difference between λ and m , the actual yearly circulation of the book in question is, in a sense, a measure of the library's failure to satisfy completely its clientele.

We call λ the demand parameter, or the expected number of persons wishing to borrow the book, because λ is a probabilistic quantity. We want λ to measure the popularity of a given book, not to try to predict exactly how many persons come in for the book in a given year; actual arrivals vary from year to year. If the popularity of the book remained the same year after year then λ would be the average value of the number of arrivals per year. But, even though its popularity varies, we can use as our measure of popularity what would be the

average number of persons wanting the book, if its popularity stayed the same.

The importance of λ arises because estimates of what happens when a change is made in use patterns, such as the retirement of part of the collection, or the buying of duplicate copies of some books, or the change in mean loan period of some or all books, depends on the demand for each book, not on the number of times it happened to be borrowed last year. Circulation can be changed by policy changes only because there are other potential users of the book, who did not have a chance to borrow it, and unless we have a means of estimating that potential demand, we cannot estimate the effect of policy changes on circulation. If our analysis depended on knowing the exact value of demand for each particular book, the analysis would be unusable, because of lack of data. We are dealing with collections of many books, however, and the important thing to know is what fraction of the collection has a particular value of λ , not which book has that value. As we shall demonstrate, this fraction can be calculated, once \underline{N} , \underline{R} and μ are known.

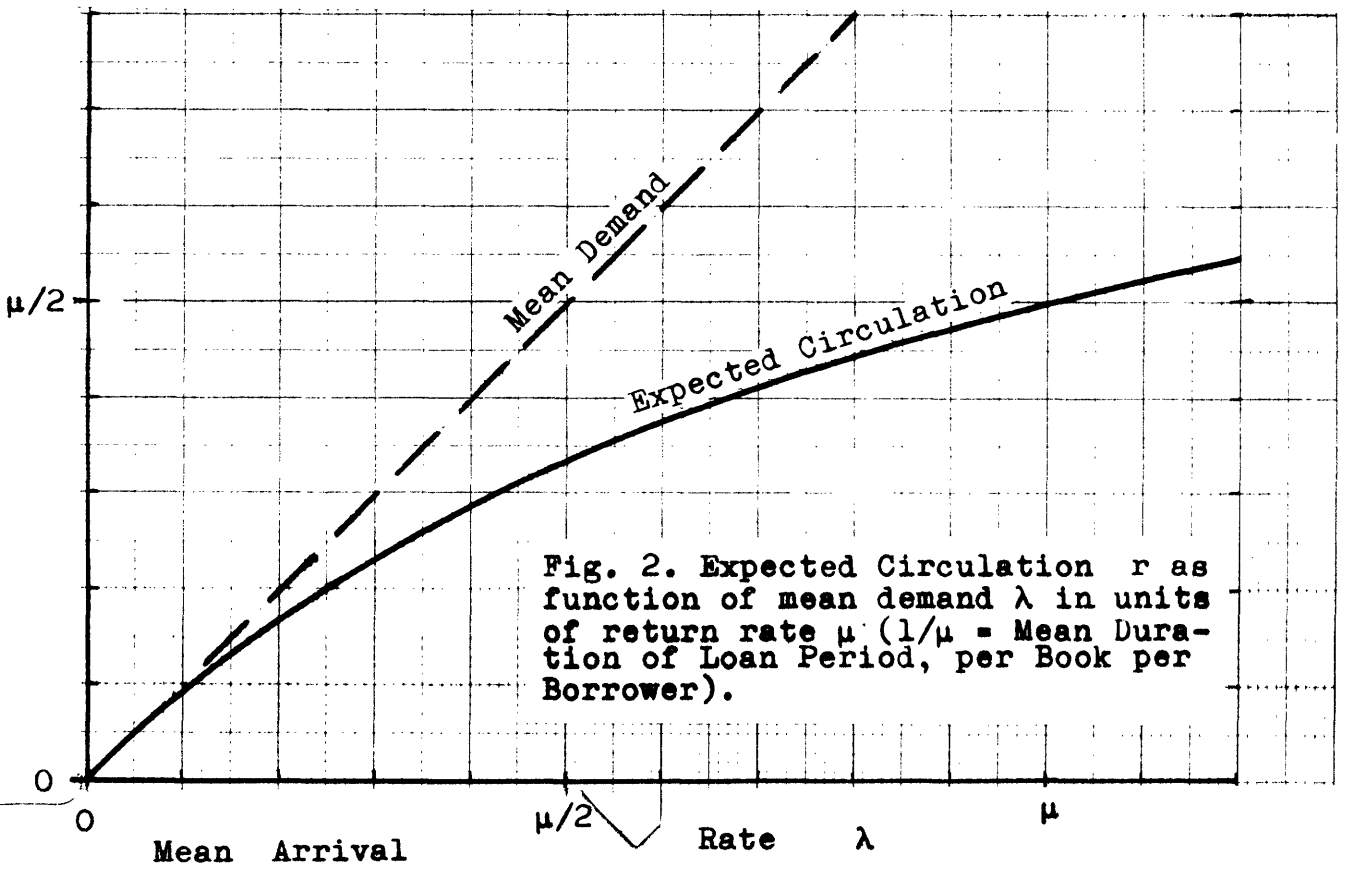
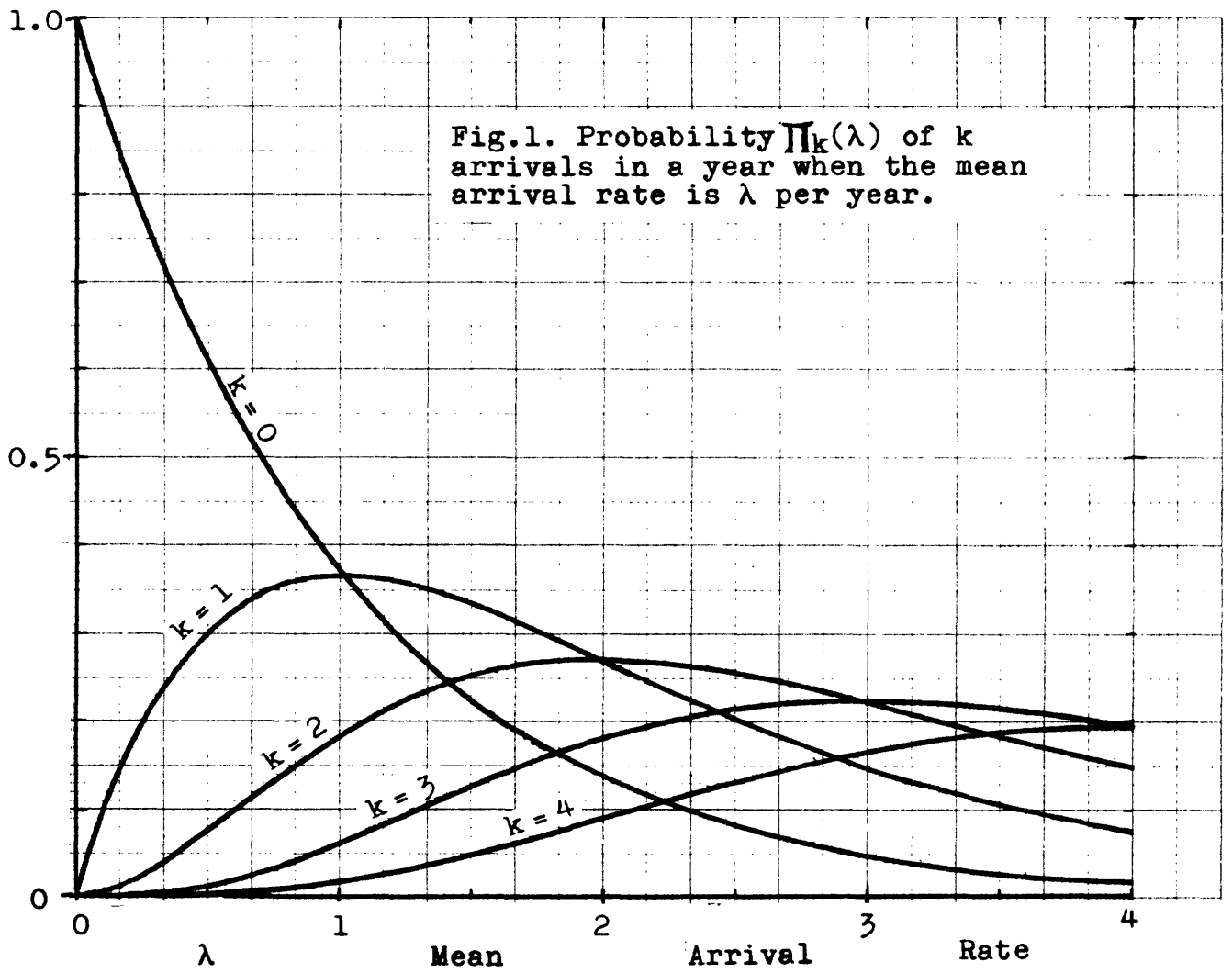
Part of the difficulty in conception comes from the fact that λ is defined statistically. Since prospective users of a book come in at random in time (that is, they are as likely to arrive one week as the next), actual arrivals per year will fluctuate from year to year. Though there may be 25 arrivals in 10 years ($\lambda = 2.5$ if demand has remained constant), arrivals in any one year are fairly likely to be either 2 or 3 but, almost as likely, may be 1 or 4 and, once in a while, may be 5 or none. All we can say, if arrivals are equally likely to come any week

during the year (or any other period of steady usage), is that the probability that \underline{k} prospective borrowers of the book come during the year is

$$\prod_{\underline{k}}(\lambda) = (\lambda^{\underline{k}}/\underline{k}!)e^{-\lambda} \quad (2)$$

Though we cannot specify, for a book with given λ , just how many persons will arrive looking for it during a given year, we can specify what fraction of all the books with a given λ will have \underline{k} persons ^{per year} looking for them.

The probability distribution of Eq.(2) is called the Poisson distribution⁷. It is characteristic of any phenomenon that occurs at random in time or in space. Figure 1 shows plots of some of these probabilities, as functions of λ , for different values of \underline{k} . It shows the wide spread of possible values of \underline{k} , for a given value of λ . For expected value $\lambda = 1$, one arrival and no arrival are equally likely, and 2, 3 etc., arrivals are large enough so that the average number arriving is just 1. Of all the books in the collection with yearly demand equal to some value of λ ($\lambda = 2.5$, for example) then the fraction of those books that would actually have \underline{k} prospective borrowers arrive during the year would be $\prod_{\underline{k}}(2.5)$; there would be a few more books with 2 arrivals than with 1 or 3, but there also would be some books, with $\lambda = 2.5$, having ^{arrivals} 5 or having no arrival. For a book to have a large value of demand parameter λ does not guarantee that, during a given year, its actual circulation will ^{not} be much smaller than λ , or will ^{not} be much larger than λ . With a large collection of books, however, we can be fairly precise as to what fraction of them have \underline{k} prospective borrowers and thus, by extension of the analysis, what fraction have yearly circulation \underline{m} .



The mean demand per book, \underline{D} , the average value of λ for the books of the collection, can be calculated if \underline{N} , \underline{R} and μ are known and if one makes use of a general form of distribution-in-circulation that has been shown² to hold in all cases where it has been tested. One can then set up a self-consistent chain of reasoning, involving a set of probability distributions that lead to greater understanding of the circulation process.

There are five probability distributions, two unconditional, two conditional and one joint, that are involved in the analysis. The first, $\underline{p}(\underline{m})$, is the probability that a book, picked at random from the collection, happens to circulate \underline{m} times during the year under study. It is the fraction of books in the collection that circulated \underline{m} times that year. As with all probability distributions, all the fractions, for all values of \underline{m} , must add up to unity, the whole collection;

$$\sum_{\underline{m}=0}^{\infty} \underline{p}(\underline{m}) = 1 \quad \text{also} \quad \underline{P}(\geq \underline{m}) = \sum_{\underline{n}=\underline{m}}^{\infty} \underline{p}(\underline{n}) \quad ; \quad \underline{P}(\geq 0) = 1 \quad (3)$$

where $\underline{P}(\geq \underline{m})$ is the cumulative probability that a book of the collection circulates \underline{m} or more times a year. These probabilities are related to the measured quantity, \underline{R} , the mean yearly circulation per book of the collection;

$$\underline{R} = \sum_{\underline{m}=1}^{\infty} \underline{m} \underline{p}(\underline{m}) = \sum_{\underline{m}=1}^{\infty} \underline{P}(\geq \underline{m}) \quad (4)$$

The general form of this distribution, that has been found to correspond to many collections, will be discussed later. Here we are only defining $\underline{p}(\underline{m})$ and $\underline{P}(\geq \underline{m})$ and giving their general properties and their relationship to \underline{R} , which must hold no matter what dependence on \underline{m} they may have.

The corresponding distribution for demand is a continuous function, since λ can take on all positive values. Thus $f(\lambda)$ is the probability density that a book of the collection has demand parameter λ , so that the chance of arrival of someone who would borrow the book if it is in, is given by Eq.(2). An alternative statement is that the expected fraction of the collection that have demand parameter between λ and $\lambda + d\lambda$ is $f(\lambda) d\lambda$. Analogous to Eqs.(3) and (4), the distribution must satisfy the usual requirements,

$$\int_0^{\infty} f(\lambda) d\lambda = 1 \quad \text{also} \quad F(\geq \lambda) = \int_{\lambda}^{\infty} f(y) dy ; F(0) = 1 \quad (5)$$

As with the relation between $p(\underline{m})$ and \underline{R} , the equation

$$\underline{D} = \int_0^{\infty} \lambda f(\lambda) d\lambda \quad (6)$$

relates $f(\lambda)$ to the mean demand per book of the collection. The relationships must hold, no matter what form $f(\lambda)$ has.

The conditional probabilities come about when we ask for the relationship between demand and circulation for any given book. In principle, any given book of the collection can be designated, in regard to its circulation potential, by two numbers, \underline{m} and λ . Taking all the books that circulate \underline{m} times a year, conditional probability $f(\lambda|\underline{m}) d\lambda$ is the expected fraction of this class of books that have demand between λ and $\lambda + d\lambda$. Since $f(\lambda|\underline{m})$ is a probability density, it must satisfy the equation

$$\int_0^{\infty} f(\lambda|\underline{m}) d\lambda = 1 \quad \text{also} \quad \delta(\underline{m}) = \int_0^{\infty} \lambda f(\lambda|\underline{m}) d\lambda \quad (7)$$

where $\delta(\underline{m})$ is the mean demand for all the books of the collection that circulate \underline{m} times per year. As shown in Appendix B, approximate values of $\delta_{\underline{m}}$ are given in Table IV.

Also one could, in principle, pick out all the books of the collection that have demand parameter equal to some value λ . The fraction of these that circulate \underline{m} times a year, $\underline{p}(\underline{m}|\lambda)$, is the second conditional probability we must consider. Since $\underline{p}(\underline{m}|\lambda)$ is a probability distribution over \underline{m} , it must satisfy the usual equation, no matter what form it has;

$$\sum_{\underline{m}=0}^{\infty} \underline{p}(\underline{m}|\lambda) = 1 \quad \text{also} \quad \sum_{\underline{m}=1}^{\infty} \underline{m} \underline{p}(\underline{m}|\lambda) = \underline{r}(\lambda) \quad (8)$$

where $\underline{r}(\lambda)$ is the average circulation of all those books that have demand parameter equal to λ .

Finally we can ask for the fraction of the collection that has both circulation \underline{m} and demand parameter λ , the joint probability $\underline{f}(\lambda \text{ and } \underline{m})$. It can be obtained from the probabilities defined earlier; it is either the conditional probability $\underline{p}(\underline{m}|\lambda)$ that the book circulates \underline{m} times a year if it has demand λ , times the probability that it does have demand λ ; or else it is the chance $\underline{f}(\lambda|\underline{m})$ that it has demand λ if it has circulation \underline{m} , times the chance $\underline{p}(\underline{m})$ that it does have circulation \underline{m} ;

$$\underline{f}(\lambda \text{ and } \underline{m}) = \underline{p}(\underline{m}|\lambda) \underline{f}(\lambda) = \underline{f}(\lambda|\underline{m}) \underline{p}(\underline{m}) \quad (9)$$

From Eqs.(3) to (8) we see that if we have a formula for this joint probability we can calculate the rest of the probabilistic quantities,

$$\sum_{\underline{m}=0}^{\infty} \underline{f}(\lambda \text{ and } \underline{m}) = \underline{f}(\lambda) \quad ; \quad \int_0^{\infty} \underline{f}(\lambda \text{ and } \underline{m}) \underline{d}\lambda = \underline{p}(\underline{m}) \quad (10)$$

and thence to the other quantities obtained from $\underline{p}(\underline{m})$ and $\underline{f}(\lambda)$.

Demand and Circulation Distributions - As indicated earlier, it is possible without undue effort, to measure the mean yearly circulation \underline{R} and the mean return rate μ for a homogeneous collection of books or periodicals in the library. What we

have now to do is to show how, from these two measured quantities plus a general knowledge of the probabilistic structure of most library circulation characteristics, we can obtain the other quantities defined in the previous section; in particular, how we can calculate the mean demand \underline{D} per book of the collection.

From Eq.(9) we see that if we know $\underline{p}(\underline{m})$ and $\underline{f}(\lambda|\underline{m})$, we could then compute $\underline{f}(\lambda \text{ and } \underline{m})$, and if we then knew $\underline{f}(\lambda)$, we could compute the rest. But, as we shall shortly demonstrate, we know $\underline{p}(\underline{m}|\lambda)$ and have a good idea as to the form of $\underline{p}(\underline{m})$, so we are not able to complete either side of Eq.(9). What has to be done in this case is to assume a form for $\underline{f}(\lambda)$ and then show that it, combined with the known probability $\underline{p}(\underline{m}|\lambda)$, results in a form for $\underline{p}(\underline{m})$, obtained from Eq.(10), that corresponds with the known form.

First, in regard to $\underline{p}(\underline{m})$, the probability that a book circulates \underline{m} times a year. As mentioned in the first section, and as demonstrated elsewhere², most homogeneous collections have circulation distributions that conform sufficiently closely to a modified geometrical form;

$$\underline{p}(\underline{m}) \simeq \begin{cases} \alpha & (\underline{m} = 0) \\ (1-\alpha)(1-\gamma) \gamma^{\underline{m}-1} & (\underline{m} > 0) \end{cases} ; \quad (11)$$

$$\underline{P}(\geq \underline{m}) \simeq \begin{cases} 1 & (\underline{m} = 0) \\ (1-\alpha) \gamma^{\underline{m}-1} & (\underline{m} > 0) \end{cases}$$

where α and γ are related to the mean circulation rate \underline{R} by the equations

$$\underline{R} = (1-\alpha)(1-\gamma) \sum_{\underline{m}=1}^{\infty} \underline{m} \gamma^{\underline{m}-1} = \frac{(1-\alpha)}{(1-\gamma)} \quad \text{or} \quad \gamma = 1 - \frac{1-\alpha}{\underline{R}} \quad (12)$$

Therefore if we know the value of α , the fraction of the collection that does not circulate, in addition to knowing \underline{R} ,

we can find the value of γ for the collection and thus determine all the elements of the distribution in circulation. This gives us further insight into the circulation process; we can, for example, estimate how many volumes in the collection circulate more than, say, five times during the year, and thus estimate how much it would cost if all books that circulate more than five times would have duplicates purchased.

But counting the volumes that do not circulate may involve more work than can be afforded; we prefer to try indirect methods that require no other measured quantities in addition to \underline{R} and μ . To do this we must assume a form for $\underline{f}(\lambda)$ that can be combined with the known form for $\underline{p}(\underline{m}|\lambda)$ to obtain a formula for the joint probability $\underline{f}(\lambda \text{ and } \underline{m})$. If this, by integration according to Eq.(10), results in a form for $\underline{p}(\underline{m})$ that conforms to Eq.(11), then we have determined the nature of the circulation process entirely in terms of mean circulation \underline{R} and mean return rate μ , as we set out to do.

First we derive the form for the conditional probability $\underline{p}(\underline{m}|\lambda)$ that a book of the collection has circulation rate \underline{m} if it has demand parameter λ . This is obtained from the simplest elements of queuing theory. As a start we note that, on the average, the fraction of time, during the year, a given book is off the shelf, borrowed by someone, is equal to the average number of circulations of the book per year, \underline{r} , times the average time $(1/\mu)$ the book is out per circulation. Of the average number λ of persons coming per year to borrow the book, those that come while the book is out, $\lambda(\underline{r}/\mu)$, do not leave with the book; the others, $\lambda - \lambda(\underline{r}/\mu)$, who come when the book is in, borrow the book, and thus must equal the number of

yearly circulations;

$$\underline{r} = \lambda - \lambda(\underline{r}/\mu) \quad \text{or} \quad \underline{r} = \frac{\lambda\mu}{\lambda + \mu} \quad (13)$$

This relationship is plotted in Fig.2. We see that, as demand λ increases, there is an increasing discrepancy between demand and circulation; the book is more and more often away from the library, fewer and fewer prospective borrowers are satisfied. By the time λ equals μ only half the demand is satisfied.

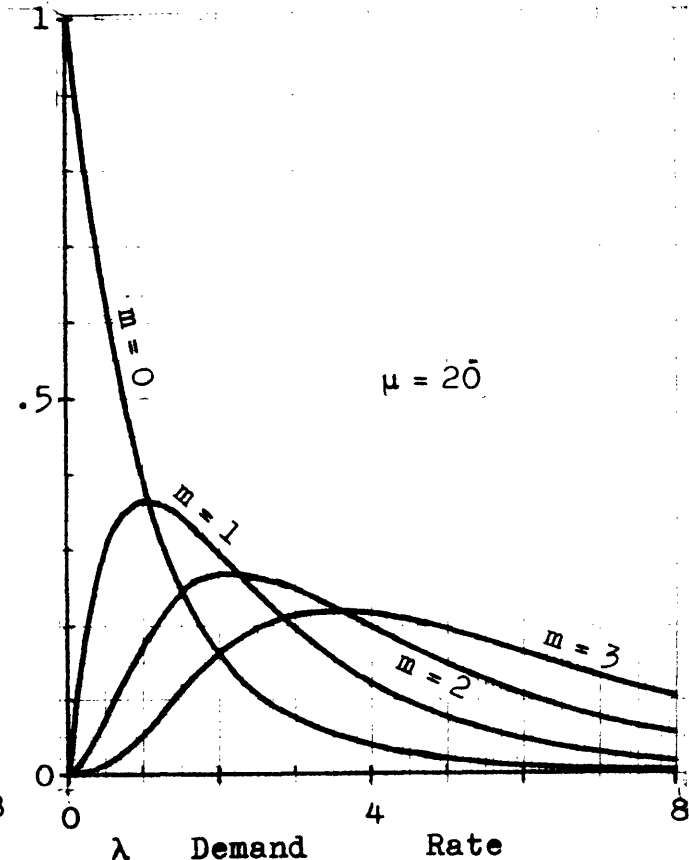
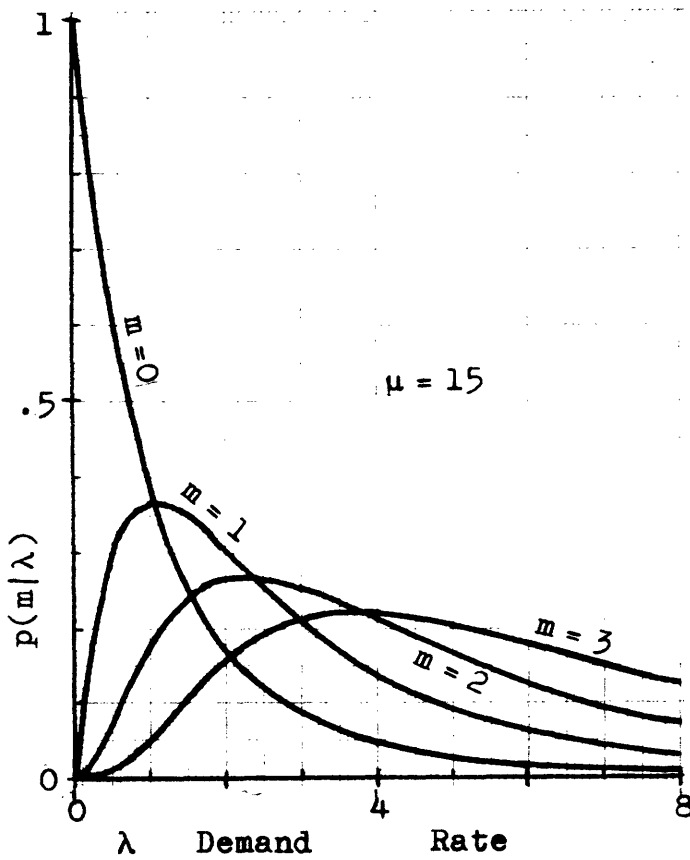
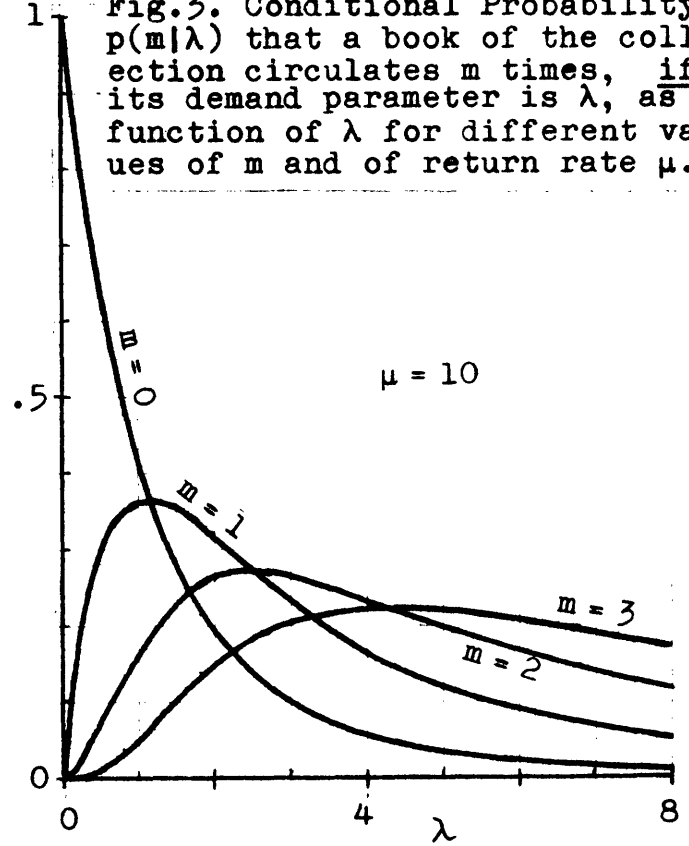
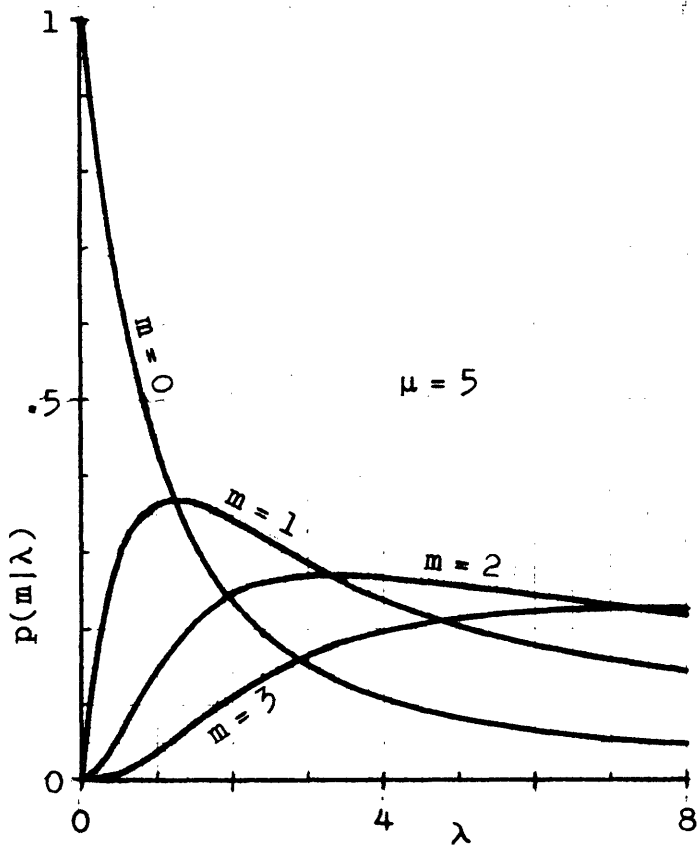
If arrivals of potential borrowers come at random in time during the year, so that actual arrivals are distributed according to Eq.(2), then actual circulations are also random in time and actual yearly circulation should also be distributed according to Eq.(2), with \underline{r} substituted for λ . Thus the conditional probability that a book of the collection circulates \underline{m} times during the year, if it has demand parameter λ , is

$$\underline{p}(\underline{m}|\lambda) = \frac{1}{\underline{m}!} \left(\frac{\lambda\mu}{\lambda + \mu} \right)^{\underline{m}} \exp\left(\frac{-\lambda\mu}{\lambda + \mu} \right) \quad (14)$$

This distribution satisfies Eq.(8) and holds as long as arrivals of would-be borrowers are equally likely to come any week of the year (or of the winter months, or any other interval during which library use is statistically steady). It assumes that there is little correlation between demand rate λ and return rate μ , so we can assume that the average value of (λ/μ) is equal to the ratio between their average values.

Plots of $\underline{p}(\underline{m}|\lambda)$, as function of λ for different values of μ are shown in Fig.3. The effect of the loan period, as might be expected, is greatest on the books in high demand. For a low demand rate $\lambda = 1$ all four values of μ show that a probability of one circulation is roughly equal to that of no circulation during the year and the chance of circulating 2 or

Fig.3. Conditional Probability $p(m|\lambda)$ that a book of the collection circulates m times, if its demand parameter is λ , as function of λ for different values of m and of return rate μ .



more times is much less. However, for $\lambda = 3$ the strict return policy ($\mu = 20$) has $m = 1, 2$ and 3 roughly equally likely, whereas a lax policy ($\mu = 5$) has $m = 3$ only half as likely as $m = 1$ and 2 . When $\mu = 5$ the probability for $m = 3$ does not reach its maximum value until demand rate equals 8 ; for $\mu = 20$ this probability reaches its maximum value by $\lambda = 4$ (and p for $m = 5$ would have its maximum at $\lambda = 8$).

The Distribution of Demand - We have thus obtained a dependable form for the conditional probability $p(m|\lambda)$. If we also knew the form of $f(\lambda)$, the probability density that a book of the collection has demand parameter λ , we would have the second half of Eq.(9) for the joint probability $f(\lambda \text{ and } m)$, that a book has both demand rate λ and circulation rate m . Furthermore, if we can assume that the circulation distribution is the modified geometric one of Eq.(11), as illustrated in Fig.4 and elsewhere², we could then use the first half of Eq.(11) to determine the conditional probability $f(\lambda|m)$, that a book has demand λ if its yearly circulation is m . We of course do not know the distribution $f(\lambda)$; demand is too hard to measure directly. What must be done is to assume a form for $f(\lambda)$ and then verify the choice by showing that the resulting form for $p(m)$, obtained via Eq. (10), corresponds to the form of Eq.(11).

If the form of $p(m)$ is geometric, it is reasonable to assume that the form of $f(\lambda)$ is an analogous exponential form,

$$f(\lambda) = \frac{1}{D} e^{-\lambda/D} \quad ; \quad F(\geq \lambda) \equiv \int_{\lambda}^{\infty} f(y) dy = e^{-\lambda/D} \quad (15)$$

so that $D \equiv \int_0^{\infty} \lambda f(\lambda) d\lambda = \int_0^{\infty} F(\geq \lambda) d\lambda$

where $F(\geq \lambda)$ is the chance that a book has demand equal to λ or

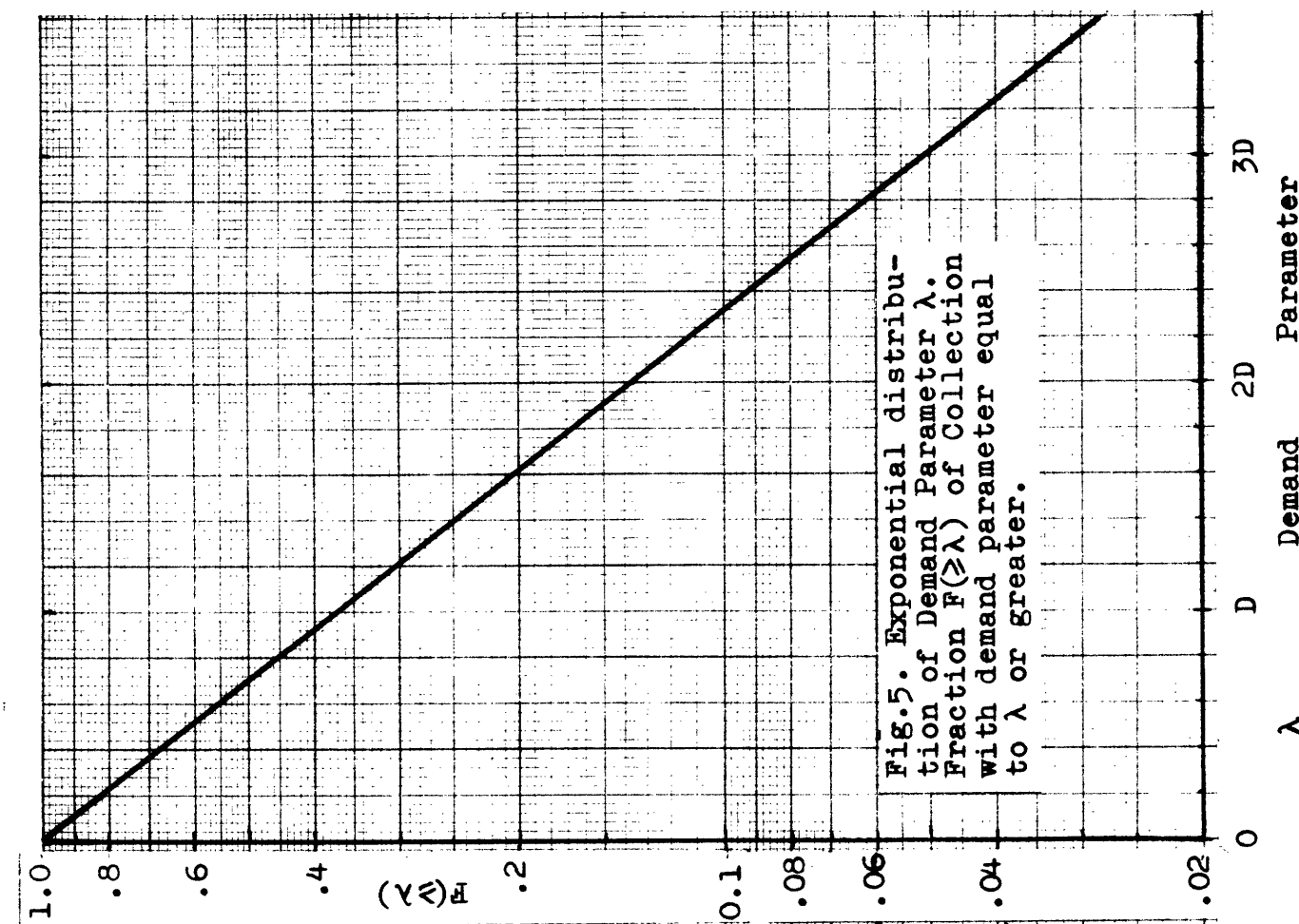


Fig.4. Probability $P(m)$ that a book of the collection circulates m or more times a year. Straight lines correspond to modified geometric distribution; circles and crosses are data from actual collections.

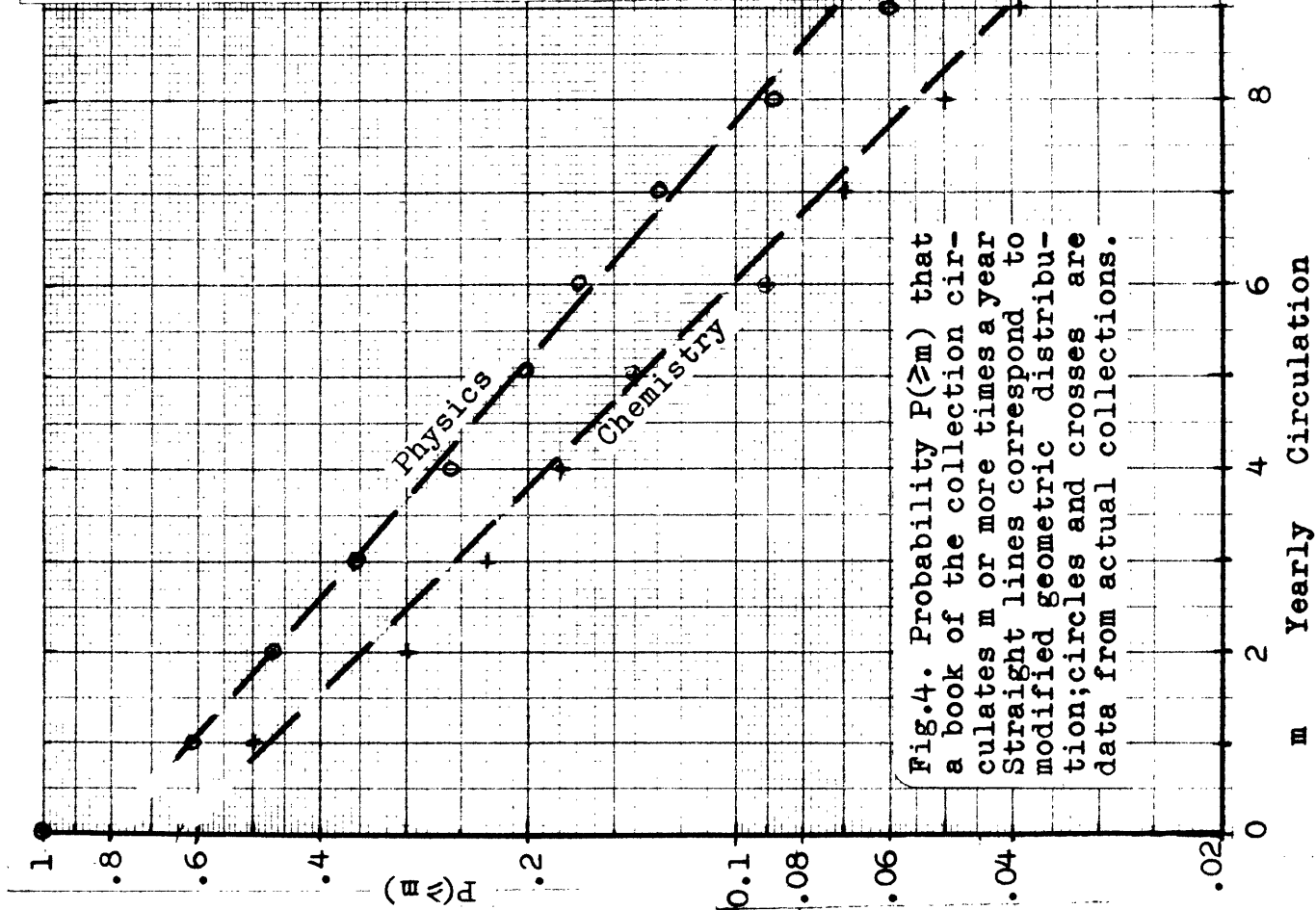


Fig.5. Exponential distribution of Demand Parameter λ . Fraction $F(\geq \lambda)$ of Collection with demand parameter equal to λ or greater.

greater and \underline{D} , as in Eq.(6), is the mean demand per book of the collection. A plot of $\underline{F}(\geq \lambda)$ is shown in Fig.5, to compare with the form of the circulation distribution, shown in Fig.4. In Appendix A we show that the resulting form for $\underline{p}(\underline{m})$, obtained via Eq.(10), does indeed approximate Eq.(11) and Fig.4, so our assumption as to the form of $\underline{f}(\lambda)$ is justified.

Therefore a good approximation to the joint probability $\underline{f}(\lambda \text{ and } \underline{m})$ that a book of the collection both circulates \underline{m} times a year and also has demand parameter λ , is

$$\underline{f}(\lambda \text{ and } \underline{m}) \equiv \underline{p}(\underline{m}|\lambda)\underline{f}(\lambda) = \frac{1}{\underline{D}\underline{m}!} \left(\frac{\lambda\underline{m}}{\lambda + \underline{m}}\right)^{\underline{m}} \exp\left(-\frac{\lambda}{\underline{D}} - \frac{\lambda\underline{m}}{\lambda + \underline{m}}\right) \quad (16)$$

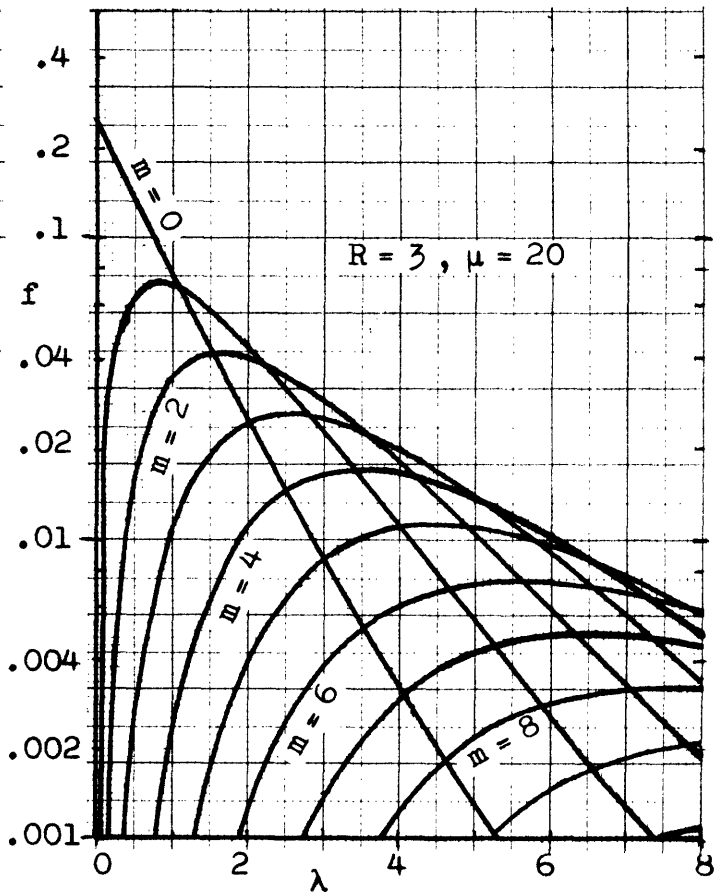
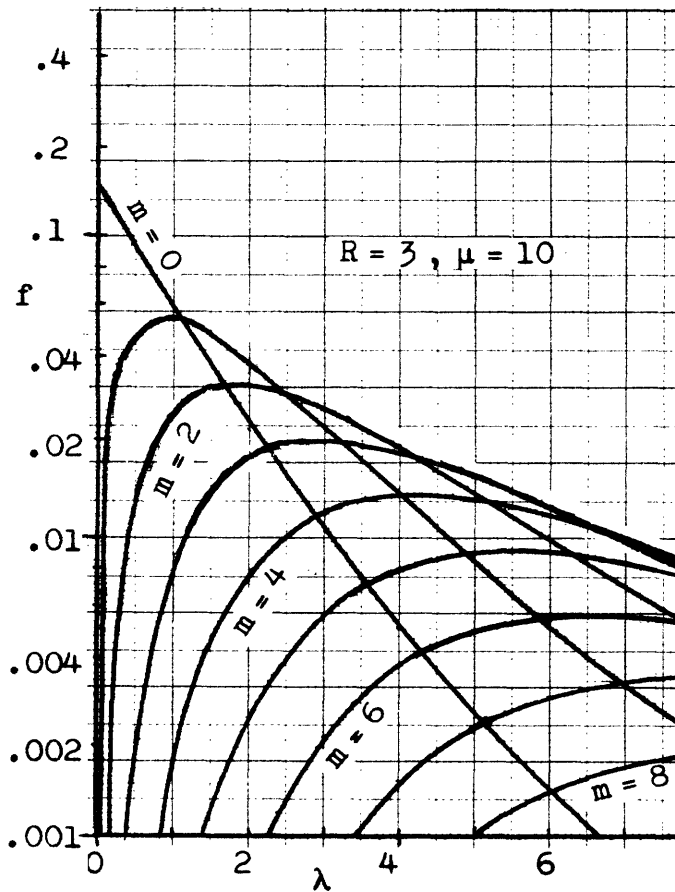
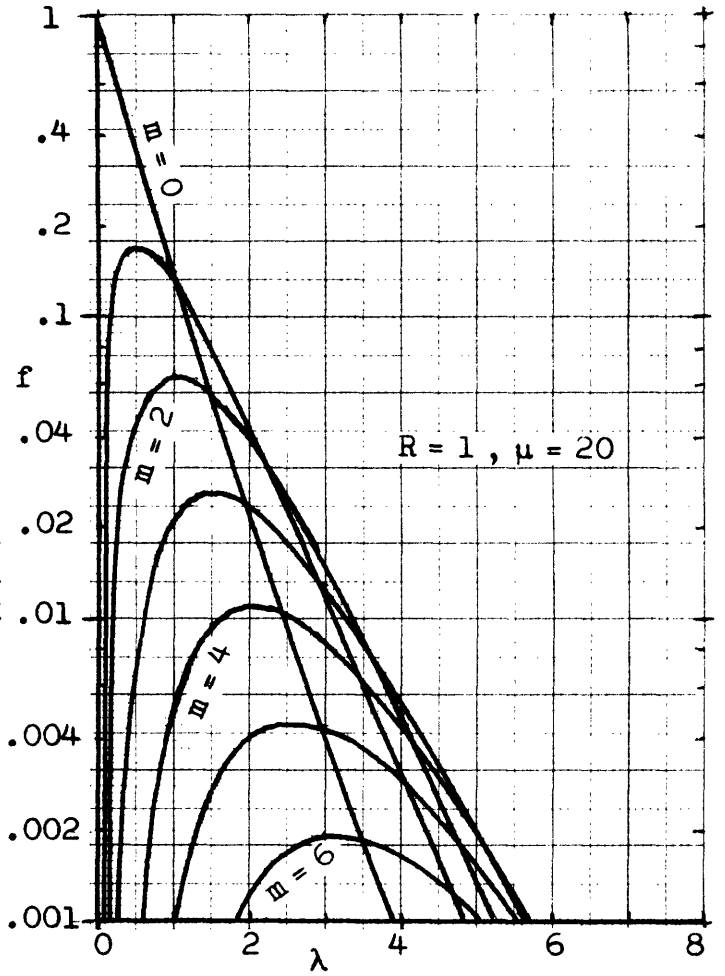
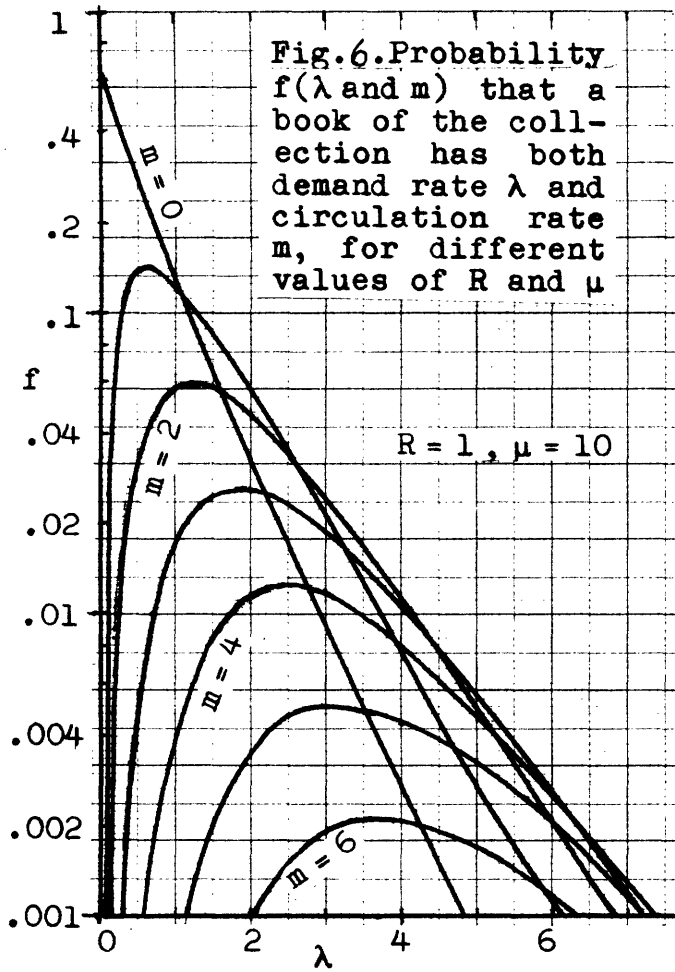
These quantities are plotted in Fig.6, as functions of λ , for a few different values of \underline{R} and of $\underline{\mu}$ (how \underline{D} depends on \underline{R} will be shown in the next section). The curves are on a logarithmic scale because $\underline{f}(\lambda \text{ and } \underline{m})$ drops off rapidly in value as λ differs more and more from its value at the peak.

The fraction of books circulating \underline{m} times a year ($\underline{m} = 4$, for example, for $\underline{R} = 1$ and $\underline{\mu} = 10$) is extremely small for λ small, rises to a maximum at a value of λ that depends on \underline{R} and on $\underline{\mu}$ (about $\lambda = 2.5$ for $\underline{R} = 1$, $\underline{\mu} = 10$) and then drops off rapidly. Values of $\lambda_{\underline{m}}/\underline{\mu}$ ($\lambda_{\underline{m}}$ is the value of λ at the maximum of \underline{f}) are given in Table IV, for different values of $\underline{m}/\underline{\mu}$ and of \underline{D} .

Likewise $\underline{f}(\lambda \text{ and } \underline{m})$ peaks sharply with circulation \underline{m} ; for example, for $\underline{R} = 3$, $\underline{\mu} = 20$ and $\lambda = 6$, \underline{f} is largest for $\underline{m} = 4$, the \underline{f} 's for $\underline{m} = 3$ and 5 are somewhat smaller, those for $\underline{m} = 2$ and 6 still smaller and those for $\underline{m} = 1$ and 8 or greater are less than one third that for $\underline{m} = 4$. As we can see from the figure and also from Table IV, the peaks come for demand λ approximately ^{proportional} λ

Insert, p. 17, p. 17

Fig. 6. Probability $f(\lambda$ and $m)$ that a book of the collection has both demand rate λ and circulation rate m , for different values of R and μ



to \underline{m} , the circulation; for $\underline{R} = 1$ the peaks are for λ about equal to $(2/3)\underline{m}$, for $\underline{R} = 3$ they are more nearly equal to \underline{m} and for \underline{R} greater than 3 they are for λ greater than \underline{m} . As \underline{m} increases the values of $\lambda_{\underline{m}}$ drop more and more below the value \underline{m} . This whole behavior is not surprising. A book can have a demand λ either greater or smaller than its actual circulation, as shown in Eq.(14) and Fig.3, because of the fluctuations in actual arrivals and circulation. For those books with circulation \underline{m} larger than the mean circulation \underline{R} of the collection, the chances are good that the circulation was somewhat better than expected; in other words the peak of the probability of demand for these books would be expected to come at a value $\lambda_{\underline{m}}$ somewhat less than \underline{m} . On the other hand, the peaks for \underline{m} about equal to \underline{R} and for \underline{R} larger than 1 would be expected to occur at demand values a bit larger than \underline{m} . In any case there always are some books, with circulation \underline{m} , that have demand λ larger than \underline{m} and others with λ smaller than \underline{m} . And, vice versa, of all the books with a given value of demand parameter λ , because of the fluctuating nature of circulation some have circulation less than λ and some happen to have \underline{m} larger than λ .

Mean Demand for the Collection - Our next task, having authenticated the form of $\underline{f}(\lambda \text{ and } \underline{m})$, given in Eq.(16), is to show how the mean demand \underline{D} , the as-yet unknown quantity in Eq. (16), can be given in terms of the measured values of \underline{R} and μ , thus closing the logical circuit. We combine Eqs.(4) and (10) with Eq.(16) and use the formula

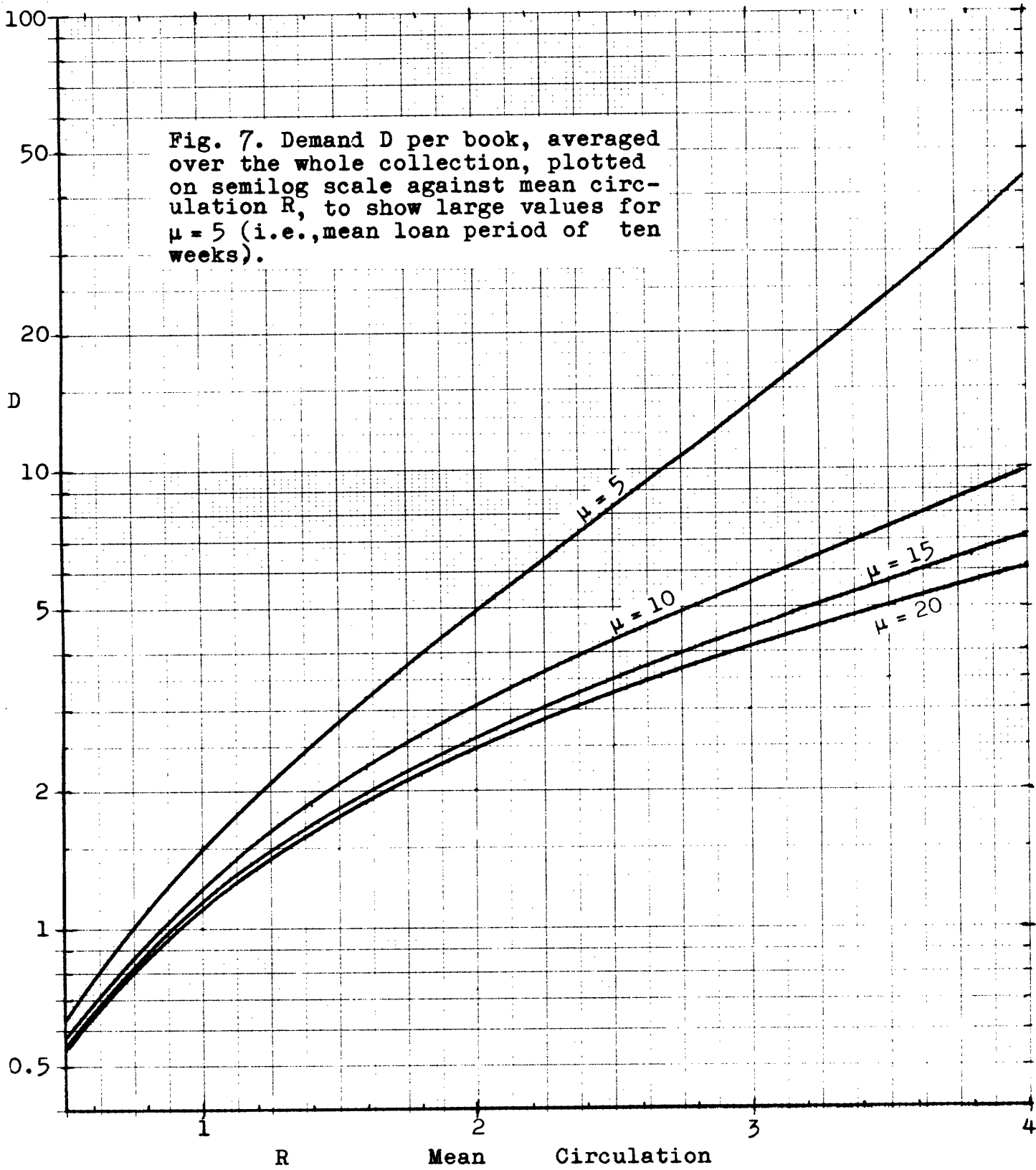
$$\sum_{\underline{m}=1}^{\infty} (\underline{m}/\underline{m}!) \underline{w}^{\underline{m}} = \underline{w} e^{\underline{w}} \quad \text{where} \quad \underline{w} = \frac{\lambda \mu}{\lambda + \mu}$$

to obtain

$$\begin{aligned}
 \underline{R} &= \sum_{\underline{m}=1}^{\infty} \underline{m} \underline{p}(\underline{m}) = \int_0^{\infty} \sum_{\underline{m}=1}^{\infty} \underline{m} \underline{f}(\lambda \text{ and } \underline{m}) \underline{d}\lambda \\
 &= \frac{1}{\underline{D}} \int_0^{\infty} \sum_{\underline{m}=1}^{\infty} \frac{\underline{m}}{\underline{m}!} \left(\frac{\lambda \underline{\mu}}{\lambda + \underline{\mu}}\right)^{\underline{m}} \exp\left(-\frac{\lambda}{\underline{D}} - \frac{\lambda \underline{\mu}}{\lambda + \underline{\mu}}\right) \underline{d}\lambda = \frac{1}{\underline{D}} \int_0^{\infty} \left(\frac{\lambda \underline{\mu}}{\lambda + \underline{\mu}}\right) e^{-\lambda/\underline{D}} \underline{d}\lambda \\
 &= \frac{\underline{\mu}}{\underline{D}} \int_0^{\infty} \left(1 - \frac{\underline{\mu}}{\lambda + \underline{\mu}}\right) e^{-\lambda/\underline{D}} \underline{d}\lambda = \underline{\mu} \left[1 - \frac{\underline{\mu}}{\underline{D}} e^{\underline{\mu}/\underline{D}} \int_{\underline{\mu}}^{\infty} \frac{1}{\underline{u}} e^{-\underline{u}/\underline{D}} \underline{d}\underline{u}\right] \\
 &= \underline{\mu} \left[1 - \frac{\underline{\mu}}{\underline{D}} e^{\underline{\mu}/\underline{D}} E_1(\underline{\mu}/\underline{D})\right] \quad (\underline{u} = \underline{\mu} + \lambda) \quad (17)
 \end{aligned}$$

where $E_1(x)$ is the exponential integral of x . From tables⁸ of this function we can find $(\underline{R}/\underline{\mu})$ as a function of $(\underline{D}/\underline{\mu})$ and then, by inversion, we can calculate the mean per book demand \underline{D} as a function of mean circulation \underline{R} , for different values of the return rate $\underline{\mu}$. The results are tabulated in Table I and are displayed in Fig.7.

The difference between the \underline{D} of Fig.7 and \underline{R} is the mean number of unsatisfied users, per book, for the collection. An estimate of the total yearly unsatisfied borrowers for the collection is thus $\underline{D} - \underline{R}$ times \underline{N} , the number of books in the collection. We see that \underline{D} rises more rapidly than mean circulation \underline{R} , particularly for long mean loan periods (small values of return rate $\underline{\mu}$). In fact for $\underline{\mu} = 5$ (mean loan period about 2 months, the mean demand rate per book is already $2\frac{1}{2}$ times the circulation rate when the mean circulation is only 2 per book. For $\underline{\mu} = 10$, which is typical of many university libraries (mean loan period, including renewals, about a month) with a mean circulation per book of 2.4 (not unusual for collections of technical books²) the mean demand per book per year is 4. If there are 4000 books in that collection, it means that about 16000 times during the year someone came to the library who



would have borrowed a book of the collection and in only 9600 cases was the person successful in borrowing it; in the other 6400 cases the desire was unsatisfied (or at least the prospective borrower did not get that book). For collections with even larger mean circulation rates the fraction of unsatisfied users is still larger.

It is now possible to estimate the effect on circulation by a change in loan period policy. Since it is unlikely that a change in such policy would make much of a change in the demand for a particular book, mean demand will remain constant as μ is changed and mean circulation \underline{R} does change; if the book is on the shelf a greater fraction of the year, more prospective borrowers will find a given book and so its circulation will increase. Although \underline{R} is the measurable quantity, \underline{D} is the more invariant quantity; hence the importance of finding a simple way to evaluate the \underline{D} for a collection.

For example, for a collection having a lax return policy ($\mu = 5$), with a mean circulation $\underline{R} = 2$ per book per year, Fig.7 shows that mean demand \underline{D} is about 5 per book per year. If now the return policy can halve the mean loan period (make $\mu = 10$) without changing \underline{D} , we see that mean circulation can be increased to 2.8. For a collection of 4000 books, therefore, the total circulation can be increased from 8000 a year to 11200 a year, without changing the books, simply by changing the loan policy. Although one can ^{without analysis,} guess, _^ that reducing the loan period will increase circulation, it is much more valuable to have a quantitative estimate of how much the circulation will increase, to assist in deciding whether a proposed change is worth while.

The Distribution in Circulation - Having now demonstrated a self-consistent chain of reasoning that allows us to obtain mean demand \underline{D} and the joint probability $\underline{p}(\lambda \text{ and } \underline{m})$, from measured values of \underline{N} , \underline{R} and $\underline{\mu}$ for a homogeneous collection, we can now investigate other ways by which the analysis can indicate how other changes in library policy would affect library use. To do this it is necessary to return to Eqs.(11) and (12), where we noted that we could calculate the circulation distribution $\underline{p}(\underline{m})$ if we could calculate the quantity $\alpha = \underline{p}(0)$ from \underline{R} and $\underline{\mu}$. We have now shown that this is possible. The details are given in Appendix A; the result is

$$\begin{aligned} \underline{p}(0) \equiv \alpha &= \int_0^{\infty} \underline{p}(\lambda \text{ and } 0) \underline{d}\lambda = \frac{1}{\underline{D}+1} \left[1 + \frac{2}{\underline{\mu}} \left(\frac{\underline{D}}{\underline{D}+1} \right)^2 + \dots \right] \\ &\simeq \left[1/(\underline{D}+1) \right] \quad (2/\underline{\mu} \text{ small}) \end{aligned} \quad (18)$$

For values of $\underline{\mu}$ as large or larger than about 8 we are justified in using the approximate formula, which means that we can rewrite Eq.(11) in terms of \underline{R} and of \underline{D} , which is determined by \underline{R} and $\underline{\mu}$,

$$\underline{p}(\underline{m}) \simeq \begin{cases} 1/(\underline{D}+1) & (\underline{m} = 0) \\ \frac{\underline{D}}{\underline{G}(\underline{D}+1)} \left(\frac{\underline{G}-1}{\underline{G}} \right)^{\underline{m}-1} & (\underline{m} > 0) \end{cases} ; \quad \underline{R} = \frac{\underline{D}\underline{G}}{\underline{D}+1} \quad (19)$$

$$\text{or } \underline{G} = \underline{R} \frac{\underline{D}+1}{\underline{D}} ; \quad \underline{\gamma} = \frac{\underline{G}-1}{\underline{G}} = 1 - \frac{\underline{D}}{\underline{R}(\underline{D}+1)}$$

Values of \underline{D} , α , \underline{Q} and $\underline{\gamma}$ are given in Table II for a few different values of \underline{R} and of $\underline{\mu}$; other values, for other values of \underline{R} and $\underline{\mu}$ can be computed by using Table I and Eqs.(19).

Values of $\underline{p}(\underline{m})$ are plotted in Fig.8 for different values of \underline{m} , $\underline{\mu}$ and \underline{R} . The smaller \underline{R} is the more rapidly $\underline{p}(\underline{m})$ reduces in value as \underline{m} increases; low probability of large \underline{m} means a small average circulation. As $\underline{\mu}$ is increased (as mean loan period is shortened), $\underline{p}(0)$, the fraction of the collection

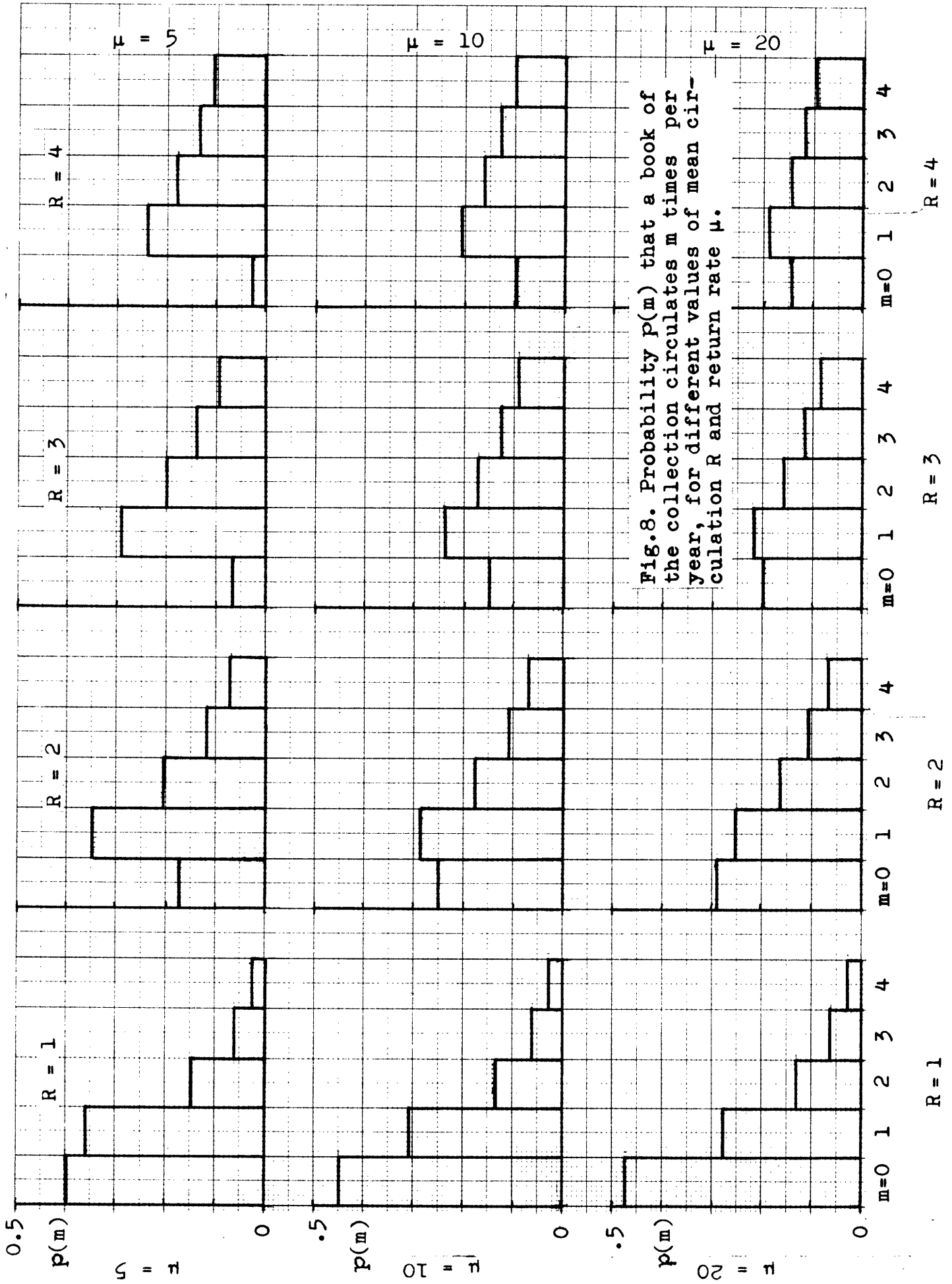


Fig.8. Probability $p(m)$ that a book of the collection circulates m times per year, for different values of mean circulation R and return rate μ .

not circulating at all increases; a point enlarged on in the next paragraph.

More extended values of $\underline{P}(\geq \underline{m})$, the fraction of the collection circulating \underline{m} or more times, are given in Table III, values of $\underline{p}(\underline{m}) = \underline{P}(\geq \underline{m}) - \underline{P}(\geq \underline{m} + 1)$ are obtained by subtraction. We note again the interesting characteristic of libraries with long loan periods; for collections with the same mean circulation, the longer the loan period (the smaller μ is) the smaller is the fraction $1 - \underline{P}(\geq 1) = \underline{p}(0)$ of books not circulating at all. The effect is not large when the mean circulation is small, but for heavily used collections the effect is noticeable. For $\underline{R} = 3$ and $\mu = 20$, one fifth of the collection does not circulate during the year; for $\underline{R} = 3$ and $\mu = 5$ only one fourteenth of the collection does not circulate. The effect is a compound one; if circulation rate is to be kept constant, lengthening the loan period (reducing the return rate μ) increases the mean demand \underline{D} and thus puts a greater pressure on all the books. The fraction of books with very high circulation ($\underline{m} = 5$ or more) cannot increase appreciably if \underline{R} is kept the same; therefore there must be fewer books with zero circulation. For a given library, of course, changing μ changes \underline{R} , not \underline{D} (as was discussed earlier), so a comparison, keeping \underline{R} constant and changing μ (and thus \underline{D}) is in effect comparing two libraries (or different collections in the same library) with the same mean circulation but with different loan policies.

The Effect of Duplicate Copies - Other changes affected by changed library policies also depend on λ rather than directly on circulation. Because the circulation interference, shown

in Eqs.(13) and (14) depends on the relation between demand and return rate, the reduction of interference caused by the presence of a duplicate copy of a popular book must be analyzed first in terms of demand; dependence on circulation is only indirect. The addition of a duplicate copy does not change the demand for the book; it increases circulation because the chance of both copies being out is less than the chance of one being off the shelf. According to queuing theory, if the demand rate for a particular book is λ , then the expected circulation rate for a single copy is, according to Eq.(13), $\lambda\mu/(\lambda+\mu)$, less than λ by the amount $\lambda^2/(\lambda+\mu)$, the measure of unsatisfied demand. Purchase of a duplicate copy brings the combined circulation up⁹ to $\mu\lambda(\lambda+\mu)/(\lambda^2+2\lambda\mu+2\mu^2)$, which is closer to λ but not yet equal to it, because there are still times when both copies are off the shelf. Therefore the increase in circulation per duplicate bought, for books with demand λ , is the difference between these,

$$\rho(\lambda) = \frac{\mu\lambda^2(\lambda+2\mu)}{(\lambda+\mu)(\lambda^2+2\mu\lambda+2\mu^2)} \quad (20)$$

If we knew the value of λ for each book it would be fairly simple to calculate, for example, how much increase in circulation could result from duplicating each book with demand greater than some value λ . But it is circulation that is known, so we must ask a different question; how much additional circulation will result if all books with circulation greater than m are duplicated? This is not the same question, nor is the procedure as effective a way to decrease unsatisfied demand as would choosing the books according to demand, if this were possible. But this is not possible, so we must answer the second question. To do this we must average $\rho(\lambda)$, weighted by

the joint probability $f(\lambda \text{ and } \underline{m})$, over all values of λ . This gives us the net gain in circulation, per book of the collection, produced by duplicating all books that circulate \underline{m} times a year.

The sharply peaked shape of the curves of the joint probabilities $f(\lambda \text{ and } \underline{m})$ enable us to make an easy transition from a quantity $\underline{S}(\lambda)$, depending on the demand for a book, on to the average value of \underline{S} for all books of the collection that have circulation rate \underline{m} . If $\underline{S}(\lambda)$ does not change much in value over the range of λ producing the peak of $f(\lambda \text{ and } \underline{m})$, then we can substitute, for $\underline{S}(\lambda)$ its value $\underline{S}(\lambda_{\underline{m}})$ at the maximum of $f(\lambda \text{ and } \underline{m})$, in the integral giving its average value,

$$\begin{aligned} \underline{s}(\underline{m}) &\equiv \int_0^{\infty} \underline{S}(\lambda) f(\lambda \text{ and } \underline{m}) d\lambda \\ &\simeq \underline{S}(\lambda_{\underline{m}}) \int_0^{\infty} f(\lambda \text{ and } \underline{m}) d\lambda = \underline{S}(\lambda_{\underline{m}}) \underline{p}(\underline{m}) \end{aligned} \quad (21)$$

where we have used Eq.(10) to take the last step. $\underline{S}(\lambda_{\underline{m}})$ is a good approximation to the average value of $\underline{S}(\lambda)$ over the books with circulation \underline{m} and $\underline{S}(\lambda_{\underline{m}})\underline{p}(\underline{m})$ is the per book contribution of \underline{S} to the whole collection, of all the books in it with circulation \underline{m} . Probability $\underline{p}(\underline{m})$ is given in Eq.(19) and some values can be obtained from Table III; $\lambda_{\underline{m}}$, the values of λ at the maximum of $f(\lambda \text{ and } \underline{m})$ are given in Table IV; Appendix B shows how they are calculated.

The gain in circulation $\rho(\lambda)$ of Eq.(20) is just such a function; its value does not change much over the range of λ where $f(\lambda \text{ and } \underline{m})$ is large. Therefore the average per book gain in circulation obtained by duplicating all books in the collection with circulation \underline{m} is

$$\underline{q}(\underline{m}) = \int_0^{\infty} \rho(\lambda) f(\lambda \text{ and } \underline{m}) d\lambda \simeq \rho(\lambda_{\underline{m}}) \underline{p}(\underline{m}) \quad (22)$$

Having values of $\rho(\lambda_{\underline{m}})$, $p(\underline{m})$ as functions of \underline{m} , for a collection with the specified values of \underline{R} and μ , we can quickly calculate

$$Q(\geq \underline{m}) = \sum_{n=\underline{m}}^{\infty} \rho(\lambda_{\underline{n}}) p(\underline{n}) \quad (23)$$

the expected increase in circulation^{per book} resulting if all the books in the collection with circulation \underline{m} or greater were to be duplicated. Table V gives a few sets of value of this increase, along with related values of $\rho(\lambda_{\underline{m}})$ and $P(\geq \underline{m})$, as functions of \underline{m} for different values of \underline{R} , and thus of \underline{D} , for $\mu = 10$ (other tabulations can be made for other values of μ and \underline{R} by use of Tables II and IV).

This Table V makes it possible to decide whether it really is worth while to buy duplicate copies of high-circulation books and, if it is, which books should be duplicated. If we decide to duplicate all books that have circulated \underline{m} times or more last year, then $P(\geq \underline{m})$ indicates what fraction of the collection will have to be duplicated and thus indicates the cost of so doing. The columns $Q(\geq \underline{m})$ show the average gain in per book circulation produced by the second copy. If now we divide $Q(\geq \underline{m})$ by $P(\geq \underline{m})$, the fraction of the collection that gets duplicate copies, the ratio is the average additional circulation produced per duplicate copy bought. If this is greater than the mean circulation \underline{R} , then every duplicate book bought is circulating more than the average book of the unduplicated collection, and therefore will have more use than the average book of the original collection. (To be quite accurate, we should take into account the fact that \underline{R} , for a given collection, gradually decreases from year to year²; however usually the change is not large and may be neglected.)

For example, we see from Table V, for $\underline{D} = 3$, $\underline{R} \approx 2$ and $\mu = 10$ (a fairly heavily used collection) that $\underline{Q}(\geq 7)/\underline{P}(\geq 7)$ is somewhat greater than 2; and from the value of $\underline{P}(\geq 7)$ that about one 22nd of the collection circulates 7 or more times. If the original collection was 1000 books (with total circulation 2000 per year) and we buy duplicate copies of the roughly 44 books that have circulated 7 or more times a year, we will have increased circulation by an additional 100-odd borrowings; not a great addition, but each of the 44 duplicates will be expected to add, on the average, more than the 2 circulations per book per year that the original collection has.

The gains are diluted by having to choose duplicatable books by circulation rather than by demand, but we have no simple, alternative way, as long as we cannot directly measure demand for a book. At least the analysis outlined in this paper shows, quantitatively, what can be done in guiding policy decisions if only circulation (and mean loan period) can be measured. And any gain in quantitative knowledge of what goes on is of value.

APPENDIX A.

Demonstration that the Analysis is Self-consistent.

According to Eq.(10), the integral of $f(\lambda \text{ and } m)$ over λ should equal the probability $p(\underline{m})$ that a book of the collection circulates m times a year. If the choice of Eq.(15), that the probability density $f(\lambda)$ that a book has demand λ , is correct, then the result of the integration should come out to conform with the modified geometrical form of Eq.(19), which corresponds fairly well with actual collections, as shown in Fig.4 and elsewhere². To check this, $f(\lambda \text{ and } m)$ was integrated over λ numerically for a whole sequence of values of \underline{m} , for several different values of μ and of \underline{R} , and thus, via Eq.(17) of D.

The results turn out to check quite well with the geometric form of Eq.(19), as well as in actual magnitude. Two examples are shown in Fig.9. The solid lines are the geometric form of Eq.(19), with \underline{G} and γ as required by the chosen values of \underline{R} and μ . The circles and crosses are the results of integrating the $f(\lambda \text{ and } m)$ of Eq.(16), with \underline{D} as given by Eq.(17) for the chosen values of \underline{R} and μ . The correspondence is at least as good as that between Eq.(19) and the data for an actual collection of books, as shown in Fig.4 and elsewhere².

The circles and crosses begin to fall below the straight lines for large values of \underline{m} . This also seems to be characteristic of the circulation distribution of actual collections, though the small numbers of high-circulation books renders the data rather uncertain. It is possible, however, that our assumption of an exponential shape for $f(\lambda)$, as given in Eq.(15), is slightly more in accord with reality than is the form of $p(\underline{m})$

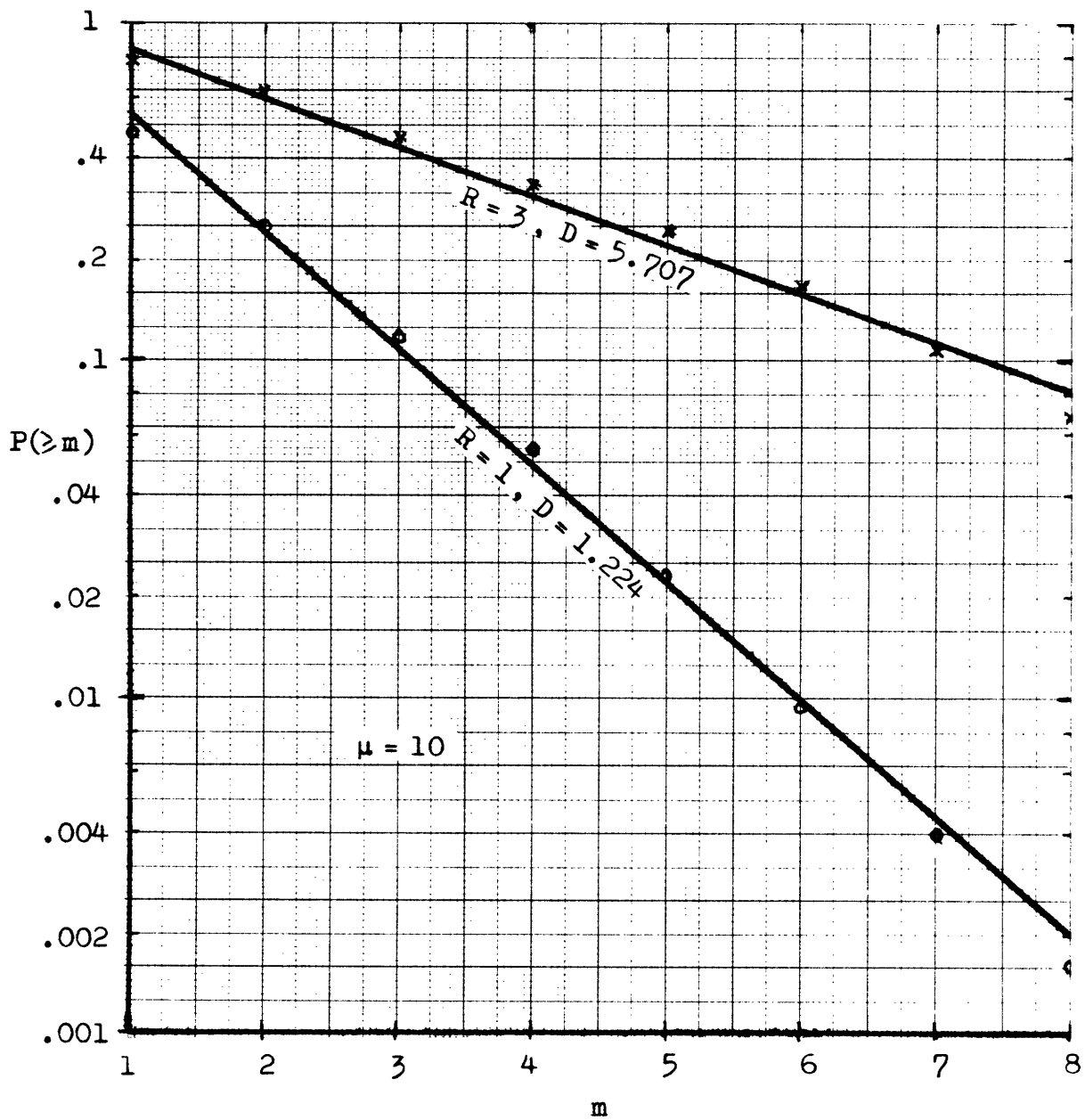


Fig.9. Correspondence of $P(\geq m)$ of Eq.(19) (solid lines) with $P(\geq m)$ obtained by numerical integration of the $f(\lambda$ and $m)$ of Eq.(16) (circles and crosses) for $\mu = 10$ and two values of R , and thus of D .

given in Eq.(19). These fine details, however, are going beyond the accuracy of the data available at present. It thus has been demonstrated that the formulas in this paper, and the graphs and tables accompanying them, are within the degree of accuracy justified by the data, and also the degree of accuracy required by the use of the results in determining policy.

APPENDIX B.

Locating the Peaks of $f(\lambda$ and $m)$.

As shown in Eq.(21) approximate values of the integral of $\underline{S}(\lambda)\underline{f}(\lambda$ and $\underline{m})$, for any slowly-varying function \underline{S} , are obtained by using the value of \underline{S} at the maximum of \underline{f} and then using the approximation of Eq.(19) for $\underline{p}(\underline{m})$. The calculation of course depends on finding the value of λ at the maximum of $\underline{f}(\lambda$ and $\underline{m})$.

We can write

$$f(\lambda \text{ and } \underline{m}) = \frac{1}{\underline{Dm}!} \left(\frac{\lambda \underline{\mu}}{\lambda + \underline{\mu}} \right)^{\underline{m}} \exp\left(-\frac{\lambda}{\underline{D}} - \frac{\lambda \underline{\mu}}{\lambda + \underline{\mu}}\right) = \frac{\underline{\mu}^{\underline{m}}}{\underline{Dm}!} e^{-\underline{\mu} \underline{F}(\underline{y})}$$

where

$$\underline{F} = -\frac{\underline{m}}{\underline{\mu}} \ln\left(\frac{\lambda}{\lambda + \underline{\mu}}\right) + \frac{\lambda}{\underline{\mu} \underline{D}} + \frac{\lambda}{\lambda + \underline{\mu}} = -\underline{a} \ln(1 - \underline{y}) + \frac{1}{\underline{D} \underline{y}} - \underline{y} + 1 - \frac{1}{\underline{D}}$$

$$\underline{y} = \frac{\underline{\mu}}{\lambda + \underline{\mu}} = 1 - \frac{\lambda}{\lambda + \underline{\mu}} ; \quad \lambda = \underline{\mu} \frac{1 - \underline{y}}{\underline{y}} ; \quad \underline{a} = \frac{\underline{m}}{\underline{\mu}}$$

The maximum of this function comes at $(\underline{dF}/\underline{dy}) = 0$, or at

$$1 + \frac{1}{\underline{D} \underline{y}^2} - \frac{\underline{a}}{1 - \underline{y}} = 0 \quad \text{or at} \quad \underline{y} = 1 - \frac{\underline{a}}{1 + (1/\underline{D} \underline{y}^2)} \quad (\text{B1})$$

This last equation can be solved numerically to find the root, $\underline{y}_{\underline{m}}$, and thence $\lambda_{\underline{m}} = (\underline{\mu}/\underline{y}_{\underline{m}})(1 - \underline{y}_{\underline{m}})$, the value of λ at the maximum of $\underline{f}(\lambda$ and $\underline{m})$. Values of these roots, for different values

of $\underline{a} = \underline{m}/\mu$ and of \underline{D} are given in Table IV. Values for intermediate values of \underline{a} or \underline{D} may be obtained by 4-point interpolation if needed.

Notice that λ itself varies relatively slowly over the range of values in which $\underline{f}(\lambda \text{ and } \underline{m})$ is non-negligible; therefore λ itself satisfies the requirements for the function $\underline{S}(\lambda)$ of Eq.(21). Consequently, as required by Eqs.(7), (9) and (10),

$$\delta(\underline{m}) \equiv \int_0^{\infty} \lambda \underline{f}(\lambda | \underline{m}) \underline{d}\lambda = \frac{1}{\underline{p}(\underline{m})} \int_0^{\infty} \lambda \underline{f}(\lambda \text{ and } \underline{m}) \underline{d}\lambda$$

and then, from Eq.(21) with $S = \lambda$,

$$\delta(\underline{m}) \simeq \frac{\lambda_{\underline{m}}}{\underline{p}(\underline{m})} \int_0^{\infty} \underline{f}(\lambda \text{ and } \underline{m}) \underline{d}\lambda = \lambda_{\underline{m}} \quad (\text{B2})$$

Thus the $\lambda_{\underline{m}}$ of Table IV is an approximate value for $\delta(\underline{m})$, the mean value of per-book demand for all those books in the collection that circulate \underline{m} times a year. As already mentioned in the discussion of Fig.6, this mean value of demand $\lambda_{\underline{m}}$ is smaller than the circulation $\frac{\underline{m}}{\underline{a}}$ of the books it represents when \underline{D} (and therefore \underline{R}) is small. As \underline{D} is increased, however, mean demand $\delta(\underline{m}) \simeq \lambda_{\underline{m}}$ rises above \underline{m} ($\lambda_{\underline{m}}/\mu$ rises above $\underline{a} = \underline{m}/\mu$) the more quickly the smaller \underline{m} is. In every case, of course, the weighted average of $\delta(\underline{m})$ over the whole collection (which is \underline{D}) is larger than the weighted average of \underline{m} over the collection (which is \underline{R}).

References.

1. Morse, Philip M., Library Effectiveness, MIT Press, Cambridge, 1968, Chapters 6 and 7.
2. Chen, Ching-Chih, Applications of Operations Research Models to Libraries, MIT Press, Cambridge, 1976, Chapter 6
See also Reference 1, Chapters 7 and 8.
3. Bradford, S.C. Documentation, Crosby Lockwood, London, 1948.
4. Morse, Philip M. The Geometric and the Bradford Distributions, a Comparison, in process of publication.
5. Morse, Philip M. Queues, Inventories and Maintenance, John Wiley and Sons, New York, 1958.
6. Reference 1, page 119.
7. See, for example, Parzen, Emanuel, Modern Probability and its Applications, John Wiley and Sons, New York, 1960, page 251.
8. See, for example, the N.B.S. Handbook of Mathematical Functions, ed. M.Abramowitz and I.Stegun, U.S.Gov.Printing Office, Washington, D.C. 1964, pg. 228 et seq.
9. Reference 1, pg. 71 et seq.

TABLE I

Values of R/μ vs. D/μ and vice versa.

D/μ	R/μ	D/μ	R/μ	D/μ	R/μ	D/μ	R/μ	D/μ	R/μ
0	0	.40	.241185	.80	.361210	1.20	.439033	1.60	.495279
.05	.045629	.45	.259934	.85	.372615	1.25	.447004	1.65	.501280
.10	.084367	.50	.277343	.90	.383465	1.30	.454671	1.70	.507094
.15	.118072	.55	.293578	.95	.393794	1.35	.462055	1.75	.512731
.20	.147889	.60	.308778	1.00	.403652	1.40	.469171	1.80	.518200
.25	.174617	.65	.323052	1.05	.413075	1.45	.476039	1.85	.523510
.30	.198814	.70	.336493	1.10	.422093	1.50	.482670	1.90	.528666
.35	.220896	.75	.349186	1.15	.430738	1.55	.489080	1.95	.533676
.40	.241185	.80	.361210	1.20	.439033	1.60	.495279	2.00	.538545

R/μ	D/μ	R/μ	D/μ	R/μ	D/μ	R/μ	D/μ	R/μ	D/μ
0	0	.10	.1224	.20	.3028	.30	.5707	.40	.9812
.02	.0208	.12	.1531	.22	.3479	.32	.6391	.42	1.0882
.04	.0433	.14	.1862	.24	.3968	.34	.7135	.44	1.2060
.06	.0677	.16	.2221	.26	.4500	.36	.7949	.46	1.3360
.08	.0940	.18	.2609	.28	.5079	.38	.8838	.48	1.4797
.10	.1224	.20	.3028	.30	.5707	.40	.9812	.50	1.6391
.12	.1531	.22	.3479	.32	.6391	.42	1.0882	.52	1.8167

See Eq.(17)

TABLE II,

Values of \underline{D} , $1-\alpha$, \underline{G} and γ for different values of \underline{R} and μ .

\underline{R}	$\mu = 5$				$\mu = 10$			
	\underline{D}	$1-\alpha$	\underline{G}	γ	\underline{D}	$1-\alpha$	\underline{G}	γ
0	0	0	1.0000	0	0	0	1.0000	0
0.2	0.216	.1776	1.1259	.1118	0.208	.1722	1.1615	.1391
0.4	.470	.3197	1.2511	.2007	.433	.3022	1.3238	.2446
0.6	.765	.4334	1.3843	.2776	.677	.4037	1.4863	.3272
0.8	1.111	.5263	1.5201	.3421	.940	.4845	1.6511	.3943
1.0	1.514	.6022	1.6605	.3978	1.224	.5504	1.8170	.4496
1.2	1.984	.6648	1.8048	.4459	1.531	.6049	1.9838	.4959
1.4	2.540	.7175	1.9512	.4875	1.862	.6506	2.1519	.5353
1.6	3.196	.7617	2.1006	.5240	2.221	.6895	2.3204	.5690
1.8	3.975	.7990	2.2528	.5561	2.609	.7229	2.4899	.5984
2.0	4.906	.8807	2.4077	.5847	3.028	.7517	2.6605	.6241
2.2	6.030	.8578	2.5648	.6101	3.479	.7767	2.8324	.6469
2.4	7.398	.8809	2.7244	.6329	3.968	.7987	3.0048	.6672
2.6	9.086	.9091	2.8862	.6535	4.500	.8182	3.1778	.6853
2.8	11.18	.9179	3.0504	.6722	5.079	.8355	3.3513	.7016
3.0	14.22	.9343	3.2110	.6886	5.707	.8509	3.5257	.7164
3.2	17.15	.9449	3.3866	.7047	6.391	.8647	3.7007	.7298
3.4	21.42	.9554	3.5587	.7190	7.135	.8771	3.8765	.7420
3.6	26.91	.9642	3.7338	.7322	7.949	.8883	4.0529	.7533
3.8	34.03	.9715	3.9117	.7444	8.838	.8984	4.2300	.7636
4.0	43.32	.9774	4.0923	.7556	9.812	.9075	4.4077	.7731

See Eqs.(17) and (19).

TABLE II, continued.

Values of \underline{D} , $1-\alpha$, \underline{G} and γ for different values of \underline{R} and μ .

\underline{R}	$\mu = 15$				$\mu = 20$			
	\underline{D}	$1-\alpha$	\underline{G}	γ	\underline{D}	$1-\alpha$	\underline{G}	γ
0	0	0	1.0000	0	0	0	1.0000	0
0.2	0.205	.1701	1.1756	.1494	0.204	.1694	1.1804	.1528
0.4	.422	.2968	1.3479	.2581	.416	.2938	1.3615	.2655
0.6	.650	.3939	1.5231	.3434	.636	.3888	1.5434	.3521
0.8	.890	.4709	1.6989	.4114	.866	.4641	1.7238	.4199
1.0	1.143	.5334	1.8749	.4666	1.105	.5249	1.9050	.4751
1.2	1.410	.5851	2.0511	.5124	1.354	.5752	2.0863	.5207
1.4	1.690	.6283	2.2284	.5512	1.612	.6172	2.2685	.5592
1.6	1.985	.6650	2.4060	.5844	1.880	.6528	2.4511	.5920
1.8	2.296	.6966	2.5840	.6130	2.156	.6831	2.6349	.6205
2.0	2.623	.7240	2.7625	.6380	2.448	.7100	2.8170	.6450
2.2	2.969	.7480	2.9410	.6600	2.749	.7333	3.0003	.6667
2.4	3.332	.7692	3.1203	.6795	3.062	.7538	3.1838	.6859
2.6	3.715	.7879	3.2999	.6970	3.386	.7720	3.3679	.7031
2.8	4.118	.8046	3.4799	.7126	3.724	.7883	3.5519	.7185
3.0	4.542	.8196	3.6605	.7268	4.076	.8030	3.7360	.7323
3.2	4.984	.8329	3.8421	.7397	4.442	.8162	3.9204	.7449
3.4	5.455	.8451	4.0233	.7514	4.822	.8282	4.1051	.7564
3.6	5.952	.8562	4.2048	.7622	5.218	.8392	4.2899	.7669
3.8	6.476	.8662	4.3868	.7720	5.629	.8491	4.4751	.7765
4.0	7.031	.8755	4.5689	.7811	6.056	.8583	4.6605	.7854

See Eqs.(17) and (19).

TABLE III.

Values of Cumulative Probability $P(\geq m)$

\underline{m}	$\mu = 5$	$\underline{R} = 1$		$\underline{R} = 2$			$\underline{R} = 3$			$\underline{R} = 4$		
		10	20	5	10	20	5	10	20	5	10	20
0	1	1	1	1	1	1	1	1	1	1	1	1
1	.602	.550	.525	.831	.752	.710	.934	.851	.803	.977	.908	.858
2	.240	.248	.249	.486	.469	.458	.643	.610	.588	.739	.702	.674
3	.095	.111	.119	.284	.293	.295	.443	.437	.431	.558	.542	.529
4	.038	.050	.056	.166	.183	.191	.305	.313	.315	.422	.419	.416
5	.015	.023	.027	.097	.114	.123	.210	.224	.231	.319	.324	.327
6	.006	.010	.013	.057	.071	.079	.145	.161	.169	.241	.251	.257
7	.002	.005	.006	.033	.044	.051	.100	.115	.124	.182	.194	.202
8	.001	.002	.003	.019	.028	.033	.069	.082	.091	.137	.150	.158
9	.000	.001	.001	.011	.017	.021	.047	.059	.066	.104	.116	.124
10	-	.000	.001	.007	.011	.014	.033	.042	.049	.079	.090	.098
11	0	0	0	.004	.007	.009	.022	.030	.036	.059	.069	.077
12				.002	.004	.006	.015	.022	.026	.045	.054	.060
13				.001	.003	.004	.011	.016	.019	.034	.041	.047
14				.001	.002	.002	.007	.011	.014	.026	.032	.037
15				.001	.001	.002	.005	.008	.010	.019	.025	.029
16				0	.001	.001	.004	.006	.008	.015	.019	.023
17					0	.001	.002	.004	.006	.011	.015	.018
18						0	.002	.003	.004	.008	.011	.014
19							.001	.002	.003	.006	.009	.011
20							.001	.002	.002	.005	.007	.009

See Eq.(19).

$$p(\underline{m}) = P(\geq \underline{m}) - P(\geq \underline{m} + 1)$$

TABLE IV.

Values λ_m of Demand λ at the maximum of $p(\lambda \text{ and } m)$, for different values of D and of $a = m/\mu$. Note that $\lambda_m \simeq \delta_m$, the average demand for a book of the collection with circulation m .

Values of λ_m/μ .

a	$D = 1$	2	3	4	5	6	8	10
0.2	0.0996	0.1382	0.1602	0.1747	0.1850	0.1928	0.2039	0.2113
.5	.2442	.3532	.4249	.4780	.5198	.5540	.6075	.6480
.8	.3801	.5629	.6927	.7954	.8812	.9555	1.0800	1.1827
1.1	.5068	.7596	.9462	1.0987	1.2298	1.3460	1.5471	1.7197
1.0	.4656	.6956	.8637	1.0000	1.1163	1.2188	1.3949	1.5445
1.5	.6625	1.0000	1.2551	1.4675	1.6529	1.8191	2.1117	2.3670
2.0	.8393	1.2695	1.5987	1.8751	2.1180	2.3369	2.7245	3.0650
3.0	1.1479	1.7321	2.1820	2.5615	2.8961	3.1985	3.7355	4.2086
4.0	1.4142	2.1249	2.6728	3.1356	3.5436	3.9126	4.5682	5.1459
5.0	1.6511	2.4709	3.1029	3.6367	4.1075	4.5334	5.2897	5.9565

Values of $\rho(\lambda_m)$, mean per-duplicate gain in circulation for books with circulation rate m , as function of D and of $a = m/\mu$.

Values of $\rho(\lambda_m)/\mu$

a	$D = 1$	2	3	4	5	6	8	10
0.2	.00857	.01562	.02036	.02373	.02626	.02821	.03160	.03302
.5	.05250	.07663	.10139	.12028	.13534	.14772	.16703	.18158
.8	.08578	.15092	.19746	.23322	.26203	.28603	.32425	.35384
1.1	.13066	.22091	.28309	.32979	.36683	.39730	.44516	.48170
1.0	.11584	.19851	.25624	.30000	.33493	.36382	.40949	.44458
1.5	.18676	.30000	.37365	.42693	.46802	.50104	.55158	.58903
2.0	.24809	.37749	.45650	.51142	.55257	.58490	.63314	.66795
3.0	.34402	.48418	.56245	.61410	.65144	.68002	.72148	.75056
4.0	.41422	.55366	.62728	.67443	.70784	.73306	.76914	.79410
5.0	.46765	.60276	.67146	.71461	.74483	.76746	.79954	.82158

See Eqs.(20), (22), (23) and (B2).

For each value of D and of μ , mean circulation R can be determined from Tables I or II.

TABLE V.

Values of $P(\geq m)$, fraction of books circulating m or more times a year ^{and} of $\rho(\lambda_m)$, mean gain in circulation per duplicated book; also of $Q(\geq m)$, expected gain in circulation per book of the collection and of $Q(\geq m)/P(\geq m)$, added circulation per duplicate, if all books circulating m or more times were to be duplicated, all as functions of m and of D (and thus of R) for $\mu = 10$.

$\mu = 10, D = 2, R = 1.4789$ $\alpha = .3333, G = 2.2183, \gamma = .5492$					$\mu = 10, D = 3, R = 1.9881$ $\alpha = .25, G = 2.65095, \gamma = .6228$				
m	$P(\geq m)$	$\rho(\lambda_m)$	$Q(\geq m)$	Q/P	$P(\geq m)$	$\rho(\lambda_m)$	$Q(\geq m)$	Q/P	
2	.3661	0.1562	.14745	0.403	.4671	0.2036	.3035	0.650	
3	.2011	.3340	.12166	.605	.2909	.4306	.2676	.920	
4	.1104	.5395	.09139	.828	.1811	.7072	.2204	1.217	
5	.06065	.7663	.06453	1.063	.1128	1.0139	.1721	1.526	
6	.03331	1.0079	.04358	1.308	.07025	1.3349	.1289	1.835	
7	.01829	1.2577	.02844	1.554	.04375	1.6581	.0935	2.137	
8	.01005	1.5092	.01807	1.798	.02725	1.9746	.0662	2.429	
9	.00552	1.7560	.01123	2.034	.01697	2.2779	.0459	2.705	
10	.00303	1.9851	.00686	2.264	.01057	2.5624	.0313	2.961	
11	.00166	2.2091	.00415	2.500	.00658	2.8309	.0211	3.207	

$\mu = 10, D = 4, R = 2.4119$ $\alpha = .20, G = 3.0148, \gamma = .6683$					$\mu = 10, D = 5, R = 2.7734$ $\alpha = .1667, G = 3.3281, \gamma = 6995$				
2	.5346	0.2373	0.4746	0.888	.5829	0.2626	.6460	1.108	
3	.3573	.5077	.4325	1.210	.4078	.5603	.6000	1.471	
4	.2388	.8377	.3723	1.559	.2853	.9339	.5314	1.863	
5	.1596	1.2028	.3060	1.917	.1997	1.3534	.4513	2.262	
6	.10665	1.5832	.2423	2.272	.1396	1.7897	.3702	2.652	
7	.07127	1.9636	.1863	2.614	.09765	2.2179	.2951	3.024	
8	.04763	2.3322	.1399	2.937	.06831	2.6203	.2300	3.367	
9	.03183	2.6810	.1030	3.236	.04778	2.9893	.1763	3.688	
10	.02127	3.0000	.0747	3.512	.03343	3.3493	.1333	3.991	
11	.01422	3.2979	.0535	3.762	.02338	3.6683	.0997	4.279	
12	.00950	3.5653	.0380	4.000	.01636	3.9727	.0739	4.517	
13	.00635	3.8183	.0267	4.205	.01144	4.2363	.0544	4.755	
14	.00424	4.0527	.0188	4.434	.00800	4.4710	.0398	4.975	
15	.00234	4.2693	.0130	4.577	.00560	4.6802	.0291	5.196	

See Eqs.(19), (20), (22) and (23).