Demand-Adjusted Shelf Availability Parameters

by Paul B. Kantor

One measure of library performance is the probability that a patron can find the book or item that he or she seeks. This parameter (document availability) is best measured by sampling the expressed demand. We describe a simple technique for adjusting data obtained by a shelf-list sample so that the adjusted parameters reflect the performance as it would be measured by sampling expressed demand. Two adjustment schemes are discussed. A scheme based upon three classes of items, and adjusted for the varying levels of demand associated with these classes, is advocated. An example is worked out in detail.

The most important results of any study of the availability of documents are parameters characterizing the causal factors which inhibit patrons from finding the items which they seek For this reason, the collection of data based upon expressed demand, as introduced by Buckland et al and elaborated by Kantor and others is the exactly appropriate tool [1] The collection of data based upon a sample drawn from the shelf list, introduced by Kaske and elaborated by Altman and de Prospo is, on the other hand, far less intrusive [2] The latter technique necessarily fails to provide data on the adequacy of the collection, and on the user's skills at the catalog and the shelf [3] More significantly, the data which it provides on circulation interference and on the "other library factors" are subject to a significant inherent bias [4] The purpose of the present article is to lay out a straightforward technique which compensates for this bias

Exposition

Parameters Correlated with Demand

It is clear, upon reflection, that not all of the materials in a circulating collection are equally in demand In particular, large old libraries are likely to have collections whose breadth and depth extend far beyond the interests and demands expressed by the current library users In this situation the probability that an item may be found on the shelf is an overestimate of the probability that an item in the special subset of materials currently in demand will be found on the shelf The materials in high demand are precisely the ones which are more likely to be checked out, or lost in some process

If we knew nothing else about the collection, analysis could not proceed any further without direct inquiry into the expressed demand Fortunately there is considerable evidence provided not only by common sense, but by the experience of many medical libraries, and the detailed data of the Pittsburgh study, suggesting that certain obvious parameters are correlated with demand [5] Three such parameters are accession number, imprint date, and time elapsed since last circulation

For each of these parameters one may establish three rough groups youngest items (by accession number, or by date, or by time spent on the shelf since last circulation), middle-aged items, and elder items The precise definition of the boundary lines need not be the same for each library, as the corrections described below are calculated for each library separately Typical groups might be less than one year, more than one year but less than five, and more than five years, for date since last check out The ranges for other variables would depend upon the history of the library's growth There is considerable freedom in the choice

Naive and Partially Adjusted Parameters

With three categories defined (to be specific, let us imagine that they are based upon accession date), one then draws a sample from the shelf list, traces the items, and records the results in a three-by-three table, as shown in Figure 1 From these data we will first calculate a "naive" branching diagram, which implicitly assumes that demand is distributed uniformly over the collection The totals are placed at the heads of the arrows, and the performance parameters are calculated from these totals, as shown in Figure 2

It is clear, from the data themselves, that the proportion of items in the first row which were found to be circulating is higher than the corresponding proportion in the other two rows This is surely a reflection of the fact that they are more heavily in demand, and suggests that the rows be weighted

Paul B Kantor is President, Tantalus, Inc , Cleveland, OH

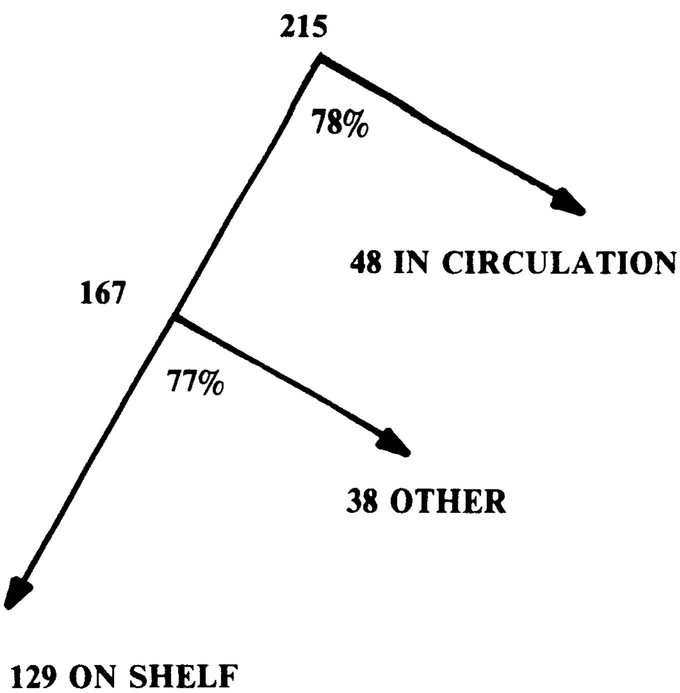
The Journal of Academic Librarianship, vol 7, no 2, p 78-82

Figure 1
Results of a Shelf List Study of Item Availability*

Group	In Circulation	On Shelf	Other
	Status		
Young	27	13	21
Middle	15	40	9
Elder	6	76	8
Totals	48	129	38

* Rows represent groups which are expected to be not all equal in demand

Figure 2
Availability Parameters
Calculated on a Branching Diagram
(The value at each vertex is found by dividing the total at it into the total one step down, e g , 77% = 129 / 167)



according to the incidence of circulation The results of such a calculation are shown in Figure 3, which represents a data analysis worksheet The weight factor for each row is simply the number in the first column, divided by the row total The adjusted results are then summed, column by column These numbers, as a result of the weighting process, are no longer integers However, their grand total will be equal to the original total in the first column

When these new column totals are transferred to a branching diagram (Figure 4) the performance parameters must be recalculated, and they are found to be significantly different from the naive results They will, in addition, be closer to the parameters measured in a study of the expressed demand Failure to make some adjustment could result in a serious bias against libraries with younger collections (which are more uniformly in demand) in favor of older collections with a large proportion of currently unused material

In view of the significant change which such a correction can produce, it is appropriate to consider whether the simple correction used here is complete It assumes that the relative demand for various items is the same as the relative chance that the items are presently checked out There is certainly a direct correlation, but there is not strict proportionality[7] This is because, when an item is heavily in demand, circulation interference becomes important, some borrowers are frustrated (the point of the whole study), and circulation does not keep up with demand

Fully Adjusted Performance Measures

A more sophisticated adjustment is possible, based on a formula which relates the demand for an item to the chance that it is checked out This formula is derived by Morse, for the specific case of a single copy and a Poisson distribution of demand [6] (These technical assumptions are probably adequate for all but a few unusual libraries) If Q represents the chance that a book is checked out, and R represents the demand rate (measured in "requests arriving during an interval equal to the average duration of a loan"), the demand can be calculated from the equation

$$R = Q/(1-Q)$$

In the present application, the fraction $Q/(1-Q)$ is given by the first entry in each row divided by the sum of the other two The calculations resulting from this more refined model are shown in Figure 5 The corresponding branching diagram is shown in Figure 6 We see that there is a further significant change in the performance parameters, although it is smaller than the change produced by the first rough adjustment

Discussion

The procedure is substantially the same when the three groups are based upon imprint date When time elapsed since last circulation is to be used several technical problems arise First, with a photorecord circulation system, the necessary data are not available Second, with a traditional card file it is not possible to know the time of last circulation unless either the book or the card is located Thus the analysis based on date since last circulation (which is attractive in principle, since this datum is clearly correlated with demand) can only be done at libraries where the circulation records are fully automated, and maintained in some journal form (or as compact histories for the individual items)

Although the number of fully automated circulation systems is not presently large, it seems clear that there will be many more in the coming years Building in the capability for a demand-adjusted availability study of the type described here would save time and effort for the library director It should be incorporated as a design or purchase constraint by those libraries which have not yet committed themselves to a specific computerized circulation system

Let us briefly review the availability data obtained by shelf-list studies in the general context of document availability The patron's quest for a specific known item is seen to be

Figure 3
Data Collection and Analysis Form for Demand-Adjusted Shelf Availability
(A better approximation of the true availability parameters is found by multiplying the entries in each row by the fraction of items in that row which is found to be checked out)

	Raw Data				Adjusted Data		
Group	In Circulation	On Shelf	Other	Adjustment Factor	In Circulation	On Shelf	Other
Young	27	13	21	27/61	11 95	5 75	9 30
Middle	15	40	9	15/64	3 52	9 38	2 11
Elder	6	76	8	6/90	40	5 07	53
Totals	48	129	38		15 87	20 20	11 94

frustrated in five possible ways which, in their logically correct order, are the library has never acquired the item, the patron cannot use the catalog successfully, the item is checked out, the item is somehow lost or misplaced or, finally, the patron does not find the item, though it is on the shelf The overall chance of success (document availability measure) is the product of the five individual factors defined here Of these five, it is only the third and fourth which can be measured in a study of items drawn from the shelf list With the corrections described in this article, the results of such a study should yield the same values for these two performance parameters as would be found in a full study based upon expressed demand Whitlatch and Kieffer[7] have demonstrated that the other parts of a study can be carried out separately

A Worked Example

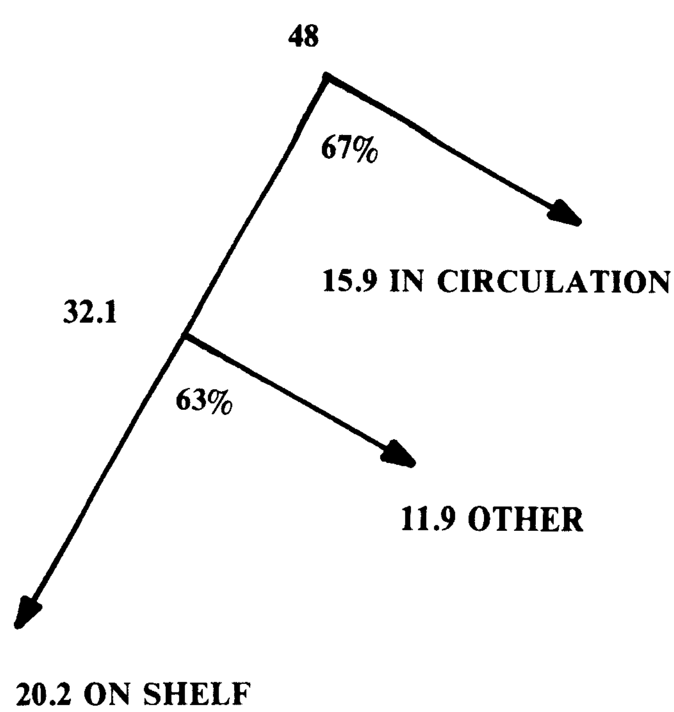
The preceding has been a theoretical discussion To see how the procedure works in practice we will follow a hypothetical study from beginning to end A library with a collection of 120,000 items decides to use a sample of 500 items in a quick study of shelf availability Staff members are aware that such a study will not isolate certain user skills (or lack thereof), nor will it signal occasions on which the items which users need are not in the collection at all However, they are interested in estimating the relative significance of circulation interference vis-a-vis other library controlled factors For these purposes, a sample drawn from the shelf list is the appropriate tool

The sample is drawn with the aid of a table of random numbers, with the first two digits indicating the catalog drawer and the next two indicating the depth (in inches) at which the card is to be picked The title and call number are noted on tally sheets, together with the acquisition number and space for recording the book status The acquisition numbers are then divided into three roughly equal groups Let us suppose they are

1 since 76-500
2 before 76-500, but more recent than 68-001
3 before 68-001

The second phase of the study requires checking to determine the status of each item For efficiency the survey workers

Figure 4
Branching Diagram
Corresponding to the Simple Correction
(Note that the availability parameters are significantly reduced)



begin at the circulation file and note each item which is found to be checked out For all other items they then proceed to the correct location on the shelves and determine that the item is either there or "elsewhere " A typical line on the tally sheet looks like this

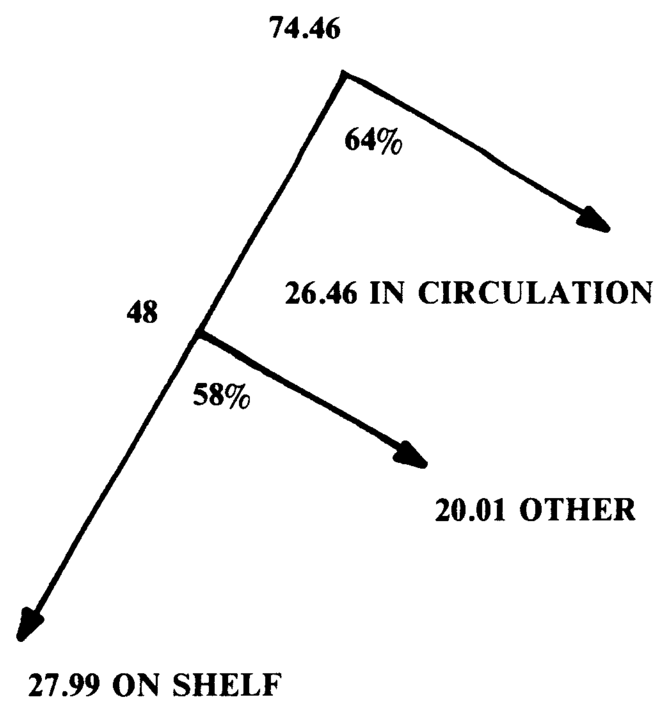
von Neumann, Mathematical Foundations of Quantum 57-130 C S O

(The code "CSO" stands for "Circulating, Shelf, or Other" and has been printed on the worksheets so that tracers may simply cross out the appropriate symbol when they determine the status of the item)

Figure 5
Calculations of a More Sophisticated Correction Which Recognizes that Demand Actually Exceeds the Circulation

	Raw Data				Adjusted Data		
Group	**In Circulation**	**On Shelf**	**Other**	**Adjustment Factor**	**In Circulation**	**On Shelf**	**Other**
Young	27	13	21	27/34	21 44	10 32	16 68
Middle	15	40	9	15/49	4 59	12 24	2 76
Elder	6	76	8	6/84	43	5 43	57
Totals	48	129	38		26 46	27 99	20 01

Figure 6
Branching Diagram for the More Sophisticated Correction
(The parameters are reduced even further)



When every item has been traced the data analysis begins The total number of items in the "young" category is found to be 183, of which 46 are circulating, 110 are on the shelf, and the rest (27) are "somewhere else " The total number of items in the "middle-aged" category is found to be 154, of which 24 are circulating, 119 are on the shelf, and the rest (11) are "somewhere else " Finally, for the third category, there are 163 items, of which only 12 are circulating, 143 are on the shelf, and 8 are "somewhere else "

These data are arranged into a table as follows

	In Circulation	On the Shelf	Other
Group 1	46	110	27
Group 2	24	119	11
Group 3	12	143	8

Following the procedure described above, each row is multiplied by a weighting factor This factor represents the relative demand for items in the corresponding row It is even greater than the fraction which is out because, when demand is heavy, check outs do not keep up with it, due to circulation interference For the first row the weighting factor is

(number out) / (number not out) = 46/137 = 0 336

For the second row it is 24/130 = 0 185, and for the third row it is 12/151 = 0 079

When each row is multiplied by its corresponding weight factor the adjusted table becomes

	In Circulation	On the Shelf	Other
Group 1	15 44	36 9	9 1
Group 2	4 43	22 0	2 0
Group 3	0 95	11 4	0 6
Totals	20 82	70 3	11 7

The entries in this table are then entered into a branching diagram with the following interpretation Of the total which are not circulating (70 3 + 11 7 = 82), 70 3 are on the shelf This represents a library performance factor of 70 3/82 = 86% Of the grand total drawn from the library collection (82 + 20 82 = 103), 82 are not checked out This represents a circulation availability factor of 82/103 = 80% The overall availability when both factors are taken into account is the product of these two numbers, or 68% Since the second of these performance factors (corresponding to circulation) is definitely lower than the first, the library decides to work on getting books back from circulation more promptly

In this hypothetical example we see that the adjusted calculation reveals a combined availability ratio of 68 3%, while, without adjustment, it would have been overstated as 372/500 = 74 4% The importance of this adjustment is underscored by the fact that the "statistical error" assigned to the naive measurement would have been plus or minus 2% In other words, the adjustment for uneven distribution of demand is three times as important as the sampling errors

To sum up, the adjustment for uneven distribution of demand can make a significant difference in the calculation of shelf availability The only difficulty is a conceptual one some may feel uncomfortable at the fact that the adjusted table contains "fractions of items " This is a natural consequence of the fact that each item traced is really a representative of its whole class, and, with a small sample, fractions must occur Although this is a hypothetical example (not based upon an actual study) it exhibits the important features of the process Although one-third of the items traced are in the oldest group, that group as a whole is less in demand, and the items in the third row of the adjusted table account for only about one-seventh of the total of all entries in that table

The problem of estimating the accuracy of these adjusted availability factors leads to some complicated formulae, which are outlined in the Appendix The central point should not, however, be obscured by any formula

Not all items are equally in demand, and ignoring that fact can lead to serious overestimates of document availability By multiplying the data gathered in each class by the demand factor

Demand factor = (number checked out) / (all others)

these overestimates can be avoided

Acknowledgments

This problem was brought to my attention by Harold Olsen, chairman of the LAMA Committee on College and University Library Statistics and Katherine Emerson, chair of the ANSI Z 39 Standards Committee Since some standard is to be adopted, it seems especially important to have one which is as realistic and free from inherent bias as possible Thanks also to Katherine Emerson and Richard Dougherty for comments on the manuscript of this article

References

[1]M K Buckland, A Hindle, A G Mackenzie, and I Woodburn, "Systems Analysis of a University Library," Final Report on a Research Project, University of Lancaster Occasional Papers, No 4, 1970, see also Buckland et al , *Book Availability and the Library User* (New York Pergamon Press, 1978), P B Kantor, "Availability Analysis," JASIS, 27, pp 100-12, T Saracevic, W Shaw, and P B Kantor, "Causes and Dynamics of User Satisfaction in an Academic Library," *College and Research Libraries* 36 (1977), 7-18, J B Whitlatch and K Kieffer, "Service at San Jose State University Survey of Document Availability," *Journal of Academic Librarianship* 4 (1978) 196-99

[2]N K Kaske, "Effectiveness of Library Operations A Management Information Systems Approach and Application," Library Science Dissertation, University of Oklahoma, 1973, E Altman, E De Prospo, P M Clark, and E C Clark, "A Data Gathering and Instructional Manual for Performance Measures in Public Libraries " (Chicago Celadon Press, 1978)

[3]Whitlatch and Kieffer, pp 196-99, B Lipetz, "Catalog Use in a Large Research Library," *Library Quarterly* 42 (1972) 129-39, J A Urquhart, and J L Schofield, "Measuring Reader's Failure at the Shelf " *Journal of Documentation* 27 (1971) 273-86

[4]P B Kantor, "Vitality, an Indirect Measure of Relevance," *Collection Management* 2 83-95

[5]Pittsburgh University, *A Cost-Benefit Model of Some Critical Library Operations in Terms of Use of Materials,* Pittsburgh University of Pittsburgh, 1978 (Allen Kent, Project Director)

[6]P M Morse, *Library Effectiveness* (Cambridge, MA MIT Press, 1968)

[7]Whitlatch and Kieffer, pp 196-99

▼

Appendix

The error analysis associated with this correction procedure could become quite complicated, since several mathematical operations are involved Careful analysis shows that the crucial source of statistical error is generally the number in the first row and column of the data table, since it contributes multiplicatively to the dominant terms of the corrected table A good rough estimate of the standard error E is given by

$$E^2 = ((2-W)W-(1-P))^2 \cdot W(1-W)A / N^2$$

In this expression W is the weight factor for the first row, A is the row total for that row, P is the adjusted chance that an item is not checked out, and N is the total number of items found to be circulating The complete error formula (for the square of E) is the sum of three such expressions, one for each row In the particular case used in the exposition of this paper, the full sum is

$$E^2 = (3\ 18 + 0\ 39 + 0\ 07) / (48 \times 48), \text{ or } E = 4\%$$

The interpretation of this standard error is that the observed value of P (the fraction not circulating) should not deviate from the true value by more than twice the standard error more than 5 percent of the time The terms within the parentheses come from the first, second, and third rows respectively, and it is clear that the first row dominates in the calculation of the error