

## On the Stability of Distributions of the Type Described by Trueswell

*Application of rules for weeding that are based upon the unequal distribution of demand over the collection require that the distribution remain stable over time. A mathematical expression is derived that tests that stability. Verification of the expression is not inordinately time consuming and is particularly easy in the case of automated circulation systems.*

TRUESWELL has introduced an interesting technique for examining the distribution of use (and/or demand) over a collection of circulating books.<sup>1</sup> Items are grouped into classes according to the time that has elapsed since their last recorded use (circulation), resulting in a distribution curve. This distribution may be studied for the items that are active during a given sample period, which may be a few days or an entire semester. It may also be studied (by sampling methods) for the collection as a whole. Making the reasonable assumption that, on the average, the items that have most recently been used satisfy most of the demand, one may select a cutoff interval, such as three years. If, say, 60 percent of the items observed at circulation have been active within the last three years while only 30 percent of the entire collection has, we may say that "30 percent of the collection satisfies 60 percent of the demand for circulating materials." In Trueswell's best-known presentation of this model, the numbers were 20 percent and 80 percent, with the result that rules of this type are referred to generally as "20/80 rules."

A recent paper in this journal, which reports on such a study, suggests that there may be some confusion as to the generality and import of rules of the 20/80 type.<sup>2</sup> The purpose of the present note is to clarify the

nature and use of such rules and to present a new test that bears on their validity and usefulness.

A specific example of a 20/80 rule for which, as mentioned above, the two parameters need not be "20" and "80" (nor need they sum to 100) will apply only to the specific library at which it was measured. Such rules are of use to the managers of crowded libraries because they indicate the potential benefits to be gained from an extended effort to weed the collection. If, for example, the parameters were 50/50, 60/60, and so on, it would indicate that all the parts of the collection are, on the average, equally in demand. In such a case there would be no point in trying to decide which parts are not heavily in demand. The decisions with regard to removal or remote storage must be made on some other basis, or at random. On the other hand, with a parameter set such as 10/90 one would be encouraged to look for the 90 percent of the collection that is in very low demand.

Thus we see that "20/80 rules" do not provide a universal numerical value, but, as a class, they provide a convenient way of characterizing the nonuniformity of the distribution of demand over the circulating items in a collection. As a practical matter, they are of little interest unless there is a crowding problem of some kind, and they are not a guide to further action unless the two numbers involved are quite far from each other.

*Paul B. Kantor is president, Tantalus, Inc., management consultants, Cleveland, Ohio.*

Implicit in the application of 20/80 rules there is an assumption that, to our knowledge, has not yet been discussed in the literature: the specific rule derived for a given library is relatively stable over time. If it is stable over time, then a remeasurement one, two, or three years hence will lead to the same parameters and the same policy conclusions. If the parameters are not stable over time, it would not be wise to base any substantial policy decisions upon the information obtained at one particular instant.

In principle, stability can be tested by repeating the determination of the parameters in several successive years. When there is a pressing need to reduce the collection, it is not practical to wait. In the following paragraphs we will outline a method for testing the stability of the 20/80 parameters in a time period that may be as short as a week. The central idea is that the chance that a book will appear at the circulation desk is related to its position along the curve describing distribution of demand in just such a way that *the distribution measured at the circulation desk is the derivative of the distribution measured in the collection as a whole*. The argument can be expressed in terms of obscure mathematical objects (the Laplace transform of the distribution of demand<sup>3</sup>), but we believe that the following less formal argument conveys the essence of the proof.

For any collection of items we may define  $F(t)$  to be the fraction of the collection that has, at this moment, been inactive for at least a time  $t$ . (For example, if  $F(\text{one week}) = 90$  percent, it means that 90 percent of the collection has been neither checked out nor acquired during the past week.) In order to include every item we treat acquisition as an "activity" of the item. Clearly  $F(0) = 100\%$  and  $F(t)$  must decrease steadily as the argument  $t$  increases.

In dealing with a circulating collection there are two such distributions to be considered:

$F_{\text{CIRC}}^d(t)$  = the distribution corresponding to books that are checked out during some sample interval of length  $d$ .

$F_{\text{COLL}}(t)$  = the distribution correspond-

ing to books in the collection as a whole.

These distributions will be different in shape, unless the present demand for an item is completely independent of its age. In making a 20/80 analysis, the first of these functions is used to define a cutoff time. This time is then used as the argument of the second function to complete the description. For example, there is some age  $t_{80}$  such that only 20 percent of the circulating items will have been inactive for longer than  $t_{80}$ . Mathematically this is expressed by the equation shown as equation 1. This means that 80 percent of all items active during the sample period have either been acquired or circulated during the most recent period of length  $t_{80}$ . One then samples the circulation history of the collection as a whole to ask what fraction of it has either been acquired or circulated during the same time period. If the inactive fraction  $F_{\text{COLL}}(t_{80})$  has the value 0.60, then the rest (40 percent) must account for 80 percent of the circulation. This is a 40/80 rule. If we make the comparison on the basis of the period in which 95 percent of the items that circulate during the sample period have had a prior circulation, we will get a "something/95" rule, and so on. The "something" in this case will, of course, be larger than 40 percent.

$$F_{\text{CIRC}}^d(t_{80}) = 0.20 = 1 - .80 \quad [1]$$

It might seem that a longer data collection period ( $d$ ) is always preferable to a shorter one. If the circulation volume is large enough, however, collection during a short period can provide not only adequate statistics but also an important test of the stability of the distribution  $F_{\text{COLL}}(t)$ .

The condition for stability is simply that  $F_{\text{COLL}}$  be the same at the end of the measurement interval as it was at the beginning. During that interval some of the items have aged gracefully (by an amount  $d$ ) while others have been active and have changed their position on the distribution curve. Mathematically the stability relation can be expressed in terms of the average demand ( $a$ ) in the form shown in equation 2. (Calculations are simplified by noting that the product of  $a$  and  $d$  is simply the number of

circulations occurring during the measurement interval.)

$$F_{\text{COLL}}^{\text{after}}(t) = F_{\text{COLL}}^{\text{before}}(t - d) - adF_{\text{CIRC}}^d(t) \quad [2]$$

If the interval  $d$  is taken to be relatively small (such as a week), then the expression  $F_{\text{COLL}}^{\text{before}}(t - d)$  can be approximated in terms of the derivative of that function (equation 3) where the prime denotes differentiation.

$$F_{\text{COLL}}^{\text{before}}(t - d) = F_{\text{COLL}}^{\text{before}}(t) - dF_{\text{COLL}}^{\text{before}\prime}(t) \quad [3]$$

Finally, imposing the stability condition that  $F_{\text{COLL}}^{\text{before}}$  and  $F_{\text{COLL}}^{\text{after}}$  are precisely the same function, we have equation 4.

$$dF_{\text{COLL}}^{\text{before}\prime}(t) = -adF_{\text{CIRC}}(t) \quad [4]$$

Thus one may test the stability of the distribution of demand in a relatively short period of time by determining  $F_{\text{CIRC}}$  directly, determining  $F_{\text{COLL}}$  by a sampling study, and comparing the former with the derivative of the latter. More important, once stability has been established, it is no longer necessary to measure the function  $F_{\text{COLL}}$  directly. It is, instead, sufficient to measure the distribution  $F_{\text{CIRC}}$  and compute the

other distribution by numerical integration. This provides the library manager with information on the two distribution curves at a substantially lower cost.\*

The process of determining  $F_{\text{CIRC}}$  is greatly simplified in the presence of a suitably designed automated circulation system, which retains the date of last activity for each item. However, the most important step, which should be taken by any library planning to use a 20/80 rule, is to establish the stability of the parameters. The mathematical relation derived above provides a particularly prompt and inexpensive means for doing so.

In order to research this question further, Tantalus, Inc., will perform the necessary mathematical test for the first five libraries that care to submit information on both  $F_{\text{CIRC}}$  and  $F_{\text{COLL}}$ .

\*The mathematically inclined reader will find a more general discussion, for long observation spans, in R. W. Trueswell and S. J. Turner, "Stimulating Circulation Use Characteristic Curves Using Circulation Data," *Journal of the American Society for Information Science* 30:83-88 (1979).

#### REFERENCES

1. Richard W. Trueswell, "A Quantitative Measure of User Circulation Requirements, and Its Possible Effect on Stack Thinning and Multiple Copy Determination," *American Documentation* 16:20-25 (1965); Richard Trueswell, "Some Behavioral Patterns of Library Users, the 80/20 Rule," *Wilson Library Bulletin* 43:458-61 (1969).
2. Seymour H. Sargent, "The Uses and Limitations of Trueswell," *College & Research Libraries* 40:416-23 (1979); see also the comment by Trueswell that immediately follows this paper (p.424-25).
3. Paul B. Kantor, "Vitality: An Indirect Measure of Relevance," *Collection Management* 2:83-95 (1978).