Library & Information Science Research 34 (2012) 317–323

Contents lists available at SciVerse ScienceDirect

Library & Information Science Research

A long-tail inspired measure to assess resource use in information services

Petros A. Kostagiolas [a,*], Nikolaos Korfiatis [b], Marios Poulos [a]

[a] Department of Archives and Library Science, Ionian University, Ioannou Theotokis 72, Corfu 49100, Greece
[b] Department of Computer Science, Institute for Informatics and Mathematics, Goethe University Frankfurt, Frankfurt am Main, Germany

ARTICLE INFO

Available online 5 August 2012

ABSTRACT

A measure for the assessment of resource use in information services provides an application within the information services area of the *long tail* concept. The analysis quantifies the long tail concept and is applicable to all libraries and information services, independent of their size and circulation properties. The paradigm of the Lorenz curve and the Gini coefficient from macroeconomics is adapted in the context of information services. The Gini coefficient examines the skewness of the Lorenz curve against an ideal cumulative distribution that is represented by a 45-degree line. In this context, the Lorenz curve depicts the cumulative percentage of collection items versus the number of loans and identifies the different levels of library resource use on the part of users by measuring how close the Gini coefficient is to the unity (i.e., the Gini of the 45-degree line is identified).

1. Introduction

A key issue for any organization to consider is whether its operation is the most efficient in relation to the goals set by the management. In the context of an organization that deals with the provision of information resources, whether in intangible or tangible form, a main objective is to make the resources available to the users, so that they can make an effective use of them in order to accomplish their tasks (Bawden, 1990). For instance, within a library, one particular objective is concerned with how effectively the library's collection items are used. A common approach in demonstrating the effectiveness of the operation of an information service such as an academic library is usually the number of people who use the library, and the number of loans recorded for items in the library's collection. Although this metric provides an overall user activity picture, it is rather naïve, and fails to take into account the demand characteristics of individual collection items, and the favorability they might experience due to seasonal popularity, for example, items used as a resource in an academic course. Therefore, a more thorough examination of the library characteristics is required (Bollen & Van de Sompel, 2008).

This becomes pertinent when decisions for funding (e.g., allocation of library resources or collection expansion) or justifying of the library to external funding bodies have to be undertaken; such an evaluation is a critical precedent of funding decisions. In fact, because of the rapid advances of information technology and the Internet, the distribution of the demand should be studied in order to devise a measure of the use of library resources.

2. Problem statement

User preferences and behavioral patterns over resource utilization affect demand in a number of ways, some of which are rather easily understood, while others require a closer investigation. Management decision-making can be supported by studying the characteristics of the library demand distribution and by associating this analysis with a number of other library specific details, including:

- collection properties (Dewey classification, collection size held locally, permanent or temporary access rights acquired for accessing remote resources, etc.);
- availability of physical and electronic facilities and resources, and
- results of surveys regarding the informational, research, educational, cultural or recreational needs of library users, and so forth.

Some approaches to this kind of analysis have emerged from the field of economics and others from management science. None of the available methods directly addresses issues related to assessing the shape of the demand distribution. The proposed method is based on the *long tail* concept and employs the econometric concepts of the Lorenz curve and the Gini coefficient. These are properly adapted to the context of academic libraries through a case study. This method goes beyond a simple representation of the distribution characteristics and provides illuminating evidence for the interpretation of collection use by patrons. Such a measure can be associated with a variety of other details in order to reflect the collection-use performance of a library; it can provide insights into core drivers for management for undertaking a dialogue with the user's demands.

The development of a quantitative measure for the assessment of the thickness of the tail of the demand distribution provides insights

* Corresponding author.
E-mail address: pkostagiolas@ionio.gr (P.A. Kostagiolas).

0740-8188/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.lisr.2012.05.002

into the use of the library's collection. Such a methodology might help management staff increase awareness of the current level of library collection use, and this could lead to actions for enhancing library efficiency and to improving total use of low-demand materials. Fig. 1 presents a visualization of the definition of short-tail and long-tail areas in the context of the demand distribution of a library. The long-tail or short-tail theoretical framework explaining the demand model may hint at a number of actions. Naturally, the library should seek to improve the current level of user choices, or create new choices for titles or information resources.

This research presents a simple but innovative way to measure and interpret (a) the use of an information-service provider's collection (here, a library), and (b) the efficiency of different information-service providers or collections using a measure that is independent of size and circulation properties.

3. Literature review

3.1. The long tail concept for libraries

Kuhlthau (2004) provided a detailed study of assessing and evaluating the services of a library, with a focus on its resource utilization and, more particularly, the importance of individual user-demand characteristics, along with information-seeking behavior in requesting and making use of a particular item in a library collection. In particular Kuhlthau (1991) argued that the demand for a resource is largely influenced by external and internal factors that direct the user toward the decision to borrow or not borrow a particular item. This can be interpreted through the Pareto distribution-inspired axiom of the 80/20 rule, in which 20% of collection items accounts for 80% of loans (Britten, 1990; Burrell, 1985; Koch, 1998; Nisonger, 2008; Trueswell, 1969); items are divided according to their popularity. Although the 80/20 rule provides an easy and understandable way of characterizing the popularity of collection items, its crisp definition creates a disadvantage for its interpretation. Specifically, the 80/20 rule generically describes only the popular items, and not the demand behavior of the remaining 80% of the items in the library collection. Furthermore, less popular items are becoming more and more popular—primarily because of the availability of digital resources and electronic information systems that provide easy access to the whole library collection—hence making the tail of the demand-distribution thicker. These increasingly popular items provide exceptions to the 80/20 axiom.

A closer look on what happens to the 80% of items that might or might not actually account for 20% of loans suggests that the concept of the long tail comes to prominence when related to the issues primarily concerning digital information resources and electronic library systems (Anderson, 2008). The long-tail theory represents a challenge for library scholars trying to develop successful user-oriented management strategies. That is because, as opposed to the 80/20 rule, the long-tail theory directs attention toward possible actions to make users aware of less popular collection items, and how the accessibility of the collection resources can be enhanced. For example, in the field of marketing, the long tail provides a valuable strategy for improving the sales of online goods and exposing these goods using niche-oriented marketing campaigns (Brynjolfsson, Hu, & Smith, 2006). This approach assumes that management should decide and then find ways to systematically channel user demand at the tail of the demand distribution. For libraries and information services, this can create mechanisms that turn the available information niches (identified by the long tail) into an economic and cultural force of an individualized user-choice capability, which in turn will make the provision of resources more distributed, and users more dependent on it.

In the analysis of the so-called long-tail economy, Anderson (2008) identified six themes of the long-tail era (far more niche goods than hits; cost of reaching those is falling; new distribution tools drive demand down the tail; the demand curve flattens; accumulation of niches comprise a market rivaling the hits; and, the natural shape of the demand is revealed) and three forces (the democratization of production, the democratization of distribution, and the connection between supply and demand). In a more profound sense, this new era requires the study of the demand curve and a subsequent quantification of collection use. Dempsey (2006) noted that, “The long tail is not simply about having a lot of material: it is about ensuring that that material is readily discoverable and accessible, and that it is seen by those who need to see it.” When associated with other library specific matters (e.g., personnel and other resource availability, and collection size), the assessment of the demand curve may actually provide an indication for the operational efficiency of information-resource utilization—that is, the minimization of resource utilization for creating maximum benefit (Hernon & McClure, 1990). Such theoretical discussions of the new techno-socioeconomic forces produce a novel challenge for library management professionals and academics within a theoretical framework. For example, the question of whether or not to retain older titles relates to the actual title usage frequency over time of publication.

In the context of printed publications, an analysis of requests for stored biomedical journal volumes published before 1993 showed that older biomedical journal literature received substantial attention (Starr & Williams, 2008). Thus, the authors concluded that retaining older printed volumes, or providing easy access to older literature in electronic form, is required for meeting user information needs. Although the long tail idea has, in one way or another, been diachronically present in libraries, some authors have demonstrated the potential of the long tail through a combination of physical and virtual services in order to further engage users and embrace new content, such as the employment of blogs and wikis in order to provide a place for news, events, and discussion, build a sense of community, increase user participation, and decrease library response time (Casey & Savastinuk, 2006). Long-tail considerations for mirroring user needs onto the decisions made about collection development and optimizing library materials access could be considered one of the most important library operations (Holt, 2007).

In a more technical sense, the long-tail concept relates to the statistical property of a demand distribution that is thickening in the tails, showing that the demand share that is generated by the large number of low demand items exceeds that of the relatively few high-demand items. The long-tail concept is applicable to most libraries not only in the demand for electronic books, journals, and other library materials, but also to library services, such as interlibrary loans and their educational implications (Brown & Adler, 2008; Dempsey, 2006). Although electronic library systems do not necessarily imply that all libraries have ideal long-tail distributions, the long-tail concept can provide a yardstick to use in identifying critical insights related to collection use by users, as well as identifying issues that the library has to assume in order to increase this use (user training, search innovations, exploitation of digital collections and technologies, etc.). Therefore, measuring the exact length of the long tail is a

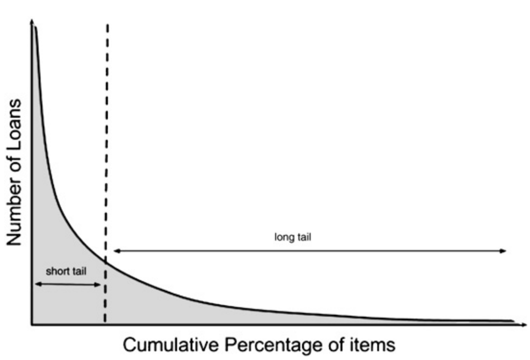

Fig. 1. Long tail versus short tail in the library and information services context.

key issue, as is how to ascertain this measurement as independently as possible from the number of users that the collection targets, and the number of items that this collection serves. All these issues serve as motivation, since a measure to assess resource use of information services is actually a measure of information-service resources exploitation, and at the same time a measure of the demand distribution tail.

3.2. Long-tail versus short-tail

The long tail is viewed as an indication of the significance of the total usage of low-demand materials (McGrath, 2004). User demand should be driven to the whole library collection through formal library policies and through the employment of electronic library services. As Storey (2005) suggested, the long tail philosophy requires libraries to make everything available by digitizing their collections to make them available both electronically and physically, and to provide a wide range of information beyond the local holdings, including holdings of other libraries and commercial publishers. Moreover, new services are developed in order to help connect the library collection with the users (supply with demand), and increase the number of user choices by thickening the tail of the demand curve.

In the digital economy, the long-tail culture seems to dominate over traditional economy, but it should be noted that for a specific library scenario, the opposite may occur. For example, one may consider a more static library culture, where the library has a print collection that concentrates only on the most popular titles. This library orientation is based on library policies that channel user demand towards a small number of print titles in order to minimize handling and cost. Hence, library services are *substitutable*, meaning that they are largely indistinguishable from one library to another, with no individual character. In this rather static sense, libraries are managed with a focus on cost and rapid response time. Limited organizational innovations are developed to reduce the time needed to respond to demand, and collection development is designed to drive users to the head of the distribution demand curve—that is, the most popular titles in the collection. For example, a short-tail library case may be found in a typical school library where students are guided toward the same books that are constantly borrowed. For the school-library scenario, items are available to a relatively small number of people, and it should be noted that “user choice” may or may not be conceived as a compulsory activity that is directly linked to current curriculum areas, or as a desired activity linked to reading for pleasure, perhaps in areas outside the current curriculum. In that context, Genoni (2007) considered space and financial limitations, and suggested that while “individual libraries might increasingly find it attractive to focus on the development and grooming of the head, this may not be the case for a system as a whole.” Digitizing material and developing repositories may appeal as viable solutions for individual library storage problems, while at the same time improving the resource use of library and information services.

3.3. Lorenz curve and demand inequality: Implications of econometrics

In macroeconometrics, the *income inequality* theory tries to describe how the distribution of wealth is depicted among households and individuals between nations or within a country (Gottschalk & Smeeding, 1997). As noted, a specific aspect of this research is to produce an easy-to-follow metric that can monitor the distribution of wealth by the number or percentage of households that belong to that category. A popular instrument for depicting income distribution is the Lorenz curve (Lorenz, 1905), which graphically represents a cumulative distribution function, so that each point represents a statement such as “20% of households have 10% of total income.” Recalling the 80/20 rule, the Lorenz curve is an ideal tool for representing that concept graphically, as depicted by the red dot in Fig. 2. Specifically, the adapted Lorenz curve in the context of a library depicts a normalized cumulative distribution (0% to 100%) of the number of loans over the number of items. The ideal Lorenz curve shows that each percentage increase in the number of loans accounts for the same percentage of the number of items. For example, 10% of the total number of loans corresponds to 10% of the total number of items. This is depicted by the 45-degree line, which corresponds to the ideal and most effective use of resources in the context of a library. On the other hand, unequal cumulative percentage distributions of the number of items and the number of loans will yield a curve, that is, the adapted Lorenz curve in Fig. 2. Therefore, a curve may result, for example, if the first, say, 25% of loans account for 5% of items, the second 25% of loans for 10% of items, the third 25% for 15% of items, and the last 25% of items account for 70% of items.

From the diagram in Fig. 2, it can also be implied that the shorter the library demand faction's long tail, the more the Lorenz curve will be directed towards the 45-degree line. Therefore, as a way of assessing the use of the items exhibited by the number of loans, it is necessary to measure the distance between the adapted Lorenz curve and the 45-degree line. On the other hand, comparisons over time of a single collection/use combination are very useful and quite valid. And yet, these comparisons do not require measures that ignore size and counts; these over-time comparisons can be made quite adequately using the raw data. And if these turn out to be the most informative types of comparisons for managers, they should not be discouraged in favor of a more statistically sophisticated measure.

The Gini index is widely used in cases of the Lorenz curve depicting income distribution (Gastwirth, 1972; Yitzhaki, 1983). The Gini index measures distances irrespective of the unit; therefore, it can also be theorized as a collection-use measure. The Gini coefficient numerically assesses the space between the 45-degree line and the adapted Lorenz curve, and its value ranges from 0 to 1. A low Gini coefficient indicates that the distance of the Lorenz curve and the ideal distribution is low, which means that inequality is also low. This implies that the long tail is nonexistent, and the short tail is large enough to absorb it. In contrast, a large Gini means that the distance is quite high, which in turn means that inequality is high, and the short-tail of the distribution is negligible. The special advantage of the Gini coefficient is that it is by definition independent of the size of the population, both in number of items and in number of loans, which makes the comparison between collections and libraries easier to understand. In essence, a measure describing the length of the long tail distribution is independent from the number of users that the collection targets, as well as the number of items of a particular collection. This measure, however, should be employed with caution when making actual comparisons among different distributions from different libraries, or among different collections in a single library

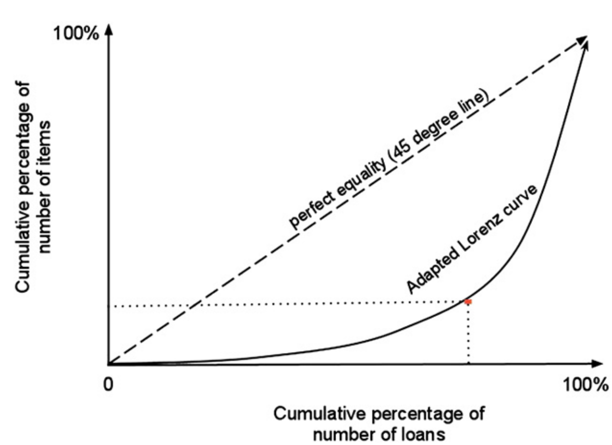

Fig. 2. The adapted Lorenz curve expressed in cumulative percentage dimensions on the library demand distribution, with the 80/20 rule represented by the red dot.

(as Table 2 illustrates). It is more straightforward within an unchanged library operational environment to make comparisons related to the degree of use over time of a single collection.

4. Estimating the Gini coefficient for the library demand curve: a case study

4.1. Dataset description

The case study dataset was collected by extracting raw data records from the library management system. Table 1 shows the variables that were included in the extracted dataset. The library collection for this case study consists of 4,125 titles, with 16,084 items physically represented in the library's collection. The library covers the needs of 2,395 users across departments and years of study. The total number of loans for the two academic years time period of this case study was 51,804. As shown in Fig. 3, the library employed is a typical academic library, where the faculties are more concentrated in arts and humanities, which are the subject of the departments listed as library users.

4.2. Estimation

Assume that the library in this case study is providing n horizontally-differentiated items to its users. The library's collection includes k popular items and $n-k$ other items. The focus is on the study of the demand-site explanation of the library users' loan behavior, which is directly linked to the library's efficiency. A user u_i who borrows an item κ for $f=(1 \ldots f_k)$ times is considered, and the loan matrix is constructed, for each item k as $\mathrm{M_k}=\{u_i f_{k(i)}\}$ where $\mathrm{f_{k(i)}}$ is the number of times the user u_i has borrowed the item k. The dimensions of M_k are 1×2. Moreover, the master matrix as $\mathrm{M} = \left(k, \sum_1^k \sum_i^i f_k(u_i)\right)$ is constructed, with the matrix ranked in a reverse order so as to be able to construct the distribution of borrowing per item. After the transformation is provided, a user count is obtained for each of the items, based on the barcode ID variable, which is unique and represents each user once. Fig. 4 illustrates the distribution of the summary counts of loans, with the cumulative percentage of the items registered in the library's collection. For reasons of clarity, the vertical axis is on logarithmic scale in order to compare orders of magnitude between loans and the cumulative percentage of items of all the collections in the library.

In Fig. 4, a power distribution of the form $F(x) = a \bullet x^k$, with $a=32.469$ and $k=-0.776$ is fitted. The $F(x)$ is increasing convex function, and x is the set of values for which $1>F(x)>0$. Furthermore, the mean μ of $F(x)$ can be considered as the integration factor, which is important for the estimation of the Lorenz curve. In order to quantify the inequalities between the distributions of item loans over the collection, the y scale (vertical axes) is normalized using the cumulative percentage of the total number of loans. In that case, the number of loans of each item is divided with the total number of loans; this distribution is considered to have the same properties of the distribution of income inequality, and thus can be represented with a Lorenz curve (Lambert & Aronson, 1993). Thus, given any degree of freedom, the Lorenz curve $L(x)$ is described by Gastwirth (1972) as:

$$L(x) = \mu^{-1} \int_0^x F^{-1}(t) \mathrm{d}t \quad (1)$$

As noted, the Lorenz curve has been extensively used in the field of econometrics and statistics to study inequalities among the income of individuals in the context of national income distribution. In essence, it is a method of graphically representing the proportion of the distribution assumed by the y percentage of the distribution values.

Fig. 5 displays the Lorenz curve of the cumulative percentage of loans versus the cumulative percentage of the items in the library's collection. Conforming to developments from the field of macroeconomics—and especially the distribution of income inequality over the Lorenz curve—it describes the 45-degree line. This fact indicates that each item in theory can be borrowed at the same rate by all the users of the library. The 45-degree line provides an angle where the distribution of loans versus items is perfectly equal, and the library is providing each item with full use. In order to quantify how far the library's operation mode is from the aforementioned ideal—and rather unrealistic—situation, the Gini Index is used (Gastwirth, 1972). The Gini index describes the ratio of the area between the 45-degree line and the Lorenz curve (L(x); Area A in Fig. 5) over the total area under the 45-degree line (Area A and Area B in Fig. 5). $L(x)$ is convex, and its derivative, $L'(x)$ (see Eq. (1)), therefore, is defined by the following formula: $x=F(\mu)$. Then, the Gini index is the area of concentration by choosing (within Area A of Fig. 5) k trapezoids of x:

$$0 = x0 < x1 < x2 < x3 < \ldots < xk < xk+1$$

Thus, this method leads to an underestimation of both A and B areas, since the straight line connecting—x^i, $L(x_i)$) to (x^{i+1}, $L(x_{i+1})$—lies above the convex curve $L(x)$, where $i=1 \ldots n$ samples

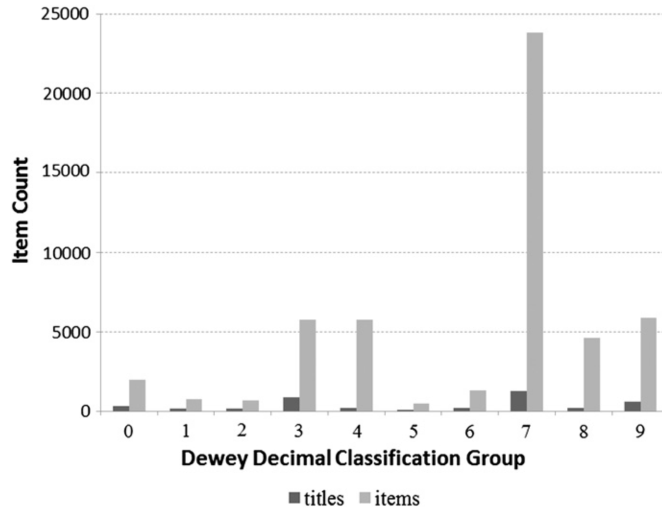

Fig. 3. Distribution of the items and titles in the library collection across the Dewey decimal classification standard.

Table 1
The variables provided in the dataset.

Variable name	Variable description	Usage in the analysis
barcode	The barcode of the user's membership card	Identification of the user u_i
usergroupid	The group that the user belongs to (e.g., the university department in which they are enrolled, as well as their year of study)	Identification of the user's affiliation and year of study
department	The department in which the user is enrolled	Identification of the department /faculty that the user belongs to
itemlcn	The location of the item in the library collection	Identification of the location of the item in the library's collection
ddc	The codification of the items' title according to the Dewey decimal classification	Estimation of the number of loans for this particular item.

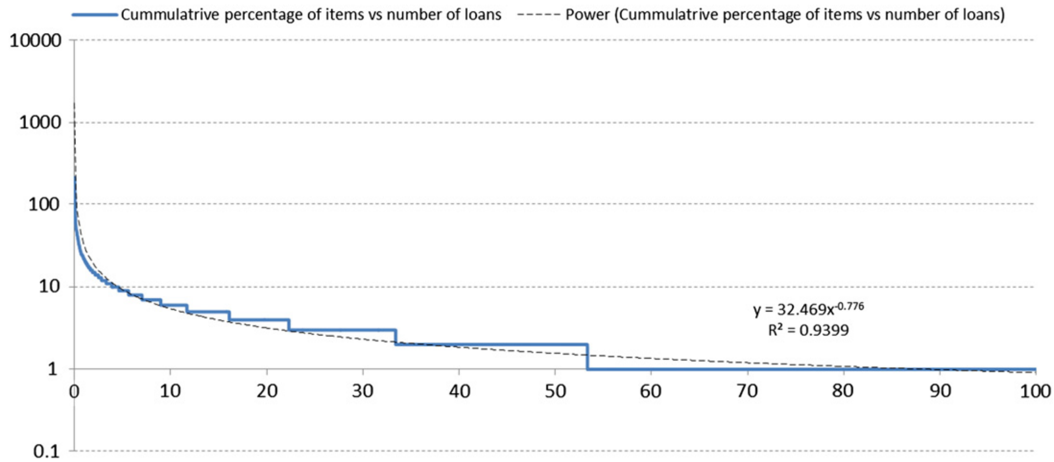

Fig. 4. Cumulative percentage of items (solid line) versus the fitted[1]

[1] The overall fit of this distribution is excellent and denotes that it follows a power law ($R^2=0.94$). power distribution (dotted line). The vertical axis is in logarithmic scale for display reasons.

of books. Then, the execution of the above-mentioned procedure yields the calculation of Gini index and is described as follows:

$$G = 1-\sum_{i=0}^{k}(x_{i+1}-x_i)\left[L(x_i)+L(x_{i+1})\right] \tag{2}$$

The smaller the value of the Gini coefficient, the higher the democratization of the item loan distribution across users. In other words, the short tail becomes larger. Moreover, a small value of Gini index does not necessary imply appropriate library collection use. The Gini index is easy to compute and interpret, and it provides a measure of inequality of the demand distribution across all library users. The shape of the Lorenz curve should also be taken into account, in view of the fact that libraries with similar collections and similar Gini coefficients can still have quite different loan distributions.

The department variable classifies users in the dataset into six different academic departments, all having access to the complete collection of the library's loan catalogue. Having this information, one is able to construct individual cumulative distributions of loans versus items for the user-loan history for all six available departments, as represented in Fig. 6. Table 2 illustrates the actual estimations of the Gini coefficient, as approximated by constructing the cumulative distribution graph for the library's items and the corresponding loan history.

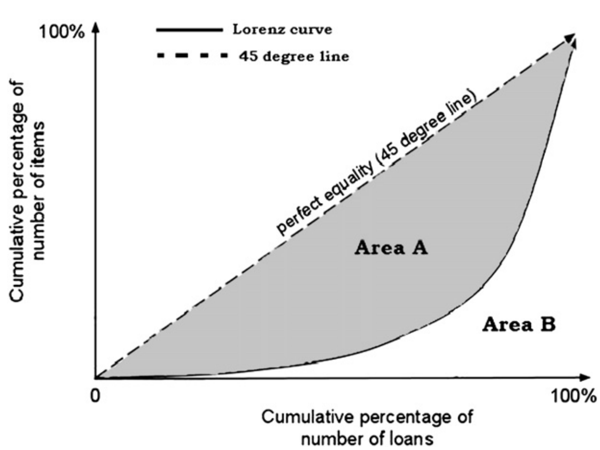

Fig. 5. The Lorenz curve (solid line) as represented by the cumulative percentage of loans versus the cumulative percentage of items in the library collection.

5. Discussion

The value of the Gini index changes across collections, and the corresponding Lorenz curves show different degrees of information resource utilization. Therefore, in support of management decisions based on the Gini coefficient values, information from the different shapes of Lorenz curve should also be obtained. Indeed, because of the varied shapes of the associated Lorenz curves, libraries with similar collections and Gini coefficients can still have quite different demand distributions. For example, consider a situation where half of potential library users do not use its resources, and the other half uses all items equally. This produces a Lorenz curve that is linear from (0, 0) to (0.5, 0), and then linear to (1, 1), with a corresponding Gini coefficient of 0.5. However, the same Gini coefficient would have been obtained for a library in which 75% of users borrow the 25% of items equally, while the remaining 25% of users share the remaining 75% of items equally. That is, a Lorenz curve that is linear from (0, 0) to (0.75, 0.25), and then linear to (1, 1). The Lorenz curve presents a multitude of such comparisons, which should be observed together with the Gini index value in order to lead library managers through the necessary steps.

On the other hand, an interpretation of the Gini coefficient value obtained through the analysis provided can be indicative of how democratized (i.e., long tail) and how concentrated (i.e., short tail) the loan history is, and hence demonstrate the degree of the library's collection penetration to users. As mentioned before, the Gini coefficient describes the area between the perfect equal cumulative distribution of loans versus items, and the current one obtained from the dataset. The smaller the value of the Gini coefficient, the more equal the distribution of loans versus items, and the longer the demand distribution at the tail. That is, the library collection is used by more people. Three different cases based solely on the value of the coefficient (from 0 to 1, where 0 indicates perfect equality and 1 indicates maximum inequality) can be identified.

5.1. Case 1: high concentration–low utilization

High concentration denotes a high value of the Gini coefficient, which suggests that the area between the Lorenz curve of the cumulative distribution of the items/user loans and the perfectly equal case (45-degree line) is very high (Gini's value in the fourth quintile). That can be interpreted as a case where a small fraction of the library's collection is borrowed by most of users, while the rest of the collection is not appealing or not accessible to users (e.g., decay of library material,

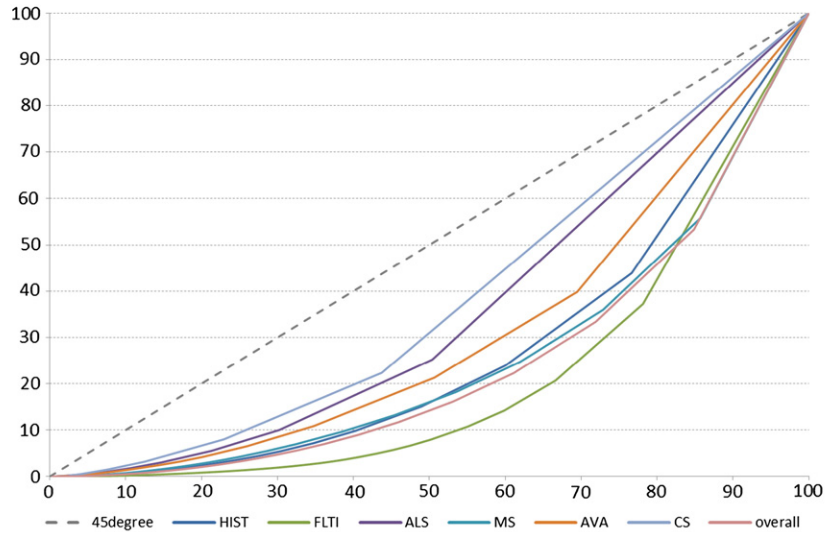

Fig. 6. Different Lorenz curves representing the collections that are part of the academic library that is described in the dataset.

change of information needs, and difficulty in accessing the collections items).

5.2. Case 2: moderate concentration–moderate utilization

This case considers a value of Gini in the second and third quintile, where an average percentage of a library's collection is exploited by the users and denotes a standard profile of operational efficiency. The library in this case has a portion of patrons who use its materials extensively, and another portion who are spontaneous users. For example, in an academic library the graduate students tend to make more use of the library's collection than undergraduate students because of their different curriculum requirements, and other information demands.

5.3. Case 3: low concentration–high utilization

In this case, library collection use is high and the value of the Gini coefficient belongs to the first quintile, where the area between the perfectly equal distribution and the Lorenz curve is minimal. This denotes that the library's collection is utilized equally by its users, and that the collection is in equal demand across all patrons. The library is efficient from the perspective that it manages to provide all the items in its collection with about the same frequency and the collection is accessible by all users. In that case, the library accomplishes its mission in the best possible way.

6. Conclusion

The case study analysis provided is based on two assumptions. First, a higher overall use of library materials is desirable because this has a positive effect on user information-needs satisfaction. And second, an inequality long-tail inspired measure of library demand characteristics, as opposed to the rather unrealistic target of equal resource use of all items by all users, is a useful standard of a library's operational efficiency.

Information-service providers such as documentation centers and academic libraries can employ this index in a number of different circumstances, and as an instrument of self-assessment. As with any measurement method, however, the method has advantages and disadvantages, and should be accompanied by other measures of collection use.

Other issues that need to be addressed for a collection use analysis relate to the identification of resource metrics, or indicators that can be used for measuring the different collection management activities, resources, contributions, or impact on users. Before launching such an analysis, one should clarify whether it is intended for internal or external use, and employ a group of metrics or indices so as to express the quantity or the quality of a library's collection resources. The overall expectations of those who conduct the analysis should not exceed the abilities of the method itself.

The long-tail inspired metric allows for the definition and quantification of management initiatives for collection utilization and development. Apart from libraries, which are the main focus of the case study, the adapted Gini index can also be used for the assessment of digital libraries and digital repositories in general, where item loans can be substituted by the request rate through the digital document's identifier. Generally speaking, administrative elements for collection development are linked to a specific management philosophy about high, moderate, or low-demand concentration; this includes a number of decisions regarding systems and services. As demonstrated, different demand segments may behave quite differently for a number of reasons when using a specific information service. Therefore, before decisions are made, an understanding of the overall business environment is required, one that includes infrastructure availability, service-provision practices, staff availability, co-operations, and competitors. Management actions linked to demand concentration may include training, investment in social networking services, hiring specialized staff, user subscription services, and so forth.

A set of open issues can be pursued from that point by relating the Gini index with other instruments for evaluating the effectiveness of information-service providers, as well as controlling external factors that affect the calculation of the Gini index. One example of this might be the task of normalizing the Gini index with the library opening hours in order to do cross-institution comparisons. Another

Table 2
The approximated value of the Gini coefficient across the different classes of users (corresponds to academic department acronyms) who use the library described in the dataset.

Category	Coefficient value	Number of affiliated users
Overall	0.51	2395
HIST	0.46	600
FLTI	0.56	575
ALS	0.28	251
MS	0.49	490
AVA	0.37	281
CS	0.23	106

approach could be to use filtering and recommendation mechanisms (Callan & Smeaton, 2003; Huang, Chung, Ong, & Chen, 2002; Torres, McNee, Abel, Konstan, & Riedl, 2004) so as to favor long-tail items over short-tail items for library users, in order to increase the demand, and thus reduce the value of the Gini index. Empirical data collection methods for the identification and classification of libraries and collections in the demand categories have been suggested. This would in turn permit a way to correlate or classify collection contents, and perhaps allow for the development of a typology of collections and user populations. It would also be quite interesting to compute Gini scores and plot the Lorenz curve for groups of libraries with varied user populations, such as undergraduate colleges versus graduate schools versus science libraries, law libraries, music libraries, and so forth. This could lead to normative data against which libraries and collections could be compared.

References

Anderson, C. (2008). *The long tail: Why the future of business is selling less of more.* New York, NY: Hyperion Books.

Bawden, D. (1990). *User-oriented evaluation of information systems and services.* London, UK: Gower Aldershot.

Bollen, J., & Van de Sompel, H. (2008). Usage impact factor: The effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology*, *59*, 136–149.

Britten, W. A. (1990). A use statistic for collection management: The 80/20 rule revisited. *Library Acquisitions: Practice & Theory*, *14*(2), 183–189.

Brown, J. S., & Adler, R. P. (2008). Minds on fire: Open education, the long tail, and learning 2.0. *Educause Review*, *43*(1), 16–32.

Brynjolfsson, E., Hu, Y. J., & Smith, M. D. (2006). From niches to riches: Anatomy of the Long Tail. *Sloan Management Review*, *47*(4), 67–71.

Burrell, Q. L. (1985). The 80/20 rule: Library lore or statistical law? *Journal of Documentation*, *41*, 24–39.

Callan, J., & Smeaton, A. (2003). Personalisation and recommender systems in digital libraries. (Joint NSF-EU DELOS Working Group Report). Retrieved from http://www.ercim.eu/publication/ws-proceedings/Delos-NSF/Personalisation.pdf.

Casey, M., & Savastinuk, L. (2006). Library 2.0: Service for the next-generation library. *Library Journal*, *131*(14), 40–42. Retrieved from http://www.libraryjournal.com/article/CA6365200.html.

Dempsey, L. (2006). Libraries and the long tail. *D-Lib Magazine*, *12*(4). Retrieved from http://www.dlib.org/dlib/april06/dempsey/04dempsey.html.

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *The Review of Economics and Statistics*, *54*(3), 306–316.

Genoni, P. (2007). Libraries, the long tail and the future of legacy print collections. *Library and Information Science Research Electronic Journal*, *17*(1). Retrieved from http://libres.curtin.edu.au/libres17n1/Genoni_Ess_Op.pdf.

Gottschalk, P., & Smeeding, T. M. (1997). Cross-national comparisons of earnings and income inequality. *Journal of Economic Literature*, *35*(2), 633–687.

Hernon, P., & McClure, C. R. (1990). *Evaluation and library decision making.* Stamford, CT: Ablex Publishing Corporation.

Holt, G. E. (2007). Economic realities in optimizing library materials access. *The Bottom Line: Managing Library Finances*, *20*(1), 45–49.

Huang, Z., Chung, W., Ong, T. -H., & Chen, H. (2002). A graph-based recommender system for digital library. *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 65–73). Portland, OR: ACM Press.

Koch, R. (1998). *The 80/20 principle: The secret to success by achieving more with less.* New York, NY: Currency Doubleday Publishing Group.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, *42*, 361–371.

Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services.* Westport, CT: Libraries Unlimited.

Lambert, P. J., & Aronson, J. R. (1993). Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal*, *103*(420), 1221–1227.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, *9*(70), 209–219.

McGrath, M. (2004). Interlending and document supply: A review of the recent literature — 50. *Interlending & Document Supply*, *32*(4), 244–254.

Nisonger, T. E. (2008). The "80/20 rule" and core journals. *The Serials Librarian*, *55*(1/2), 62–84.

Starr, S., & Williams, J. (2008). The long tail: A usage analysis of pre-1993 print biomedical journal literature. *Journal of the Medical Library Association*, *96*(1), 20–27.

Storey, T. (2005). The long tail and libraries. *OCLC Newsletter*, *268*. Retrieved from http://www.oclc.be/us/en/news/publications/newsletters/oclc/2005/268/downloads/thelongtail.pdf.

Torres, R., McNee, S., Abel, M., Konstan, J., & Riedl, J. (2004). Enhancing digital libraries with TechLens+. *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries* (pp. 228–236). Tuscon, AZ: ACM Press.

Trueswell, R. L. (1969). Some behavioral patterns of library users: The 80/20 rule. *Wilson Library Bulletin*, *43*(5), 458–461.

Yitzhaki, S. (1983). On an extension of the Gini inequality index. *International Economic Review*, *24*(3), 617–628.

Petros A. Kostagiolas holds a Ph.D. from the University of Birmingham, England. Currently, he is a lecturer in information services management in the Department of Archive and Library Science, Ionian University, where he teaches library management, quality management, and research methodology.

Nikolaos Korfiatis holds a Ph.D. from Department of Informatics at the Copenhagen Business School (CBS) in Denmark. He received his B.Sc. in information systems from the Athens University of Economics and Business and his diploma in engineering of interactive systems with focus on information retrieval from the Royal Institute of Technology (KTH) in Stockholm, Sweden. His research interests span the fields of information science and economic sociology, with a particular emphasis on the analysis of online social networks and evaluation mechanisms.

Marios Poulos is an assistant professor at the Department of Archives and Library Sciences at the Ionian University in Corfu, Greece. He received his Ph.D. degree from the University of Piraeus in the field of biological information processing. His research interests are in metadata and semantics research with an emphasis on intelligent methods for the extraction of semantic information from various sources.