Duality revisited: Construction of fractional frequency distributions based on two dual Lotka laws

EGGHE, Leo & RAO, Ravichandra (2002) Duality revisited: Construction of fractional frequency distributions based on two dual Lotka laws. In: JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 53(10). p. 789-801.

DOI: 10.1002/asi.10103
Handle: http://hdl.handle.net/1942/770

DUALITY REVISITED : CONSTRUCTION OF FRACTIONAL FREQUENCY DISTRIBUTIONS BASED ON TWO DUAL LOTKA LAWS

by

L. Egghe, LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium[1]
and
UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium

and

I.K.R. Ravichandra Rao,
DRTC, ISI, 8th Mile, Mysore Road, R.V. College P.O., Bangalore, India[1]
and
LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium[2]

ABSTRACT

Fractional frequency distributions of e.g. authors with a certain (fractional) number of papers are very irregular and, therefore, not easy to model or to explain. This paper gives a first attempt to this by assuming two simple Lotka laws (with exponent 2) : one for the number of authors with n papers (total count here) and one for the number of papers with n authors, $n \in \mathbb{N}$. Based on an earlier made convolution model of Egghe, interpreted and reworked now for discrete scores, we are able to produce theoretical fractional frequency

---

[1] Permanent address.

[2] Research on this paper has been executed while this author was a visiting professor in LUC. He is grateful to LUC for financial support.

distributions with only one parameter which are in very close agreement with the practical ones as found in a large dataset produced earlier by Rao. The paper hence also shows that (irregular) fractional frequency distributions are a consequence of Lotka's law and are not examples of breakdowns of this famous historical law.

I. Introduction

In Rousseau (1992), the discussion started on the fractional frequency distribution of e.g. authors and the distribution of their fractional scores in a bibliography. Fractional scores means that if an author has published a paper in which there are i (i=1,2,3,...) authors in total, then this author (and all the other authors in this paper) receives a score $\frac{1}{i}$. This is different from the total scoring system in which every author receives a score 1 in such a paper, hence here scores are always entire numbers, as opposed to the fractional scoring system. One then wonders how the fractional frequency distribution of author scores looks like in a given bibliography. To be precisely clear about the score of an author in such a bibliography, let us give an example. Suppose an author in this bibliography has a paper where he/she is the only author, has another paper where there are 2 authors in total and has 3 other papers where there are 3 authors in total. Then the overall fractional score of this author is the sum of the fractional scores per paper, i.e. $1+\frac{1}{2}+3.\frac{1}{3}=2.5$. Note that we avoided to use the term "total fractional score" which could be confusing with the overall score in the total counting system of this author which would be 5 in this case.

While Lotka's law (or distribution) very well applies to any scoring system where entire numbers are used - we can even go back to the historical paper Lotka (1926) for this - it is clear that this is not the case anymore for the fractional scoring system. As Rousseau (1992) points out, a fractional score of $\frac{1}{8}$ probably will occur less frequently than a score $\frac{1}{4}$ since we can assume that there are less 8-authored papers than 4-authored ones. Even if this would not be the case we can go further to $\frac{1}{16}$, $\frac{1}{32}$, ... scores which will not occur very frequently. In addition to this, a score $\frac{1}{4}$ need not only come from authorship in a 4-authored paper but is also obtained in the case the author has 2 8-authored papers. To go back to the score of 2.5 above the number of possibilities is (theoretically) unlimited.

Indeed, an overall score of 2.5 can also be reached via 2 1-authored papers and 1 2-authored paper, but also via 5 2-authored papers, but also via 3 2-authored papers and 4 4-authored papers and so on.

Burrell and Rousseau (1995) present numerical simulations of fractional frequency distributions showing their irregular shapes.

From this it is very clear that calculating the fractional frequency distribution is very difficult - if not impossible. Indeed, in any theoretical work, any positive rational number $q=\frac{p}{r}$ $(p,r\in\mathbb{N})$ is a possible fractional frequency (indeed **one** of the many ways to obtain q as an overall fractional score is by having p r-authored papers). So, if we want to determine the fractional frequency distribution one requires a formula for f(q), the probabilty (or fraction) to have an overall fractional score $q\in Q^+$, the positive rational numbers. This is virtually impossible since there are an infinite number of these and since we expect, due to the very irregular shape (see the examples above and also further, where we discuss the Rao data-set) of such a distribution, that, for each $q\in Q^+$, a different formula for f(q) is needed, i.e. we do not expect to be able to produce one analytical function for f(q) where q appears as a parameter.

Therefore, in Egghe (1993) we decided to tackle the problem, where the rational number q is replaced by any real number $z\in\mathbb{R}^+$. The argument - briefly - was as follows. Let $\varphi(i)$ denote Lotka's law (any exponent) being the fraction of authors with i papers in the total scoring system (i.e. independent of the number of co-authors in these papers), $i\in\mathbb{N}$. Let $f_1(z)$ be the fraction of authors with a fractional score z in $\frac{1}{}$ paper (hence $z=1,\frac{1}{2},\frac{1}{3},\frac{1}{4},\ldots$). Then the overall fractional frequency distribution f is given by $(z\in Q^+)$

$$f(z) = \sum_{i=1}^{\infty} \underbrace{(f_1\circledast\ldots\circledast f_1)}_{i \text{ times}}(z)\varphi(i) \qquad (1)$$

where $\circledast$ denotes convolution, applied here i times in every term of the sum. Indeed, (1) follows from the Theorem of Total Probability (also called Partition Theorem) and the fact

that $(f_1 \circledast \ldots \circledast f_1)(z)$, where we have i times f_1, is the distribution of the fractional frequencies of authors, given one has published i papers (see e.g. Chung (1974) or Blom (1989) or virtually any good text book on probability theory).

As said, in Egghe (1993) we were unable to use (1) in the discrete case ($z \in Q^+$) but we studied the continuous case ($z \in \mathbb{R}^+$). In Egghe (1993) we could indeed show that (1) is not a decreasing function anymore but is increasing up to $z=1$ from where it starts decreasing. This explained the "overall" view of a fractional frequency distribution but not at all its values for every rational z. We refer to the Table in the Appendix for a very large fractional data-set, collected earlier in Rao (1995), with accompanying graph (Figure I in the Appendix). There the "overall" view is clear but also the irregularity of the individual data is evident. The graph in the Appendix clearly shows that as fractions of papers (q) increases, f(q) tends to zero. Further f(q) is not a smooth curve and it moves up and down frequently even for large values of q. The mode of the distribution is 1 and also for $q=0.5$, f(q) is 11,673 indicating that the distribution has two modal values. The mean and median are 1.121 and 0.99252 respectively. The variance and standard deviation are 1,1721 and 1.0826 respectively.

We close this overview of existing results by remarking that the model (1) follows from a dual approach of informetrics (cf. Egghe (1989), (1990)), since f_1 is derived from the Lotka law ψ, the dual of the Lotka law φ : $\psi(j)$ denotes the fraction of papers with j authors ($j \in \mathbb{N}$). Hence model (1) involves (and only involves) the two dual Lotka laws φ and ψ. It is, therefore, that we argued in Egghe (1993) that modelling fractional counting does not prove a breakdown of Lotka's law (as argued in Rousseau (1992)) but, on the contrary, Lotka's laws φ and its dual ψ explain the fractional frequency distribution. We therefore consider (1) to have a high informetric explanatory value.

For this reason, we continue to use (a variant of) model (1), in the attack of the problem of explaining the irregular shape of the fractional frequency distribution f(q), $q \in Q^+$. As explained above, this is not possible for every $q \in Q^+$. In this paper, therefore, we apply a variant of (1) to grouped data, where we only allow for a few fractional scores q. Let us

explain this in more detail. Let $i \in \mathbb{N}$ be fixed. A fractional scoring model which is between the classical one and the total scoring system is the following :

1. If we have a paper with j authors, $j=1,\ldots,i-1$, each author receives a score of $\frac{1}{j}$,
2. If we have a paper with j authors, $j=i,i+1,\ldots$, each author receives a score of $\frac{1}{i}$.

Note that for $i=1$ we have the total scoring system, which we will not use here. If i increases we move closer and closer to the fractional scoring system. Since now, for each fixed $i \in \mathbb{N}$, only the scores $1,\ldots,\frac{1}{i}$ per paper are possible, we are in a position to calculate all possible fractional frequency probabilities, where we will limit ourselves to total scores $q \leq 2$. One reason for this is that the most interesting part of scores is in the interval [0,2] (the highest values belong to that interval and, although irregularity persists after $q=2$, the overall tendency is a decreasing function). Another reason is the increased difficulty in calculating f(q) for higher q. Indeed, the higher q, the more possibilities there are of patterns of publications to reach this value. This will also be illustrated in the sequel.

Section III will deal with $i=2$. Here the only possible values for q are $\frac{1}{2},1,\frac{3}{2},2$ and this case is too rough in comparison with the detailed Rao data (Rao (1995)). Appropriate grouping in these data gives a first (rough) comparison between the theoretically obtained fractional frequency distribution and the experimental one.

The fourth section deals with $i=3$. Here we have possible fractional scores q : $\frac{1}{3},\frac{1}{2},\frac{2}{3},\frac{5}{6}$, $1,\frac{7}{6},\frac{4}{3},\frac{3}{2},\frac{5}{3},\frac{11}{6},2$ (we stop at 2), and for each of them we present an analytical formula for its probability f(q). Corresponding groupings in the Rao data now gives a remarkable agreement between the theoretical and experimental fractional frequency distribution.

The fifth section deals with $i=4$. Possible fractional scores q here are : $\frac{1}{4},\frac{1}{3},\frac{1}{2},\frac{7}{12},\frac{2}{3}$, $\frac{3}{4},\frac{5}{6},\frac{11}{12},1,\frac{13}{12},\frac{7}{6},\frac{5}{4},\frac{4}{3},\frac{17}{12},\frac{3}{2},\frac{19}{12},\frac{5}{3},\frac{7}{4},\frac{11}{6},\frac{23}{12},2$. Again, for each of them we give an analytical formula for its probability f(q) and corresponding groupings in the Rao data again gives a remarkable agreement between the theoretical and experimental fractional frequency distribution.

The case i=5 is also elaborated (in section VI), now yielding formulae for not less than 83 fractional scores $q \leq 2$. The model is very good but the only problem is that groupings in the Rao data often are not necessary since the obtained intervals are too small. In this way, many very low probabilities are compared. This phenomenon could be compared with the choice - in statistics - of how many bars one uses in a histogram : usually a computer program gives an optimal choice (see also Egghe and Rousseau (2001)). If we take less bars the graph becomes too rough ; if we use more bars, we risk to have too many intervals so that reasonable groupings are not possible anymore.

For the Rao data the cases i=3 and 4 are the best but we stress the fact that the very detailed cases i=5 has its application in even larger data-sets, yet to be constructed. We therefore plead to construct very large fractional data-sets (i.e. fractional frequency scores of authors in vast domains) so that our model for i=5 is more suited.

In the next section we present the general theoretical model which is a discrete (exact) version of the continuous (approximated) version in Egghe (1993).

II. Exact, discrete, theoretical model for the fractional frequency distribution, derived from two dual Lotka laws.

Let $\varphi(n)$ denote the fraction of authors with n papers ($n \in \mathbb{N}$, total counts). Its dual analogue (cf. Egghe (1989), (1990)) is the function ψ where $\psi(n)$ denotes the fraction of papers with n authors ($n \in \mathbb{N}$). In the sequel we will use concrete Lotka laws for these functions but the following important result is independent of the choice of function for φ and ψ.

Lemma II.1. Let $f_1(z)$ denote the fraction of the authorships (i.e. author occurences) with fractional score z in 1 paper. Then

$$f_1(z) = \frac{\psi\left(\frac{1}{z}\right)}{\mu z} \qquad (2)$$

where μ denotes the average number of authors per paper.

Proof :

$$f_1(z) = \frac{\text{total } \sharp \text{ authorships with fractional score z in 1 paper}}{\text{total } \sharp \text{ authorships}}$$

$$f_1(z) = \frac{\frac{1}{z}(\text{total } \sharp \text{ papers with } \frac{1}{z} \text{ authors})}{\text{total } \sharp \text{ authorships}}$$

$$f_1(z) = \frac{1}{z} \frac{\text{total } \sharp \text{ papers}}{\text{total } \sharp \text{ authorships}} \frac{\text{total } \sharp \text{ papers with } \frac{1}{z} \text{ authors}}{\text{total } \sharp \text{ papers}}$$

$$f_1(z) = \frac{1}{z} \frac{1}{\mu} \cdot \text{fraction of papers with } \frac{1}{z} \text{ authors}$$

$$f_1(z) = \frac{\psi\left(\frac{1}{z}\right)}{\mu z}. \qquad \square$$

Note that $z = \frac{1}{n}$, $n \in \mathbb{N}$, necessarily. This shows that the fractional score of an author in 1 paper is directly derivable from ψ. Together with its dual function φ we will be able to derive the theoretical, discrete fractional frequency distribution. We thank one of the referees to complete our proof of an earlier version.

Proposition II.2. : Let f(z) denote the fractional frequency distribution of a bibliography (more generally an IPP - see Egghe (1989), (1990)) for which φ and ψ are the valid, dual, (entire) frequency functions as described above. Then, for every $z \in Q^+$:

$$f(z) = \sum_{i=1}^{\infty} (\underbrace{f_1 \circledast \ldots \circledast f_1}_{i \text{ times}})(z)\varphi(i), \qquad (3)$$

where $\circledast$ denotes convolution. Here it is assumed that the fractional scores distributions in a paper are independent and identically distributed (i.i.d.), being the distribution f_1.

Proof : Let N be the (random variable of the) number of papers ($i \in \mathbb{N}$) and Y(j) the fractional score from the j^{th} paper. Then

$$f(z) = P(\text{overall fraction} = z)$$

$$= P(Y(1) + Y(2) + \ldots + Y(N) = z)$$

$$= \sum_{i=1}^{\infty} P(Y(1) + Y(2) + \ldots + Y(N) = z \mid N=i)P(N=i)$$

by the Theorem of Total Probability. So

$$f(z) = \sum_{i=1}^{\infty} P(Y(1)+Y(2)+\ldots+Y(i)=z)P(N=i)$$

(independence of N w.r.t.the Ys)

$$f(z) = \sum_{i=1}^{\infty} P(Y(1) + Y(2) + \ldots + Y(i) = z)\, \varphi(i).$$

The distribution of $Y(1)+Y(2)+\ldots+Y(i)$ is given by the convolution of the individual distributions since we assumed independence of the Ys (cf. Chung (1974), Blom (1989)) and this becomes the i-fold convolution of f_1, using the assumption of identical distributions for the Ys. Hence we proved (3).

Note : Although it is relatively easy to accept that all Ys have the same distribution, the fact that they are independent is less sure. It is indeed not certain that a certain score in paper 2 (say) is independent of the score in paper 1, due to collaboration habits. However, we have to suppose independence for the intricate model (see (3) and further on !) to work. The assumption can be considered as a simplification which is acceptable in this first attempt to model fractional frequencies. It will be clear in the sequel that our model fits real data very well which is a (post factum !) argument for the acceptance of this simplification.

From the above it is hence clear, at least theoretically, that the very irregular fractional frequency distribution is determined by the entire frequency distribution φ and its dual ψ. In the sequel we will use the simplest (and in informetrics most important) distribution : the Lotka distribution with exponent 2. So we will use ($n \in \mathbb{N}$)

$$\varphi(n) = \frac{6}{\pi^2 n^2}. \tag{4}$$

Here $\frac{6}{\pi^2} \approx 0.6079271$ is the normalizing constant, assuring that

$$\sum_{n=1}^{\infty} \varphi(n) = 1 \tag{5}$$

as is required for a discrete distribution (cf. also Egghe and Rousseau (1990)).

For the dual analogue ψ of φ, we use the same simple function

$$\psi(n) = \frac{6}{\pi^2 n^2}, \tag{6}$$

$n \in \mathbb{N}$. Hence $\varphi = \psi$, mathematically, but φ and ψ have dual interpretations. Note that (2) and (6) imply that

$$f_1(z) = \frac{6z}{\pi^2 \mu} \tag{7}$$

Note on the use of the law of Lotka

The use of the law of Lotka for the distribution φ is undisputed (although other distributions can be used, of course). As one of the referees points out this has been confirmed hundreds of times (going even back to Lotka (1926) itself). However the same referee disputes the use of Lotka's law for distribution ψ. As he/she rightly points out, Lotka's law is not fitting well the distribution of number of authors per paper in the cases Ajiferuke (1991) and Rousseau (1994). It is fitting well in case there are more single-authored papers than 2-authored ones but in several cases there are more 2-authored papers than single-authored ones. Nevertheless we used Lotka's faw for ψ for the following reasons :

(i) We wanted to develop further a "Lotka-type" informetrics theory as we did before in several papers. Hereby we want to show the interaction of the two dual laws and also that nothing else is needed to explain the fractional frequency distributions. Of course, basic for the model is proposition II.2 and - as indicated by the same referee - the model (3) is probably very robust in the sense that it will not matter very much what are the exact distributions that are used for φ and ψ. Another referee even advocates to use the real experimental data for φ and ψ. While this has value in the testing of the validity of model (3), this methodology would not shed light on the dual mechanism that is explained in this paper (via formula (2)). In short, we want to investigate how "far" we can go with "Lotka-type informetrics". It would then also be interesting to develop - in a consequent way - other informetric theories, based on other frequency distributions. Note that Egghe (2000) and Egghe and Rao (2002) are examples of an explanation of the first citation distribution and of the most-recent-reference distribution, where (especially in the latter paper) it was made clear that the distribution of the number of references does not follow Lotka's law.

(ii) Using Lotka's law for φ and ψ is easy and much more easy than using distributions of the lognormal type (which would have been more exact, certainly in the case of ψ).

(iii) In a forthcoming paper we intend to investigate other distributions for φ and ψ. Even the use of the uniform distribution for ψ (which is clearly not the correct one!) could be considered, thereby showing the "power in itself" of the methodology that is developed here (robustness).

(iv) Last but not least : our data show an approximate Lotka law for ψ. Only the cases of 1 and 2 authors per paper yield a more or less equal number of papers, contrary to Lotka's law. The reason that we encounter a distribution close to Lotka's law is that we have papers in mathematics where collaboration is less than in some other disciplines (such as e.g. chemistry,...).

The parameter μ (the mean of ψ) will be determined by an ad hoc method (see further). We prefer it this way rather than calculating the mean of ψ, for n limited to a finite number of

values (as in practise). Allowing an infinite number of n in ψ yields a distribution with infinite mean.

As explained in the previous section, formula (3), interpreted for continuous $z \in \mathbb{R}^+$ yields in Egghe (1993) an explanation of the overall behavior of f : increasing in [0,1] and decreasing beyond 1. However it does not give an explanation for the many irregularities in rational points $z \in Q^+$. This will be done in the rest of this paper. We will be able to inspect the properties of f in (3) by adapting it a bit by restricting the number of possible fractional scores (per paper) from below as explained in the introduction (and repeated further on). Analogous groupings in the experimental data will then allow for the comparison between the theoretical model with the experimental fractional frequency distribution.

III. The case i=2 : allowing an author score of 1/2 or 1 in 1 paper.

Although too rough, this case will very simply illustrate the methodology that we will apply in this paper to yield discrete fractional frequency distributions. Note that in all our studies in this paper we will limit ourselves to fractional scores $q \leq 2$, as explained in the introduction.

In this simple model, an author receives a score 1 if he/she is an author in a single-authored paper. If he/she is author in a multi-authored paper, this author receives a score $\frac{1}{2}$. Let us call g_1 the author distribution of fractional scores in 1 paper. By definition

$$g_1(1) = f_1(1) \tag{8}$$

$$g_1\left(\frac{1}{2}\right) = \sum_{i=2}^{\infty} f_1\left(\frac{1}{i}\right) \tag{9}$$

$$g_1\left(\frac{1}{2}\right) = 1 - f_1(1), \tag{10}$$

since f_1 is a distribution. Using (7) this yields

$$g_1(1) = \frac{6}{\pi^2\mu} \tag{11}$$

$$g_1\left(\frac{1}{2}\right) = 1 - \frac{6}{\pi^2\mu} \tag{12}$$

We apply now (3) but with f_1 replaced by g_1 and for the values $z=\frac{1}{2},1,\frac{3}{2},2$, the only possible scores (inferior to 2). This gives

$$f\left(\frac{1}{2}\right) = g_1\left(\frac{1}{2}\right)\varphi(1) \tag{13}$$

$$f(1) = g_1(1)\varphi(1) + \left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(2) \tag{14}$$

$$f\left(\frac{3}{2}\right) = 2\, g_1\left(\frac{1}{2}\right)g_1(1)\varphi(2) + \left(g_1\left(\frac{1}{2}\right)\right)^3\varphi(3) \tag{15}$$

$$f(2) = (g_1(1))^2\varphi(2) + 3\left(g_1\left(\frac{1}{2}\right)\right)^2 g_1(1)\varphi(3) + \left(g_1\left(\frac{1}{2}\right)\right)^4\varphi(4), \tag{16}$$

where φ is given by (4).

These values are then compared by the corresponding grouped data from Rao's table in the Appendix, grouped as follows :

- *score* $\frac{1}{2}$ *corresponds to grouping the data in the interval*]0,0.75]
- score 1 corresponds to grouping in]0.75,1.25]
- score $\frac{3}{2}$ corresponds to]1.25,1.75]
- score 2 corresponds to]1.75,2.25].

It is clear that we take the interval]0,0.75] for $\frac{1}{2}$ (and not]0.25,0.75]) since in our model, all fractional scores (in one paper), smaller than $\frac{1}{2}$, are transformed into $\frac{1}{2}$.

Of course, these groupings are not a perfect analogue of our simplified model since an author with an overall score of $\frac{1}{4}$, being the result of participation in 2 8-authored papers is classified into the score 1 in the model while it is classified in the interval]0,0.75] in the grouping. This difference is there but will diminish in the next cases where we allow for smaller fractions. Also, if we find good results in this setting, this will indicate that the above difference is not destroying the similar nature of both simplifications.

The (only) parameter μ is determined by requiring $f\left(\frac{1}{2}\right)$ to be exact :

$$f\left(\frac{1}{2}\right) = \left(1 - \frac{6}{\pi^2\mu}\right)\frac{6}{\pi^2} = \frac{18,892}{46,853}$$

$$= \frac{\#\ \text{in}\ ]0,0.75]}{\text{total}\ \#}$$

(see the table in the Appendix). This yields $\mu = 1.80537576$. We have now the following table and graph, comparing theoretical and experimental fractional frequency distributions.

Table 1. Distribution of overall fractional scores (case of i=2)

q	Theoretical (f)	Experimental
1/2	0.4033047	0.4033047
1	0.2715154	0.345911681
3/2	0.0875965	0.079546667
2	0.0545977	0.081360852

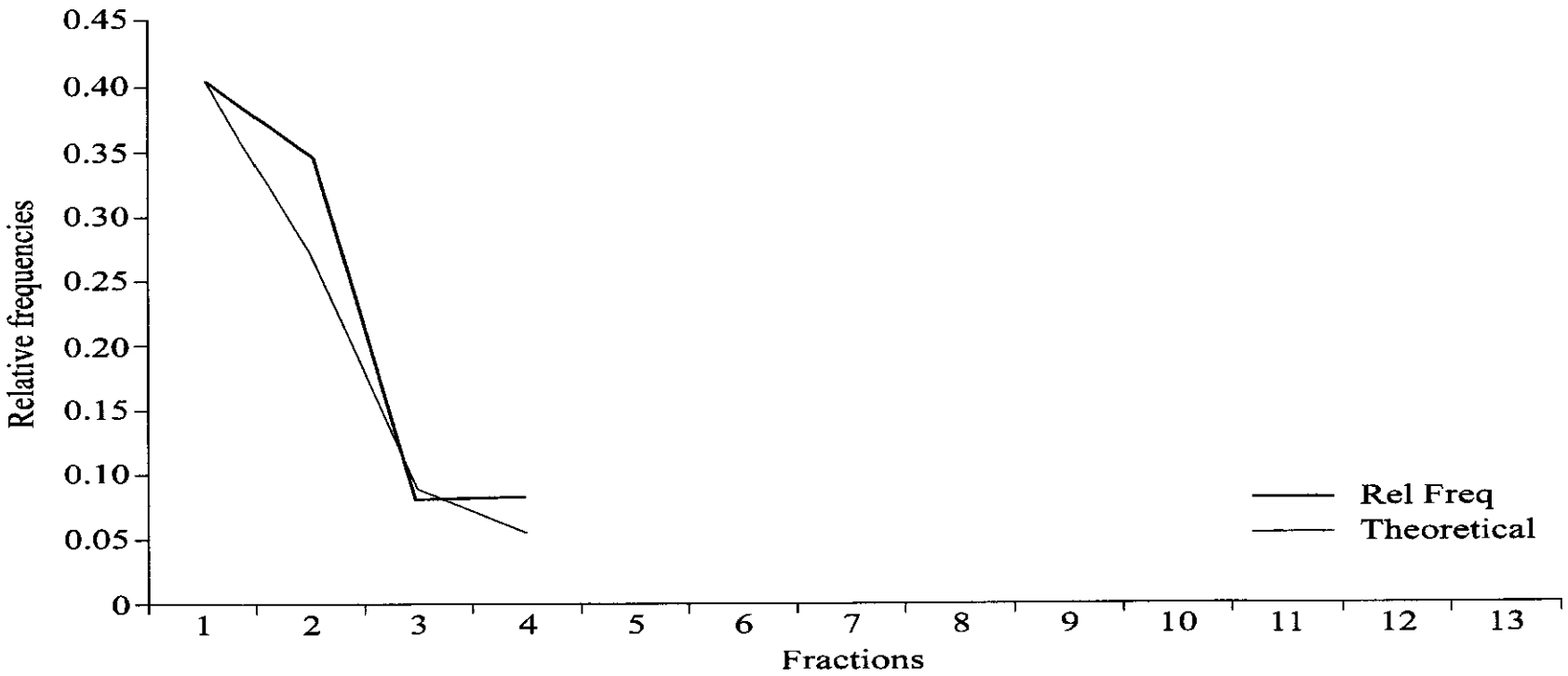

Fig. 1 Theoretical and experimental fractional frequency distributions (case of i=2).

From the above we see, although we only compare 4 fractions, both theoretical and experimental graphs are following the same pattern. This will become more clear in the next (more important and more interesting) cases.

IV. The case i=3 : allowing an author score of 1/3, 1/2 or 1 in 1 paper.

This will be the first interesting case. Here an author receives a score 1 if he/she is an author in a single-authored paper, a score $\frac{1}{2}$ if he/she is an author in a 2-authored paper and a score $\frac{1}{3}$ if he/she is an author in a j-authored paper, for all $j \geq 3$. Now we have

$$g_1(1) = f_1(1) = \frac{6}{\pi^2\mu} \tag{17}$$

$$g_1\left(\frac{1}{2}\right) = f_1\left(\frac{1}{2}\right) = \frac{3}{\pi^2\mu} \tag{18}$$

$$g_1\left(\frac{1}{3}\right) = 1 - \left(f_1(1) + f_1\left(\frac{1}{2}\right)\right)$$

$$= 1 - \frac{9}{\pi^2\mu} \tag{19}$$

Possible overall fractional scores (in $]0,2]$) are $\frac{1}{3},\frac{1}{2},\frac{2}{3},\frac{5}{6},1,\frac{7}{6},\frac{4}{3},\frac{3}{2},\frac{5}{3},\frac{11}{6},2$ (theoretical) to be compared with grouped data (from the Table in the Appendix) in the intervals $]0,\frac{5}{12}]$, $]\frac{5}{12},\frac{7}{12}]$, $]\frac{7}{12},\frac{9}{12}]$, $]\frac{9}{12},\frac{11}{12}]$, $]\frac{11}{12},\frac{13}{12}]$, $]\frac{13}{12},\frac{15}{12}]$, $]\frac{15}{12},\frac{17}{12}]$, $]\frac{17}{12},\frac{19}{12}]$, $]\frac{19}{12},\frac{21}{12}]$, $]\frac{21}{12},\frac{23}{12}]$, $]\frac{23}{12},\frac{25}{12}]$.

We have the following formulas from which the theortical fractional frequency distribution can be calculated.

$$f\left(\frac{1}{3}\right) = g_1\left(\frac{1}{3}\right)\varphi(1) \tag{20}$$

$$f\left(\frac{1}{2}\right) = g_1\left(\frac{1}{2}\right)\varphi(1) \tag{21}$$

$$f\left(\frac{2}{3}\right) = \left(g_1\left(\frac{1}{3}\right)\right)^2\varphi(2) \tag{22}$$

$$f\left(\frac{5}{6}\right) = 2\, g_1\left(\frac{1}{3}\right)g_1\left(\frac{1}{2}\right)\varphi(2) \tag{23}$$

$$f(1) = g_1(1)\varphi(1) + \left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(2) + \left(g_1\left(\frac{1}{3}\right)\right)^3\varphi(3) \tag{24}$$

$$f\left(\frac{7}{6}\right) = 3\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1\left(\frac{1}{2}\right)\varphi(3) \tag{25}$$

$$f\left(\frac{4}{3}\right) = 2\, g_1(1)g_1\left(\frac{1}{3}\right)\varphi(2) + 3\left(g_1\left(\frac{1}{2}\right)\right)^2 g_1\left(\frac{1}{3}\right)\varphi(3) + \left(g_1\left(\frac{1}{3}\right)\right)^4\varphi(4) \tag{26}$$

$$f\left(\frac{3}{2}\right) = 2\, g_1\left(\frac{1}{2}\right)g_1(1)\varphi(2) + \left(g_1\left(\frac{1}{2}\right)\right)^3\varphi(3) + 4\left(g_1\left(\frac{1}{3}\right)\right)^3 g_1\left(\frac{1}{2}\right)\varphi(4) \tag{27}$$

$$f\left(\frac{5}{3}\right) = 3\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1(1)\varphi(3) + 6\left(g_1\left(\frac{1}{3}\right)\right)^2\left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(4) + \left(g_1\left(\frac{1}{3}\right)\right)^5\varphi(5) \tag{28}$$

$$f\left(\frac{11}{6}\right) = 6\, g_1\left(\frac{1}{3}\right)g_1\left(\frac{1}{2}\right)g_1(1)\varphi(3) + 4\, g_1\left(\frac{1}{3}\right)\left(g_1\left(\frac{1}{2}\right)\right)^3\varphi(4) + 5\left(g_1\left(\frac{1}{3}\right)\right)^4 g_1\left(\frac{1}{2}\right)\varphi(5) \tag{29}$$

$$f(2) = (g_1(1))^2\varphi(2) + 3\left(g_1\left(\frac{1}{2}\right)\right)^2 g_1(1)\varphi(3) + 4\left(g_1\left(\frac{1}{3}\right)\right)^3 g_1(1)\varphi(4) + \left(g_1\left(\frac{1}{2}\right)\right)^4\varphi(4) + 10\left(g_1\left(\frac{1}{3}\right)\right)^3\left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(5) + \left(g_1\left(\frac{1}{3}\right)\right)^6\varphi(6). \tag{30}$$

Again, the parameter μ is determined by

$$f\left(\frac{1}{3}\right) = \left(1 - \frac{9}{\pi^2\mu}\right)\frac{6}{\pi^2} = \frac{\sharp \text{ in }]0,\frac{5}{12}]}{\text{total } \sharp}$$

$$= \frac{6{,}508}{46{,}853}$$

yielding $\mu = 1.1819488123$. We obtain the following remarkable table and graph.

Table 2. Distribution of overall fractional scores (case of i=3)

q	Theoretical (f)	Experimental
1/3	0.1389322	0.1389322
1/2	0.1562373	0.2530468
2/3	0.0079701	0.0112693
5/6	0.0294265	0.0236911
1	0.323324	0.3125520
7/6	0.0027311	0.0097112
4/3	0.0389478	0.0192304
3/2	0.0417687	0.0340426
5/3	0.0059575	0.0049303
11/6	0.0129368	0.0072781
2	0.0483318	0.070959

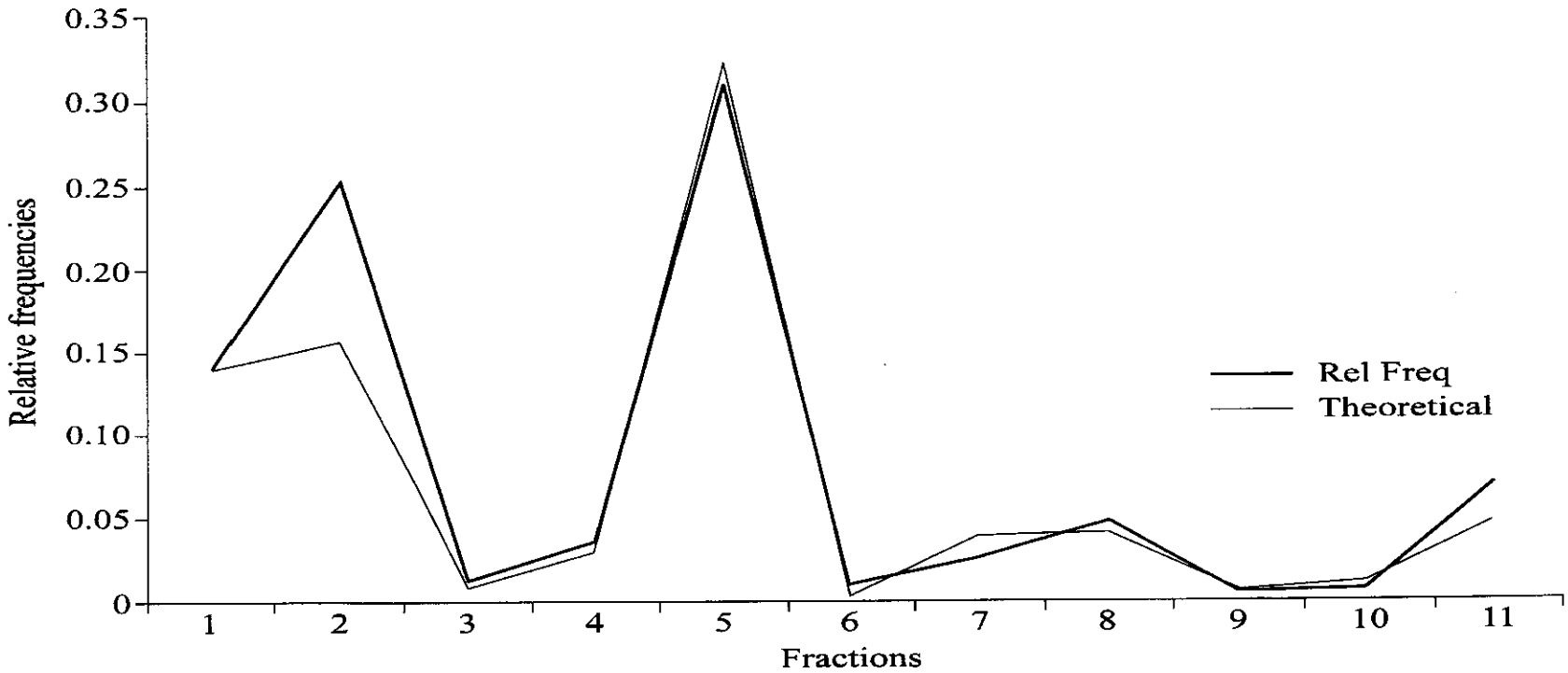

Fig 2 Theoretical and experimental fractional frequency distributions (case of i=3).

The agreement between the theoretical and experimental results is remarkable. This proves that the two dual Lotka laws are capable of modelling fractional frequency distributions (except, maybe, for $q=\frac{1}{2}$, on which we will comment at the end of the paper). The model also allows to prove the following inequalities (non-exhaustive list, proofs are left to the reader) :

1. $f\left(\frac{1}{3}\right) < f\left(\frac{1}{2}\right) \Leftrightarrow \mu < 1.216$ (as in our case)

2. $f\left(\frac{2}{3}\right) < \frac{1}{4} f\left(\frac{1}{3}\right)$

3. $f(1) > 2f\left(\frac{1}{2}\right)$

4. $f\left(\frac{5}{6}\right) < \min\left(\frac{1}{2} f\left(\frac{1}{2}\right), \frac{3}{2\pi^2} f\left(\frac{1}{3}\right)\right)$

5. $f\left(\frac{4}{3}\right) > 2f\left(\frac{5}{6}\right)$

6. $f(2) > f\left(\frac{3}{2}\right)$

7. $f\left(\frac{5}{3}\right) > 2f\left(\frac{7}{6}\right)$

8. $f(1) > f\left(\frac{3}{2}\right)$

9. $f\left(\frac{2}{3}\right) < f\left(\frac{5}{6}\right) \Leftrightarrow \mu < 1.52$ $\left(\text{and hence } f\left(\frac{1}{3}\right) < f\left(\frac{1}{2}\right) \Rightarrow f\left(\frac{2}{3}\right) < f\left(\frac{5}{6}\right)\right)$.

10. $\lim_{\mu \to \infty} f(q) \begin{cases} = 0, & q \neq \frac{n}{3},\ n \in \mathbb{N} \\ = \frac{\pi^2}{6n^2}, & q = \frac{n}{3} \end{cases}$

The explanation for this last regularity is as follows : for extremally high μ, the change to have a paper with less than 3 authors is very small. In this case one can only receive a fractional score of $\frac{1}{3}$ per paper. Hence the only overall scores that are possible are $q=\frac{n}{3}$ if an author has n papers. The probability for this last event is $\varphi(n)=\frac{\pi^2}{6n^2}$, $n\in\mathbb{N}$.

V. The case i=4 : allowing an author score of 1/4, 1/3, 1/2 or 1 in 1 paper.

We are heading now towards increasing refinement : more fractional scores are obtained (and hence their probability must be determined) and - on the corresponding experimental side - more but smaller intervals are used to group data. So, for any data set, there comes a time where extra refinements lead to too few data in the groupings and hence to the comparison of many very small numbers. This will be experienced from i=5 on (see next section). It is our feeling that for the Rao data, the present case i=4 is the most interesting one.

In this case an author receives a score 1 if he/she is an author in a single-authored paper, a score $\frac{1}{2}$ if he/she is an author in a 2-authored paper, a score $\frac{1}{3}$ if he/she is an author in a 3-authored paper and a score $\frac{1}{4}$ if he/she is an author in a j-authored paper, for all $j\geq 4$. Now we have

$$g_1(1) = f_1(1) = \frac{6}{\pi^2\mu} \tag{31}$$

$$g_1\left(\frac{1}{2}\right) = f_1\left(\frac{1}{2}\right) = \frac{3}{\pi^2\mu} \tag{32}$$

$$g_1\left(\frac{1}{3}\right) = f_1\left(\frac{1}{3}\right) = \frac{2}{\pi^2\mu} \tag{33}$$

$$g_1\left(\frac{1}{4}\right) = 1 - \left(f_1(1) + f_1\left(\frac{1}{2}\right) + f_1\left(\frac{1}{3}\right)\right)$$

$$= 1 - \frac{11}{\pi^2\mu} \tag{34}$$

Possible overall fractional scores (in]0,2]) are $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{7}{12}$, $\frac{2}{3}$, $\frac{3}{4}$, $\frac{5}{6}$, $\frac{11}{12}$, 1, $\frac{13}{12}$, $\frac{7}{6}$, $\frac{5}{4}$, $\frac{4}{3}$, $\frac{17}{12}$, $\frac{3}{2}$, $\frac{19}{12}$, $\frac{5}{3}$, $\frac{7}{4}$, $\frac{11}{6}$, $\frac{23}{12}$, 2 (theoretical) to be compared with grouped data (from the Table in the Appendix) in the intervals $]0,\frac{7}{24}]$, $]\frac{7}{24},\frac{9}{24}]$, $]\frac{9}{24},\frac{13}{24}]$, $]\frac{13}{24},\frac{15}{24}]$, $]\frac{15}{24},\frac{17}{24}]$, $]\frac{17}{24},\frac{19}{24}]$, $]\frac{19}{24},\frac{21}{24}]$, $]\frac{21}{24},\frac{23}{24}]$, $]\frac{23}{24},\frac{25}{24}]$, $]\frac{25}{24},\frac{27}{24}]$, $]\frac{27}{24},\frac{29}{24}]$, $]\frac{29}{24},\frac{31}{24}]$, $]\frac{31}{24},\frac{33}{24}]$, $]\frac{33}{24},\frac{35}{24}]$, $]\frac{35}{24},\frac{37}{24}]$, $]\frac{37}{24},\frac{39}{24}]$, $]\frac{39}{24},\frac{41}{24}]$, $]\frac{41}{24},\frac{43}{24}]$, $]\frac{43}{24},\frac{45}{24}]$, $]\frac{45}{24},\frac{47}{24}]$, $]\frac{47}{24},\frac{49}{24}]$.

We have the following formulas from which the theoretical fractional frequency distribution can be calculated.

$$f\left(\frac{1}{4}\right) = g_1\left(\frac{1}{4}\right)\varphi(1) \tag{35}$$

$$f\left(\frac{1}{3}\right) = g_1\left(\frac{1}{3}\right)\varphi(1) \tag{36}$$

$$f\left(\frac{1}{2}\right) = g_1\left(\frac{1}{2}\right)\varphi(1) + \left(g_1\left(\frac{1}{4}\right)\right)^2\varphi(2) \tag{37}$$

$$f\left(\frac{7}{12}\right) = 2\, g_1\left(\frac{1}{4}\right)g_1\left(\frac{1}{3}\right)\varphi(2) \tag{38}$$

$$f\left(\frac{2}{3}\right) = \left(g_1\left(\frac{1}{3}\right)\right)^2\varphi(2) \tag{39}$$

$$f\left(\frac{3}{4}\right) = 2\, g_1\left(\frac{1}{4}\right)g_1\left(\frac{1}{2}\right)\varphi(2) + \left(g_1\left(\frac{1}{4}\right)\right)^3\varphi(3) \tag{40}$$

$$f\left(\frac{5}{6}\right) = 2\, g_1\left(\frac{1}{3}\right)g_1\left(\frac{1}{2}\right)\varphi(2) + 3\left(g_1\left(\frac{1}{4}\right)\right)^2 g_1\left(\frac{1}{3}\right)\varphi(3) \tag{41}$$

$$f\left(\frac{11}{12}\right) = 3\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1\left(\frac{1}{4}\right)\varphi(3) \tag{42}$$

$$f(1) = g_1(1)\varphi(1) + \left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(2) + 3\left(g_1\left(\frac{1}{4}\right)\right)^2 g_1\left(\frac{1}{2}\right)\varphi(3) + \left(g_1\left(\frac{1}{4}\right)\right)^4\varphi(4) \tag{43}$$

$$f\left(\frac{13}{12}\right) = 6\, g_1\left(\frac{1}{4}\right) g_1\left(\frac{1}{3}\right) g_1\left(\frac{1}{2}\right)\varphi(3) + 4\left(g_1\left(\frac{1}{4}\right)\right)^3 g_1\left(\frac{1}{3}\right)\varphi(4) \tag{44}$$

$$f\left(\frac{7}{6}\right) = 3\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1\left(\frac{1}{2}\right)\varphi(3) + 6\left(g_1\left(\frac{1}{3}\right)\right)^2\left(g_1\left(\frac{1}{4}\right)\right)^2\varphi(4) \tag{45}$$

$$\begin{aligned} f\left(\frac{5}{4}\right) &= 2\, g_1(1) g_1\left(\frac{1}{4}\right)\varphi(2) + 3\left(g_1\left(\frac{1}{2}\right)\right)^2 g_1\left(\frac{1}{4}\right)\varphi(3) \\ &\quad + 4\, g_1\left(\frac{1}{2}\right)\left(g_1\left(\frac{1}{4}\right)\right)^3\varphi(4) + 4\, g_1\left(\frac{1}{4}\right)\left(g_1\left(\frac{1}{3}\right)\right)^3\varphi(4) \\ &\quad + \left(g_1\left(\frac{1}{4}\right)\right)^5\varphi(5) \end{aligned} \tag{46}$$

$$\begin{aligned} f\left(\frac{4}{3}\right) &= 2\, g_1(1) g_1\left(\frac{1}{3}\right)\varphi(2) + 3\, g_1\left(\frac{1}{3}\right)\left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(3) \\ &\quad + 12\left(g_1\left(\frac{1}{4}\right)\right)^2 g_1\left(\frac{1}{3}\right) g_1\left(\frac{1}{2}\right)\varphi(4) + \left(g_1\left(\frac{1}{3}\right)\right)^4\varphi(4) \\ &\quad + 5\left(g_1\left(\frac{1}{4}\right)\right)^4 g_1\left(\frac{1}{3}\right)\varphi(5) \end{aligned} \tag{47}$$

$$f\left(\frac{17}{12}\right) = 12\, g_1\left(\frac{1}{4}\right)\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1\left(\frac{1}{2}\right)\varphi(4) + 10\left(g_1\left(\frac{1}{4}\right)\right)^3\left(g_1\left(\frac{1}{3}\right)\right)^2\varphi(5) \tag{48}$$

$$\begin{aligned} f\left(\frac{3}{2}\right) &= 2\, g_1(1) g_1\left(\frac{1}{2}\right)\varphi(2) + 3\left(g_1\left(\frac{1}{4}\right)\right)^2 g_1(1)\varphi(3) + \left(g_1\left(\frac{1}{2}\right)\right)^3\varphi(3) \\ &\quad + 6\left(g_1\left(\frac{1}{4}\right)\right)^2\left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(4) + 4\left(g_1\left(\frac{1}{3}\right)\right)^3 g_1\left(\frac{1}{2}\right)\varphi(4) \\ &\quad + 5\left(g_1\left(\frac{1}{4}\right)\right)^4 g_1\left(\frac{1}{2}\right)\varphi(5) + 10\left(g_1\left(\frac{1}{4}\right)\right)^2\left(g_1\left(\frac{1}{3}\right)\right)^3\varphi(5) \\ &\quad + \left(g_1\left(\frac{1}{4}\right)\right)^6\varphi(6) \end{aligned} \tag{49}$$

$$f\left(\frac{19}{12}\right) = 6\, g_1\left(\frac{1}{4}\right) g_1\left(\frac{1}{3}\right) g_1(1)\varphi(3) + 12\, g_1\left(\frac{1}{4}\right) g_1\left(\frac{1}{3}\right)\left(g_1\left(\frac{1}{2}\right)\right)^2 \varphi(4)$$
$$+ 20\left(g_1\left(\frac{1}{4}\right)\right)^3 g_1\left(\frac{1}{3}\right) g_1\left(\frac{1}{2}\right)\varphi(5) + 5\, g_1\left(\frac{1}{4}\right)\left(g_1\left(\frac{1}{3}\right)\right)^4 \varphi(5)$$
$$+ 6\left(g_1\left(\frac{1}{4}\right)\right)^5 g_1\left(\frac{1}{3}\right)\varphi(6) \tag{50}$$

$$f\left(\frac{5}{3}\right) = 3\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1(1)\varphi(3) + 6\left(g_1\left(\frac{1}{3}\right)\right)^2\left(g_1\left(\frac{1}{2}\right)\right)^2 \varphi(4)$$
$$+ 30\left(g_1\left(\frac{1}{4}\right)\right)^2\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1\left(\frac{1}{2}\right)\varphi(5) + \left(g_1\left(\frac{1}{3}\right)\right)^5 \varphi(5)$$
$$+ 15\left(g_1\left(\frac{1}{4}\right)\right)^4\left(g_1\left(\frac{1}{3}\right)\right)^2 \varphi(6) \tag{51}$$

$$f\left(\frac{7}{4}\right) = 6\, g_1\left(\frac{1}{4}\right) g_1\left(\frac{1}{2}\right) g_1(1)\varphi(3) + 4\left(g_1\left(\frac{1}{4}\right)\right)^3 g_1(1)\varphi(4)$$
$$+ 4\, g_1\left(\frac{1}{4}\right)\left(g_1\left(\frac{1}{2}\right)\right)^3 \varphi(4) + 10\left(g_1\left(\frac{1}{4}\right)\right)^3\left(g_1\left(\frac{1}{2}\right)\right)^2 \varphi(5)$$
$$+ 20\, g_1\left(\frac{1}{4}\right)\left(g_1\left(\frac{1}{3}\right)\right)^3 g_1\left(\frac{1}{2}\right)\varphi(5) + 20\left(g_1\left(\frac{1}{4}\right)\right)^3\left(g_1\left(\frac{1}{3}\right)\right)^3 \varphi(6)$$
$$+ 6\left(g_1\left(\frac{1}{4}\right)\right)^5 g_1\left(\frac{1}{2}\right)\varphi(6) + \left(g_1\left(\frac{1}{4}\right)\right)^7 \varphi(7) \tag{52}$$

$$f\left(\frac{11}{6}\right) = 6\, g_1\left(\frac{1}{3}\right) g_1\left(\frac{1}{2}\right) g_1(1)\varphi(3) + 12\left(g_1\left(\frac{1}{4}\right)\right)^2 g_1\left(\frac{1}{3}\right) g_1(1)\varphi(4)$$
$$+ 4\, g_1\left(\frac{1}{3}\right)\left(g_1\left(\frac{1}{2}\right)\right)^3 \varphi(4) + 30\left(g_1\left(\frac{1}{4}\right)\right)^2 g_1\left(\frac{1}{3}\right)\left(g_1\left(\frac{1}{2}\right)\right)^2 \varphi(5)$$
$$+ 5\left(g_1\left(\frac{1}{3}\right)\right)^4 g_1\left(\frac{1}{2}\right)\varphi(5) + 30\left(g_1\left(\frac{1}{4}\right)\right)^4 g_1\left(\frac{1}{3}\right) g_1\left(\frac{1}{2}\right)\varphi(6)$$
$$+ 15\left(g_1\left(\frac{1}{4}\right)\right)^2\left(g_1\left(\frac{1}{3}\right)\right)^4 \varphi(6) + 7\left(g_1\left(\frac{1}{4}\right)\right)^6 g_1\left(\frac{1}{3}\right)\varphi(7) \tag{53}$$

$$f\left(\frac{23}{12}\right) = 12\, g_1\left(\frac{1}{4}\right)\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1(1)\varphi(4) + 30\, g_1\left(\frac{1}{4}\right)\left(g_1\left(\frac{1}{3}\right)\right)^2\left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(5)$$
$$+ 60\left(g_1\left(\frac{1}{4}\right)\right)^3\left(g_1\left(\frac{1}{3}\right)\right)^2 g_1\left(\frac{1}{2}\right)\varphi(6) + 6\, g_1\left(\frac{1}{4}\right)\left(g_1\left(\frac{1}{3}\right)\right)^5\varphi(6)$$
$$+ 21\left(g_1\left(\frac{1}{4}\right)\right)^5\left(g_1\left(\frac{1}{3}\right)\right)^2\varphi(7) \qquad (54)$$

$$f(2) = (g_1(1))^2\varphi(2) + 3\left(g_1\left(\frac{1}{2}\right)\right)^2 g_1(1)\varphi(3) + 12\left(g_1\left(\frac{1}{4}\right)\right)^2 g_1\left(\frac{1}{3}\right)g_1(1)\varphi(4)$$
$$+ 4\left(g_1\left(\frac{1}{3}\right)\right)^3 g_1(1)\varphi(4) + \left(g_1\left(\frac{1}{2}\right)\right)^4\varphi(4) + 5\left(g_1\left(\frac{1}{4}\right)\right)^4 g_1(1)\varphi(5)$$
$$+ 10\left(g_1\left(\frac{1}{4}\right)\right)^2\left(g_1\left(\frac{1}{2}\right)\right)^3\varphi(5) + 10\left(g_1\left(\frac{1}{3}\right)\right)^3\left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(5)$$
$$+ 15\left(g_1\left(\frac{1}{4}\right)\right)^4\left(g_1\left(\frac{1}{2}\right)\right)^2\varphi(6) + 60\left(g_1\left(\frac{1}{4}\right)\right)^2\left(g_1\left(\frac{1}{3}\right)\right)^3 g_1\left(\frac{1}{2}\right)\varphi(6)$$
$$+ \left(g_1\left(\frac{1}{3}\right)\right)^6\varphi(6) + 7\left(g_1\left(\frac{1}{4}\right)\right)^6 g_1\left(\frac{1}{2}\right)\varphi(7)$$
$$+ 35\left(g_1\left(\frac{1}{4}\right)\right)^4\left(g_1\left(\frac{1}{3}\right)\right)^3\varphi(7) + \left(g_1\left(\frac{1}{4}\right)\right)^8\varphi(8). \qquad (55)$$

The parameter μ is determined by

$$f\left(\frac{1}{4}\right) = \left(1 - \frac{11}{\pi^2\mu}\right)\frac{6}{\pi^2} = \frac{\sharp \text{ in }]0,\frac{7}{24}]}{\text{total } \sharp}$$
$$= \frac{1{,}698}{46{,}853}$$

yielding $\mu = 1.1851868311$. We obtain the following - again remarkable - table and graph.

Table 3. Distribution of overall fractional scores (case of i=4)

q	Theoretical (f)	Experimental
1/4	0.036241009	0.036241009
1/3	0.1039555	0.102233051
1/2	0.1564714	0.250570935
7/12	0.0030927	0.003052099
2/3	0.0044441	0.01274198
3/4	0.0046533	0.00420464
5/6	0.013455	0.019827973
11/12	0.0003526	0.001323288
1	0.3220503	0.309905449
13/12	0.0010631	0.001686125
7/6	0.0015435	0.005698675
5/4	0.0100909	0.003905833
4/3	0.0290481	0.018654088
17/12	0.000205	0.001899558
3/2	0.0417575	0.053358376
19/12	0.0024309	0.000661644
5/3	0.0035013	0.003478966
7/4	0.0033829	0.001430004
11/6	0.0093664	0.007000619
23/12	0.0004929	0.000277464
2	0.0476329	0.0697428726

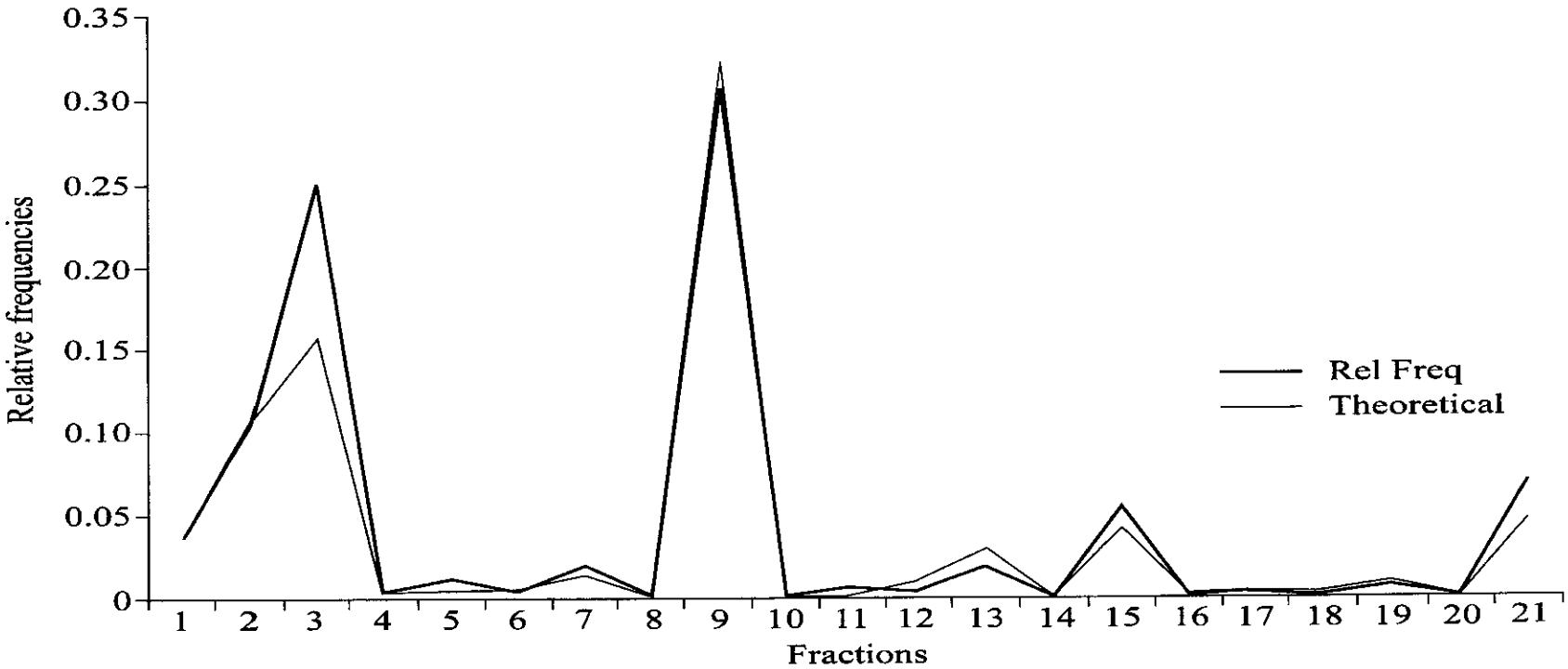

Fig .3 Theoretical and experimental fractional frequency distributions (case of i=4).

We can say that, again, the agreement between the theoretical and experimental results is remarkable. In this model we can prove the following inequalities (the proofs are straightforward and left to the reader) (non-exhaustive list).

1. $f\left(\frac{1}{2}\right) > \frac{3}{2} f\left(\frac{1}{3}\right)$

2. $f\left(\frac{7}{12}\right) \leq \frac{1}{\pi^2} f\left(\frac{1}{4}\right)$ and equality is valid $\Leftrightarrow \mu=1$ (i.e. $\Leftrightarrow$ every paper has one author)

3. $f\left(\frac{2}{3}\right) < \frac{1}{2\pi^2} f\left(\frac{1}{3}\right)$

4. $f(1) > 3\, f\left(\frac{1}{3}\right)$

5. $f\left(\frac{3}{2}\right) > \frac{9}{2} f\left(\frac{2}{3}\right)$

6. $f\left(\frac{7}{6}\right) < \frac{6}{\pi^4} f\left(\frac{1}{2}\right)$

7. $f\left(\frac{5}{4}\right) > 3\, f\left(\frac{7}{12}\right)$

8. $f\left(\frac{13}{12}\right) > 3\, f\left(\frac{11}{12}\right)$

9. $f\left(\frac{11}{12}\right) < \frac{4}{3\pi^2} f\left(\frac{7}{12}\right)$

10. $f\left(\frac{3}{4}\right) > \frac{3}{2} f\left(\frac{7}{12}\right)$

11. $f\left(\frac{3}{4}\right) < \frac{1}{2} f\left(\frac{1}{2}\right)$

12. $f\left(\frac{17}{12}\right) < \frac{16}{5\pi^2} f\left(\frac{13}{12}\right)$

13. $f\left(\frac{13}{12}\right) < \frac{9}{2\pi^2} f\left(\frac{3}{4}\right)$ and hence

14. $f\left(\frac{17}{12}\right) < \frac{72}{5\pi^4} f\left(\frac{3}{4}\right)$.

15. $\lim_{\mu\to\infty} f(q) \begin{cases} = 0, \ q \neq \frac{n}{4}, \ n \in \mathbb{N} \\ = \frac{\pi^2}{6n^2}, \ q = \frac{n}{4} \end{cases}$

The argument given in section IV for the similar result also applies here.

VI. The case i=5 : allowing an author score of 1/5, 1/4, 1/3, 1/2 or 1 in 1 paper.

This case is very complex as we will describe below. Happily it turns out that this case is a bit "overkill" w.r.t. to the Rao data in the Appendix. Nevertheless we describe it here (briefly) so that it can be used for larger datasets, to be produced in the future. That this case is a bit "overkill" for our data is not a drawback for the model : one could compare this with the case in statistics where one is graphing a set of continuous data, be means of a histogram where too many bars are used (i.e. where the abscissa intervals are too small w.r.t. the number of data that one has).

We repeat that in this case an author receives a score $\frac{1}{j}$ if he/she is an author in an j-authored paper ($j \leq 5$) and where an author receives a score $\frac{1}{5}$ if he/she is an author in an j-authored paper ($j \geq 5$). Now we have

$$g_1(1) = f_1(1) = \frac{6}{\pi^2\mu} \tag{56}$$

$$g_1\left(\frac{1}{2}\right) = f_1\left(\frac{1}{2}\right) = \frac{3}{\pi^2\mu} \tag{57}$$

$$g_1\left(\frac{1}{3}\right) = f_1\left(\frac{1}{3}\right) = \frac{2}{\pi^2\mu} \tag{58}$$

$$g_1\left(\frac{1}{4}\right) = f_1\left(\frac{1}{4}\right) = \frac{3}{2\pi^2\mu} \tag{59}$$

$$g_1\left(\frac{1}{5}\right) = 1 - \left(f_1(1) + f_1\left(\frac{1}{2}\right) + f_1\left(\frac{1}{3}\right) + f_1\left(\frac{1}{4}\right)\right)$$

$$= 1 - \frac{25}{2\pi^2\mu} \tag{60}$$

The complete list of possible fractional scores comprises 83 rational numbers (in]0,2]) (for a list, see Table 4, first column) and the same number of half-open intervals for the groupings in the experimental data set in the Appendix. For each possible q we determined analytical formulae for f(q), based again on the g_1-variant of formula (3). They can be calculated as before, but the 83 formulae are, relatively speaking, more intricate. Nevertheless, they still contain only one parameter μ which we determined as μ=1.2899744688 based on our requirement that the value $f\left(\frac{1}{5}\right)$ must be exact. We can provide the reader with the formulae for f(q) and with the necessary intervals for the groupings. We omit it here since it would consume several pages. For the same reason we omit the Table giving the experimental and theoretical values for the case of i=5. We only provide the experimental and theoretical curves. Experimental and theoretical curves are so similar that they practically overlap if they are shown on a single XY-plane ; it is thus difficult to indentify the two different curves in a single XY-plane. Therefore, experimental and theoretical frequency distributions are shown in Figures 4 and 5 seperately. However, an attempt has been made to show both the curves on a single XY-plane in Figure 6. One can see from Figure 6 that the model fits the experimental data remarkably well.

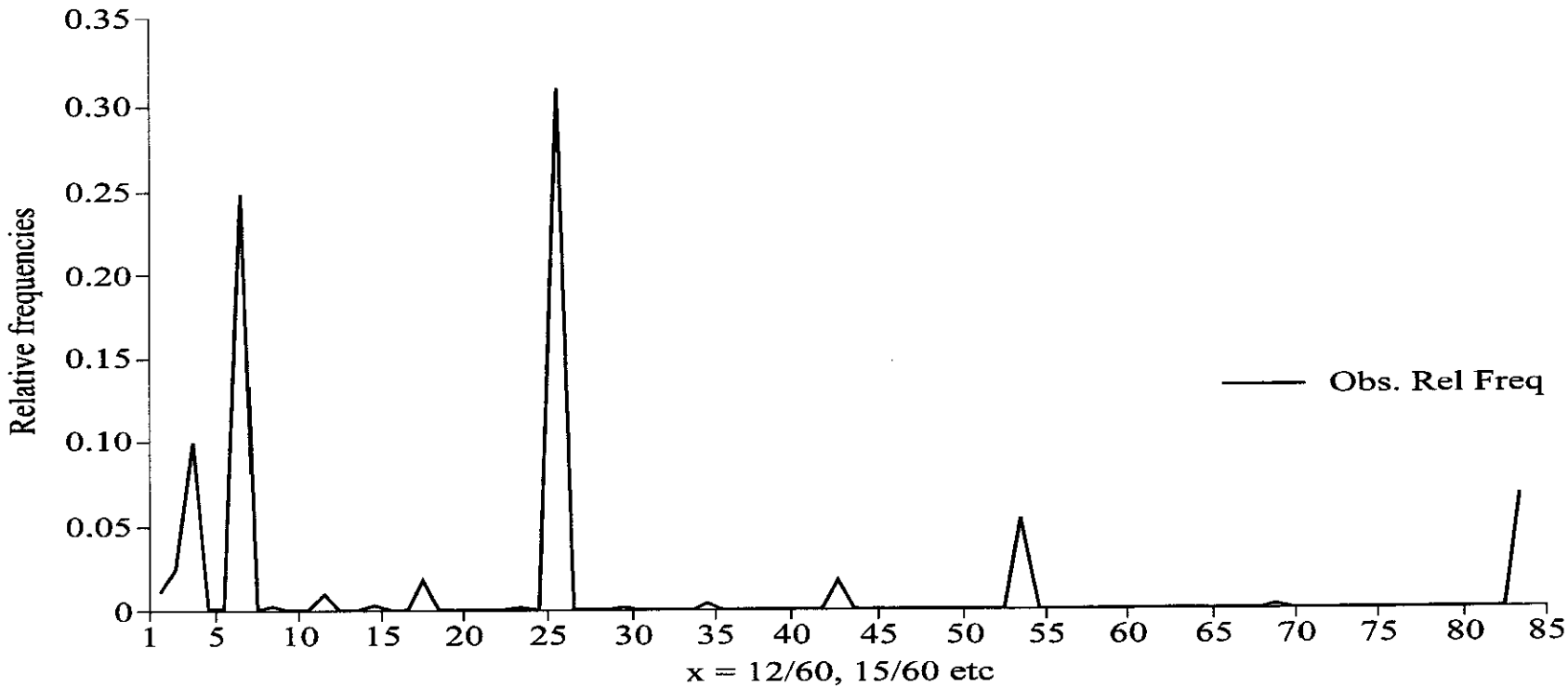

Fig. 4 Frequency curve for the experimental data (case of i=5)

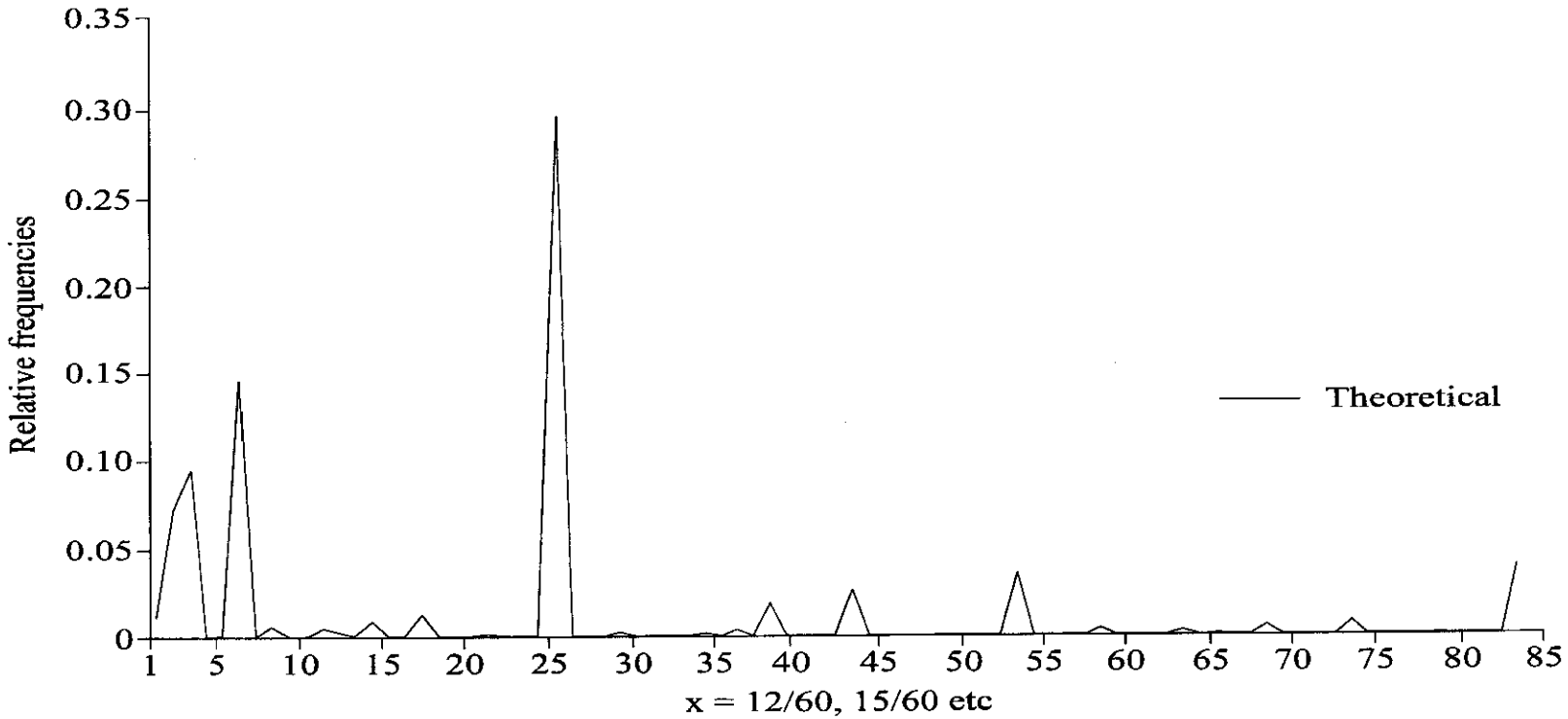

Fig. 5 Frequency curve for the theoretical values (case of i=5)

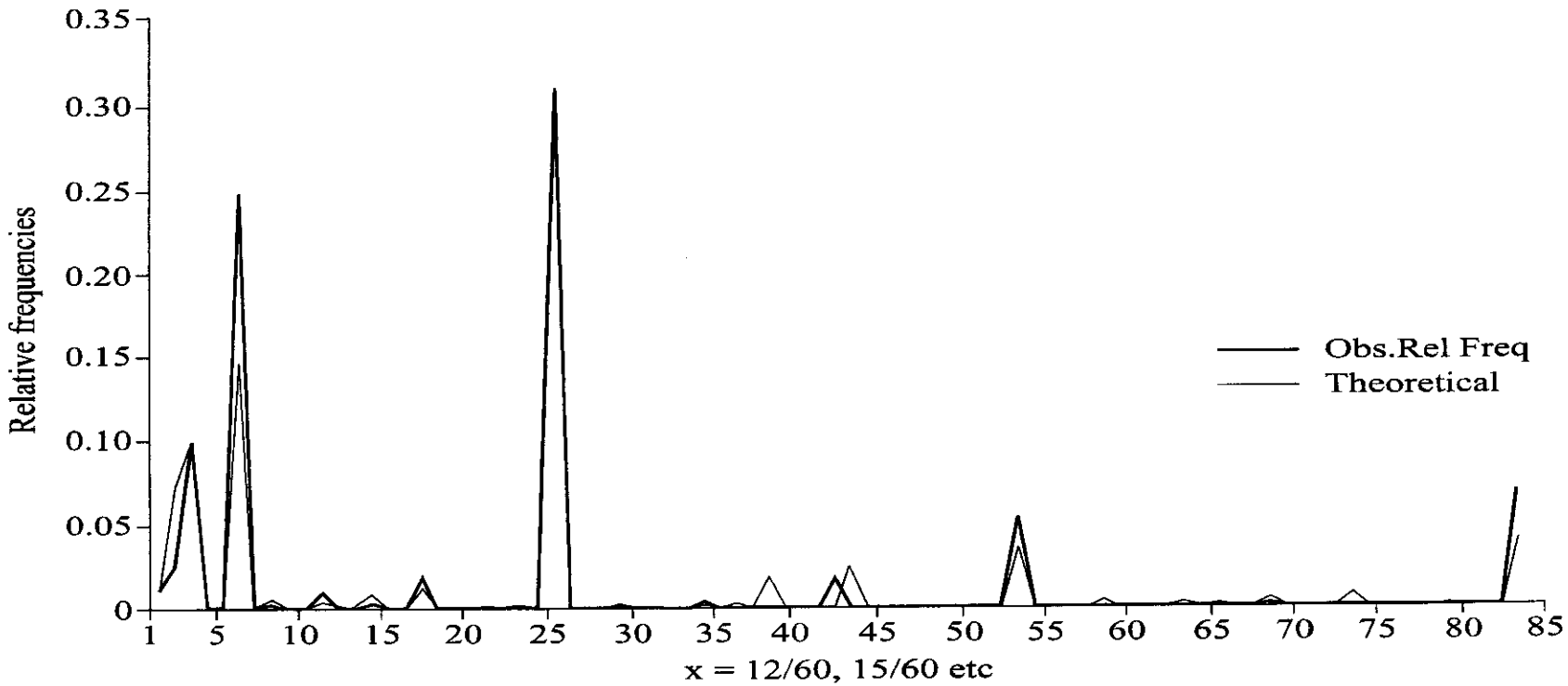

Fig . 6 Frequency curves for experimental and theoretical values (case of i=5)

VII. Conclusions and a remark on the case q=1/2

We applied the model

$$f(z) = \sum_{i=1}^{\infty} \underbrace{(f_1 \circledast \ldots \circledast f_1)}_{i \text{ times}}(z)\varphi(i) \qquad (3)$$

of Egghe (1993), where f is the overall fractional frequency distribution, f_1 is the fractional frequency distribution in 1 paper and φ is the distribution of the number of papers per author (total count). We obtained the exact discrete result that

$$f_1(z) = \frac{\psi\left(\frac{1}{z}\right)}{\mu z}, \qquad (2)$$

where ψ is the distribution of the number of authors per paper. The distributions φ and ψ are each others dual (cf. Egghe (1989), (1990)) and in this paper we use the simplest frequency distribution, known in informetrics, namely the discrete Lotka law with exponent 2 :

$$\varphi(n) = \psi(n) = \frac{6}{\pi^2 n^2} \qquad (4),(6)$$

(cf. Egghe and Rousseau (1990)).

In order to model the very irregular fractional frequency distributions (of the overall fractional scores of authors in a (large) bibliography), we use a variant of the fractional scoring system : fix $i \in \mathbb{N}$. An author receives a score of $\frac{1}{j}$ if he/she has a paper with j authors in total ($j \leq i$) and receives a score of $\frac{1}{i}$ if he/she has a paper with j authors in total ($j \geq i$). Per fixed i, a new fractional scoring distibution g_1 in 1 paper is derived (based on f_1) that can be used in our scoring system.

For i=2, 3, 4 and 5 we have determined the overall theoretical fractional frequency distribution which has the advantage that it contains only one parameter (which we estimated in each case). We then compared with the corresponding experimental fractional frequency graph of the (accordingly) grouped data, based on the data of Rao (1995), reproduced in the Appendix. The agreement is remarkable. For the Rao data-set the cases i=3 and i=4 appear to be best. The case i=5 is a good model but requires a grouping of data in very small intervals so that, for the Rao-data, this case is a bit "overkill", comparable with the case of the use of a histogram in statistics with too many bars w.r.t. the given data.

One remark on the fraction $q=\frac{1}{2}$ is in order. As is clear from the graphs in Figs. 2, 3, 6, the agreement between the theoretical and the experimental graphs is the poorest in $q=\frac{1}{2}$. This is -most probably - due to the fact that we used Lotka's law for ψ, the fraction of papers with a certain number n of authors :

$$\psi(n) = \frac{6}{\pi^2 n^2} \tag{6}$$

Although such a choice is good for its dual φ, there are many cases where there are relatively more papers with 2 authors than given by (6) (R. Rousseau - oral communication). In fact (6) gives $\psi(2)=\frac{1}{4}\psi(1)$ and is, most probably, the reason for our underestimation of $f(\frac{1}{2})$ in all cases i=3, 4, 5, since in our data - as mentioned in section II - we have $\psi(1)\approx\psi(2)$.

In a forthcoming paper we will investigate this further but we can report here on first attempts by replacing ψ in (6) by a Poisson distribution. We noticed already that a Poisson distribution, if the parameter λ in chosen in the appropriate way, is better capable of describing the distribution of the number of authors per paper. Using this ψ in our fractional frequency model we indeed obtained an improvement in $q=\frac{1}{2}$ and even in q=2 although now q=1 is more poorly modelled. We will investigate this further and see whether an overall improvement of the model obtained in this paper can be obtained.

We further conclude that two features of our model are :

- The resulting theoretical frequency curve clearly shows several ups and downs which are exactly similar to the ups and downs in the experimental frequency curve ; both the curves are so similar that it is difficult to distinguish between each other if they are shown on a single XY-plane.
- As i increases, the required number of formulae to compute theoretical values increases considerably and it may become difficult to compute the theoretical calues. In such circumstances an easy approach to derive the formulae to compute the theoretical values is absent. In order to get the solution for these problems, further investigation is required.

References

I. Ajiferuke (1991). A probabilistic model for the distribution of authorships. Journal of the American Society for Information Science 42(4), 279-289.

G. Blom (1989). Probability and Statistics, Theory and Applications, Springer-Verlag, Berlin.

Q. Burrell and R. Rousseau (1995). Fractional counts for authorship attribution : a numerical study. Journal of the American Society for Information Science 46(2), 97-102.

K.L. Chung (1974). A Course in Probability Theory. Academic Press, New York.

L. Egghe (1989). The Duality of informetric Systems with Applications to the empirical Laws. Ph. D. Thesis, City University, London (UK).

L. Egghe (1990). The duality of informetric systems with applications to the empirical laws. Journal of Information Science 16(1), 17-27.

L. Egghe (1993). Consequences of Lotka's law in the case of fractional counting of authorship and of first author counts. Mathematical and Computer Modelling 18(9), 63-77.

L. Egghe (2000). A heuristic study of the first-citation distribution. Scientometrics 48(3), 345-359.

L. Egghe and I.K.R. Rao (2002). Theory and experimentation on the most-recent-reference distribution. Scientometrics 53(3), to appear.

L. Egghe and R. Rousseau (1990). Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science, Elsevier, Amsterdam.

L. Egghe and R. Rousseau (2001). Elementary Statistics for effective Library and Information Service Management. Aslib, London (UK).

A.J. Lotka (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16, 317-323.

I.K. Ravichandra Rao (1995). A stochastic approach to analysis of distribution of papers in mathematics : Lotka's law revisited. Proceedings of the fifth international Conference of the international Society for Scientometrics and Informetrics, Rosary

College, 1995 (M.E.D. Koenig and A. Bookstein, eds.), 455-464. Learned Information, Medford, NJ, USA.

R. Rousseau (1992). Breakdown of the robustness property of Lotka's law : the case of adjusted counts for multi-authorship attribution. Journal of the American Society for Information Science 43(10), 645-647.

R. Rousseau (1994). The number of authors per article in library and information science can often be described by a simple probability distribution. Journal of Documentation 50(2), 134-141.

Appendix

Table of experimental fractional frequency distribution of author scores in mathematics

X : Fraction of papers	No. of authors
0.02	1
0.04	1
0.06	9
0.07	2
0.08	1
0.09	13
0.1	17
0.13	12
0.14	35
0.16	3
0.17	93
0.2	341
0.25	1,166
0.29	4
0.33	4,772
0.35	16
0.38	1
0.39	2
0.41	17
0.42	2
0.43	1
0.45	6
0.46	1
0.48	2
0.5	11,673

0.53	36
0.56	1
0.58	136
0.59	3
0.62	2
0.63	1
0.64	5
0.65	2
0.66	17
0.69	474
0.72	23
0.74	1
0.76	170
0.78	3
0.82	2
0.84	919
0.86	7
0.88	1
0.9	6
0.92	2
0.94	48
0.96	6
0.98	1
0.99	1
1	14,507
1.02	9
1.04	2
1.06	2
1.08	66
1.09	3
1.1	2

1.11	2
1.12	2
1.13	2
1.14	22
1.15	2
1.16	15
1.17	189
1.2	37
1.23	2
1.25	177
1.28	3
1.29	1
1.3	1
1.32	2
1.33	868
1.37	2
1.38	1
1.4	2
1.41	2
1.42	19
1.45	6
1.48	4
1.5	2,547
1.53	9
1.55	1
1.57	1
1.58	27
1.6	1
1.62	1
1.64	1
1.65	1

1.66	1
1.67	149
1.7	11
1.73	1
1.75	65
1.79	1
1.83	328
1.9	2
1.91	1
1.92	9
1.95	1
1.98	6
2	3,255
2.03	6
2.06	1
2.07	1
2.08	47
2.09	1
2.12	2
2.14	2
2.15	1
2.16	2
2.17	70
2.2	26
2.25	50
2.28	2
2.29	1
2.33	288
2.36	1
2.37	1
2.38	1

2.4	2
2.41	3
2.42	20
2.44	1
2.45	4
2.5	789
2.55	2
2.58	34
2.63	1
2.66	6
2.67	55
2.7	9
2.73	1
2.75	25
2.83	130
2.86	1
2.87	3
2.9	1
2.91	1
2.92	9
2.95	1
3	1,191
3.03	3
3.08	10
3.1	1
3.11	1
3.17	42
3.2	9
3.23	1
3.25	29
3.33	126

3.36	1
3.42	4
3.45	2
3.48	1
3.5	33
3.53	1
3.58	5
3.6	2
3.66	2
3.67	33
3.7	2
3.72	1
3.75	13
3.78	1
3.8	1
3.83	80
3.91	1
3.92	1
3.95	1
3.98	1
4	386
4.08	9
4.12	2
4.14	1
4.16	2
4.17	20
4.24	1
4.25	1
4.33	3
4.4	1
4.41	1

4.42	5
4.5	143
4.51	1
4.53	2
4.58	2
4.67	22
4.7	2
4.72	1
4.75	7
4.83	37
7.87	1
4.95	1
5	117
5.04	1
5.08	4
5.09	1
5.14	1
5.17	8
5.2	2
5.25	4
5.26	1
5.33	27
5.4	1
5.42	2
5.5	60
5.58	4
5.67	14
5.7	1
5.75	4
5.83	17
5.88	1

5.92	1
6	70
6.08	2
6.17	2
6.2	4
6.25	3
6.28	1
6.33	14
6.5	26
6.58	1
6.67	8
6.75	1
6.83	8
6.87	3
6.92	1
7	18
7.08	2
7.17	3
7.2	1
7.33	2
7.5	12
7.53	1
7.58	1
7.67	2
7.83	4
8	22
8.11	2
8.33	5
8.5	6
8.58	1
8.67	1

8.83	6
9	7
9.08	1
9.17	1
9.53	4
9.5	7
9.83	3
10	4
10.03	1
10.33	1
10.5	1
10.57	1
10.67	1
11	4
11.33	1
11.42	1
11.5	1
11.67	1
11.75	1
11.83	1
12	1
12.17	1
12.23	1
12.5	1
12.67	1
12.83	11
13	1
13.67	1
14.34	1
15	1
15.5	1

15.83	1
16	1
17.06	1
17.2	1
17.58	1
18	1
Total	46,853

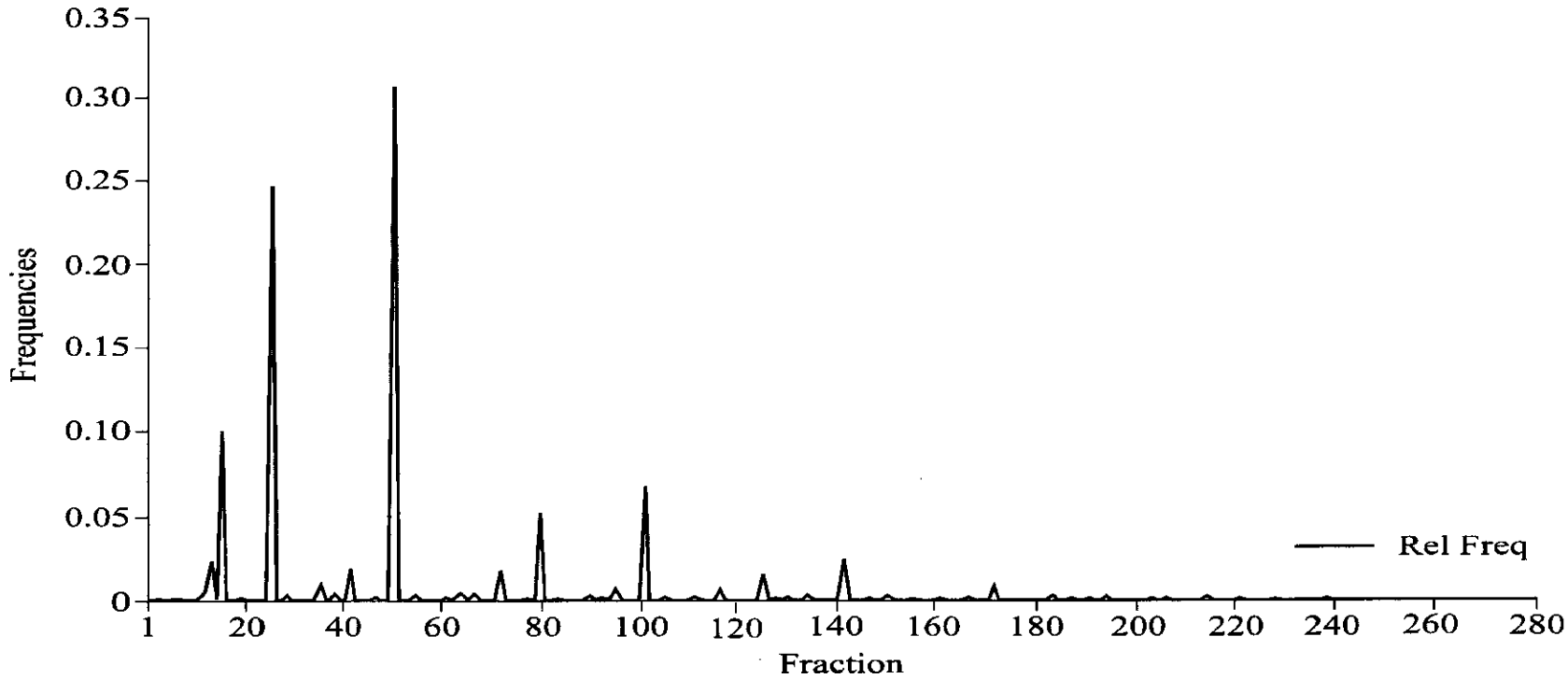

Fig. I. A fractional frequency curve (experimental data of previous table)