Scientometrics, Vol. 20. No. 1 (1991) 93-106

THE EXACT PLACE OF ZIPF'S AND PARETO'S LAW AMONGST THE CLASSICAL INFORMETRIC LAWS‡

L. EGGHE

*LUC, Universitaire Campus, B-3590 Diepenbeek (Belgium)**
UIA, Universiteitsplein 1, B-2610 Wilrijk (Belgium)

(Received October 4, 1989)

In this paper, the special place of Zipf's law and Pareto's law amongst other classical informetric laws (such as Bradford's graphical and verbal law, Weber-Fechner's or Brookes', Leimkuhler's and Mandelbrot's) is revealed and explained. Equivalencies amongst some of these laws are proved. We also determine the conditions under which Bradford's graphical law is a special case of Bradford's verbal law.

I. Introduction and definitions

In the sequel we will consider information production processes (IPP) (such as bibliographies, linguistical texts, ..., cf. Refs 7-9) that are continuous, i.e. in which we consider all the functions to be defined on intervals of the form [0, x], $x > 0$, as will be indicated below. We will use the terminology of sources "having" or "producing" items (such as bibliographies in which one considers journals having articles on a certain topic.)

We suppose that the sources are arranged in decreasing order of the number of items they have (or, more exactly, consider the dual system as explained in Refs 7 and 8). The source set is associated (because of the continuous setting) with the interval [0, T] and the items set with the interval [0, A]. We repeat the definitions of the classical laws we want to study here: the laws of Zipf or Pareto, Mandelbrot, Bradford (graphical and verbal, group dependent and group free), Brookes or Weber-Fechner and Leimkuhler.

‡ Dedicated to the memory of Michael J. Moravcsik

* Permanent address

Elsevier, Amsterdam – Oxford – New York – Tokyo
Akadémiai Kiadó

I.1. The verbal formulation of Bradford's law, group-dependent and group-free

I.1.1. The group-dependent verbal form of Bradford's law

This formulation is in fact the original one as given by Bradford.[2]

Given any IPP, we say that this IPP satisfies Bradford's verbal law with p groups ($p \in N$ fixed, but arbitrary), if we can divide the set I into q equal parts each containing $y_0 > 0$ items such that, following the ordering on S, we have a corresponding number of sources equal to (respectively):

$$r_0, r_0k, r_0k^2, \ldots, r_0k^{p-1}$$

for a certain $r_0 > 0$ and $k > 1$.

The number k is called the Bradford factor (or multiplicator) and is, of course, dependent on $p : k = k(p)$.

I.1.2. The group-free verbal form of Bradford's law

This form was defined in Ref. 7 (see also Refs 8 and 9) as follows: Let $\Sigma(i)$ denote the cumulative number of sources up to the item coordinate $i \in [0, A]$ and suppose Σ to be differentiable. Let σ be defined as :

$$\sigma = \Sigma'$$

We say that the IPP satisfies the *group-free verbal law of Bradford* if, for every $i \in I$,

$$\sigma(i) = M.K^i \qquad (1)$$

where $M > 0$ and $K > 1$ are constants.

Formula (1) is called the group-free Bradford function.

The number K is called the group-free Bradford factor and, of course, is independent of p in the previous section (p does not exist here!). This definition allows us to recognise Bradford's law as a function just like the other informetric laws that will be defined in the sequel.

In Refs 7 and 9 it is shown that the formulations in I.1.1. and I.1.2. are mathematically equivalent when we allow any $p \in N$ in I.1.1. This result is not needed in the sequel.

I.2. The graphical formulation of Bradford's law, group-dependent and group-free

I.2.1. The group-dependent graphical formulation of Bradford's law

Due to an apparent mis-interpretation of Bradford's original definition I.1.1., one can find the following law in the literature.[16]

Fix $p \in N$. We say that our IPP satisfies the graphical formulation of Bradford's law with p groups ($p \in N$ fixed but arbitrary) if we can divide the set [0, A] into p equal parts, each containing $y_0 > 0$ items, such that, following the order as defined above, we have, for the first y_0 items, the first $r_1 > 0$ sources, for the first $2\, y_0$ items, the first $r_1 k_1$ sources ($k_1 > 1$), for the first $3\, y_0$ items, the first $r_1 k_1^2$ sources, and so on until: for the first (p-1) y_0 items, the first $r_1 k_1^{p-2}$ sources and finally, the $p\, y_0 = A$ items stand for $r_1 k_1^{p-1} = T$ sources.

I.2.2. The group-free graphical formulation on Bradford's law

In view of I.2.1., one can define the following group-free analogue of I.1.2.

Let Σ be as in I.1.2. Then :

$$\Sigma(i) = M_1 K_1^{i} \tag{2}$$

where M_1 and $K_1 > 1$ are constants.

I.3. Leimkuhler's law

Let R(r) denote the cumulative number of items in the sources $s \in [0, r]$, for every $r \in [0, T]$. Then

$$R(r) = a \log (1 + br) \tag{3}$$

where a and b are positive constants. R is the corresponding Leimkuhler function. This definition can be found in Ref.12.

I.4. Brookes' law or the law of Weber-Fechner

Let R be as in I.3. Then

$$R(r) = \alpha \log [\beta(1+r)], \tag{4}$$

where α and β are positive constants. R is the corresponding Brookes (or Weber-Fechner) function. See Ref.3, for a reference.

I.5. Mandelbrot's law

Let g(r) denote the density of the numbers of items in $r \in [0, T]$.

Then, for every $r \in [0, T]$[13]:

$$g(r) = G/(1+Hr), \tag{5}$$

where G and H are constants and $r \in [0, T]$.

I.6. Zipf's law or Pareto's law

In the notation of I.5., we have here :

$$g(r) = F/(1 + r) \tag{6}$$

where F is a constant[17,15] (we restrict our attention to the power 1 in the denominator of (6). g is called the Zipf (or Pareto) function. Classically Zipf's function is considered as defined only for discrete values of r while Pareto's is defined for continuous values of r, but the functions themselves are the same.

Note that usually formula (6) is defined without the 1 in the denominator but then the ranks start in 1. Hence our approach (using $1 + r$ and $r \in [0, T]$) is in fact the same.

We start by the obvious remark that, if we take $H = 1$ in (5), then we find (4); hence Zipf's law (Pareto's law) is a special case of Mandelbrot's law. In a less obvious context this has also been observed in Ref. 14.

In Refs. 7 and 9 and partially in Refs. 14 and 1 the following theorem has been proved :

Theorem: The following assertions are equivalent for an IPP :

(i) It satisfies Bradford's law (verbal), group-dependent, for every $p \in \mathbf{N}$.

(ii) It satisfies Bradford's law (verbal), group-free.

(iii) It satisfies Leimkuhler's law.

(iv) It satisfies Mandelbrot's law.

Furthermore, we could show the following relations :

$$a = (y_0)/(\log k) = 1/(\log K) \tag{7}$$

$$b = (k-1)/r_0 = (\log K)/M \tag{8}$$

$$G = ab \tag{9}$$

$$H = b \tag{10}$$

$$K = k^{p/A} \tag{11}$$

It is our purpose to make a second "closed" circuit of equivalencies between Zipf's law, Bradford's (graphical) laws and Brookes' law. As defined above, such a closed circuit cannot be proved as will be explained in the next (second) section. In the third section we will remedy the problems and in the fourth section we will prove our second closed circuit of equivalent informetric laws, incorporating Zipf's law. Since Zipf's law is a special case of Mandelbrot's ($H=1$, as explained above), this second circuit is hence a special case of the first circuit.

The results and proofs of this article are a more accurate version of the earlier attempts in Ref.6. There we dealt with the notion of "asymptotic equivalence" in order to get rid of some problems (to be mentioned in the next section). The solution presented in this paper is mathematically correct and yields more insight in Zipf's and Bradford's graphical law.

II. Problems in classification

The fact that Zipf's law is equivalent with Brookes' law causes no problems. The reader can check the easy proof in the next section. In order, however, to show that Zipf's law is also equivalent to Bradford's graphical laws, we need, in view of the previous section that, if $H=1$, Bradford's graphical laws are the same as Bradford's verbal laws (since the latter are equivalent with Mandelbrot's law and since Zipf's law is Mandelbrot's law if $H=1$).

As defined above (which is the classical way), Bradford's graphical laws *cannot* be considered the same as Bradford's verbal laws (and this is true for the group-dependent as well as the group-free versions). This will be shown now.

Theorem II.1: Bradford's verbal law with p groups ($p \in N$ fixed but arbitrary) is never the same as Bradford's graphical law with p groups.

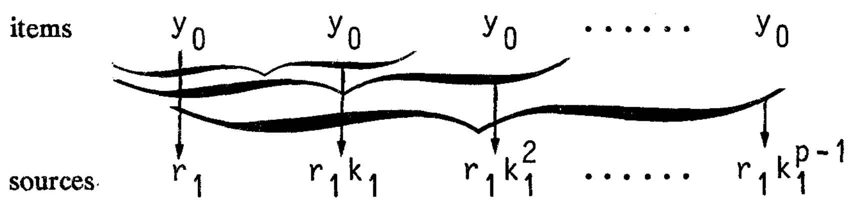
Proof:

The verbal Bradford law for p groups yields r_0, k and $y_0=A/p$ such that (schematically) we have for every group:

items	y_0	y_0	y_0	......	y_0
	$\downarrow$	$\downarrow$	$\downarrow$		$\downarrow$
sources	r_0	r_0k	r_0k^2	......	r_0k^{p-1}

Situation I

Suppose this situation is also describable via the graphical law of Bradford with p groups in its classical formulation. Then each group still has y_0 items. We now have r_1 and k_1 such that



Situation II

From this viewpoint, we never have the two situations occurring together. Indeed, to have both situations we need to have $r_1 = r_0$ and $r_1k_1 = r_0 + r_0k$; hence

$$k_1 = 1 + k \tag{12}$$

But then the third group contains

$$r_1k_1^2 = r_0(1+k)^2 \tag{13}$$

sources; while in the first case this group has r_0k^2 sources. Since both groups must be equal (since they are made that way) we conclude : The above situations are never the same.

Theorem II.2 : Bradford's verbal group-free law is never the same as Bradford's graphical group-free law.

Proof :

The verbal group-free Bradford law gives the regularity (1) :

$$\sigma(i) = MK^i \tag{1}$$

for every $i \in [0, A]$. Hence

$$\Sigma(i) = \int_0^i \sigma(i')\, di'$$

$$\Sigma(i) = [M/(\log K)]\, K^i - M/(\log K) \tag{14}$$

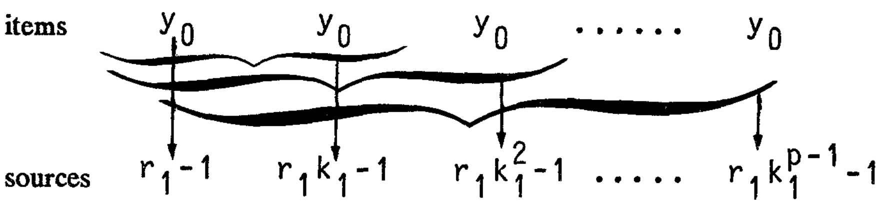
No values for M and K can be found in order that (14) should be the same as (2).

III. Remedy

So, we cannot use the graphical form of Bradford's law, as defined in Section I.2. However, in this section we will show that, if all the ranks in the definitions in I.2. are lowered with one, then we can considered Bradford's graphical laws as the same as Bradford's verbal laws, if and only if $H=1$. Since this operation is such a minor deviation of the original definitions we keep the same names. Therefore we redefine.

Definition III.1 : We say that an IPP satisfies the graphical form of Bradford's law with p groups ($p \in N$ fixed but arbitrary) if we can divide the set [0, A] into p equal parts, each containing $y_0 > 0$ items such that we have, following the order as defined in section I, for the first y_0 items, the first r_1-1 > 0 sources, for the first 2 y_0 items, the first r_1k_1 -1 sources ($k_1 > 1$), for the first 3 y_0 items, the first $r_1k_1^2$ -1 sources, and so on until: for the first (p-1)y_0 items, the first $r_1k_1^{p-2}$ -1 sources and finally, the p y_0=A items stand for $r_1k_1^{p-1}$ -1=T sources.

This situation is depicted as in Situation III :



Situation III

Definition III.2 : We say that an IPP satisfies the graphical form of Bradford's law, group-free, if (with Σ as in I.1.2.):

$$\Sigma(i)=M_1K_1^i - 1, \tag{15}$$

where M_1 and $K_1 > 1$ are constants.

We can now show that, if $H=1$ (hence, in the case that Zipf's and Mandelbrot's law are the same), then Bradford's graphical and verbal laws are the same. Furthermore, surprisingly, also the converse is true : only if $H=1$, Bradford's graphical and verbal laws are the same. This is proved now.

Theorem III.3 : The following assertions are equivalent :

(i) $H=1$.

(ii) Bradford's graphical law for p groups ($p \in N$ fixed but arbitrary) is the same as Bradford's verbal law for p groups.

Proof : *(i) ⇒ (ii)*

If $H=1$ or, equivalently [by (10)] $b=1$, then the following relations suffice :

$$r_1 - 1=r_0 \tag{16}$$

$$k_1=k=1 + r_0 \tag{17}$$

[$k=1+r_0$ by (8)].

The reader can verify that, with these definitions, both the graphical and the verbal formulations of Bradford's law are valid. We just check it for the first three groups :

and
$$r_0=r_1-1$$
$$r_0+r_0k=r_1k_1-1$$

by the very definitions (16) and (17).

But then also (for the third group) :

$$r_0+r_0k+r_0k^2=r_1k_1^2-1$$

since both sides equal $3r_0+3r_0^2+r_0^3$.

(ii) ⇒ *(i)*

Here we have that, since necessarily $p \geq 3$ (otherwise there is no "law"), we must have :

$$r_0=r_1-1 \tag{18}$$
$$r_0+r_0k=r_1k_1-1 \tag{19}$$
$$r_0+r_0k+r_0k^2=r_1k_1^2-1 \tag{20}$$

From (20) we find:

$$k_1=(1+r_0+r_0k+r_0k^2)/r_1k_1,$$

which is, by (19), equal to :

$$k_1=(1+r_0+r_0k+r_0k^2)/(1+r_0+r_0k) \tag{21}$$

Also, by (19),

$$k_1=(1+r_0+r_0k)/r_1,$$

which is, by (18), equal to :

$$k_1=(1+r_0+r_0k)/(1+r_0) \tag{22}$$

(21) and (22) now yield :

$$k=1+r_0$$

Hence, using (8) and (10) gives $b=H=1$.

Remark that in the above case $r_1=k_1$ necessarily.

Theorem III.4: The following assertions are equivalent :

(i) $H=1$

(ii) Bradford's graphical law (group-free) is the same as Bradford's verbal law (group-free).

Proof: *(i) ⇒ (ii)*

If $H=1$, then, by (8) and (10) :

$$M = \log K \tag{23}$$

This relation guarantees that :

$$\sigma(i) = MK^i$$

if and only if:

$$\Sigma(i) = K^i - 1 \tag{24}$$

for every $i \in [0, A]$, which shows that both Bradford's graphical and verbal laws are valid (the former with $M_1 = 1$).

(ii) ⇒ (i)

If we have

$$\sigma(i) = MK^i$$

and

$$\Sigma(i) = M_1K^i - 1 \tag{25}$$

for every $i \in [0, A]$, then, since also

$$\Sigma(i) = \int_0^i \sigma(i')\, di'$$

$$\Sigma(i) = [M/(\log K)]\, K^i - M/(\log K) \tag{26}$$

we find, from (25) and (26) that

$$M_1 = 1$$

and

$$M = \log K$$

(8) and (10) then yield $b = H = 1$.

The above regularities are very coincidental. But the results of the next section are even more surprising. In case $H=1$, the graphical laws of Bradford as defined in III.1. and III.2. are the equivalents of Zipf's law.

IV. The equivalents of Zipf's law

Theorem IV.1 : For an IPP with continuous Σ, the following assertions are equivalent :

(i) The IPP satisfies the graphical formulation of Bradford's law group-dependent, for every $p \in N$, but with the relation $r_1 = k_1$.

(ii) The IPP satisfies the graphical group-free Bradford function, with $M_1 = 1$.

(iii) The IPP satisfies Brookes' law with $\beta = 1$.

(iv) The IPP satisfies Zipf's (or Pareto's) law.

In this case, we have the following relations between the parameters:

$$\alpha = 1/(\log K_1) = F \tag{27}$$

$$K_1 = k_1^{p/A} \tag{28}$$

$$r_1 = K_1^{y_0} = k_1 \tag{29}$$

Proof :

Proof of the equivalence of (i) and (ii)

(a) (i) implies (ii)

Let first $i \in [0, A]$ be such that $i = qA/p$ where $q \leq p$, $q, p \in N$, $q \geq 1$. By (i) we have, with p groups:

$$\Sigma(i) = r_1 k_1^{q-1} - 1$$

for a certain r_1, $k_1 > 1$, with $r_1 = k_1$

$$\Sigma(i) = r_1 k_1^{(pi/A)-1}$$

$$\Sigma(i) = (r_1/k_1)\ (k_1^{p/A})^i - 1$$

$$\Sigma(i) = K_1^i - 1,$$

with

$$K_1 = k_1^{p/A}. \tag{30}$$

Now k_1 is p dependent but k_1^p is p-independent. This can be seen as follows. Take two values $p \neq p'$, $p, p' \in N$.

Let $r_1 = k_1$ correspond with p and $r'_1 = k'_1$ correspond with p', according to (i). Hence:

$$r_1 k_1^{p-1} - 1 = T$$

$$r_1 k_1^{p'-1} - 1 = T$$

Hence:

$$k_1^p - 1 = T = k_1'^{p'} - 1$$

Hence:

$$k_1^p = k_1'^{p'}.$$

Consequently, K_1 as defined in (30) is independent on the particular choice of $p \in$ N. So:

$$\Sigma(i) = K_1^i - 1$$

is a fixed function of i in the set

$$\{qA/p \quad q \leq p, q,p \in N\}$$

which is dense in [0, A]. Since we suppose Σ continuous and since the function

$$i \rightarrow k_1^i - 1$$

is already a continuous extension of Σ on [0, A], we conclude that (cf. Ref. 18, where more details on functions on dense sets can be found)

$$\Sigma(i) = K_1^i - 1$$

for every $i \in [0, A]$.

(b)(ii) implies (i)

Let $p \in N$ be arbitrary. Let $y_0 = y_0(p) = A/p$ and $r_1 - 1 = r_1(p) - 1 = \Sigma(y_0) = K_1^{y_0} - 1$. Then:

$$\Sigma(2y_0) = K_1^{2y_0} - 1$$
$$= K_1^{y_0} \cdot K_1^{y_0} - 1$$

and more generally, for every $i = 2,...,p$:

$$\Sigma(iy_0) = K_1^{iy_0} - 1$$
$$= (K_1^{y_0})\,(K_1^{y_0})^{i-1} - 1$$

Hence, putting

$$r_1 = K_1^{y_0} = k_1$$

we have (i) for every $p \in N$.

Proof of the equivalence of (ii) and (iii)

(a) (ii) implies (iii)

Since

$$\Sigma(i) = K_1^i - 1,$$

for every $i \in [0, A]$, we have, using $i = R(r)$, and $\Sigma(i) = r$

$$r = K_1^{R(r)} - 1$$

Hence

$$R(r) = [1/(\log K_1)] \log (r+1),$$

being Brookes' law, for every $r \in [0, T]$, but with $\beta = 1$.

Here

$$\alpha = 1/(\log K_1)$$

(b) (iii) implies (ii)

Given

$$R(r) = \alpha \log (1+r)$$

for every $r \in [0, T]$, we have trivially :

$$R^{-1}(i) = r = e^{i/\alpha} - 1 .$$

Hence, using $\Sigma(i) = r$ again

$$\Sigma(i) = K_1^{i} - 1$$

for every $i \in [0, A]$. Here

$$K_1 = e^{1/\alpha}$$

Proof of the equivalence of (iii) and (iv)

This proof is executed using the general formula (definition of R and g) :

$$R(r) = \quad g(r')\, dr'$$

for every $r \in [0, T]$.

(a) (iii) implies (iv)

Since from the previous formula one also has $g(r) = R'(r)$, we find, using

$$R(r) = \alpha \log (1+r),$$

$$g(r) = R'(r) = \alpha/(1 + r),$$

being Zipf's law.

(b) (iv) implies (iii)

Now we have

$$R(r) = \int_0^r \alpha/(1 + r')\, dr'$$
$$R(r) = \alpha \log(1+r)$$

Here we have the relation: $\alpha = F$

This completes the proof of the theorem.

Remark

Although Zipf's law (hence also Pareto's law) is a special case of the classical informetric laws (such as Mandelbrot), it is clear that we only encounter Zipf's law in highly concentrated(cf. Refs. 4, 10) cases. Indeed, from (8) and (10) it follows that if $H=1$,

$$k = 1 + r_0 \qquad (31)$$

From this we can draw the conclusion that we are dealing here with a highly concentrated situation in the sense that *or* r_0 is small, which is a way of saying that (take $p=3$ to fix the ideas) the core group of highly produced sources is small, *or* r_0 is large but then, according to (31), k must be large and hence, the core group of r_0 sources is nevertheless small w.r.t. the other groups r_0k, r_0k^2 and so on.

We conclude that linguistics-cf Zipf's law (or econometrics - cf Pareto's law) can be viewed as part of classical informetrics, but in practice there is a separation since

1. In most informetric examples we have $b=H < 1$ and indeed $b=H << 1$.
2. In linguistics and econometrics one often finds $b=1$.

That $b<<1$ in most informetric examples can be seen in Ref 5. I know of only one bibliography with $b \approx 1$: the ORSA bibliography (cf. Ref. 11. or see again Ref. 5., where b has been calculated).

A philosophical explanation why Zipf's law represents more concentrated situations can be found in Ref. 6. Practical evidence of this is included in Ref. 4.

Conclusion

Zipf's law nor Pareto's law fit in the classical definition of the graphical law of Bradford. If, however we lower the ranks in the latter law with one we found that:

- Zipf's law (or Pareto's law) fits in the graphical law of Bradford as well as in the verbal law of Bradford;
- and that the latter property is only valid if Zipf's law (or Pareto's law) is valid.

This shows the unique and surprising place of Zipf's and Pareto's law.

References

1. A. BOOKSTEIN, Robustness properties of the bibliometric distributions, I and II, to appear in *Journal of the American Society for Information Science.*
2. S.C. BRADFORD, Sources of information on specific subjects. *Engineering*, 137 (1934) 85-88. (Reprinted in *Journal Information Science* 10 (4) (1985) 176-180.
3. B.C. BROOKES, Numerical methods of bibliographical analysis, *Journal of Documentation*, 29 (1973) 18-43.
4. L. EGGHE, Pratt's measure for some bibliometric distributions and its relation with the 80/20 rule, *Journal of the American Society for Information Sciences*, 38 (4) (1987) 288-297.
5. L. EGGHE, Applications of the theory of Bradford's law to the calculation of Leimkuhler's law and to the completion of bibliographies. To appear in *J. Amer. Soc. Inf. Sci.*
6. L. EGGHE, On the classification of the classical bibliometric laws, *Journal of Documentation*, 44 (1) (1988) 53-62.
7. L. EGGHE, *The Duality of Informetric Systems with Applications to the Empirical Laws*, Ph. D. Thesis, City University, London (UK), 1989.
8. L. EGGHE, The duality of informetric systems with applications to the empirical laws, *Journal of Information Science*, 16 (1990) 17-27.
9. L. EGGHE, A note on different Bradford multipliers, *Journal of the American Society for Information Science*, 41 (1990) 204-209.
10. L. EGGHE, R. ROUSSEAU, Elements of concentration theory. In: *Informetrics 89/90. Proceedings of the 2nd International Conference on Bibliometrics, Scientometrics and Informetrics*, London (Canada), 1989, L. EGGHE, R. ROUSSEAU (Eds), Elsevier, Amsterdam, 1990, pp. 97-137.
11. M.G. KENDALL, The bibliography of operational research. *Operational Research Quarterly*, 11 (1960) 31-36.
12. F.F. LEIMKUHLER, The Bradford distribution. *Journal of Documentation*; 23 (3) (1967) 197-207.
13. B.B. MANDELBROT, *The Fractal Geometry of Nature*, W.H. Freeman and Cy, New York, 1983.
14. R. ROUSSEAU, Relations between continuous versions of bibliometric laws. To appear in *J. Amer. Soc. Inf. Sci.*
15. H. THEIL, *Economics and Information Theory*, North-Holland, Amsterdam, 1967.
16. E.A. WILKINSON, The Bradford-Zipf distribution. OSTI-report # 5172. University College, London (UK), 1973.
17. G.K. ZIPF, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge, 1949 (reprinted: Hafner, New York, 1965).
18. T.M. APOSTOL, *Mathematical Analysis*, Addison-Wesley, Reading, MA, 1974.