AN ATTEMPT TO MEASURE THE NUMBER OF THE TOTAL VOCABULARY OF THE INDIVIDUAL BY A STATISTICAL ANALYSIS OF A SAMPLE OF SPEECH OR WRITING

David Sydney Mackay Kerr

A Dissertation Submitted to the Faculty of Education
University of the Witwatersrand, Johannesburg
for the Degree of Master of Education

Johannesburg 1980

ABSTRACT

The development of vocabulary tests is analysed and criticised.

A new method of measurement is proposed based on an analysis of a sample of recorded speech or writing from an individual.

The individual's vocabulary is defined as a fixed set of words having a probability of occurrence associated with each member of the set, and distributed according to a mathematical function, not necessarily specified.

A theorem is derived which expresses the vocabulary contained in a text as a function of the frequency distribution and the length of text.

Four forms of the text are quoted depending on whether the frequency distribution is taken as continuous or discrete and the text length as a natural or a real number.

A variety of functions was used in an attempt to estimate the number of words known by the individual using a running text length on the vocabulary contained in that text. Composition books from six high school children were used. No success was obtained.

A better theorem was obtained from Sichel[9] which related the number of words appearing r times each in a text N words long with the frequency distribution of occurrence and the length of text.

Two large texts from two children, two and a half years old each, and one very long text more than 13 000 words long, from a fourteen year old girl were analysed at accumulative intervals. The analysis gave length of text analysed, number of words appearing once, twice, etc. and vocabulary contained in text.

Both the texts by one of the younger children and one of the texts of the other younger child were analysed in the forward direction, i.e. chronological order, then in reverse. They were randomised then analysed in the forward and backward directions.

Estimates of the number of the individual's vocabulary were made using the number of words appearing once and the average number of times the words, in the sample, appeared (first cell and mean method).

Acceptable estimates resulted but a drawback was that the estimate of vocabulary showed an upward drift as a sample size (i.e. text length) increased. This was corrected by making an asymptotic fit.

The method has definite possibilities as a psychometric tool.

TO MY WIFE SHIRLEY AND MY CHILDREN ANGUS, LAURIAN AND DUNCAN WHOM I ADORE.

Acknowledgements

The author wishes to express his grateful thanks to his supervisors Professor Gourlay and Professor Kerrich, both of whom have retired during the course of the project. The author is especially grateful to his current supervisor, Professor Sichel of the Graduate School of Business Administration who not only knew what the author was talking about but even more so.

To Roy and Connie Page-Shipp and Peter and June Mitchell for recording their daughters' conversations under what must have been very trying circumstances.

To Sandré Cruse for writing more than thirteen thousand words.

To the late Mr R P Rees for allowing his pupils' composition books to be used.

Lastly to Mrs El[illegible] Pearson and Mrs Annette van der Walt for typing the script.

_ABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 – INTRODUCTION

1.1 The Start

LECTURER: 'This system of teaching reading is based on the fact that a child of five and a half, when he comes to school for the first time has a vocabulary of not more than five hundred words.'

STUDENT: 'Can you give me any reference as to how the figure of five hundred was arrived at?'.

LECTURER: 'When you've had as much experience of little children as we have, you will know that a child of five and a half does not have a vocabulary of more than five hundred words.'

This is a verbatim quotation of what took place during a lecture at a College of Education in 1965.

The conversation is significant in that it illustrates a common phenomenon in Education, that is the quotation of numbers without any mention of the modus mensurandi; indeed in the case of vocabulary, an attempt to discover the modus mensurandi led to the present investigation.

Two methods have been used by Educationists and Psychologists. They are:

1.2 Sampling the dictionary

Terman[1] used a dictionary containing 18 000 words and selected the last word in each successive sixth column; this resulted in a sample of 100 words.

Terman's test required the tester to interview the testee and ask the testee what the word meant. He gives a selection of acceptable and non-acceptable definitions for the guidance of testers, including 'full credit' and 'half credit'.

The estimate of the testee's vocabulary was then made using the equation:

$$\frac{\text{no. of correct answers}}{\text{no. of words in test}} = \frac{\text{vocabulary}}{\text{no. of words in dictionary}} \quad \ldots\ldots\ldots \ (1)$$

Substituting Terman's values in (1) gives

$$\frac{\text{no. of correct answers}}{100} = \frac{\text{vocabulary}}{18\ 000}$$

$\therefore$ Vocabulary = no. of correct answers x 180

Later the multiple choice synonym test was developed and is the most common in current use.

In this type of test the word appears underlined in a short sentence and alongside it appears a list of four or five possible synonyms. The person under test then has to select the synonym which he thinks is correct.

Generally therefore, if the individual scored v words correct in a sample of N words taken from a dictionary containing M words then by simple proportion:

$$\frac{v}{N} = \frac{x}{M} \qquad \ldots\ldots\ldots\ldots\ (2)$$

where x = the individual's vocabulary.

Here it must be noted that this method cannot estimate the vocabulary as a list of words i.e. the individual's own dictionary but it is purely and simply a number.

Solving equation (2) for x gives:

$$x = v\,\frac{M}{N} \qquad \ldots\ldots\ldots\ldots\ (3)$$

and since v is a natural number, the possible values of x are natural multiples of $\frac{M}{N}$.

1.2.1 Defects of Sampling

This method has the following inherent defects:

(a) Satisfying the tester in a multiple choice test does not guarantee that the person under test knows the word sufficiently well to be able to use it correctly.

(b) The multiple choice method of assessment requires that the testee knows the word, the synonym and that they are synonyms.

(c) The multiple choice method can be used just as effectively to test the synonym in the list as it is to test the word underlined in the sentence.

(d) Using a large dictionary and a number of words in the test of a few hundred will result in a population with vocabularies which are multiples of say a thousand or more.

(e) The method is of no use with illiterates and pre-school children.

(f) If the words in the test are graded for difficulty, testees may either stop when they have reached their limit or may continue by guessing until they finish the test. In this case the question of answers correct due to guessing must be raised.

(g) The compilers of the dictionary are taken as the arbiters of which words are permitted the individual. No allowance can be made for special 'family' words or words from another language.

(h) Depending on which dictionary is used, Hartman[2] 1941 found that current methods of estimating vocabulary size 'fit only the dictionaries upon which they are based'.
Four lists from different dictionaries administered to 100 college students showed large variations in vocabulary size.

Four lists from a single dictionary administered to normal school graduates yielded similar results and serial order lists administered as a check showed a high degree of agreement all indicating mean vocabularies in excess of 200 000 words.

'There is only one conclusion possible - the average undergraduate has an actual vocabulary of a magnitude conventionally deemed unbelievable.'

1.2.2 Advantages of Sampling

The multiple choice method has the following advantages:

(a) It is easy to administer and mark. It does not require any special skill or training on the part of the testor.
(b) Large numbers of testees can be tested at the same time.
(c) The test may be administered in a short time.
(d) The same test can be used over a wide range of vocabularies.
(e) Using a separate answer sheet, test papers or booklets may be re-used.

1.3 Exhaustion

The second method is an attempt to 'record to exhaustion'. In this method the individual's utterances are recorded over a finite period of time and it is assumed that during the recording period the individual exhausts his store of words.

Due to the obvious difficulty of determining the point of exhaustion the method has been adapted so as to record the common vocabulary of a group of children. Burroughs[5] recorded children's conversations either by hidden microphones on school playgrounds or by interviews with individual children.

After recording, the words were then ranked in popularity, as reflected by the number of times they appeared in the sample.

A criterion was then established for discarding words assumed not to be common to the group, the usual criterion being that words used by two members or less were deemed to be not common to the group and were discarded.

The obvious defect of this method is that the probability of a rare word appearing increases with the size of the sample of words taken from the individual and it is therefore extremely unreliable to set up such arbitrary criteria as flagging interest on the part of the children. This method of recording, although of interest, is not relevant to this study.

1.4 <u>Others</u>

Other recorders (4) and (5) took down every word used by a child during a week or a day but did not use the data to try and estimate the child's total vocabulary.

CHAPTER II - A NEW APPROACH

2.1 A New Approach

In view of the disadvantages and unreliability of current vocabulary tests it was decided that a new approach be explored.

2.2 Criteria

The following criteria were assumed:

(i) One individual at a time should be investigated.

(ii) Nothing extraneous to the testee's experience or lifestyle should be used.

(iii) The testee should not know that he/she was being tested or even studied.

(iv) The words recorded, tested or used should be the testee's own.

It was assumed, concerning any individual, that at any one instant

(i) he/she possessed

a finite vocabulary

(ii) with each item of vocabulary was associated a finite probability of occurrence

(iii) in any finite sample of words recorded from the individual, in order, repetitions included, lay the evidence for an estimate of the total content of vocabulary.

From the foregoing it is obvious that as an individual produces words from his/her vocabulary, he/she naturally repeats more and more of the words already used, i.e. that the probability of the next word being a new appearance is monotonic decreasing with increasing sample size.

2.3 Two Variables

Even simpler, instead of thinking about rates it was decided to define two variables:

(i) total words used, t

(ii) the vocabulary contained in t, $v(t)$

Again it is obvious that as the individual uses more and more words so the vocabulary contained in the words used can either increase or remain the same.

Thus $v(t)$ is a positive monotonic increasing function of t, bounded above, because even if the individual under test knew all the words in several languages and several dictionaries in each language, his/her vocabulary would yet be finite.

Expressed mathematically: $v(t) = f(t)$ and $\lim_{t \to \infty} v(t) = V$

See Figure 1 on page 9.

Figure 1 shows the graph of $v(t)$ vs t as a series of discrete points. Bearing in mind that t and $v(t)$ are both natural numbers this is a true representation of the co-ordinates. For convenience, later, a smooth curve will be used.

2.4 Recording

As a pilot survey, essay books were borrowed from six high school pupils. The pupils were selected so as to give the greatest possible spread of results. One normal, average, well adjusted, middle of the road, steady worker, one highly intelligent but unpredictable quasi-neurotic and four adaptation class members, selected to contrast with the other two.

English composition books were borrowed with the pupil's consent and the individual's dictionary was compiled. Each word was checked with the dictionary to see whether it, th[illegible] word, had appeared before or was a first appearance. If the word had been used before one was added to t and $v(t)$ remained constant. If the word was a first appearance, one was added to t and one to $v(t)$.

2.5 Lemmatisation

The recordings were made in 1958 and at that date no rules for lemmatisation were available and arbitrary decisions were made, the general rule being 'new idea = new word'. Regular derivations by the addition of ed, ing, ation etc., were not taken as 'new words' unless some special meaning was usually attached to the derived form of the word.

At this stage it is not really important whether the words are lemmatised or not because each person will have his/her own lemma ratio and after lemmatising the sample, a more or less fixed ratio of vocabulary to lemmas will be found. In fact this lemma ratio may well provide an interesting field for research.

The results of the word counts at intervals of 100 upwards and a description of the subjects follows: See Table 1 on pages 7 and 8.

2.6 Data

TABLE 1. 6 CHILDREN v(t) vs (t)

André : Age 14 years
Background : Lower middle class
Environment : Suburban
IQ : Normal
Personality : Friendly, reliable (class monitor)

t	100	200	300	400	500	600	700	800	900	1000
v(t)	71	124	154	198	[illegible]	[illegible]	305	325	360	389
t	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
v(t)	411	445	477	511	536	556	581	599	612	630

Jane : Age 14 years
Background : Professional
Environment : Suburban
IQ : Well above normal
Personality : Highly strung, unreliable

t	100	200	300	400	500	600	700	800	900	1000
v(t)	66	107	149	201	246	284	324	356	380	411
t	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
v(t)	440	466	501	531	560	590	614	647	671	689
t	2100	2200	2300	2400	2500	2600	2700	2800	2900	
v(t)	707	734	754	779	794	803	847	874	898	

Noel : Age 13 years
Background : Suburban Institution
IQ : Below normal
Personality : Diffident, defensive

t	100	200	300	400	500	600	700
v(t)	61	112	160	202	238	257	287

[illegible]ABLE 1 [illegible]

[illegible] : [illegible]

[illegible] : Suburban Institution

[illegible] : Below normal

[illegible] : [illegible]

[illegible]	100	[illegible]	300	[illegible]	[illegible]	600	700	800
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	256	287	311

[illegible] : [illegible]

[illegible] : [illegible]

[illegible] : [illegible]

[illegible] : [illegible]

[illegible] : [illegible]

[illegible]	100	200	300	400	500	600	700	800	[illegible]	1000
[illegible](t)	67	118	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	430

[illegible]	1100	1200
[illegible]	[illegible]	[illegible]

[illegible] : [illegible]

[illegible] : [illegible]

[illegible] : [illegible]

[illegible] : [illegible]

[illegible]	100	200	300	400	500	600	700
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible] Composite Graph

[illegible] on the same set of axes [illegible]

[illegible] curves after only [illegible]

[illegible] curve was given by Malcolm [illegible] enquiries revealed [illegible] for sociological

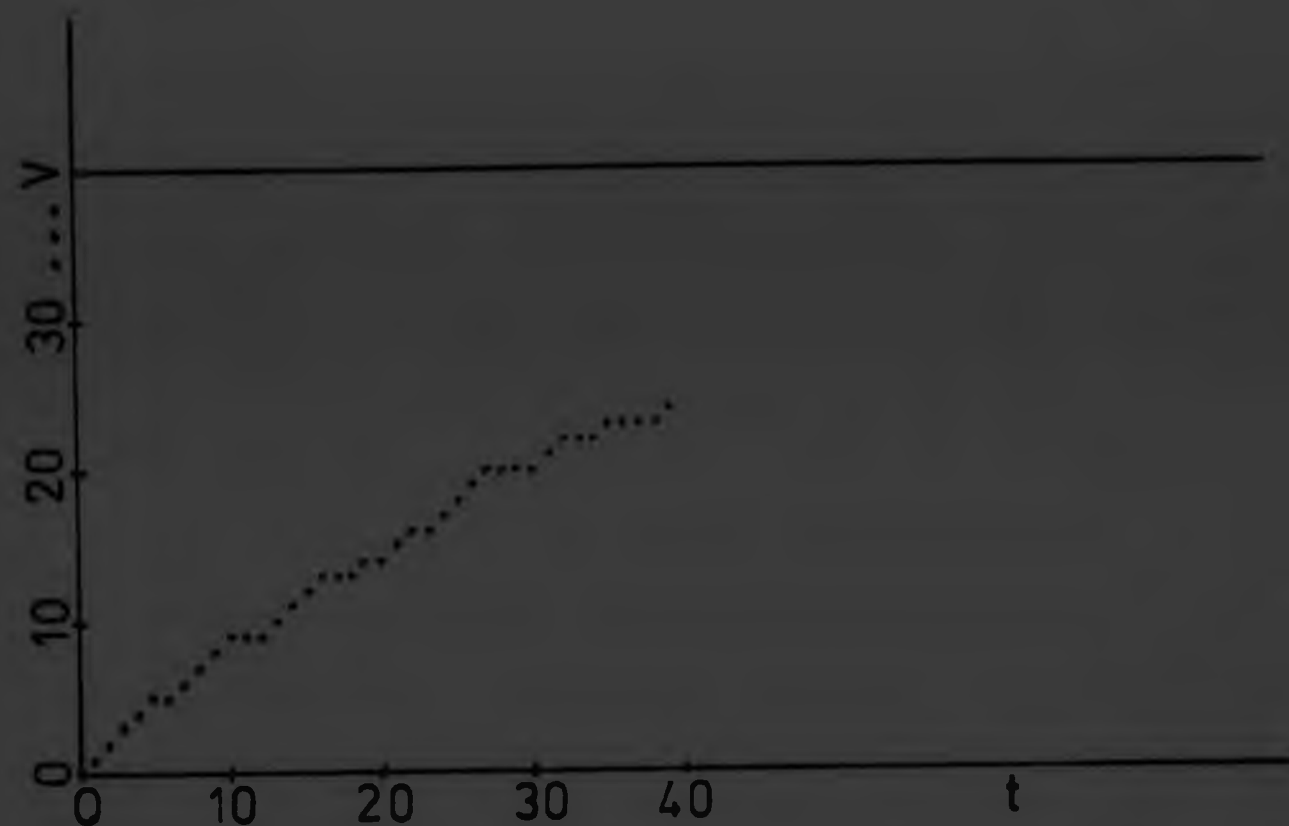


FIG 1 Graph of v(t) vs t (discrete points)

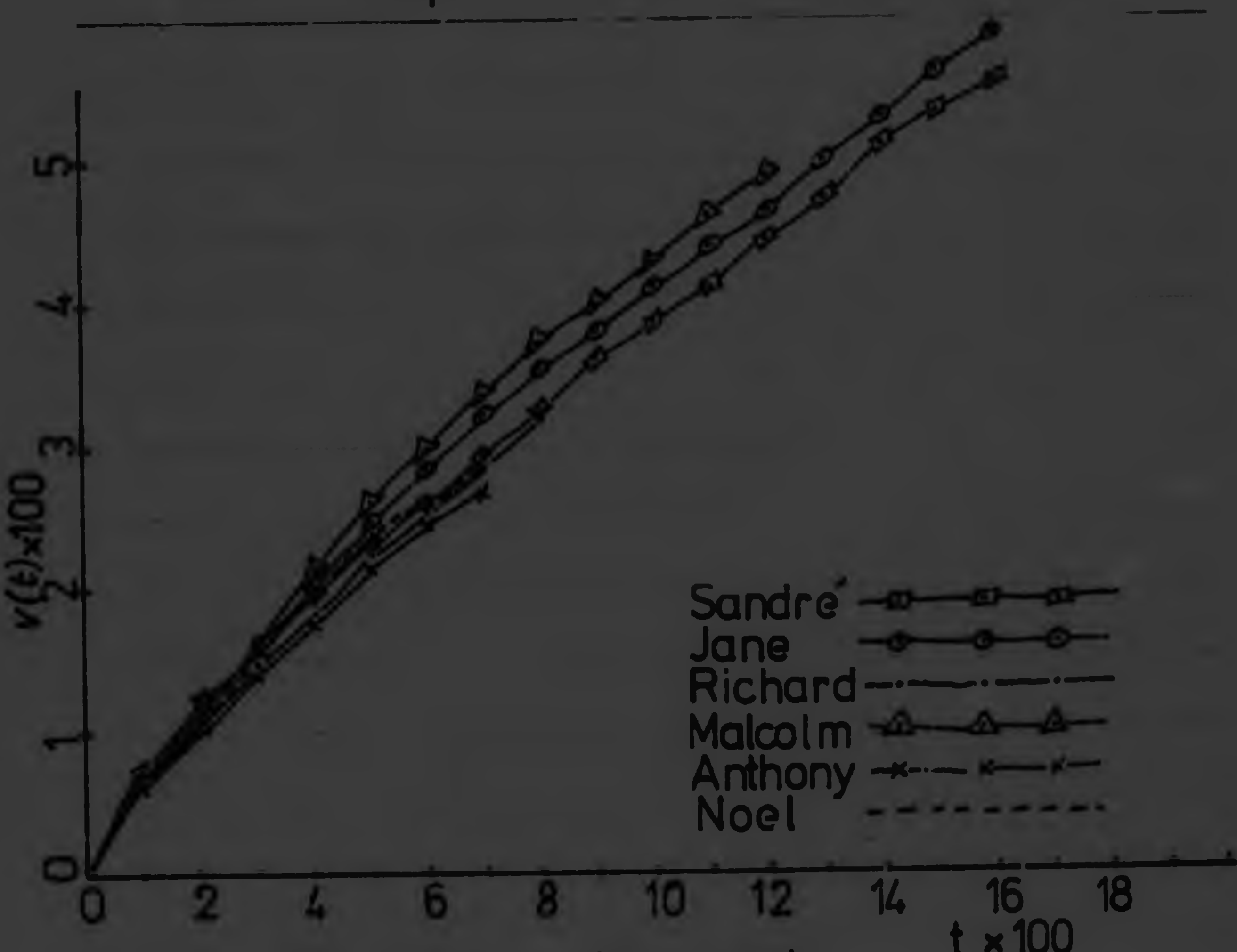


FIG 2 Composite graph

reasons.

As a preliminary investigation the results of this pilot study showed sufficient promise to permit further work.

2.8 No Model

The absence of a mathematical model to describe the data caused considerable delay because it was realized that as the graphs would have to be extrapolated, the criterion of mere 'goodness of fit' would be, in fact, no criterion at all.

To quote Kerrich J.E.[6] "Give me any set of points, boy, and I'll fit you any one of [illegible] models, each one as good a fit as the others; the question is 'which one and why'?"

2.9 General Properties of any Model

The following general properties of any model must be adhered to:

1. $v(t)$ as a function of t, must be bounded above.
2. The gradient of $v(t)$ vs t must be monotonic decreasing with limit zero.
3. $v(1) = 1$, i.e. the first word used must be a first appearance, for $t = 1$, $v(1) = 1$.

2.10 Probabilities

In search for a suitable model it was realised that the individual's vocabulary was purely and simply a set of probabilities and that when a person spoke or wrote, he/she was simply sampling from his/her set of probabilities.

The details of the derivation are given in the next chapter.

CHAPTER III

THE FUNDAMENTAL VOCABULARY THEOREM

3.1 Finite Number of Words

The v bulary of an individual is a finite number of words V.

3.2 A Set of Probabilities

Each word appears with its own probability x, but any attempt to measure this probability for any given word, used by the individual, using the ratio

$$x = \frac{\text{no. of appearances}}{\text{no. of words in sample}}$$

will give a set of varying results distributed about a mean. In the case of words appearing only once in a sample, the true probability of occurr can only be guessed. For this reason a continuous distribution of probabilities is assumed, as an adequate approximation.

3.3 Distribution of Frequencies

Such a frequency distribution is shown in Figure 3 on page 13.

The frequency function $f(x)$ is defined as follows:

$f(x)\,\delta x$ = the number of words having frequency between x and $x+\delta x$

See Figure 4 on page 13.

Any such distribution must have the following properties:

3.3.1 $\int_0^1 f(x)dx = V$ .... (4)

The area under the $f(x)$ curve: i.e.

$\int^1 f(x)dx = V$ = total vocabulary

3.3.2 $\int_0^1 xf(x)dx = 1$

The sum of the products of the probabilities and the number of words having those probabilities, i.e.

$\int_0^1 xf(x)dx = 1$ .... (5)

3.4 v(t) A Function of t - Derivation

After a total of t words has been used, this sample, t, will contain a number $v(t)$ of words from the vocabulary. The vocabulary $v(t)$ contained in t will consist of words of varying probability.

$v(x,t)\delta x$ is defined to be the number of words with probability between x and $x+\delta x$, brought out in a sample of t words.

In general, the vocabulary $v(t)$, which is the total of all different words in t, i.e. $v(t) = \int_0^1 v(x,t)dx$, will be drawn initially from:

1. Words of higher probabilities i.e. large x, and
2. Those δ intervals with a large ordinate i.e. populous intervals.

See Figure 5 on page 13.

v(x) and f(x) are plotted together in Figure 6 on page 21.

For the delta strip shown the shaded area represents the number of words having probability between x and $x+\delta x$, remaining in the individual's vocabulary (or word store).

Taking a further δt words from the individual will bring out more words from the delta strip.

The increase in $v(x,t)\delta x$ namely $\delta\left[v(x,t)\ \delta x\right]$ will depend on:

1. The number of words in the delta strip not yet used, i.e. $\left[f(x) - v(x,t)\right]\delta x$
2. The probability, x, with which each of the words in the delta strip appears.
3. δt, provided δt is small.

This gives the difference equation

$$\delta\left[v(x,t)\delta x\right] = x\delta t\left[f(x)\delta x - v(x,t)\delta x\right] \qquad \text{........ (6)}$$

but by definition

$$\delta\left[v(x,t)\delta x\right] = v(x,t+\delta t)\delta x - v(x,t)\delta x \qquad \text{........ (7)}$$

$$\therefore\ v(x,t+\delta t)\delta x - v(x,t)\delta x = x\delta t\left[f(x)\delta x - v(x,t)\delta x\right]$$

For the delta strip shown the following are constants:

1. x
2. δx
3. $f(x)\delta x$
4. $xf(x)\delta x$

The variables are:

1. t
2. δt
3. $v(x,t)$

Transposing and collecting gives:

$$v(x,t+\delta t)\delta x - v(x,t)\delta x + v(x,t)x\delta x\delta t = xf(x)\delta x\delta t \qquad \text{......... (8)}$$

Since t is a natural number, the lowest value of δt is 1

$\therefore$ (8) becomes

$$v(x,t+1)\delta x = v(x,t)\delta x\ (1-x) + xf(x)\delta x \qquad \text{......... (9)}$$

putting t = 0 in (9) gives $v(x,0)\delta x\ (1-x) = 0$

$$\therefore\ v(x,1)\delta x = xf(x)\delta x \qquad \text{........ (10)}$$

Which is not surprising because this is a statement of the probability that the first word will appear from the delta strip, f(x), x

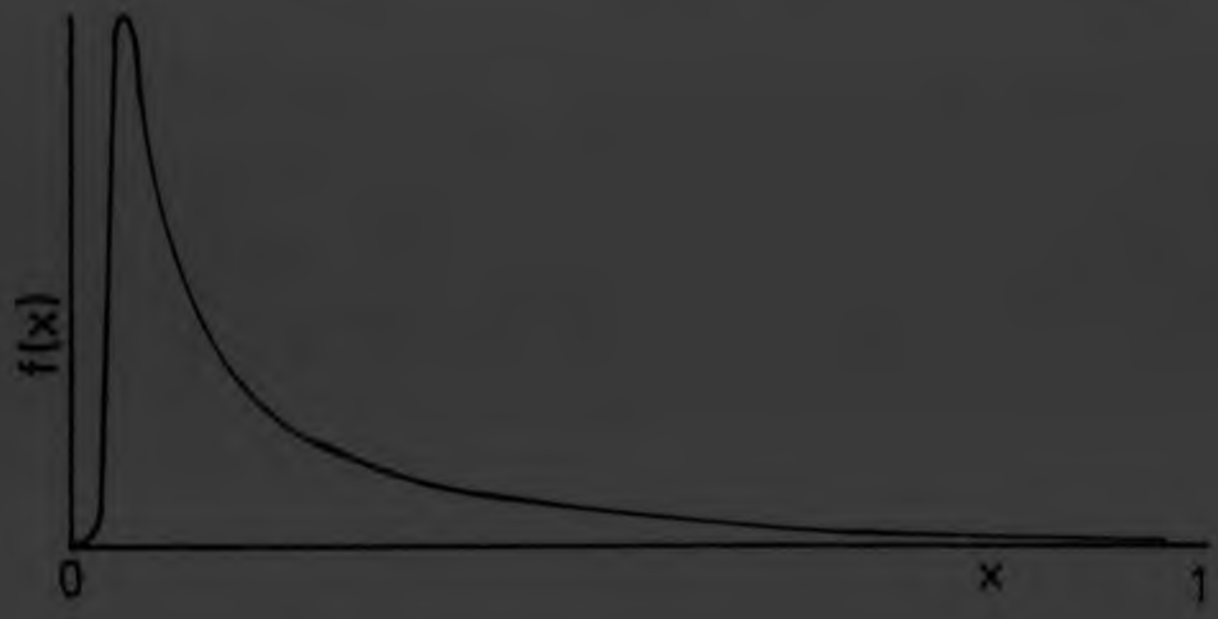


FIG 3 Distribution of word occurrence probabilities

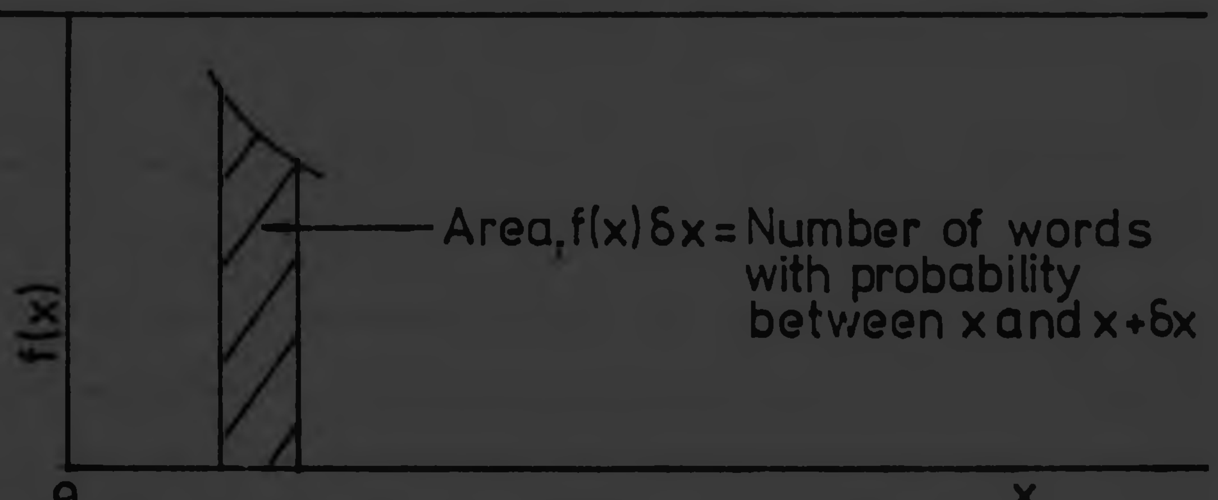


FIG 4 Delta strip of probabilities

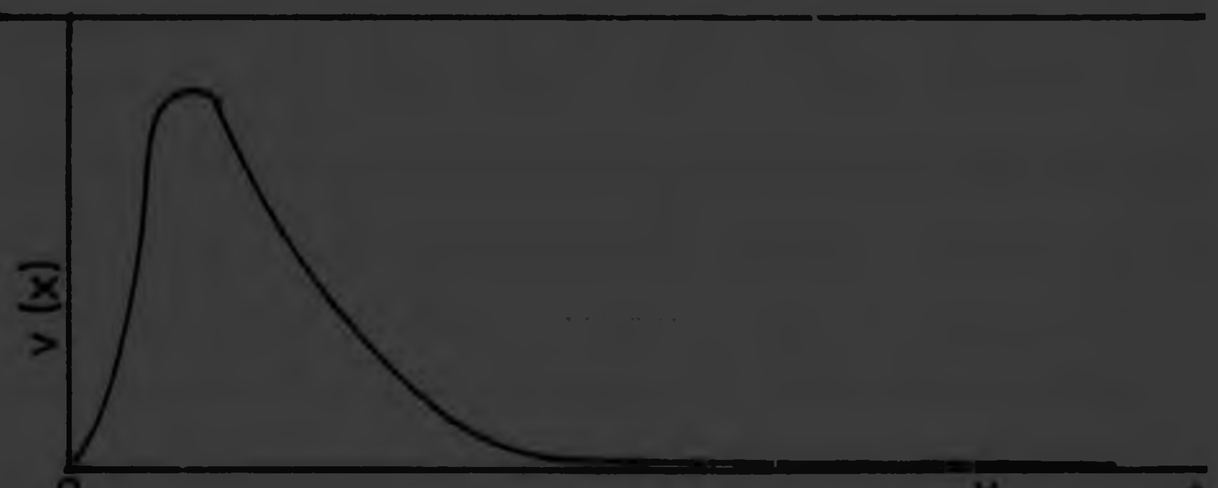


FIG 5 Graph of v(x) vs x

Putting t = 1 in (9) gives

$$v(x,2)\delta x = v(x,1)\delta x(1-x) + xf(x)\delta x$$

substituting for $v(x,1)\delta x$ from (10) gives

$$v(x,2)\delta x = xf(x)\delta x(1-x) + xf(x)\delta x \qquad \ldots\ldots\ldots (11)$$

putting t = 2 in (9) gives

$$v(x,3)\delta x = v(x,2)\delta x(1-x) + xf(x)\delta x$$

substituting $v(x,2)\delta x$ from (11) gives

$$v(x,3)\delta x = \left[xf(x)(1-x)\delta x + xf(x)\delta x\right](1-x) + xf(x)\delta x \qquad \ldots\ldots (12)$$

$$v(x,3)\delta x = xf(x)\delta x\left[(1-x)^2 + (1-x) + 1\right]$$

continuing to increase t gives, for t = t

$$v(x,t)\delta x = xf(x)\delta x\left[(1-x)^{t-1} + \ldots\ldots (1-x) + 1\right] \qquad \ldots\ldots\ldots (13)$$

Summing the terms of the G.P. in the square brackets gives:

$$v(x,t)\delta x = xf(x)\delta x\,\frac{(1-x)^t-1}{(1-x)-1}$$

$$\text{and } v(x,t)\delta x = f(x)\delta x\left[-(1-x)^t + 1\right]$$

$$= f(x)\delta x - (1-x)^t f(x)\delta x$$

Now letting $\delta x \rightarrow 0$ and summing for all the delta strips, gives:

$$\int v(x,t)dx = \int f(x)dx - \int (1-x)^t f(x)x$$

$$\therefore \quad v(t) = V - \int(1-x)^t f(x)dx \qquad \ldots\ldots\ldots (14)$$

3.4.1 t, A Continuous Variable

If instead of taking t as a natural number, t is taken as a continuous variable then (6) i.e.

$$\delta\left[v(x,t)\delta x\right] = x\delta t\left[f(x)\delta x - v(x,t)\delta x\right] \qquad \ldots\ldots\ldots (15)$$

can be solved by integration as follows:

$$\frac{-\delta\left[v(x,t)\delta x\right]}{f(x)\delta x - v(x,t)\delta x} = -x\delta t \qquad \ldots\ldots\ldots (16)$$

and since for any given delta strip, x, f(x) and δx are constant, equation (16) becomes:

$$\int\frac{-d\left[v(x,t)\delta x\right]}{f(x)\delta x - v(x,t)\delta x} = -x\int dt \qquad \ldots\ldots\ldots (17)$$

$$\therefore \ln\left[f(x)\delta x - v(x,t)\delta x\right] = -xt + C \qquad \ldots\ldots\ldots (18)$$

C is evaluated by putting t = 0 in (18)

$$\therefore \ln(f(x)\delta x) = C$$

and substituting this value for C in (18) gives:

$$\ln\left[f(x)\delta x - v(x,t)\delta x\right] = -xt + \ln\left[f(x)\delta t\right]$$

$$\ln\left[\frac{f(x)\delta x - v(x,t)\delta x}{f(x)\delta x}\right] = -xt$$

$$\therefore \frac{f(x)\delta x - v(x,t)\delta x}{f(x)\delta x} = e^{-xt}$$

$f(x)\,\delta x - v(x,t)\delta x \;=\; f(x)\delta x e^{-xt}$

Letting $\delta x \to 0$ and summing for all the delta strips

$$\int_0^1 f(x)dx - v(t) \;=\; \int_0^1 e^{-xt} f(x)dx$$

$$\text{i.e. } v(t) \;=\; V - \int_0^1 e^{-xt} f(x)dx \qquad \dots\dots\dots (19)$$

3.4.2 Two Forms of the Theorem. The theorem thus takes two forms.

a. For t a natural number

$$v(t) = V - \int_0^1 (1-x)^t f(x)dx \qquad \dots\dots\dots (20)$$

and

b. For t a continuous variable

$$v(t) = V - \int_0^1 e^{-xt} f(x)dx \qquad \dots\dots\dots (21)$$

3.4.3 Two Further Forms of the Theorem

But should the correct probability of occurrence for each word in the individual's vocabulary be known, the integral signs would be replaced by sigma summation signs and two further forms of the theorem be obtained.

c. For t a natural number and a discrete distribution of probability

$$v(t) = V - \sum_{i=1}^{V} (1-x_i)^t n(x_i) \qquad \dots\dots\dots (22)$$

where $n(x_i)$ = number of words having probability x_i .

d. For t a continuous variable and a discrete distribution of probability

$$v(t) = V - \sum_{i=1}^{V} e^{-x_i t} n(x_i) \qquad \dots\dots\dots (23)$$

If it is assumed that each word has its own unique probability of occurrence, then $n(x_i) = 1$ for all x_i

and $\sum_{i=1}^{V} (x_i) = 1$

Equation (22) becomes, for t a natural number,

$$v(t) = V - \sum_{i=1}^{V} (1-x_i)^t \qquad \dots\dots\dots (24)$$

and equation (23) becomes, for t a continuous variable,

$$v(t) = V - \sum_{i=1}^{V} e^{-x_i t} \qquad \dots\dots\dots (25)$$

Equation (24) was suggested by Kerrich[6] in 1969 and independently by Ridley[7] in 1970.

3.5 The Limiting cases of the Theorem

3.5.1 f(x) Continuous and t a Natural Number

Applying form (a) of the theorem to any practical case requires that:

$$v(0) = 0$$

$$v(1) = 1$$

and $$v(\infty) = V$$

We get $$v(0) = V - \int_0^1 f(x)(1-x)^0 dx$$

$$= V - \int_0^1 f(x)dx$$

$$= V - V = 0 \text{ by definition of } \int_0^1 f(x)dx$$

$$v(1) = V - \int_0^1 f(x) - \int_0^1 xf(x)dx$$

$$= V - V + \int_0^1 xf(x)dx$$

$$= 1 \text{ by definition of } \int_0^1 xf(x)dx$$

$$v(\infty) = V - \int_0^1 f(x)(1-x)^{\infty} dx$$

$$= V - \int_0^1 f(x)\, 0\, dx$$

$$= V \text{ which is as required}$$

3.5.2 x Discrete, t a Natural Number

The same is true for form (c) of the theorem, namely equation (22)

$$v(t) = V - \sum_{i=1}^{V} (1-x_i)^t \; n(x_i)$$

$$v(0) = V - \sum_{i=1}^{V} n(x_i)$$

$$= V - V = 0 \text{ which is as required}$$

$$v(1) = V - \sum_{i=1}^{V} (1-x_i)n(x_i)$$

$$= V - \sum_{i=1}^{V} n(x_i) + \sum_{i=1}^{V} x_i n(x_i)$$

$$= V - V + 1$$

$$= 1 \text{ which is as required}$$

$$V(\infty) = V - \sum_{i=1}^{V} (1-x_i)^{\infty} n(t_i)$$

$$= V$$

which is as required.

3.5.3 The Two Exponential Forms

However, in the forms (b) and (d), namely equations (21) and (23)

$$v(t) = V - \int_0^1 f(x)e^{-xt}dx$$

and $$v(t) = V - \sum_{i=1}^{V} e^{-x_i t}\, n(x_i)$$

giving $v(0) = 0$ which is as required

and $v(\infty) = V$ which is as required

but $$v(1) = V - \int_0^1 f(x)e^{-x}dx$$

and $$v(1) = V - \sum_{i=1}^{V} e^{-x_i} n(x_i)$$

which is not exactly 1 but as is usual in these cases the difference is minimal and the integration is not impossible.

3.6 General Applicability of the Theorem

The theorem is not only true for words but applies to any set of probabilities.

In general, let the number of members of the set be N, and the number revealed or exposed, be n (i.e. the expectation value).

Three well-known examples immediately come to mind.

3.6.1 Radio-active Decay

In this case the set of elements is a number, N, of atoms which can all decay with identical probability λ

(i) A discrete distribution,

(ii) t, a continuous variable.

$$\text{i.e. } v(t) = V - \sum_{i=1}^{V} e^{-x_i} n(x_i) \qquad \text{..........} \quad (26)$$

which then appears as

$$n(t) = N - \sum_{i=1}^{N} n(\lambda)\, e^{-\lambda t} \qquad \text{..........} \quad (27)$$

$$= N - Ne^{-\lambda t}$$

Transposing gives:

$$N - n(t) = Ne^{-\lambda t} \qquad \text{..........} \quad (28)$$

The **left hand side of (28) namely N - n(t),** the difference between the total number of nuclei involved and the number decayed, gives the number of nuclei undecayed.

Let the number undecayed be:

$$N_t = N - n(t)$$

thus

$$N_t = Ne^{-\lambda t} \qquad \text{..........} \quad (29)$$

which is a well-known result in Physics.

3.6.2 Discharge of a Capacitor

In this case we have a number of electrons N (each with a charge q) which can either leave the capacitor with probability k in any one second or remain in the capacitor with probability (1 - k).

Applying form (d) of the theorem we get:

$$n(t) = N - \sum_{i=1}^{N} n(k)\, e^{-kt} \qquad \text{.......... (30)}$$

$$= N - Ne^{-kt}$$

(because k is constant for all electrons of which there are N.)

$n(t)$ = the number of electrons which have left the capacitor.

N = total number of electrons involved.

Multiplying through by q gives:

$$n(t)q = Nq - Nq\, e^{-l}$$

now,

$\therefore$ $n(t)q$ = charge which has left capacitor

$Nq = Q$ = original charge on capacitor

thus, the charge remaining on the capacitor,

$$q_t = Q - n(t)q = Qe^{-kt} \qquad \text{.......... (31)}$$

which is a result well known in Electrical Engineering.

3.6.3 <u>Number of Faces shown by a Die</u>

Suppose a die has N faces each of which appears with probability $\frac{1}{N}$ per throw.

Form (c) of the theorem thus applies, namely:

$$v(t) = V - \sum_{i=1}^{V} (1 - x_i)^t\, n(x_i) \qquad \text{.......... (32)}$$

which becomes,

$$n(t) = N - \left(1 - \frac{1}{N}\right)^t N \qquad \text{.......... (33)}$$

(again there being only one probability, $\frac{1}{N}$ and N elements,

thus $\sum_{i=1}^{N} n(x_i) = N$

Simplifying (33) gives:

$$n(t) = N\left(1 - \left(1 - \frac{1}{N}\right)^t\right)$$

$n(t)$ being, in other words, the expected number of different faces showing after t throws. In the case of a six sided die, after no throw

$$n(o) = 6\left(1 - \left(\tfrac{5}{6}\right)^o\right)$$

$$= 6\ (1 - 1)$$

$$= 6 \times 0$$

$$= 0, \text{ as expected}$$

after one throw

$$n(1) = 6\ (1 - \tfrac{5}{6})$$

$$= 6 \times (\tfrac{1}{6})$$

$$= 1, \text{ as expected}$$

after two throws

$$n(2) = 6\ (1 - (\tfrac{5}{6})^2)$$

$$= 6 - 6 \times \tfrac{25}{36}$$

$$= 1\tfrac{5}{6}\ , \text{ as expected}$$

All these examples are special cases of the theorem, in particular where all members of the set have equal probability.

3.7 <u>Various Attempts to find an Algebraic expression for f(x)</u>

Taking the theorem in the form of equation (20) various functions were tried for f(x) namely

1. the log-normal
2. the Beta function
3. the distribution function $y = ae^{-.x-\frac{\cdot}{x}}$

3.7.1 <u>Criteria for Acceptability of Fit</u>

The criterion for acceptability could not be taken as goodness of fit because as has been pointed out in Chapter II, it is relatively easy to get a good fit to curves as smooth as those in Figure 2 with a large number of functions. It was therefore decided to estimate V by an iterative least squares method, not for the complete sample but for a partial sample of the first few hundred words and then re-estimate V for a larger and larger fraction of the total sample.

3.7.2 <u>No Drift</u>

If the function f(x) is to be accepted, then for increasing t, there should be no drift in the estimate of V, but the successive estimates should approach closer and closer to the true V. This should also be true for any other parameters of the distribution function. See Figure 7 on page 21.
By this criterion all three distributions proved to be completely unacceptable.

3.7.3. A Sudden Influx

Sichel[8] suggested that, because of the non-reversibility of the v(t) vs t observed function, if a sudden influx of new word took place due to a topic change for example, then the sudden increase in v(t) would raise v(t) and keep it raised, and only slowly would the graph return to normal.

See Figure 8 on page 22.

A sudden influx of new words at A would quasi-permanently raise the graph.

3.7.4. Differentiation

Sichel[8] suggested that if the function be differentiated and fitted to the differences between the v(t) for successive hundreds of t, then the sudden influx would be localised and its influence minimised.

See figure 9 on page 23.

This procedure proved equally unsatisfactory.

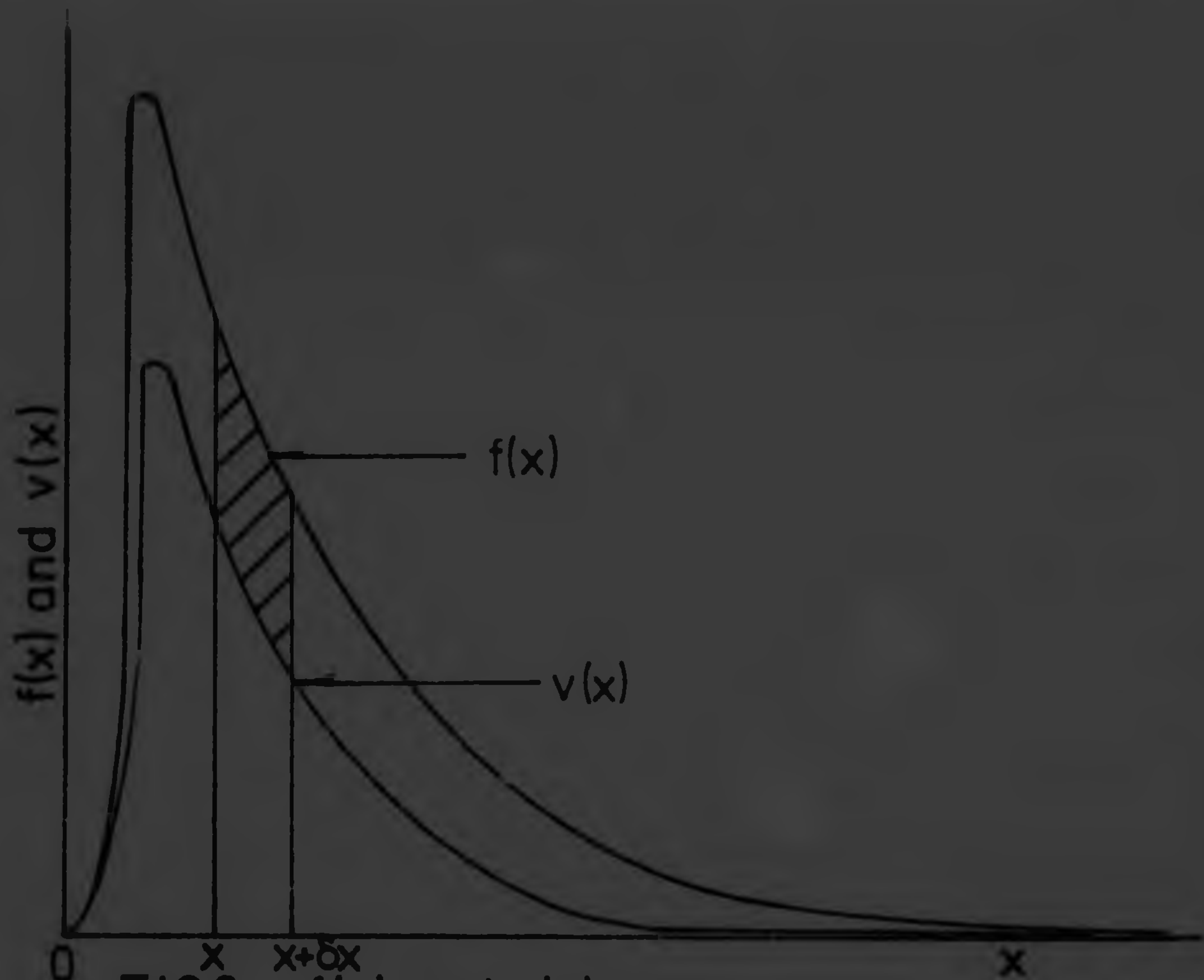


FIG 6 f(x) and v(x) vs x

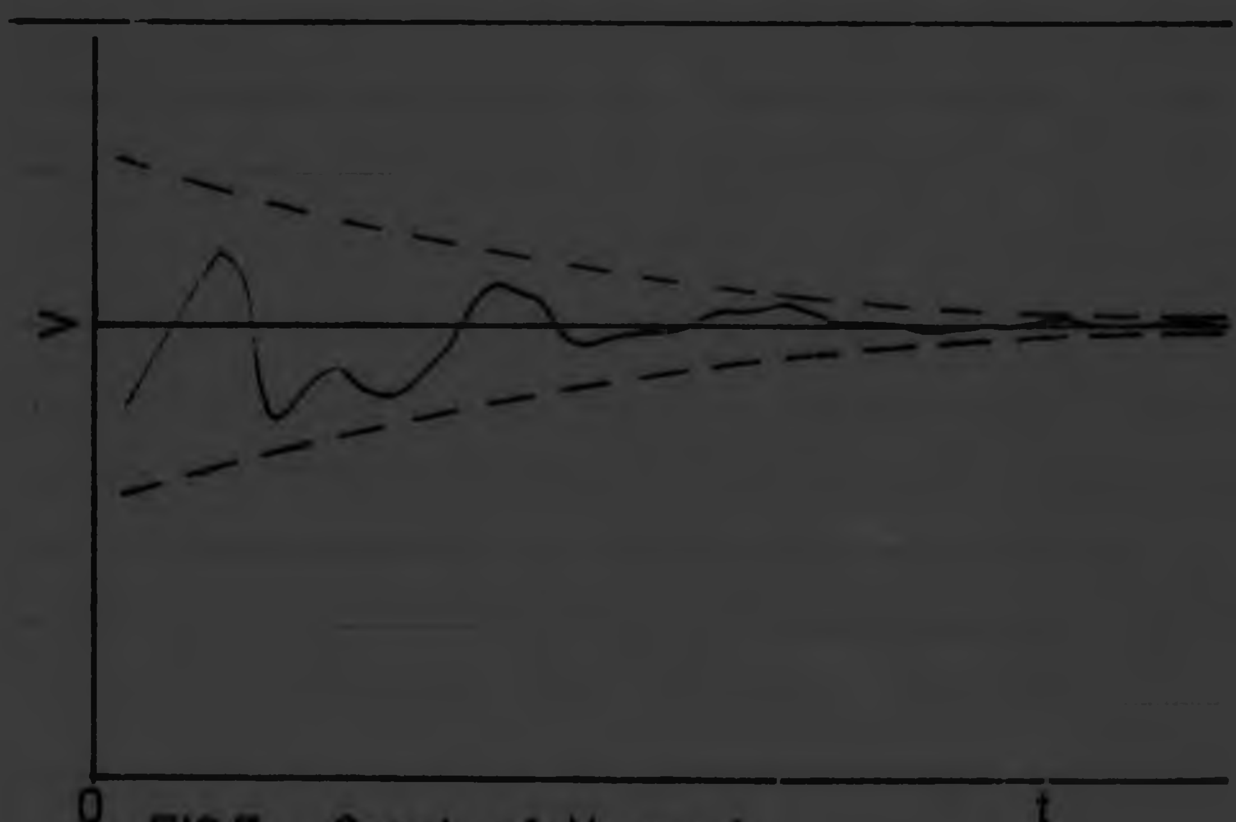


FIG 7 Graph of V vs t

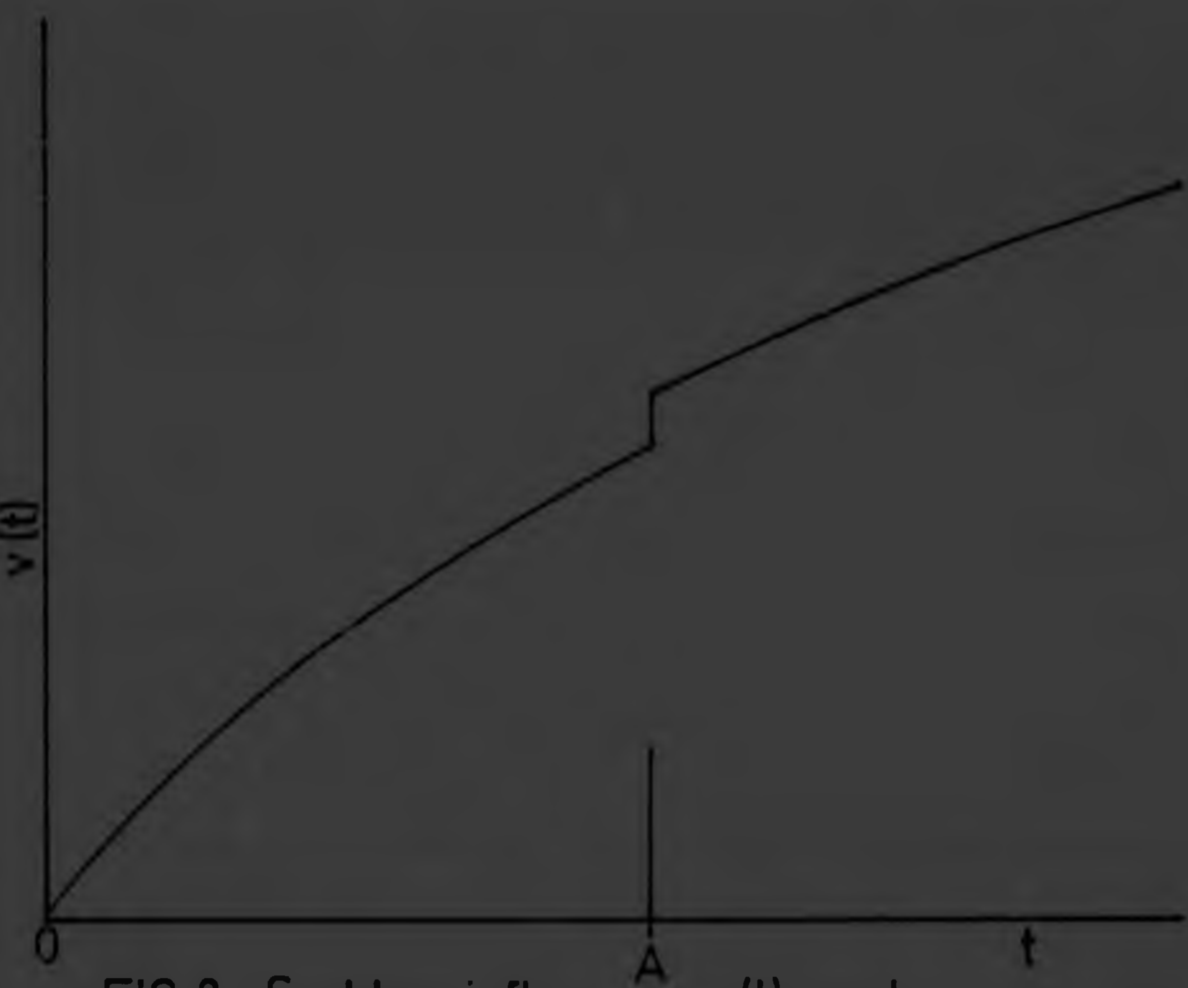


FIG 8 Sudden influx v(t) vs t

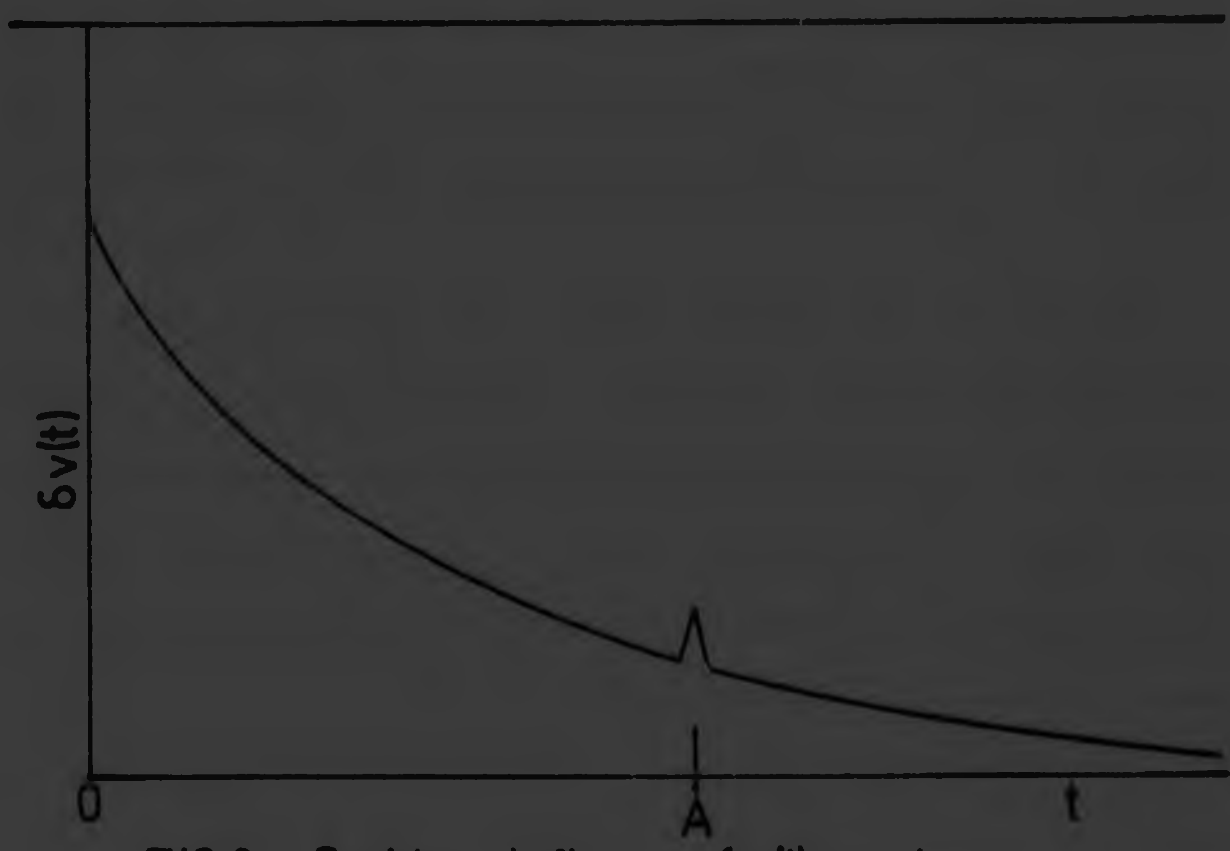


FIG 9 Sudden influx δv(t) vs t

CHAPTER IV

THE THEOREM

4.1 Conservation of Vocabulary

V is a fixed number.

Any individual has at his disposal a finite number of words V, which he can use or not use during the course of any piece of writing or speech, N words long. The number of words N, is known as the text length (whether written or spoken).

Some of the words used will appear once, others twice, thrice, etc. Other words will not appear at all.

If the words are grouped according to the number of times they appear, r, in a text of length N, we define $f_1(r|N)$, (where $0 \leq r \leq N$), as the number of words appearing r times in a text N words long.

The subscript 1 is used because the observations start at $r = 1$. This function $f_1(r|N)$ is a frequency distribution and must not be confused with the probability distribution $f(r|N)$ (without the subscript) as defined later on page 24.

$f_1(1|N)$ is the number of words appearing once in text of length N, known as HAPAX LEGOMENA.

$f_1(2|N)$ is known as HAPAX DISLEGOMENA.

$f_1(3|N)$ is known as HAPAX TRILEGOMENA etc.

Independently of any sample of whatever size, the following statement is always true.

$$V = f_1(0|N) + f_1(1|N) + \ldots\ldots f_1(r|N) + \ldots\ldots f_1(N|N) \qquad \ldots\ldots (34)$$

Of all these terms, only two are not observable namely V and $f_1(0|N)$. Thus if we transpose in (34) we get

$$V - f_1(0|N) = f_1(1|N) + \ldots\ldots f_1(r|N) + \ldots\ldots f_1(N|N)$$

$$= \sum_{r=1}^{N} f_1(r|N) = n_N \qquad \ldots\ldots (35)$$

n_N is thus the vocabulary contained in N.

Dividing $f_1(r|N)$ by n_N and $V-f_1(0|N)$ gives

$$\frac{f_1(r|N)}{n_N} = \frac{f_1(r|N)}{V-f_1(0|N)} \qquad \ldots\ldots (36)$$

$\dfrac{f_1(r|N)}{n_N}$ is measurable and is the proportion of words appearing r times in a text N words long.

$\phi(r|N)$ is now defined as:

$$\phi(r|N) = \frac{f_1(r|N)}{n_N} \qquad \dots\dots\dots\dots \quad (37)$$

If the vocabulary is considered to be a unit, no matter what its actual size we can adapt equation (34) by dividing each term by V which gives

$$1 = \frac{f_1(0|N)}{V} + \frac{f_1(1|N)}{V} + \dots\dots \frac{f_1(r|N)}{V} + \dots\dots \frac{f(N|N)}{V}$$

$\frac{f_1(r|N)}{V}$ is thus that fraction of the vocabulary appearing r times in text N words long.

$f(r|N)$ is now defined as:

$$f(r|N) = \frac{f_1(r|N)}{V}$$

Thus dividing the numerator and denominator of the right hand side of (37) by V gives

$$\phi(r|N) = \frac{f_1(r|N)}{n_N} = \frac{\frac{f_1(r|N)}{V}}{\frac{V-f_1(0|N)}{V}} = \frac{\frac{f_1(r|N)}{V}}{1-\frac{f_1(r|N)}{V}}$$

which leads to

$$\phi(r|N) = \frac{f(r|N)}{1-f(0|N)} \qquad \dots\dots\dots \quad (38)$$

The left hand side of equation (37) i.e.

$$\phi(r|N) = \frac{f_1(r|N)}{n_N}$$

is completely observable and measurable.

The right hand side of equation (38) namely:

$$\frac{f(r|N)}{1-f(0|N)}$$

requires a theoretical model which is derived as follows.

4.2 <u>A set of Probabilities</u>

Sichel[8] defines a vocabulary as follows:

" Let each word in the vocabulary V have a long-term probability of occurrence $\Pi_1, \Pi_2, \Pi_3 \ldots\ldots \Pi_i \ldots\ldots \Pi_v$.

Π_1 refers to the word with lowest probability, i.e., a word employed very rarely, and Π_v is the probability of the word most frequently used.
In the English language $\Pi_v = 0{,}07$, is associated with the word 'the'.
The other probabilities Π_i are a good deal smaller. As the vocabulary V consists of several thousand distinct words (types) with $\Pi \leq 0{,}07$, we may for the sake of mathematical convenience look at the Π_is as a continuous variable distributed as $\Psi(\Pi)$.
Theoretically we have

$$0 < \Pi_i < 1 \text{ and } \sum_{i=1}^{V} \Pi_i = 1 \text{ . "}$$

The function $\Psi(\Pi)$ is defined as is usual in statistics, i.e.

$\Psi(\Pi)\delta\Pi$ is that fraction of the vocabulary appearing with probability between Π and $\Pi + \delta\Pi$ and

$$\int_0^1 \Psi(\Pi)d\Pi = 1 \qquad \ldots\ldots\ldots (39)$$

For clarity and convenience the following function is defined:

$g(\Pi)$ where $g(\Pi)\delta\Pi$ is the number of words in the vocabulary appearing with probability between Π and $\Pi + \delta\Pi$,

thus $$\int_0^1 g(\Pi)d\Pi = V \qquad \ldots\ldots\ldots (40)$$

and therefore $g(\Pi) = V\,\Psi(\Pi)$

4.3 Binomial Distribution

Any single specific word in the vocabulary will thus occupy a unit area of the delta strip Π to $\Pi + \delta\Pi$ under the $g(\Pi)$ curve. See Figure 10 on page 26.

In any text N words long the specific word may appear once, twice, ... r times ..... or not at all.

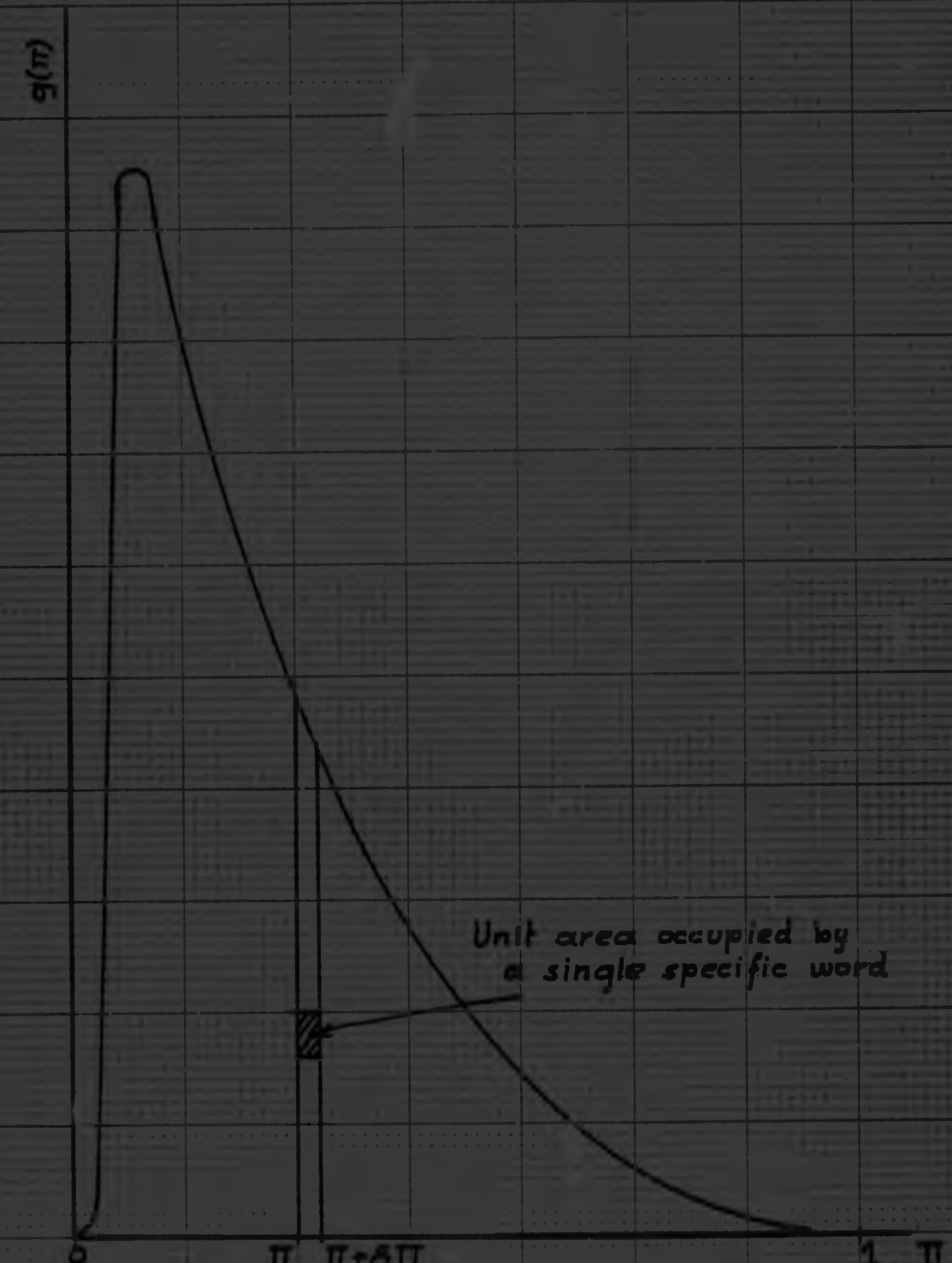


FIG. 10 Graph of g(π) vs π

The probability will be binomially distributed with the general term being

$$p(r|N) = \binom{N}{r} \Pi^r (1-\Pi)^{N-r}$$

Since there are $g(\Pi)\delta\Pi$ words in the delta strip each appearing with the same probability Π, the number of words from the delta strip appearing r times in text of length N, is thus

$$f_1(r|N,\Pi) = \binom{N}{r} \Pi^r (1-\Pi)^{N-r} g(\Pi)\delta\Pi$$

Summing for all delta strips gives

$$f_1(r|N) = \binom{N}{r} \int_0^1 \Pi^r (1-\Pi)^{N-r} g(\Pi)d\Pi \qquad \text{..........} \quad (41)$$

which is the number of words appearing r times in a text N words long.

A special case of the mixed binomial is obtained by putting $r = 0$ i.e.

$$f_1(0|N) = \int_0^1 (1-\Pi)^N \quad g(\Pi)d\Pi$$

and substituting in equation (35) gives

$$n_N = V - \int_0^1 (1-\Pi)^N \quad g(\Pi)d\Pi \qquad \text{.........} \quad (42)$$

which is identical with form (a) of the theorem as derived by a much longer argument in chapter III.

Dividing equation (41) by V gives

$$\frac{f_1(r|N)}{V} = f(r|N) = \binom{N}{r} \int_0^1 \Pi^r (1-\Pi)^{N-r} \Psi(\Pi)d\Pi \qquad \text{.........} \quad (43)$$

which is, that fraction of the vocabulary appearing r times in a text N words long.

4.4 The Mixing Distribution

Sichel(10) chooses the generalised inverse Gaussian distribution

$$\Psi(\Pi) = C\,\Pi^{\gamma-1}\,e^{-\left(\frac{1}{\theta}-1\right)\Pi-\frac{\alpha^2\theta}{4\Pi}} \quad \ldots\ldots\ldots \quad (44)$$

where $-\infty < \gamma < \infty$, $0 < \theta < 1$, and $\alpha > 0$, $C = \dfrac{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma}}{2K_{\gamma}(\alpha\sqrt{1-\theta})}$
and justifies it by stating:

> "The choice of the mixing distribution $\Psi(\Pi)$ is crucial. As the observed word frequency distributions are invariably skewed in the extreme, it is necessary to have a mixing distribution which can take on very skewed shapes".

4.5 Poisson Distribution

Sichel(11) having stated the mixed binomial (43) does not pursue it. "As the Π's are small, we may replace the binomial by the Poisson distribution". Sichel does not give a theoretical justification or derivation of the Poisson distribution.

The usual derivation of the Poisson as the limiting case of the Binomial does not apply to word frequencies.

The derivation starts with the Binomial distribution as stated

$$\phi(r|N) = \frac{N!}{r!(N-r)!} \Pi^r (1-\Pi)^{N-r}$$

where $r = 0, 1, 2, \ldots\ldots\ldots N$

and $0 < \Pi < 1$

The equation requires modification by putting $\Pi = \frac{\lambda}{N}$

Thus

$$\phi(r|N) = \frac{N!}{r!(N-r)!}\left(\frac{\lambda}{N}\right)^r\left(1-\frac{\lambda}{N}\right)^{N-r}$$

But the derivation requires that λ remain constant while $N \to \infty$ or at least the $\frac{\lambda}{N}$ be monotonic decreasing with increasing N, with limit zero.

Unfortunately $\frac{\lambda}{N}$ being constant renders the derivation invalid.

The following derivation is suggested.

4.6 Poisson Derivation

If the variable N is assumed to be a continuous variable (although it clearly is not) then we may take the lower limit of any increment in N, δN as zero (and not 1 as in the case of the Binomial). The following assumptions are made:

(i) With each word is associated a probability, or relative frequency of occurrence, Π.

(ii) If δN is small then the probability of the word occurring during δN is proportional to δN. $P(1|\delta N) = \Pi\delta N$ and the probability of the word not occurring during δN is thus $P(0|N) = 1 - \Pi\delta N$. .. (45)

(iii) It is also assumed that if δN is small enough only one event can occur during δN i.e. the probability of two events or more occurring during δN is zero. Certainly in the case of δN being a number of words, then when $\delta N \leq 1$ the condition is met.

Consider a sample of text N words long followed by an increment in N, δN.

If the event does not occur in $N + \delta N$ then it does not occur in N or in δN,

$$\text{thus } P(0|N + \delta N) = P(0|N)\,P(0|\delta N)$$

$$= P(0|N)\,(1 - \Pi\delta N) \text{ from (45)}$$

$$= P(0|N) - P(0|N)\,\Pi\delta N$$

$$\therefore P(0|N+\delta N) - P(0|N) = -P(0|N)\,\Pi\delta N$$

$$\therefore \frac{P(0|N+\delta N) - P(0|N)}{P(0|N)} = -\Pi\delta N$$

$$\text{which is } \frac{\delta P(0|N)}{P(0|N)} = -\Pi\delta N$$

letting $\delta N \to 0$ and integrating we get

$$\int \frac{dP(0|N)}{P(0|N)} = -\Pi \int dN$$

$$\text{i.e. } \ln P(0|N) = -\Pi N + C \qquad \text{.......... (46)}$$

C is evaluated by putting N = 0 which gives

$$P(0|0) = 1$$

(the probability of the word not occurring in a text zero words long is a certainty)

$$\text{and } \therefore \ln P(0|0) = 0$$

substituting in (46) gives

$$\ln P(0|0) = -\Pi \times 0 + C$$

$$\therefore C = 0$$

$$\text{and thus } P(0|N) = e^{-\Pi N}$$

Now considering the probability of the word occurring once in text $N + \delta N$ we get

$$P(1|N+\delta N) = P(1|N)\,P(0|\delta N) + P(0|N)\,P(1|\delta N)$$
$$= P(1|N)(1-\Pi\delta N) + e^{-\Pi N}\,\Pi\delta N$$
$$= P(1|N) - P(1|N)\Pi\delta N + e^{-\Pi N}\,\Pi\delta N$$
$$\frac{P(1|N+\delta N) - P(1|N)}{\delta N} = \Pi e^{-\Pi N} - P(1|N)\,\Pi$$

letting $\delta N \to 0$ and transposing gives

$$\frac{d}{dN} P(1|N) + \Pi P(1|N) = \Pi e^{-\Pi N} \quad \text{........... (47)}$$

multiplying (47) by the Integrating Factor $e^{\Pi N}$

gives $e^{\Pi N} \frac{d}{dN} P(1|N) + \Pi e^{\Pi N} P(1|N) = \Pi$

which is $\frac{d}{dN}\left[e^{\Pi N}\, P(1|N)\right] = \Pi$

Integrating with respect to N gives

$$\int \frac{d}{dN}\, e^{\Pi N}\, P(1|N)\,dN = \int \Pi\,dN$$
$$\therefore e^{\Pi N}\, P(1|N) = \Pi N + C$$

and again putting $N = 0$

gives $C = 0$ [because $P(1|0) = 0$]

thus $P(1|N) = \Pi N e^{-\Pi N}$

Further, considering the probability of the word occurring twice in text $N + \delta N$ we get

$$P(2|N+\delta N) = P(2|N)\,P(0|\delta N) + P(1|N)\,P(1|\delta N) + P(0|N)\,P(2|\delta N)$$

The third term $P(0|N)\,P(2|\delta N)$ vanishes since δN is small enough for the word to be unlikely to occur twice in it.

This gives $\frac{P(2|N+\delta N) - P(2|N)}{\delta N} = - P(2|N)\,\Pi\delta N + \Pi N e^{-\Pi N}\,\Pi\delta N$

and $\frac{d}{dN}\left[P(2|N) + \Pi P(2|N)\right] = \Pi^2 N e^{-\Pi N}$

Multiplying by the Integrating Factor $e^{\Pi N}$ and integrating gives

$$P(2|N)\, e^{\Pi N} = \frac{\Pi^2 N^2}{2}$$

$$\text{and } P(2|N) = \frac{(\Pi N)^2}{2}\, e^{-\Pi N}$$

Continuing this process i.e. substituting $\frac{(\Pi N)^2}{2} e^{-\Pi N}$ for $P(2|N)$ in

$$P(3|N+\delta N) = P(3|N)\,P(0|\delta N) + P(2|N)\,P(1|\delta N) \text{ etc.}$$

leads finally to Poisson's Equation

$$P(r|N) = \frac{(\Pi N)^r}{r!}\, e^{-\Pi N} \quad \text{........... (48)}$$

<u>NOTE</u>: **Using the notation of Chapter III i.e. putting v=x, N=t in (48) and taking the case where r=0, we get:**

$$P(0/t) = e^{-xt}$$

and since by the notation of Chapter III there are $f(x)\delta x$ words which all appear with probability x we get by multiplying e^{-xt} by $f(x)\delta x$ and summing for all delta strips,
the number of words not appearing in sample

$$V = v(t) = \int_0^1 e^{-xt} f(x)dx$$

and thus

$$v(t) = V - \int_0^1 e^{-xt} f(x)dx$$

which is form b of the theorem as derived in Chapter III.

Calculation of the Binomial and of the Poisson probabilities for

$$r = 1$$
$$\Pi \leq 0{,}01$$
$$\text{and } N \geq 100$$

showed negligible differences and the Poisson distribution has been accepted.

4.7 Mixed Poisson

Since $\Psi(\Pi)\delta\Pi$ of the vocabulary appears with probability Π the proportion of words appearing r times each in a text N words long is then given by the mixed Poisson distribution

$$f(r|N) = \frac{1}{r!}\int_0^1 e^{-\Pi N}(\Pi N)^r\,\Psi(\Pi)d\Pi \qquad \ldots\ldots\ldots (49)$$

substituting for $\Psi(\Pi)$ from (44) gives

$$f(r|N) = \frac{C}{r!}\int_0^1 e^{-N\Pi}(N\Pi)^r\,\Pi^{\gamma-1}\,e^{-\left(\frac{1}{\theta}-1\right)\Pi-\frac{\alpha^2\theta}{4\Pi}}\,d\Pi \qquad \ldots\ldots (50)$$

Integration of (50) is [illegible] by Sichel[12] as follows:

firstly since $\int_1^\infty e^{-N\Pi}$ ([illegible] $\Psi(\Pi)\, d\Pi \doteq 0$

we may replace the upper limit of the integral by ∞ giving

$$f(r|N) = \frac{C}{r!}N^r\int_0^\infty \Pi^{r+\gamma-1}\,e^{-\left(\frac{1}{\theta}+N-1\right)\Pi-\frac{\alpha^2\theta}{4\Pi}}\,d\Pi \qquad \ldots\ldots\ldots\ldots (51)$$

and secondly

putting $\Pi = \dfrac{\alpha\sqrt{\frac{\theta}{2}}}{\sqrt{\frac{1}{\theta}+(N-1)2}}\,e^u$ when $\Pi = 0,\ u = -\infty$ and when $\Pi = \infty,\ u = \infty$

$$\text{and } d\Pi = \frac{\alpha\sqrt{\theta/2}}{\sqrt{\frac{1}{\theta}+(N-1)2}}\,e^u\,du \qquad \ldots\ldots\ldots\ldots (52)$$

we get

$$f(r|N)=\frac{CN^r}{r!}\int_{-\infty}^{+\infty}\left(\frac{\alpha\sqrt{\frac{\theta}{2}}\,e^u}{\sqrt{\frac{2}{\theta}+(N-1)2}}\right)^{r+\gamma-1} e^{-\frac{(\frac{1}{\theta}+N-1)\alpha\sqrt{\frac{\theta}{2}}\,e^u}{\sqrt{\frac{2}{\theta}+(N-1)2}}-\frac{\sqrt{\frac{2}{\theta}+(N-1)2\alpha^2\theta}}{\alpha\sqrt{\frac{\theta}{2}}}e^{-u}}\;\frac{\alpha\sqrt{\frac{\theta}{2}}\,e^u}{\sqrt{\frac{\theta}{2}+(N-1)2}}\,du$$

$$=\frac{CN^r}{r!}\int_{-\infty}^{+\infty}\left(\frac{\alpha\sqrt{\frac{\theta}{2}}\,e^u}{\sqrt{\frac{2}{\theta}+(N-1)2}}\right)^{r+\gamma-1} e^{-\sqrt{\frac{2}{\theta}+(N-1)2}\;\alpha\sqrt{\frac{\theta}{2}}\,e^u-\alpha\sqrt{\frac{\theta}{2}}\sqrt{\frac{2}{\theta}+(N-1)2}\;e^{-u}}\;\frac{\alpha\sqrt{\frac{\theta}{2}}\,e^u}{\sqrt{\frac{\theta}{2}+(N-1)2}}\,du$$

$$=\frac{CN^r}{r!}\int_{-\infty}^{+\infty}\left(\frac{\alpha\sqrt{\frac{\theta}{2}}\,e^u}{\sqrt{\frac{2}{\theta}+2(N-1)}}\right)^{r+\gamma} e^{-\alpha\sqrt{\frac{\theta}{2}}\sqrt{\frac{2}{\theta}+(N-1)2}\,\left(\frac{e^u+e^{-u}}{2}\right)}\,du$$

$$=\frac{CN^r}{r!}\int_{-\infty}^{+\infty}\left(\frac{\frac{\alpha\theta}{2}\,e^u}{\sqrt{1+(N-1)\theta}}\right)^{r+\gamma} e^{-\alpha\sqrt{1+(N-1)\theta}\,(\cosh u)}\,du$$

$$=\frac{CN^r}{r!}\left(\frac{\frac{\alpha\theta}{2}}{\sqrt{1+(N-1)\theta}}\right)^{r+\gamma}\int_{-\infty}^{+\infty}e^{(r+\gamma)u}\,e^{-\alpha\sqrt{1+(N-1)\theta}\,\cosh u}\,du$$

$$=\frac{CN^r}{r!}\left(\frac{\frac{\alpha\theta}{2}}{\sqrt{1+(N-1)\theta}}\right)^{r+\gamma}\int_{-\infty}^{+\infty}\{\cosh[u(\gamma+r)]+\mathrm{Sinh}[u(\gamma+r)]\}\,e^{-\alpha\sqrt{1+(N-1)\theta}\,\cosh u}\,du$$

Now since $\sinh u(\gamma+r)$ is antisymmetric about the origin and $e^{-\sqrt{1+(n-1)\theta}\cosh u}$ is symmetric about the origin it followr that

$$\int_{-\infty}^{+\infty}\mathrm{Sinh}\,u(\gamma+r)\;e^{-\alpha\sqrt{1+(N-1)\theta}\,\cosh u}\,du=0$$

an' therefore

$$f(r|N)=2\frac{CN^r}{r!}\left(\frac{\frac{\alpha\theta}{2}}{\sqrt{1+(N-1)\theta}}\right)^{\gamma+r}\int_0^{+\infty}\cosh[u(\gamma+r)]\;e^{-\alpha\sqrt{1+(N-1)\theta}\,\cosh u}\,du \quad \ldots\ldots\ldots\ (53)$$

4.7 The Modified Bessel Function of Second Kind

Sichel[13] shows that (53) is a well known Bessel function integral namely

$$f(r|N)=2\frac{CN^r}{r!}\left(\frac{\frac{\alpha\theta}{2}}{\sqrt{1+(N-1)\theta}}\right)^{\gamma+r}K_{\gamma+r}\left(\alpha\sqrt{1+(N-1)\theta}\right) \quad \ldots\ldots\ldots\ (54)$$

where K is the modified Bessel function of second kind of order $r + \gamma$ and argument $\alpha\sqrt{1+(N-1)\theta}$

The normalising constant C can now be evaluated by putting $r = 0$, $N = 0$.

This gives the proportion of the vocabulary not appearing in a sample of no size at all. Thus

$$f(0|0) = 1$$

i.e.
$$2C\left(\frac{\frac{\alpha\theta}{2}}{\sqrt{1-\theta}}\right)^{\gamma} K_{\gamma}(\alpha\sqrt{1-\theta}) = 1$$

$$\therefore \quad C = \left(\frac{\sqrt{1-\theta}}{\alpha\frac{\theta}{2}}\right)^{\gamma} \frac{1}{2K_{\gamma}(\alpha\sqrt{1-\theta})} \qquad \text{.........} \quad (55)$$

$\therefore$ by substituting for C in (44)

$$\phi(\Pi) = \left(\frac{2\sqrt{1-\theta}}{\alpha\theta}\right)^{\gamma} \frac{1}{2K_{\gamma}(\alpha\sqrt{1-\theta})} \Pi^{\gamma-1} e^{-\left(\frac{1}{\theta}-1\right)\Pi - \frac{\alpha^2\theta}{4\Pi}} \qquad \text{.........} \quad (56)$$

and again, substituting for C in (54) gives

$$f(r|N) = \left[\frac{1-\theta}{1+(N-1)\theta}\right]^{\frac{\gamma}{2}} \frac{1}{K_{\gamma}(\alpha\sqrt{1-\theta})} \left(\frac{N\alpha\theta}{2\sqrt{1+(N-1)\theta}}\right)^{r} K_{r+\gamma}\left(\alpha\sqrt{1+(N-1)\theta}\right)\frac{1}{r!} \quad (57)$$

Sichel[14] transformed equation (49) i.e.

$$f(r|N) = \frac{1}{r!}\int_0^1 e^{-N\Pi} (N\Pi)^r \phi(\Pi)\,d\Pi$$

by putting $\lambda = N\Pi$ and $d\lambda = N\,d\Pi$

Taking ∞ as the upper limit of integration, gives

$$f(r|N) = \frac{1}{r!}\int_0^{\infty} e^{-\lambda} \lambda^r f(\lambda)\,d\lambda \qquad \text{..........} \quad (58)$$

where

$$f(\lambda)=\frac{1}{2}\,\frac{(2\sqrt{1-\theta}/\alpha\theta)^{\gamma}\lambda^{\gamma-1}}{K_{\gamma}(\alpha\sqrt{1-\theta})}\,e^{-\left(\frac{1}{\theta}-1\right)\lambda-\frac{\alpha^{2}\theta}{4\lambda}}$$

by the same reasoning as shown in equations (50) to (57),
(and with $\lambda = N\Pi$)
we get

$$f(r|N)=\frac{(\sqrt{1-\theta_N})^{\gamma}(\alpha_N\theta_N/2)^{r}}{K_{\gamma}(\alpha\sqrt{1-\theta_N})\ r!}\ K_{r+\gamma}(\alpha_N) \qquad \text{.......... (59)}$$

The symbols θ_N and α_N are chosen because having made the substitution $\lambda = \Pi N$ the parameters will be dependent on N. Comparing equations (57) and (59) gives

$$\alpha_N = \alpha\sqrt{1+(N-1)\theta} \qquad \text{......... (60)}$$

and

$$\alpha_N\theta_N = \frac{N\ \alpha\ \theta}{\sqrt{1+(N-1)\theta}} \qquad \text{.......... (61)}$$

from which we get

$$\theta_N = \frac{N\theta}{1+(N-1)\theta} \qquad \text{.......... (62)}$$

and

$$1-\theta_N = \frac{1-\theta}{1+(N-1)\theta} \qquad \text{.......... (63)}$$

which then lead to the equations

$$\alpha_N\sqrt{1-\theta_N} = \alpha\sqrt{1-\theta} \qquad \text{.......... (64)}$$

and

$$\frac{\theta_N}{(1-\theta_N)N} = \frac{\theta}{1-\theta} \qquad \text{.......... (65)}$$

and as can be seen from the right hand sides of the equations (64) and (65),

$\alpha_N\sqrt{1-\theta_N}$ and $\dfrac{\theta_N}{(1-\theta_N)N}$ are INVARIANT with N.

In addition it should be noted that γ has not changed and thus

$\gamma_N = \gamma$ — INVARIANT with N

4.8 Zero Truncation

From equation (38) we have

$$\phi(r|N) = \frac{f(r|N)}{1-f(0|N)}$$

This equation is arrived at by Sichel[14] by applying the process known as zero truncation.

Substituting for $f(r|N)$ and $f(0|N)$ from equation (59) gives

$$\phi(r|N) = \frac{\dfrac{(\sqrt{1-\theta_N})^{\gamma}(\alpha_N\theta_N/2)^r}{K_\gamma\left(\alpha_N\sqrt{1-\theta_N}\right)} K_{r+\gamma}(\alpha_N)}{1-\dfrac{(1-\theta_N)^{\frac{\gamma}{2}} K_\gamma(\alpha_N)}{K_\gamma\left(\alpha_N\sqrt{1-\theta_N}\right)}} \frac{1}{r!}$$

$$= \frac{(\alpha_N\theta_N/2)^r K_{r+\gamma}(\alpha_N)}{\dfrac{K_\gamma(\alpha\sqrt{1+(N-1)\theta})}{(\sqrt{1-\theta_N})^\gamma} - K_\gamma(\alpha_N)} \frac{1}{r!} \qquad \ldots\ldots\ldots\ldots (66)$$

putting $\gamma = -\frac{1}{2}$ "a priori" and $r = 1$

$$\text{gives } \phi(1|N) = \frac{K_{1-\frac{1}{2}}(\alpha_N)\left(\alpha_N\theta_N/2\right)}{K_{-\frac{1}{2}}(\alpha_N\sqrt{1-\theta_N})(1-\theta_N)^{-\frac{\gamma}{2}} - K_{-\frac{1}{2}}(\alpha_N)} \qquad \ldots\ldots\ldots\ldots (67)$$

4.9 Modified Bessel Function Equations

Applying the modified Bessel function symmetry equation

$$K_\nu(Z) = K_{-\nu}(Z)$$

and the elementary function equation for half integer orders,

$$K_{n+\frac{1}{2}}(Z) = \sqrt{\frac{\pi}{2Z}}\, e^{-Z} \sum_{r=0}^{N} \frac{(n+r)!}{r!(n-r)!(2Z)^r}$$

gives

$$\phi(1|N) = \frac{K_{\frac{1}{2}}(\alpha_N)\left(\alpha_N\theta_N/2\right)}{K_{\frac{1}{2}}(\alpha_N\sqrt{1-\theta_N})(1-\theta_N)^{-\gamma/2} \quad K_{\frac{1}{2}}(\alpha_N)}$$

$$= \frac{e^{-\alpha_N}\sqrt{\frac{\pi}{2\alpha_N}}\left(\alpha_N \theta_N/2\right)}{e^{-\alpha_N\sqrt{1-\theta_N}}\sqrt{\frac{\pi}{2\alpha_N}\frac{1}{\sqrt{1-\theta_N}}}\,(1-\theta_1)^{\frac{1}{4}}-\sqrt{\frac{\pi}{2\alpha_N}}\,e^{-\alpha_N}}$$

$$= \frac{e^{-\alpha_N}\left(\alpha_N \theta_N/2\right)}{e^{-\alpha_N\sqrt{1-\theta_N}} - e^{-\alpha_N}}$$

$$= \left(\alpha_N\theta_N/2\right)\left[e^{-\alpha_N\sqrt{1-\theta_N}+\alpha_N} - 1\right]^{-1}$$

$$= \frac{\alpha_N\theta_N}{2}\left[e^{\alpha_N\left(1-(1-\theta_N)^{\frac{1}{2}}\right)} - 1\right]^{-1} \qquad \text{.......... (68)}$$

4.10 Estimators

Sichel[13] contends that "as the observed proportion of words which are used once only is usually large, fairly efficient estimators may be obtained from $\hat{\phi}(1)$ and the arithmetic mean $\bar{r}$ of the sample".

From (66) we get

$$E(r|N) = \frac{\alpha_N\theta_N/2}{(\sqrt{1-\theta_N})^{1+N}}\;\frac{K_{1+\gamma}(\alpha_N\sqrt{1-\theta_N})}{\dfrac{K_\gamma(\alpha_N\sqrt{1-\theta_N})}{(\sqrt{1-\theta_N})^{\gamma}} - K_\gamma(\alpha_N)}$$

putting $\gamma = -\frac{1}{2}$ 'a priori' gives

$$E(r|N) = \frac{\alpha_N\theta_N}{2}\left[(1-\theta_N)^{\frac{1}{2}}\left(1-e^{-\alpha_N(1-\sqrt{1-\theta_N})}\right)\right]^{-1} \qquad \text{.......... (69)}$$

Dividing (69) by (68) gives

$$\frac{E(r|N)}{\phi(1|N)} = \frac{e^{\alpha_N(1-\sqrt{1-\theta_N})}}{\sqrt{1-\theta_N}} \qquad \text{.......... (70)}$$

If $E(r|N)$, $\phi(1|N)$, α_N and θ_N are replaced by statistics and estimates $\bar{r}_N$, $\hat{\phi}(1|N)$, $\hat{\alpha}_N$ and $\hat{\theta}_N$, respectively, and if we write

$\hat{g}_N = \sqrt{1-\hat{\theta}_N}$, we obtain from (70)

$$\hat{\alpha}_N(1-\hat{g}_N) = \ln\left[\hat{g}_N\bar{r}_N/\hat{\phi}(1|N)\right] \qquad \text{.......... (71)}$$

and from (68)

$$\hat{\phi}(1|N) = \hat{a}_N(1-\hat{g}_N)(1+\hat{g}_N)\left[2\left(e^{\hat{a}_N(1-\hat{g}_N)}-1\right)\right]^{-1} \qquad \text{.......... } 72)$$

Substitution of (71) into (72) leads to

$$(1+\hat{g}_N)\ln\hat{g}_N - a\hat{g}_N + b = 0 \qquad \text{.......... } (73)$$

$$\text{where } a = 2\bar{r}_N - \ln\left[\bar{r}_N/\hat{\phi}(1|N)\right] \qquad \text{.......... } (74)$$

$$\text{and } b = 2\hat{\phi}(1|N) + \ln\left[\bar{r}_N/\hat{\phi}(1|N)\right] \qquad \text{.......... } (75)$$

Equation (73) can be solved by iteration

$$\text{whence } \hat{\theta}_N = 1-\hat{g}_N^{2} \qquad \text{.......... } (76)$$

$$\text{and } \hat{\alpha}_N = (1-\hat{g}_N)^{-1}\ln\left[\hat{g}_N\bar{r}_N/\hat{\phi}(1|N)\right] \qquad \text{.......... } (77)$$

4.11 Estimate of V

From equation (57) namely

$$f(r|N) = \left(\frac{\sqrt{1-\theta}}{\sqrt{1+(N-1)\theta}}\right)^{\gamma} \frac{1}{K_\gamma(\alpha\sqrt{1-\theta})} \left(\frac{N\alpha\theta}{2\sqrt{1+(N-1)\theta}}\right)^{r} K_{r+\gamma}\left(\alpha\sqrt{1+(N-1)\theta}\right)\frac{1}{r!} \quad \text{...(78)}$$

multiplying by V, and summing we get

$$V\sum_{r=0}^{\infty} f(r|N) = V\sum_{r=o}^{\infty} f(r|N) = V$$

$$\therefore V\sum_{r=1}^{N} \phi(r|N) = V - V\,\phi(0|N)$$

$$\text{i.e. } n_N = V\left[1-\phi(0|N)\right]$$

$$= V\left[1-\left(\frac{\sqrt{1-\theta}}{\sqrt{1+(N-1)\theta}}\right)^{\gamma} \frac{K_\gamma(\alpha\sqrt{1+(N-1)\theta})}{K_\gamma(\alpha\sqrt{1-\theta})}\right] \qquad \text{.......... } (79)$$

Experience has shown that it is very convenient to put $\gamma = -\frac{1}{2}$ "a priori". It makes the mathematics much more tractable and in the case of word frequencies it has been shown that estimating γ produces values so close to $-\frac{1}{2}$ as to make $-\frac{1}{2}$ acceptable.

Sichel[16] shows that $\gamma = -\frac{1}{2}$ gives excellent agreement between theory and observation.

Thus with $\gamma = -\frac{1}{2}$ "a priori".

$$n_N = V\left[1-\left(\frac{\sqrt{1-\theta}}{\sqrt{1+(N-1)\theta}}\right)^{-1} \frac{e^{-\alpha\sqrt{1+(N-1)\theta}}\sqrt{\frac{\Pi}{2\sqrt{1+(N-1)\theta}}}}{e^{-\alpha\sqrt{1-\theta}}\sqrt{\frac{\Pi}{\sqrt{1-\theta}}}}\right] \quad ..(80)$$

$$= V\left[1-e^{+\alpha}\left(\sqrt{1-\theta}-\sqrt{1+(N-1)\theta}\right)\right] \quad .... (81)$$

$$= V\left[1-e^{-c\sqrt{1-\theta}}\left(\sqrt{1+\left(\frac{\theta}{1-\theta}\right)N}-1\right)\right] \quad ...... (82)$$

putting $c = \alpha\sqrt{1-\theta} = \alpha_N\sqrt{1-\theta_N}$ ...... (83)

and $d = \dfrac{\theta}{1-\theta} = \dfrac{\theta_N}{N(1-\theta_N)}$ ...... (84)

gives $n_N = V\left[1-e^{-c(\sqrt{1+dN}-1)}\right]$

$$\therefore V = \frac{n_N}{1-e^{-c\left(\sqrt{1+dN}-1\right)}} = \frac{n_N}{1-e^{-\alpha_N(1-\sqrt{1-\theta_N})}}$$

Sichel(17) has shown that $V = \frac{2}{cd}$ by the following argument:
Applying the notation of Chapter IV to equation 20 on page 15 which reads

$$v(t) = V - \int_0^1 (1-x)^t f(x)dx$$

we get

$$n_N = V\left[1-\int_0^1 (1-\Pi)^N \Psi(\Pi)d\Pi\right]$$

putting N = 1

gives

$$n_1 = V - V\int_0^1 (1-\Pi)\,\Psi(\Pi)d\Pi$$

$$= V - V\int_0^1 \Psi(\Pi)d\Pi + V\int_0^1 \Pi\Psi(\Pi)d\Pi$$

$$= V\int_0^1 \Pi\Psi(\Pi)d\Pi$$

But by definition and by nature

$$n_1 = 1$$

$$\therefore V = \frac{1}{\int_0^1 \Pi\Psi(\Pi)d\Pi} \quad .... (85)$$

The denominator of (85) gives the mean of the probability distribution $\Psi(\Pi)$ for word probabilities and the total vocabulory 'V' is the inverse of the mean of the probability distribution given by (44)

$$f(\Pi) = \frac{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma}}{2K_{\gamma}(\alpha\sqrt{1-\theta})} \Pi^{\gamma-1} e^{-(\frac{1}{\theta} - 1)\Pi - \frac{\alpha^2\theta}{4\Pi}}$$

where $(0 \leq \Pi \leq 1)$ and with $C = \frac{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma}}{2K_{\gamma}(\alpha\sqrt{1-\theta}}$

Bearing in mind that $\int_{1}^{\infty} f(\Pi)d\Pi = 0$

we get

$$\int_{0}^{\infty} f(\Pi)d\Pi = 1 = \frac{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma}}{2K_{\gamma}(\alpha\sqrt{1-\theta}} \int_{0}^{\infty} \Pi^{\gamma-1} e^{-(\frac{1}{\theta} - 1)\Pi - \frac{\alpha^2\theta}{4\Pi}} \qquad \text{.........} (86)$$

hence

$$\int_{0}^{\infty} \Pi^{\gamma-1} e^{-(\frac{1}{\theta} -1)\Pi - 4\frac{\alpha^2\theta}{4\Pi}} d\Pi = \frac{2K_{\gamma}(\alpha\sqrt{1-\theta})}{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma}} \qquad \text{.........} (87)$$

The required mean of $f(\Pi)$ is

$$\int_{0}^{1} \Pi f(\Pi)d\Pi = \int_{0}^{\infty} \Pi f(\Pi)d\Pi$$

$$= \frac{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma}}{2K_{\gamma}(\alpha\sqrt{1-\theta})} \int_{0}^{\infty} \Pi^{\gamma} e^{-(\frac{1}{\theta} -1)\Pi - \frac{\alpha^2\theta}{4\Pi}} d\Pi \qquad \text{.........} (88)$$

Hence by comparing indices in equations (86) and (88) we see that we must now replace γ by $\gamma+1$

Consequently

$$\frac{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma}}{2K_{\gamma}(\alpha\sqrt{1-\theta}} \int_{0}^{\infty} \Pi^{\gamma} e^{-(\frac{1}{\theta} -1)\Pi - \frac{\alpha^2\theta}{4\Pi}} d\Pi$$

$$= \frac{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma}}{2K_{\gamma}(\alpha\sqrt{1-\theta}} \; \frac{2K_{\gamma+1}(\alpha\sqrt{1-\theta}}{\left[2\sqrt{1-\theta}/\alpha\theta\right]^{\gamma+1}}$$

$$= \frac{1}{2\sqrt{1-\theta}/\alpha\theta} \; \frac{K_{\gamma+1}(\alpha\sqrt{1-\theta})}{K_{\gamma}(\alpha\sqrt{1-\theta})}$$

$$+ \frac{\alpha\Theta}{2\sqrt{1-\theta}} \quad \frac{K_{\gamma+1}(\alpha\sqrt{1-\theta})}{K_{\gamma}\ (\alpha\sqrt{1-\theta})}$$

and thus by equation (83)

$$V = \frac{2\sqrt{1-\theta}}{\alpha\Theta} \quad \frac{K_{\gamma}\ (\alpha\sqrt{1-\theta})}{K_{\gamma+1}(\alpha\sqrt{1-\theta})}$$

Now putting $\gamma = -\frac{1}{2}$

gives

$$V = \frac{2\sqrt{1-\theta}}{\alpha\Theta} \quad \frac{K_{-\frac{1}{2}}(\alpha\sqrt{1-\theta})}{K_{+\frac{1}{2}}(\alpha\sqrt{1-\theta})}$$

By the symmetry relations of Modified Bessel Functions we get:

$$V = \frac{2\sqrt{1-\theta}}{\alpha\Theta}$$

But as previously defined in (83) and (84)

$$c = \alpha\sqrt{1-\Theta} \quad \text{and } d = \frac{\theta}{1-\theta}$$

$$\therefore \quad V = \frac{2(1-6\ )}{\alpha\Theta\sqrt{1-\theta}} = \frac{2}{cd} \qquad \text{..........} \quad (89)$$

V can now be estimated from $\hat{\phi}(1|N)$ and $\bar{r}_N$ by the following steps:

(1) $a = 2\bar{r}_N - \ln|\bar{r}_N/\hat{\phi}(1|N)|$ i.e. equation (74)

$b = 2\hat{\phi}(1|N) + \ln(\bar{r}_N/\hat{\phi}(1|N))$ i.e. equation (75)

(2) a and b are substituted into (73)

i.e. $(1+\hat{g}_N)\ln\hat{g}_N - a\hat{g}_N + b = 0$

whereupon $\hat{g}_N$ is calculated by iteration

(3) $\hat{\theta}_N = 1 - \hat{g}_N^2$ i.e. equation (76)

and $\hat{\alpha}_N = (1-\hat{g}_N)^{-1} \ln|\hat{g}_N\bar{r}_N/\hat{\phi}(1|N)|$ i.e. equation (77)

(4) $\hat{c} = \hat{\alpha}_N\sqrt{1-\hat{\theta}_N}$ i.e. equation (83)

and $\hat{d} = \dfrac{\hat{\theta}_N}{N(1-\hat{\theta}_N)}$ i.e. equation (84)

and finally

(5) $\hat{V} = \dfrac{2}{\hat{c}\hat{d}}$ i.e. equation (89)

CHAPTER V - NEW DATA

5 New Data

Having acquired an improved method of estimating V, most of the data acquired in 1958 were inadequate because only n_1 and N had been recorded and the texts were no longer available. Sandré had written in excess of 13 000 words in the form of essays on topics of her own choice between May and August 1958, and the text was still extant.

5.1 Exhaustion

From the evidence in the preliminary experiments it was seen that at the end of a text a few thousand words long, the rate at which new words were being produced indicated that the subject had come nowhere near to exhausting his/her vocabulary. The asymptote was still a long way off. It was, therefore, decided to record a large text from an individual possessing a limited vocabulary. Very young children would be useful for this purpose. By chance two acquaintances had daughters aged two years and a few months, Kerry and Bronwyn.

5.2 Kerry

Age : 2 years
Background : Middle class
Environment : Peri-urban
Personality : Happy, normal child

Kerry's father was a small independent businessman (joinery and shopfitting), her mother was a secretary but was not in employment. She chose to stay at home for the sake of Kerry and her younger sister. Kerry's mother, being a trained secretary, was able to record Kerry's conversation in shorthand and type it later when time allowed. Recordings were made in two sessions per day, usually during meals and play times. The objectivity of the recording technique can be seen by details of toilet training and such everyday childhood disasters as 'Baba sick me teddy - me wash teddy'.

The words were lemmatised before being processed, e.g. raining = rain. But it was noticed that invariably, only one form of the word was known which rendered lemmatisation largely unnecessary.

Two sets of data were collected and named Kerry 1 and Kerry 2.

Kerry 1 consisted of 2 132 words, recorded from 17 February 1975 to 24 February 1975.

Kerry 2 consisted of 2 151 words, recorded from 25 February 1975 to 22 March 1975.

5.3 Bronwyn

Age : 2 years

Background : Middle class

Environment : Suburban

Personality : Happy, normal child

Bronwyn's father was a professional physicist and her mother a high school mathematics teacher, but was not employed, having decided to devote her time to bringing up Bronwyn and her older sister. Bronwyn spent a lot of time accompanying her suburban taxi driver mother conveying big sister to and from nursery school, supermarket and all the other trips, which suburban consumerism is heir to. Consequently Bronwyn was exposed to big sister and lift club mates' conversation which must inevitably have enlarged her vocabulary. This process occurs for some but not for others and it is doubtful whether the effect would be permanent.

Two sets of data were recorded; Bronwyn 1 and Bronwyn 2.

Bronwyn 1 consisted of 8 037 words recorded from 11 February 1975 to 28 February 1975.

Bronwyn 2 consisted of 2 848 words recorded from 10 March 1975 to 17 March 1975.

Bronwyn's mother did not know shorthand but by writing fast was able to 'get it all down - somehow'.

Recording times were snatched at odd moments during play sessions and during, and after meals.

Bronwyn knew a great many songs and rhymes but mostly they were not recorded except by title e.g. 'song .... '.

Children's songs vary from province to province and vary also with the cultural origin of the singer. Therefore, the songs were not included with the data.

5.4 Sandré again : A large text

Sandré had written more than 13 000 words in the form of essays on topics of her own choice during the original experiment. The text was still extant.

5.5 Running Asymptote

errich[18] contended that the failure to achieve an acceptable fit for the v vs t curves (alternatively $\hat{n}$. vs N) might have been due to the child's vocabulary growing during the test period. "You may be chasing an asymptote which is running away from you". It must be noted that both the younger children were going through their vocabulary-cum-language-explosion where within the space of a few months a child's language ability progresses from the use of a few words to long conversations.

5.6 Hysteresis

It was therefore decided to analyse the text in both the forward and reverse directions. The argument was that if the asymptote was running away with increasing N the graph of $\hat{n}$. vs N would show less curvature than for a static asymptote. Conversely if the text was reversed the graph of $\hat{n}$. vs N should show greater curvature. Since the total text length for the sample and the vocabulary contained therein must be identical for both directions, the forward and backward graphs would have the same initial and final points, namely 0,0 and N, $\hat{n}$. respectively but should show an hysteresis loop between their extremeties.

5.6.1 Randomising the text and hysteresis

To put the matter of hysteresis beyond doubt it was decided to randomise the data and analyse them in both directions, where no hysteresis would be expected.

5.6.2 Randomising the Data in general

Randomising the data would also verify whether the assumption that a vocabulary was merely a set of probabilities was tenable. The data extracted from a randomised text would not be expected to differ markedly from that obtained from the natural text.

The randomisation procedure was applied in the cases of Kerry 1, Kerry 2 and Bronwyn 1.

Having provided satisfactory results the process was no longer used and natural data only were used in the case of Bronwyn 2 and Sandré.

Reversing and randomising resulted in the large number of tables and graphs referred to.

5.7 The Data

Before being analysed the words were lemmatised using the rule stated in Chapter 2.5. In the case of the younger children almost in-

[illegible] the form of the word was known which rendered lemmatisation largely unnecessary.

The text was punched on the computer cards and an average of about [illegible] words per card was recorded.

The text was analysed at intervals of one hundred words but each analysis was made from the beginning of the text so that the first sample involved text words (tokens) one to one hundred, the second involved text words one to two hundred. In this way a larger and larger text, increasing by one hundred words at a time, was analysed.

The computer was programmed to read each word, check whether it had appeared already, and how many times. The number of appearances of each word was printed together with the text length analysed, the number of words appearing once, twice, etc. and the vocabulary contained within the text. Thus for each successive hundred words a complete word frequency distribution was prepared and printed by the computer. The complete data are shown only when testing the model in Chapter VII (see tables 33, 35 and 36 on pages 225, 227 and 228).

In the case of the data in reverse order the pile of data cards was simply reversed and the computer was required to read each word in reverse order. Thus each "word" became "drow", each "was" became "saw" and each "saw" became "was". The program proceeded without a hitch.

The data were randomised using a random number table generated by the computer. The cards were then run through the computer in both directions.

4.7.1 [illegible]

As a result of the reversing and randomising of the data a total of [illegible] was analysed. They are shown as:

Kerry (1) forward
Kerry (1) backward
Kerry (1) random forward
Kerry (1) random backward
Kerry (2) forward
Kerry (2) backward
Kerry (2) random forward
Kerry (2) random backward
[illegible] (1) forward
[illegible] (1) backward

Bronwyn (1) random forward

Bronwyn (1) random backward

Bronwyn (2) forward

Sandré forward

5.8 THE TABLES: TYPE 1. DATA, TABLES 2 TO 14 ON PAGES 48 TO 71

Column 1. Heading N. Number of words in the text analysed. In the case of the younger children the text was analysed at 100 word intervals, but in Sandré's case, with a total text length of 13 000 words, it was decided that 200 word intervals would be adequate.

Column 2. Heading n_N. Number of different words (types) contained in the text.

Column 3. Heading $\hat{f}_1(1|N)$. Hapax Legomena, i.e. number of words appearing once in the portion of text, length N.

Column 4. Heading $\hat{f}_1(2|N)$. Hapax Dislegomena, i.e. number of words appearing twice in the portion of text, length N.

Column 5. Heading $\hat{f}_1(3|N)$. Hapax Trilegomena, i.e. number of words appearing three times in the portion of text, length N.

Column 6. Heading $\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$. The ratio, to three figures, of Hapax Dislegomena to Hapax Legomena.

Column 7. Heading $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$. The ratio, to three figures of Hapax Trilegomena to Hapax Dislegomena.

Column 8. Heading $\frac{\hat{f}_1(1|N)}{n_N}$. The proportion of Hapax Legomena, to three figures, also designated by $\phi(1|N)$.

Column 9. Heading $\frac{\hat{f}_1(2|N)}{n_N}$. The proportion of Hapax Dislegomena, to three figures, also designated by $\phi(2|N)$.

Column 10. Heading $\frac{\hat{f}_1(3|N)}{n_N}$. The proportion of Hapax Trilegomena, to three figures, also designated by $\phi(3|N)$.

Column 11. Heading $\frac{N}{n_N}$. The average number of times the words appear, to three figures. This is also known as $\bar{f}_N$.

5.9 The Data Trends

$\bar{n}_N$: As expected, $\bar{n}_N$ (the number of types) is monotonic increasing. The rate of increase $\frac{\Delta\bar{n}_N}{\Delta N}$ is monotonic decreasing (as can be expected) but after an initial rapid decrease, decreases only slowly. The trend is general but local deviations occur.

$\bar{f}_1(1|N)$ increases rapidly at first but after N = 600 or thereabouts the rate of increase decreases markedly. In the case of the younger children the rate of increase decreases to a barely perceptible upward trend with many local inversions.

In the case of Bronwyn 1 the large text N > 8000 words shows the steady upward trend very clearly.

Since all observed word frequency distributions so far produced show that more words appear once than any other number of times, Hapax Legomena ($\bar{f}_1(1|N)$ is a very important constituent of any word frequency distribution.

Since any individual's vocabulary is finite and the number of words in a sample is not limited (the individual can always say or write a few more words), it follows that eventually, with a large enough sample, the individual must reach the stage where he/she has used more words twice than once. Indeed Hapax Legomena having risen with increasing text size, must eventually decline as the individual exhausts his/her vocabulary. Ultimately the stage must be reached where the text is so large that the individual tends to have used every word in his/her vocabulary at least twice and then Hapax Legomena must tend to zero.

The data show that before this stage is reached there is a long, very long interval during which $\bar{f}_1(1|N)$ remains almost static, the rate at which new words appear being equal to the rate at which words, having appeared once, are repeated.

The evidence in the tables shows that in no case was any decrease in $\bar{f}_1(1|N)$ apparent.

It can only be concluded therefore that in no case was the individual anywhere near exhausting her vocabulary.

Even in the case of Bronwyn a sample of 8 037 words is still far too small.

$\bar{f}_1(2|N)$ showed, in all cases, a steady increase with increasing N but the many local inversions make the trend difficult to read.

$\bar{f}_1(3|N)$ also showed a continuing upward trend although in the case of Bronwyn 1 the upward trend, towards the end, had all but petered out.

The ratio $\dfrac{\bar{f}_1(2|N)}{\bar{f}_1(1|N)}$ showed wild fluctuations for small N i.e. $N \leq 500$. Thereafter it rose and finally drifted slightly downwards. The many deviations from the trend are due to the small numbers involved.

The ratio $\dfrac{\bar{f}_1(3|N)}{\bar{f}_1(2|N)}$ showed a general upward trend with two exceptions. Bronwyn 1 forward shows a slight drop from a plateau between N = 5 000 and N = 7 000 towards the end of the sample. However, the variations are large and the trend may be due to chance only. Sandré showed no trend at all, only local variations.

The proportions $\hat{\phi}(1|N) = \dfrac{\bar{f}_1(1|N)}{\bar{n}_N}$ showed a general decrease in all cases, with fluctuations and a tendency for the rate of decrease to decrease with increasing N.

The proportion $\bar{\phi}(2|N) = \dfrac{\bar{f}_1(2|N)}{\bar{n}_N}$ showed no trend in any case at all, only variations about a mean and as expected with steadily rising $\bar{n}_N$ the variations decrease.

The proportion $\phi(3|N) = \dfrac{\bar{f}_1(3|N)}{\bar{n}_N}$ also showed no trend, only variations about a mean.

The average number of times the words appeared $\bar{r}_N = \dfrac{N}{\bar{n}_N}$ showed, as was to be expected, a general upward trend with local inversions.

5.10 <u>The Graphs</u>: <u>Type 1, Data</u>

Graphs of the data appear in figures 11 to 62 on pages 74 to 141.

The graphs are grouped according to vertical scale. Thus the graphs of $\bar{n}_N$, $\bar{f}_1(1|N)$, $\bar{f}_1(2|N)$ and $\bar{f}_1(3|N)$ are shown on the same set of axes because they are measured in whole numbers which do not differ from each other by much more than one order of magnitude.

TABLE 2 KERRY (1) FORWARD

N	$\bar{n}_N$	$\bar{E}_1(1\|N)$	$\bar{E}_1(2\|N)$	$\bar{E}_1(3\|N)$	$\frac{\bar{E}_1(2\|N)}{\bar{E}_1(1\|N)}$	$\frac{\bar{E}_1(3\|N)}{\bar{E}_1(2\|N)}$	$\bar{\phi}(1\|N)$ $=\frac{\bar{E}_1(1\|N)}{\bar{n}_N}$	$\bar{\phi}(2\|N)$ $=\frac{\bar{E}_1(2\|N)}{\bar{n}_N}$	$\bar{\phi}(3\|N)$ $=\frac{\bar{E}_1(3\|N)}{\bar{n}_N}$	$\bar{r}_N$ $=\frac{N}{\bar{n}_N}$
100	44	24	6	3	0,250	0,500	0,545	0,136	0,068	2,27
200	74	35	14	7	0,400	0,500	0,473	0,189	0,095	2,70
300	92	40	20	8	0,500	0,400	0,435	0,217	0,087	3,26
400	108	44	20	15	0,453	0,750	0,407	0,185	0,139	3,70
500	128	56	22	12	0,395	0,545	0,438	0,172	0,094	3,91
600	144	62	27	10	0,435	0,370	0,431	0,188	0,069	4,17
700	155	64	28	14	0,438	0,500	0,413	0,181	0,090	4,52
800	161	58	26	20	0,448	0,769	0,360	0,161	0,124	4,97
900	170	62	25	22	0,403	0,880	0,365	0,147	0,129	5,29
1000	180	65	29	20	0,446	0,690	0,361	0,161	0,111	5,56
1100	189	62	30	23	0,484	0,767	0,328	0,159	0,122	5,82
1200	197	60	34	26	0,567	0,765	0,305	0,173	0,132	6,09
1300	212	62	37	29	0,597	0,784	0,292	0,175	0,137	6,13
1400	215	60	36	26	0,600	0,722	0,279	0,167	0,121	6,51
1500	229	67	41	23	0,612	0,561	0,295	0,179	0,100	6,55
1600	235	69	42	25	0,609	0,595	0,294	0,179	0,106	6,81
1700	239	72	39	26	0,542	0,667	0,301	0,163	0,109	7,11
1800	245	66	46	27	0,697	0,587	0,269	0,188	0,110	7,35
1900	249	64	50	26	0,781	0,520	0,257	0,201	0,104	7,63
2000	256	66	49	28	0,742	0,571	0,258	0,191	0,109	7,81
2100	263	71	49	25	0,690	0,510	0,270	0,186	0,096	7,98
2132	266	74	47	25	0,635	0,532	0,278	0,177	0,094	8,02

TABLE 3 KERRY (1) BACKWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N) = \frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N) = \frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N) = \frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\hat{r}_N = \frac{N}{\hat{n}_N}$
100	51	55	17	1	0,564	0,083	0,547	0,256	0,020	1,961
200	78	49	16	11	0,490	0,688	0,515	0,205	0,141	2,564
300	96	42	23	10	0,548	0,435	0,438	0,240	0,104	3,125
400	113	47	20	16	0,426	0,800	0,416	0,177	0,142	3,540
500	129	55	21	17	0,382	0,810	0,426	0,163	0,132	3,876
600	145	58	28	17	0,483	0,607	0,400	0,193	0,117	4,138
700	157	59	31	20	0,525	0,645	0,376	0,197	0,127	4,459
800	170	64	31	23	0,484	0,742	0,376	0,182	0,135	4,706
900	182	66	32	22	0,485	0,688	0,363	0,176	0,121	4,945
1000	191	70	34	19	0,486	0,559	0,366	0,178	0,099	5,236
1100	206	76	37	19	0,487	0,514	0,369	0,180	0,092	5,340
1200	219	77	39	21	0,506	0,538	0,350	0,181	0,098	5,481
1300	222	75	45	21	0,600	0,467	0,338	0,203	0,095	5,856
1400	227	74	45	21	0,608	0,467	0,326	0,198	0,093	6,167
1500	231	71	48	22	0,676	0,458	0,307	0,208	0,096	6,494
1600	233	67	49	23	0,731	0,469	0,288	0,210	0,099	6,867
1700	243	69	48	27	0,696	0,563	0,284	0,198	0,111	6,996
1800	254	76	46	28	0,605	0,608	0,299	0,181	0,110	7,087
1900	258	75	48	30	0,641	0,625	0,291	0,186	0,116	7,364
2000	263	74	50	27	0,676	0,540	0,281	0,190	0,103	7,605
2100	266	74	48	27	0,649	0,563	0,278	0,180	0,102	7,895
2152	266	74	47	25	0,635	0,552	0,278	0,177	0,094	8,015

TABLE 4 KERRY (1) RANDOM FORWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $= \frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $= \frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $= \frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N$ $= \frac{N}{\hat{n}_N}$
1	1	1	0	0	0	0	1	0	0	1,00
100	49	27	13	2	0,482	0,154	0,551	0,265	0,041	2,04
200	82	48	12	8	0,250	0,667	0,585	0,147	0,0976	2,42
300	114	59	25	8	0,424	0,320	0,516	0,219	0,0702	2,63
400	132	65	30	10	0,462	0,333	0,492	0,227	0,0758	3,03
500	150	67	33	20	0,493	0,606	0,441	0,220	0,133	3,33
600	165	69	37	21	0,554	0,568	0,418	0,224	0,127	3,64
700	179	74	40	19	0,541	0,475	0,414	0,223	0,106	3,91
800	191	71	46	22	0,648	0,478	0,372	0,241	0,115	4,18
900	201	75	43	20	0,574	0,465	0,374	0,214	0,0977	4,49
1000	208	72	46	21	0,640	0,457	0,346	0,221	0,101	4,80
1100	213	74	45	18	0,609	0,400	0,347	0,211	0,0845	5,15
1200	219	74	45	20	0,609	0,444	0,338	0,206	0,0914	5,48
1300	227	73	48	20	0,658	0,417	0,321	0,211	0,088	5,74
1400	232	69	52	22	0,754	0,424	0,297	0,224	0,0949	6,04
1500	239	72	53	20	0,736	0,377	0,314	0,231	0,0836	6,28
1600	241	70	51	23	0,730	0,451	0,290	0,212	0,0956	6,64
1700	243	68	48	25	0,705	0,520	0,280	0,187	0,103	6,99
1800	245	66	47	27	0,712	0,575	0,269	0,182	0,110	7,35
1900	252	71	45	27	0,634	0,600	0,282	0,178	0,107	7,54
2000	258	71	48	23	0,675	0,480	0,276	0,186	0,0893	7,75
2100	263	71	48	24	0,675	0,500	0,270	0,182	0,0911	8,00
2132	266	74	47	25	0,636	0,532	0,278	0,177	0,0940	8,03

TABLE 5 KERRY (1) RANDOM BACKWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{f_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N) = \frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N) = \frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N) = \frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
100	53	36	10	3	,278	,3	,68	,189	,0569	1,89
200	89	51	20	7	,392	,35	,574	,225	,0786	2,25
300	107	57	18	7	,316	,389	,534	,168	,0655	2,80
400	124	62	20	11	,322	,55	,500	,161	,089	3,22
500	136	63	22	13	,35	,591	,464	,162	,096	3,68
600	150	67	24	18	,359	,667	,446	,160	,120	4,00
700	166	76	27	13	,357	,482	,459	,163	,078	4,21
800	177	79	28	15	,356	,536	,446	,158	,0848	4,53
900	185	76	32	17	,421	,530	,411	,174	,092	4,86
1000	195	77	34	20	,441	,589	,395	,174	,103	5,13
1100	197	70	33	23	,471	,698	,356	,168	,117	5,60
1200	204	70	33	25	,471	,760	,342	,162	,123	5,88
1300	210	72	34	25	,472	,736	,343	,162	,119	6,19
1400	217	72	33	24	,457	,729	,332	,152	,111	6,44
1500	226	73	39	23	,535	,591	,323	,172	,104	6,64
1600	234	79	34	24	,430	,706	,337	,145	,102	6,83
1700	243	81	37	22	,457	,595	,334	,153	,091	7,00
1800	247	80	37	25	,464	,676	,324	,150	,101	7,30
1900	251	74	40	28	,540	,700	,295	,159	,112	7,55
2000	258	74	43	24	,583	,558	,287	,167	,093	7,75
2100	265	76	44	26	,579	,592	,286	,166	,098	7,94
2132	266	74	47	25	,634	,532	,278	,177	,094	8,00

TABLE 6 KERRY (2) FORWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N) = \frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N) = \frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N) = \frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
100	55	36	11	4	0,306	0,364	0,655	0,200	0,073	1,82
200	89	52	13	11	0,250	0,846	0,584	0,146	0,124	2,25
300	115	63	20	11	0,317	0,550	0,548	0,174	0,096	2,61
400	134	70	27	9	0,386	0,333	0,522	0,201	0,067	2,[illegible]
500	154	78	30	12	0,[illegible]85	0,400	0,506	0,195	0,078	3,25
600	162	73	34	17	0,466	0,500	0,451	0,21[illegible]	0,105	3,[illegible]
700	185	84	37	22	0,440	0,595	0,454	0,200	0,119	3,78
800	193	85	36	21	0,424	0,583	0,440	0,1[illegible]7	0,109	4,15
900	204	81	40	25	0,494	0,625	0,397	0,196	0,123	4,4[illegible]
1000	217	84	44	27	0,524	0,614	0,387	0,20[illegible]	0,124	4,61
1100	232	89	44	31	0,494	0,705	0,[illegible]04	0,190	0,134	4,74
1200	246	97	45	31	0,464	0,689	0,394	0,18[illegible]	0,126	4,[illegible]
1300	253	99	43	28	0,434	0,651	0,391	0,170	0,111	5,14
1400	262	104	40	28	0,385	0,700	0,397	0,153	0,107	5,34
1500	270	100	45	30	0,450	0,667	0,370	0,167	0,111	5,56
1600	281	103	45	31	0,437	0,689	0,367	0,160	0,110	5,69
1700	288	102	49	31	0,451	0,633	0,354	0,170	0,110	5,90
1800	299	106	53	31	0,500	0,585	0,355	0,177	0,104	6,02
1900	307	105	56	30	0,533	0,536	0,342	0,18[illegible]	0,098	6,19
2000	315	106	57	33	0,538	0,579	0,337	0,181	0,105	6,35
2100	324	108	59	37	0,546	0,627	0,333	0,182	0,114	6,48
2151	326	107	61	36	0,570	0,590	0,328	0,187	0,110	6,60

TABLE 7 KERKY (2) BACKWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N) = \frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N) = \frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N) = \frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
100	47	28	7	7	0,250	1,000	0,596	0,149	0,149	2,128
200	83	50	12	8	0,240	0,667	0,602	0,145	0,096	2,410
300	116	68	18	10	0,265	0,556	0,586	0,155	0,086	2,586
400	142	77	29	10	0,377	0,345	0,542	0,204	0,070	2,817
500	165	88	30	18	0,341	0,600	0,533	0,182	0,109	3,030
600	183	95	32	16	0,368	0,500	0,519	0,175	0,087	3,279
700	204	103	35	15	0,340	0,429	0,505	0,172	0,074	3,431
800	216	105	38	17	0,362	0,447	0,486	0,176	0,079	3,704
900	228	106	42	18	0,396	0,429	0,465	0,184	0,079	3,947
1000	236	106	42	21	0,396	0,500	0,449	0,178	0,089	4,237
1100	249	107	48	24	0,449	0,500	0,430	0,193	0,096	4,418
1200	260	108	56	20	0,519	0,357	0,415	0,215	0,077	4,615
1300	267	101	62	22	0,614	0,355	0,378	0,232	0,082	4,869
1400	274	101	65	22	0,644	0,355	0,369	0,237	0,080	5,109
1500	282	101	62	26	0,614	0,419	0,358	0,220	0,092	5,319
1600	285	97	61	29	0,629	0,475	0,340	0,214	0,102	5,614
1700	295	101	62	28	0,614	0,452	0,342	0,210	0,095	5,763
1800	300	98	65	31	0,663	0,477	0,327	0,217	0,103	6,000
1900	304	99	51	35	0,616	0,574	0,326	0,201	0,115	6,250
2000	315	103	64	33	0,621	0,516	0,327	0,203	0,105	6,349
2100	325	108	64	33	0,593	0,516	0,332	0,197	0,102	6,462
2151	326	107	61	36	0,570	0,590	0,328	0,187	0,110	6,598

TABLE 8 KERRY (2) RANDOM FORWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N) = \frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N) = \frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N) = \frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
100	56	36	14	2	0,389	0,143	0,643	0,250	0,036	1,79
200	86	44	22	8	0,500	0,364	0,512	0,256	0,093	2,33
300	120	65	21	11	0,323	0,524	0,542	0,175	0,092	2,50
400	149	79	30	11	0,380	0,367	0,530	0,201	0,074	2,68
500	165	87	32	13	0,368	0,406	0,527	0,194	0,079	3,03
600	189	99	38	12	0,384	0,316	0,524	0,201	0,063	3,17
700	203	105	38	17	0,362	0,447	0,517	0,187	0,084	3,45
800	218	107	44	18	0,411	0,409	0,491	0,202	0,083	3,67
900	232	110	44	24	0,400	0,545	0,474	0,190	0,103	3,88
1000	244	110	50	26	0,455	0.520	0,451	0,205	0,107	4,10
1100	255	111	53	30	0,477	,566	0,435	0,208	0,118	4,31
1200	265	115	50	33	0,435	0,660	0,434	0,189	0,125	4,53
1300	275	114	56	31	0,491	0,554	0,415	0,204	0,113	4,73
1400	280	106	60	34	0,566	0,567	0,379	0,214	0,121	5,00
1500	286	107	58	34	0,542	0,586	0,374	0,203	0,119	5,24
1600	294	108	61	29	0,565	0,475	0,367	0,207	0,099	5,44
1700	302	107	62	29	0,579	0,468	0,354	0,205	0,096	5,63
1800	304	102	63	29	0,618	0,460	0,336	0,207	0,095	5,92
1900	310	103	61	31	0,592	0,508	0,332	0,197	0,100	6,13
2000	315	102	60	34	0,588	0,567	0,324	0,190	0,108	6,35
2100	320	103	60	37	0,583	0,617	0,322	0,188	0,116	6,56
2151	326	107	61	36	0,570	0,590	0,328	0,183	0,110	6,60

TABLE 9 KERRY (2) RANDOM BACKWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N) = \frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N) = \frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N) = \frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
100	54	34	11	6	0,324	0,546	0,630	0,204	,1110	1,85
200	87	52	16	7	0,308	0,438	0,598	0,184	0,0805	2,30
300	116	67	19	11	0,284	0,579	0,578	0,164	0,095	2,59
400	130	74	23	13	0,311	0,565	0,532	0,165	0,094	2,88
500	159	86	25	12	0,291	0,480	0,541	0,157	0,075	3,14
600	179	93	29	12	0,312	0,414	0,520	0,162	0,067	3,35
700	195	99	30	12	0,303	0,400	0,508	0,154	0,062	3,59
800	205	93	38	16	0,409	0,421	0,454	0,185	0,078	3,90
900	218	90	45	21	0,500	0,467	0,413	0,206	0,096	4,13
1000	232	97	47	19	0,485	0,404	0,418	0,203	0,082	4,31
1 00	245	106	43	22	0,406	0,524	0,433	0,176	0,090	4,49
1200	253	111	39	23	0,351	0,590	0,439	0,154	0,091	4,74
1300	262	115	39	20	0,339	0,513	0,439	0,149	0,076	4,96
1400	271	114	45	20	0,395	0,444	0,451	0,166	0,074	5,17
1500	279	113	50	21	0,442	0,450	0,405	0,179	0,075	5,38
1600	289	116	49	29	0,422	0,592	0,401	0,170	0,100	5,54
1700	297	113	54	31	0,478	0,574	0,380	0,182	0,104	5,72
1800	306	113	57	33	0,504	0,579	0,369	0,186	0,108	5,88
1900	317	119	55	39	0,462	0,709	0,375	0,174	0,123	5,99
2000	323	115	61	35	0,530	0,574	0,356	0,189	0,108	6,19
2100	324	107	62	38	0,579	0,613	0,321	0,191	0,117	6,48
2151	326	107	61	36	0,570	0,590	0,328	0,187	0,110	6,60

TABLE 10 BROWN (1) FORWARD

N	n_N	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N) = \frac{\hat{f}_1(1\|N)}{n_N}$	$\hat{\phi}(2\|N) = \frac{\hat{f}_1(2\|N)}{n_N}$	$\hat{\phi}(3\|N) = \frac{\hat{f}_1(3\|N)}{n_N}$	$\bar{r}_N = \frac{N}{n_N}$
100	57	33	15	4	0,455	0,267	0,579	0,263	0,070	1,754
200	96	57	16	10	0,281	0,626	0,594	0,167	0,104	2,083
300	113	55	22	14	0,400	0,636	0,486	0,196	0,124	2,655
400	142	71	22	16	0,310	0,727	0,500	0,155	0,113	2,817
500	168	85	29	11	0,341	0,379	0,506	0,173	0,065	2,976
600	194	102	34	12	0,333	0,353	0,526	0,175	0,062	3,093
700	212	107	39	17	0,364	0,436	0,505	0,184	0,080	3,302
800	235	121	45	14	0,372	0,311	0,515	0,191	0,060	3,404
900	263	142	48	17	0,338	0,354	0,540	0,183	0,065	3,422
1000	280	141	62	19	0,440	0,306	0,504	0,221	0,068	3,571
1100	290	140	64	24	0,457	0,375	0,483	0,221	0,083	3,793
1200	312	146	74	24	0,507	0,324	0,468	0,237	0,077	3,846
1300	325	148	78	23	0,527	0,295	0,455	0,240	0,071	4,000
1400	341	159	76	27	0,478	0,360	0,466	0,223	0,079	4,106
1500	353	165	78	28	0,473	0,359	0,467	0,221	0,079	4,248
1600	367	168	80	33	0,476	0,413	0,458	0,218	0,090	4,360
1700	381	175	83	30	0,474	0,361	0,459	0,218	0,079	4,462
1800	389	174	84	32	0,483	0,381	0,447	0,216	0,082	4,627
1900	396	175	85	34	0,486	0,400	0,442	0,215	0,086	4,798
2000	403	173	90	34	0,520	0,378	0,429	0,223	0,084	4,963
2100	416	173	96	38	0,555	0,396	0,416	0,231	0,091	5,048
2200	425	177	95	44	0,537	0,463	0,416	0,224	0,104	5,176
2300	435	177	97	50	0,548	0,515	0,407	0,223	0,115	5,287
2400	444	179	99	53	0,553	0,535	0,403	0,223	0,119	5,405

TABLE 10 BRONWYN (1) FORWARD (CONTD.)

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N) = \frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N) = \frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N) = \frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
2500	455	180	103	55	0,572	0,534	0,396	0,226	0,121	5,495
2600	462	183	103	54	0,563	0,524	0,396	0,223	0,117	5,628
2700	468	182	106	55	0,582	0,519	0,389	0,226	0,118	5,769
2800	479	183	108	60	0,590	0,556	0,382	0,225	0,125	5,846
2900	487	183	108	61	0,590	0,565	0,376	0,222	0,126	5,955
3000	496	186	107	63	0,575	0,589	0,375	0,216	0,127	6,048
3100	505	188	108	60	0,574	0,556	0,372	0,214	0,119	6,139
3200	511	190	104	59	0,547	0,567	0,372	0,204	0,115	6,262
3300	516	194	102	61	0,526	0,598	0,376	0,198	0,118	6,395
3400	519	195	98	60	0,503	0,612	0,376	0,189	0,116	6,551
3500	527	196	100	60	0,510	0,600	0,372	0,190	0,114	6,641
3600	532	194	98	63	0,505	0,643	0,365	0,184	0,118	6,767
3700	539	197	96	61	0,497	0,622	0,365	0,182	0,113	6,865
3800	542	192	101	64	0,526	0,634	0,354	0,186	0,118	7,011
3900	552	198	102	64	0,515	0,527	0,359	0,185	0,115	7,065
4000	563	193	110	62	0,553	0,564	0,353	0,196	0,110	7,105
4100	567	195	113	63	0,579	0,558	0,344	0,198	0,111	7,231
4200	579	202	116	61	0,574	0,526	0,349	0,200	0,106	7,254
4300	587	199	120	65	0,603	0,542	0,339	0,204	0,111	7,325
4400	594	199	122	67	0,613	0,549	0,335	0,205	0,113	7,407
4500	601	201	123	67	0,612	0,545	0,334	0,205	0,111	7,488
4600	612	207	121	69	0,585	0,570	0,338	0,198	0,113	7,516
4700	620	209	122	71	0,584	0,582	0,337	0,197	0,115	7,581
4800	636	219	126	68	0,575	0,540	0,344	0,198	0,107	7,547
4900	644	221	128	69	0,578	0,539	0,343	0,199	0,107	7,609
5000	653	227	123	73	0,542	0,593	0,348	0,188	0,112	7,657

TABLE 10 BRONWYN (1) FORWARD(CONT.)

							$\hat{\phi}(1\|N)$	$\hat{\phi}(2\|N)$	$\hat{\phi}(3\|N)$	r_N
N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\frac{N}{\hat{n}_N}$
5100	657	229	122	72	0,533	0,590	0,349	0,186	0,110	7,763
5200	669	239	117	74	0,490	0,632	0,357	0,175	0,111	7,713
5300	680	245	117	74	0,478	0,632	0,360	0,172	0,109	7,794
5400	689	246	119	75	0,484	0,630	0,357	0,173	0,109	7,837
5500	697	245	123	78	0,502	0,634	0,352	0,176	0,112	7,891
5600	711	257	122	77	0,475	0,631	0,361	0,172	0,108	7,876
5700	718	257	125	78	0,486	0,624	0,358	0,174	0,109	7,939
5800	727	260	126	82	0,485	0,651	0,358	0,173	0,113	7,978
5900	736	263	128	80	0,487	0,625	0,357	0,174	0,109	8,016
6000	745	268	128	79	0,478	0,617	0,360	0,172	0,106	8,054
6100	749	270	127	80	0,470	0,630	0,360	0,170	0,107	8,144
6200	762	274	132	81	0,482	0,614	0,360	0,173	0,106	8,136
6300	762	271	131	79	0,483	0,603	0,356	0,172	0,104	8,268
6400	765	269	130	80	0,483	0,615	0,352	0,170	0,105	8,3[illegible]
6500	774	272	132	80	0,485	0,606	0,351	0,171	0,103	8,398
6600	778	272	129	83	0,474	0,643	0,350	0,166	0,107	8,483
6700	785	273	132	81	0,484	0,614	0,348	0,168	0,103	8,6[illegible]2
6800	792	276	132	81	0,478	0,614	0,348	0,165	0,102	8,586
6900	798	273	136	80	0,498	0,588	0,342	0,170	0,100	8,647
7000	801	272	136	79	0,500	0,581	0,340	0,170	0,099	8,864
7100	807	273	141	77	0,516	0,546	0,338	0,175	0,095	8,798
7200	814	275	143	77	0,520	0,538	0,338	0,176	0,095	8,845
7300	823	281	143	74	0,509	0,517	0,341	0,174	0,090	8,870
7400	827	277	148	73	0,536	0,493	0,335	0,179	0,088	8,948
7500	833	277	151	72	0,545	0,477	0,333	0,181	0,086	9,004
7600	840	277	152	77	0,549	0,507	0,330	0,181	0,092	9,048
7700	844	278	151	77	0,543	0,510	0,329	0,179	0,091	9,123
7800	852	282	151	79	0,535	0,523	0,331	0,177	0,093	9,155
7900	859	284	154	80	0,542	0,519	0,331	0,179	0,093	9,197
8000	868	291	154	78	0,529	0,506	0,335	0,177	0,090	9,217
8057	871	291	156	79						

TABLE 11 BRONWYN (1) BACKWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{f}_N=\frac{N}{\hat{n}_N}$
100	58	36	12	4	0,3333	0,3333	0,6207	0,2069	0,0690	1,7241
200	99	63	17	6	0,2698	0,3529	0,6364	0,1717	0,0606	2,0202
300	135	81	23	10	0,2840	0,4348	0,6000	0,1704	0,0741	2,2222
400	165	95	26	16	0,2737	0,6154	0,5758	0,1576	0,0970	2,4242
500	190	110	29	12	0,2636	0,4138	0,5789	0,1526	0,0632	2,6316
600	218	126	35	12	0,2778	0,3429	0,5780	0,1606	0,0550	2,7523
700	239	134	42	14	0,3134	0,3333	0,5607	0,1757	0,0586	2,9289
800	266	151	42	17	0,2781	0,4048	0,5677	0,1579	0,0639	3,0075
900	287	160	51	19	0,3188	0,3725	0,5575	0,1777	0,0662	3,1359
1000	302	163	58	19	0,3558	0,3276	0,5397	0,1921	0,0629	3,3113
1100	317	165	62	23	0,3758	0,3710	0,5205	0,1956	0,0726	3,4700
1200	341	171	74	23	0,4327	0,3108	0,5015	0,2170	0,0674	3,5191
1300	364	185	75	27	0,4054	0,3600	0,5082	0,2060	0,0742	3,5714
1400	377	190	75	31	0,3947	0,4133	0,5040	0,1989	0,0822	3,7135
1500	390	194	79	31	0,4072	0,3924	0,4974	0,2026	0,0795	3,8462
1600	404	195	82	39	0,4205	0,4756	0,4827	0,2030	0,0965	3,9604
1700	417	202	79	44	0,3911	0,5570	0,4844	0,1894	0,1055	4,0767
1800	431	212	78	47	0,3679	0,6026	0,4919	0,1810	0,1090	4,1763
1900	442	216	81	49	0,3750	0,6049	0,4887	0,1833	0,1109	4,2986
2000	456	217	85	55	0,3917	0,6471	0,4759	0,1864	0,1206	4,3860
2100	464	213	90	57	0,4225	0,6333	0,4591	0,1940	0,1228	4,5259
2200	478	217	93	58	0,4286	0,6237	0,4540	0,1946	0,1213	4,6025
2300	496	228	94	61	0,4123	0,6489	0,4597	0,1895	0,1230	4,6371

TABLE 11 BRONWYN (I) BACKWARD (CONTD.)

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
2400	511	237	91	67	0,3840	0,7363	0,4638	0,1781	0,1311	4,6967
2500	519	237	91	67	0,3840	0,7363	0,4566	0,1753	0,1291	4,8170
2600	536	248	88	72	0,3548	0,8182	0,4627	0,1642	0,1343	4,8507
2700	552	253	90	79	0,3557	0,8778	0,4583	0,1630	0,1431	4,8913
2800	564	259	91	79	0,3514	0,8681	0,4592	0,1613	0,1401	4,9645
2900	576	260	98	78	0,3769	0,7959	0,4514	0,1701	0,1354	5,0347
3000	586	261	104	75	0,3985	0,7212	0,4454	0,1775	0,1280	5,1195
3100	594	264	103	77	0,3902	0,7476	0,4444	0,1734	0,1498	5,2189
3200	606	267	109	76	0,4082	0,6972	0,4406	0,1799	0,1254	5,2805
3300	619	271	114	74	0,4207	0,6491	0,4378	0,1842	0,1195	5,3312
3400	634	281	116	73	0,4128	0,6293	0,4432	0,1830	0,1151	5,3628
3500	646	285	120	72	0,4211	0,6000	0,4412	0,1858	0,1115	5,4180
3600	655	285	125	70	0,4386	0,5600	0,4351	0,1908	0,1069	5,4962
3700	665	291	125	72	0,4296	0,5760	0,4376	0,1880	0,1083	5,5639
3800	673	291	130	69	0,4467	0,5308	0,4568	0,1932	0,1025	5,6464
3900	683	296	129	70	0,4358	0,5426	0,4334	0,1889	0,1025	5,7101
4000	689	301	121	76	0,4020	0,6281	0,4369	0,17[illegible]6	0,1103	5,8055
4100	695	302	123	75	0,4073	0,6098	0,4345	0,1770	0,1079	5,8993
4200	703	302	125	77	0,4139	0,6160	0,4296	0,1778	0,1095	5,9744
4300	707	303	122	76	0,4026	0,6230	0,4286	0,1726	0,1075	6,0820
4400	713	303	125	77	0,4125	0,6160	0,4250	0,1753	0,1080	6,1711
4500	717	300	126	77	0,4200	0,6111	0,4184	0,1757	0,1074	6,2762

TABLE 11 BRONWYN (1) BACKWARD (CONTD.)

							$\hat{\phi}(1\|N)$	$\hat{\phi}(2\|N)$	$\hat{\phi}(3\|N)$	
N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N=\frac{N}{\hat{n}_N}$
4600	726	305	125	78	0,4098	0,6240	0,4201	0,1722	0,1074	6,3361
4700	728	305	126	77	0,4131	0,6111	0,4190	0,1731	0,1058	6,4560
4800	728	301	129	77	0,4286	0,5969	0,4135	0,1772	0,1058	6,5934
4900	731	299	128	78	0,4281	0,6094	0,4090	0,1751	0,1067	6,7031
5000	736	300	126	80	0,4200	0,6349	0,4104	0,1712	0,1087	6,7935
5100	741	297	131	79	0,4411	0,6031	0,4008	0,1768	0,1065	6,8826
5200	747	297	130	79	0,4377	0,6077	0,3976	0,1740	0,1058	6,9612
5300	749	292	133	79	0,4555	0,5940	[illegible]	0,1776	0,1055	7,0761
5400	755	294	132	79	0,4490	0,5985	0,3894	0,1748	0,1046	7,1523
5500	759	293	136	75	0,4642	0,5515	0,3860	0,1792	0,0988	7,2464
5600	764	294	137	74	0,4660	0,5401	0,3848	0,1793	0,0969	7,3298
5700	772	299	139	71	0,4649	0,5108	0,3873	0,1801	0,0920	7,3834
5800	775	296	139	75	0,4696	[illegible]	0,3819	0,1794	0,0968	7,4839
5900	784	300	138	77	0,4600	0,5580	0,3827	0,1760	0,0982	7,5255
6000	790	300	142	78	0,4733	0,5493	0,3797	0,1797	0,0987	7,5949
6100	793	299	141	80	0,4716	0,5674	0,3770	0,1778	0,1009	7,6923
6200	796	299	138	83	0,4615	0,6014	0,3756	0,1734	0,1043	7,7889
6300	797	298	137	80	0,4597	0,5839	0,3739	0,1719	0,1004	7,9046
6400	802	298	139	81	0,4664	0,5827	0,3716	0,1733	0,1010	7,9800
6500	805	296	140	81	0,4730	0,5786	0,3677	0,1739	0,1006	8,0745
6600	813	299	142	81	0,4749	0,5704	0,3678	0,1747	0,0996	8,1181
6700	815	296	143	83	0,4831	0,5804	0,3632	0,1755	0,1018	8,2209
6800	818	292	143	87	0,4897	0,6084	0,3570	0,1748	0,1064	8,3130

TABLE 11 BRONWYN (1) BACKWARD (CONTD.)

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\bar{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\bar{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
6900	827	296	143	97	0,4831	0,6084	0,3579	0,1729	0,1052	8,3434
7000	831	297	143	85	0,4815	0,5944	0,3574	0,1721	0,1023	8,4236
7100	832	293	145	86	0,4849	0,5931	0,3522	0,1743	0,1034	8,5337
7200	836	293	149	80	0,5085	0,5369	0,3505	0,1782	0,0957	8,6124
7300	841	291	149	85	0,5120	0,5705	0,3460	0,1772	0,1011	8,6801
7400	845	293	148	82	0,5051	0,5541	0,3467	0,1751	0,0970	8,7574
7500	853	295	149	80	0,5051	0,5369	0,3458	0,1747	0,0938	8,7925
7600	855	291	154	78	0,5292	[illegible]	0,3404	0,1801	0,0912	8,8889
7700	859	290	157	74	0,5414	[illegible]	0,3376	0,1828	0,0871	8,9639
7800	863	292	155	78	0,5308	[illegible]	0,3384	0,1796	0,0904	9,1224
7900	866	292	155	77	0,5308	0,4968	0,3372	0,1790	0,0889	9,0382
8000	869	290	157	78	0,5414	0,4968	0,3337	0,1807	0,0898	9,2060
8037	871	291	156	79	0,5361	0,5064	0,3341	0,1791	0,0907	9,2273

TABLE 12 BRONWYN (1) RANDOM FORWARD

							$\hat{\phi}(1\|N)$	$\hat{\phi}(2\|N)$	$\hat{\phi}(3\|N)$	
N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
100	70	53	10	4	0,1887	0,4000	0,7571	0,1429	0,0571	1,4286
200	112	75	21	5	0,2800	0,2381	0,6696	0,1875	0,0446	1,7857
300	155	100	24	15	0,2400	0,6250	0,6452	0,1548	0,0968	1,9355
400	179	110	25	16	0,2273	0,6400	0,6145	0,1397	0,0894	2,2346
500	199	115	32	15	0,2783	0,4688	0,5779	0,1608	0,0754	2,5126
600	224	129	38	16	0,2946	0,4211	0,5759	0,1696	0,0714	2,6786
700	247	143	37	23	0,2587	0,6216	0,5789	0,1498	0,0931	2,8340
800	271	157	42	26	0,2675	0,6190	0,5793	0,1550	0,0959	2,9520
900	296	170	51	23	0,3000	0,4510	0,5743	0,1723	0,0777	3,0405
1000	310	173	57	20	0,3295	0,3509	0,5581	0,1839	0,0645	3,2258
1100	329	179	63	24	0,3520	0,3810	0,5441	0,1915	0,0729	3,3435
1200	344	178	75	24	0,4213	0,3200	0,5174	0,2180	0,0698	3,4884
1300	359	184	76	28	0,4130	0,3684	0,5125	0,2117	0,0780	3,6212
1400	376	196	74	31	0,3776	0,4189	0,5213	0,1968	0,0824	3,7234
1500	395	208	76	33	0,3654	0,4342	0,5266	0,1924	0,0835	3,7975
1600	411	216	77	37	0,3565	0,4805	0,5255	0,1873	0,0900	3,8929
1700	422	222	75	39	0,3378	0,5200	0,5261	0,1777	0,0920	4,0284
1800	436	226	79	42	0,3496	0,5316	0,5183	0,1812	0,0963	4,1284
1900	444	221	88	41	0,3982	0,4659	0,4977	0,1982	0,0923	4,2793
2000	461	227	93	41	0,4097	0,4409	0,4924	0,2017	0,0889	4,3384
2100	473	229	93	45	0,4061	0,4839	0,4841	0,1966	0,0951	4,4397
2200	483	232	90	51	0,3879	0,5667	0,4803	0,1863	0,1056	4,5549
2300	490	227	97	50	0,4273	0,5155	0,4633	0,1980	0,1020	4,6939

- 2 -

TABLE 12 BRONWYN (1) RANDOM FORWARD (CONTD.)

							$\hat{\phi}(1\|N)$	$\hat{\phi}(2\|N)$	$\hat{\phi}(3\|N)$	
	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
2400	500	227	101	51	0,4449	0,5050	0,4540	0,2020	0,1020	4,8000
2500	511	230	104	51	0,4522	0,4904	0,4501	0,2035	0,0998	4,8924
2600	519	232	98	57	0,4224	0,5816	0,4470	0,1888	0,1098	5,0096
2700	529	234	102	56	0,4359	0,5490	0,4423	0,1928	0,1059	5,1040
2800	538	235	107	54	0,4553	0,5047	0,4368	0,1989	0,1004	5,2045
2900	547	238	111	51	0,4664	0,4595	0,4351	0,2029	0,0932	5,3016
3000	559	244	108	56	0,4426	0,5185	0,4365	0,1932	0,1002	5,3667
3100	569	245	114	54	0,4653	0,4737	0,4306	0,2004	0,0949	5,4482
3200	574	244	110	61	0,4508	0,5545	0,4251	0,1916	0,1063	5,5749
3300	581	244	113	61	0,4631	0,5398	0,4200	0,1945	0,1050	5,6799
3400	594	252	116	59	0,4603	0,5086	0,4242	0,1953	0,0993	5,7239
3500	600	252	116	62	0,4603	0,5345	0,4200	0,1933	0,1033	5,8333
3600	611	256	117	66	0,4570	0,5641	0,4190	0,1915	0,1080	5,8920
3700	618	257	117	65	0,4553	0,5556	0,4159	0,1893	0,1052	5,9871
3800	621	250	124	64	0,4960	0,5161	0,4026	0,1997	0,1031	6,1192
3900	636	260	124	64	0,4769	0,5161	0,4088	0,1950	0,1006	6,1321
4000	647	262	132	64	0,5038	0,5203	0,4049	0,2040	0,0989	6,1824
4100	651	262	127	72	0,4847	0,5669	0,4025	0,1951	0,1106	6,2980
4200	658	265	122	81	0,4604	0,6639	0,4027	0,1854	0,1231	6,3830
4300	665	264	125	76	0,4735	0,6080	0,3970	0,1880	0,1143	6,4662
4400	671	264	125	80	0,4735	0,6400	0,3934	0,1863	0,1192	6,5574
4500	674	261	122	86	0,4674	0,7049	0,3872	0,1810	0,1276	6,6766
4600	679	261	124	82	0,4751	0,6613	0,3844	0,1826	0,1208	6,7747

- 3 -

TABLE 12 BRONWYN (1) RANDOM FORWARD (CONTD.)

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
4700	684	261	125	82	0,4789	0,6560	0,3816	0,1827	0,1199	6,8713
4800	690	261	126	82	0,4828	0,6508	0,3783	0,1826	0,1188	6,9565
4900	699	264	127	84	0,4811	0,6614	0,3777	0,1817	0,1202	7,0100
5000	708	265	130	80	0,4906	0,6154	0,3743	0,1836	0,1130	7,0621
5100	714	263	134	84	0,5095	0,6269	0,3683	0,1877	0,1176	7,1429
5200	725	268	138	82	0,5149	0,5942	0,3697	0,1903	0,1131	7,1724
5300	732	272	134	86	0,4926	0,6418	0,3716	0,1831	0,1176	7,2404
5400	734	269	136	83	0,5056	0,6103	0,3665	0,1853	0,1131	7,3569
5500	741	270	141	81	0,5222	0,5745	0,3644	0,1903	0,1093	7,4224
5600	750	276	141	82	0,5109	0,5816	0,3680	0,1880	0,1093	7,4667
5700	754	274	145	79	0,5292	0,5448	0,3634	0,1923	0,1048	7,5597
5800	760	278	141	82	0,5072	0,5816	0,3658	0,1855	0,1079	7,6316
5900	763	277	142	78	0,5126	0,5493	0,3630	0,1861	0,1022	7,7326
6000	772	285	139	78	0,4877	0,5612	0,3692	0,1801	0,1010	7,7720
6100	776	283	142	78	0,5018	0,5493	0,3647	0,1830	0,1005	7,8608
6200	781	283	147	74	0,5194	0,5034	0,3624	0,1882	0,0948	7,9385
6300	788	284	150	72	0,5282	0,4800	0,3604	0,1904	0,0914	7,9949
6400	796	289	151	72	0,5225	0,4768	0,3631	0,1897	0,0905	8,0402
6500	798	289	148	75	0,5121	0,5068	0,3622	0,1855	0,0940	8,1454
6600	801	291	145	78	0,4983	0,5379	0,3633	0,1810	0,0974	8,2397
6700	805	292	146	76	0,5000	0,5205	0,3627	0,1814	0,0944	8,3230
6800	812	292	146	79	0,5000	0,5411	0,3596	0,1798	0,0973	8,3744
6900	818	289	146	80	0,5052	0,5479	0,3533	0,1785	0,0978	8,4352

TABLE 12 BRONWYN (1) RANDOM FORWARD (CONTD.)

							$\hat{\phi}(1\|N)$	$\hat{\phi}(2\|N)$	$\hat{\phi}(3\|N)$	
N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
7000	824	290	146	83	0,5034	0,5685	0,3519	0,1772	0,1007	8,4951
7100	826	289	143	85	0,4948	0,5944	0,3499	0,1732	0,1029	8,5956
7200	832	291	142	82	0,4880	0,5775	0,3498	0,1707	0,0985	8,6538
7300	835	287	147	82	0,5122	0,5578	0,3437	0,1760	0,0982	8,7425
7400	842	288	149	83	0,5174	0,5570	0,3420	0,1770	0,0986	8,7886
7500	846	289	149	85	0,5156	0,5705	0,3416	0,1761	0,1005	8,8652
7600	854	290	150	88	0,5172	0,5867	0,3396	0,1756	0,1030	8,8993
7700	856	289	150	88	0,5190	0,5867	0,3376	0,1752	0,1029	8,9953
7800	863	293	151	85	0,5154	0,5629	0,3395	0,1750	0,0985	9,0382
7900	866	289	156	80	0,5398	0,5128	0,3337	0,1801	0,0924	9,1224
8000	869	289	157	78	0,5433	0,4968	0,3326	0,1807	0,0898	9,2060
8037	871	291	156	79	0,5361	0,5064	0,3341	0,1791	0,0907	9,2273

TABLE 13 BRONWYN (1) RANDOM BACKWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
100	65	41	17	5	0,4146	0,2941	0,6308	0,2615	0,0769	1,5385
200	112	77	14	10	0,1818	0,7143	0,6875	0,1250	0,0893	1,7857
300	143	91	19	10	0,2088	0,5263	0,6276	0,1329	0,0699	2,0978
400	173	102	30	13	0,2941	0,4333	0,5896	0,1734	0,0751	2,3121
500	216	136	30	18	0,2206	0,6000	0,6296	0,1389	0,0833	2,3148
600	240	149	31	20	0,2081	0,6452	0,6208	0, 292	0,0833	2,5000
700	263	160	40	16	0,2500	0,4000	0,6084	0,1521	0,0608	2,6616
800	288	177	40	16	0,2260	0,4000	0,6146	0,1389	0,0556	2,7778
900	314	195	42	19	0,2154	0,4524	0,6210	0,1338	0,0605	2,8662
1000	331	197	51	23	0,2589	0,4510	0,5952	0,1541	0,0695	3,0211
1100	351	206	54	27	0,2621	0,5000	0,5869	0,1538	0,0769	3,1338
1200	369	212	58	31	0,2736	0,5345	0,5745	0,1572	0,0840	3,2520
1300	383	212	62	36	0,2925	0,5806	0,5535	0,1619	0,0940	3,3943
1400	394	214	65	33	0,3037	0,5077	0,5431	0,1650	0,0838	3,5533
1500	404	211	71	39	0,3365	0,5493	0,5223	0,1757	0,0965	3,7129
1600	415	215	71	41	0,3302	0,5775	0,5181	0,1711	0,0988	3,8554
1700	429	218	77	33	0,3532	0,4286	0,5082	0,1795	0,0769	3,9627
1800	445	225	80	45	0,3556	0,5625	0,5056	0,1798	0,1011	4,0448
1900	461	232	84	45	0,3621	0,5357	0,5033	0,1822	0,0976	4,1215
2000	471	232	88	46	0,3793	0,5227	0,4926	0,1868	0,0977	4,2463
2100	481	229	94	46	0,4105	0,4894	0,4761	0,1954	0,0956	4,3659
2200	494	236	96	48	0,4068	0,5000	0,4777	0,1943	0,0972	4,4534
2300	502	239	93	52	0,3831	0,5591	0,4761	0,1853	0,1036	4,5817

TABLE 13 BRONWYN (1) RANDOM BACKWARD (CONTD.)

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
2400	513	243	93	52	0,3827	0,5591	0,4737	0,1813	[illegible]	4,6784
2500	526	248	94	54	0,3790	0,5745	0,4715	0,1787	[illegible]	4,7529
2600	538	253	96	51	0,3794	0,5313	0,4703	0,1784	0,0948	4,8327
2700	550	255	101	53	0,3961	0,5248	0,4636	0,1836	0,0964	4,9091
2800	554	251	103	54	0,4104	0,5243	0,4531	0,1859	[illegible]	5,0542
2900	567	253	108	59	0,4269	0,5463	0,4462	0,1905	0,1041	5,1146
3000	580	259	111	61	0,4286	0,5495	0,4466	0,1914	0,1052	5,1724
3100	592	264	112	64	0,4242	0,5714	0,4459	0,1892	0,1081	5,2365
3200	607	272	113	66	0,4154	0,5841	0,4481	0,1862	0,1087	5,2718
3300	615	269	122	65	0,4535	0,5328	0,4374	0,1984	0,1057	5,3659
3400	624	273	119	70	0,4359	0,5882	0,4375	0,1907	0,1122	5,4487
3500	629	271	123	70	0,4539	0,5691	0,4308	0,1955	0,1113	5,5644
3600	637	273	126	66	0,4615	0,5238	0,4285	0,1978	0,1036	5,6515
3700	644	277	126	64	0,4549	0,5079	0,4301	0,1957	0,0994	5,7453
3800	651	280	123	68	0,4393	0,5528	0,4301	0,1889	0,1045	5,8372
3900	660	282	125	64	0,4433	0,5120	0,4273	0,1894	0,0970	5,9091
4000	664	278	130	62	0,4676	0,4769	0,4187	0,1958	0,0934	6,0241
4100	669	276	133	64	0,4819	[illegible]	0,4126	0,1988	0,0957	6,1286
4200	682	283	132	68	0,4664	0,5152	0,4150	[illegible]	0,0997	6,1584
4300	690	288	129	71	0,4479	0,5504	0,4174	0,1870	0,1029	6,2319
4400	696	290	127	68	0,4379	0,5354	0,4176	0,1825	0,0977	6,3218
4500	703	288	132	67	0,4583	0,5076	0,4097	0,1878	0,0953	6,4011
4600	710	291	131	66	0,4502	0,5038	0,4099	0,1845	0,0930	6,4789

TABLE 13 BRONWYN (1) RANDOM BACKWARD (CONTD.)

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_{\bar{N}} = \frac{N}{\hat{n}_N}$
4700	718	296	130	67	0,4392	0,5154	0,4123	0,1811	0,0933	6,5460
4800	724	294	134	71	0,4558	0,5299	0,4061	0,1851	0,0981	6,6298
4900	728	294	134	73	0,4558	0,5448	0,4038	0,1841	0,1003	6,7308
5000	739	300	135	69	0,4500	0,5111	0,4050	0,1827	0,0934	6,7659
5100	744	296	139	72	0,4696	0,518	0,3978	0,1868	0,0968	6,8548
5200	750	297	140	73	0,4714	0,5214	0,3960	0,1867	0,0973	6,9333
5300	756	300	139	74	0,4633	0,5324	0,3968	0,1839	0,0979	7,0106
5400	761	300	141	74	0,4700	0,5248	0,3942	0,1853	0,0972	7,0959
5500	764	298	140	75	0,4698	0,5357	0,3901	0,1832	0,0982	7,1990
5600	770	300	138	77	0,4600	0,5580	0,3896	0,1792	0,1000	7,2727
5700	774	300	135	80	0,4500	0,5926	0,3876	0,1615	0,1034	7,3643
5800	779	300	139	80	0,4633	0,5755	0,3851	0,1784	0,1027	7,4454
5900	786	304	140	78	0,4605	0,5571	0,3868	0,1781	0,0992	7,5064
6000	790	305	138	81	0,4525	0,5870	0,3861	0,1747	0,1025	7,5949
6100	795	307	135	84	0,4397	0,6222	0,3862	0,1698	0,1057	7,6730
6200	799	303	136	86	0,4488	0,6324	0,3797	0,1704	0,1078	7,7694
6300	803	303	138	88	0,4554	0,6377	0,3773	0,171	0,1096	7,8456
6400	808	305	137	87	0,4492	0,6350	0,3775	0,1696	0,1077	7,9208
6500	814	307	140	85	0,4560	0,6071	0,3771	0,1720	0,1044	7,9853
6600	819	308	142	81	0,4610	0,5704	0,3761	0,1734	0,0989	8,0586
6700	826	308	146	81	0,4740	0,5548	0,3729	0,1768	0,0981	8,1114
6800	830	307	148	82	0,4821	0,5541	0,3699	0,1783	0,0988	8,1928
6900	837	311	145	86	0,4662	0,5931	0,3716	0,1732	0,1027	8,2437

TABLE 13 BRONWYN (1) RANDOM BACKWARD (CONTD.)

							$\hat{\phi}(1\|N)$	$\hat{\phi}(2\|N)$	$\hat{\phi}(3\|N)$	
N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N = \frac{N}{\hat{n}_N}$
7000	838	308	146	85	0,4740	0,5822	0,3675	0,1742	0,1014	8,3532
7100	840	308	144	84	0,4675	0,5833	0,3667	0,1714	0,1000	8,4524
7200	845	306	149	81	0,4869	0,5436	0,3621	0,1763	0,0959	8,5207
7300	850	308	150	79	0,4870	0,5267	0,3624	0,1765	0,0929	8,5882
7400	854	304	156	76	0,5132	0,4872	0,3560	0,1827	0,0890	8,6651
7500	858	305	155	76	0,5082	0,4903	0,3555	0,1807	0,0886	8,7413
7600	861	305	156	75	0,5115	0,4608	0,3542	0,1812	0,0871	8.8269
7700	862	301	158	72	[illegible]5249	0,4557	0,3492	0,1833	0,0835	8,9327
7800	865	300	154	76	[illegible]133	0,4939	0,3468	0,1780	0,0879	9,0173
7900	866	297	155	71	[illegible]219	0,4581	0,3430	0,1790	0,0930	9,1224
8000	869	293	154	77	0,5256	0,5000	0,3372	0,1772	0,0826	9,2060
8037	871	291	156	79	0,5361	0,5064	0,3341	0,1791	0,09[illegible]	9,2273

TABLE 14 BRONWYN (2) FORWARD

N	n_N	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\bar{f}_1(2\|N)}{\bar{f}_1(1\|N)}$	$\frac{\bar{f}_1(3\|N)}{\bar{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}(3\|N)}{\hat{n}_N}$	$\bar{r}_N$ $\frac{N}{\hat{n}_N}$
100	64	46	9	6	0,130	0,667	0,719	0,141	0,094	1,563
200	97	62	12	9	0,194	0,750	0,639	0,124	0,093	2,062
300	123	72	22	7	0,306	0,318	0,585	0,179	0,057	2,439
400	156	91	30	11	0,330	0,367	0,583	0,192	0,071	2,564
500	186	106	36	17	0,340	0,472	0,570	0,194	0,091	2,688
600	205	114	36	21	0,316	0,583	0,556	0,176	0,102	2,927
700	226	119	45	22	0,378	0,489	0,527	0,199	0,097	3,097
800	251	136	46	21	0,338	0,457	0,542	0,183	0,084	3,187
900	272	150	50	20	0,333	0,400	0,551	0,184	0,074	3,309
1000	294	161	55	23	0,342	0,418	0,548	0,187	0,078	3,401
1100	307	164	59	26	0,360	0,441	0,534	0,192	0,085	3,583
1200	321	164	67	28	0,405	0,418	0,511	0,209	0,087	3,738
1300	326	165	75	28	0,455	0,373	0,506	0,230	0,086	3,988
1400	351	174	74	31	0,425	0,419	0,496	0,211	0,088	3,989
1500	359	172	77	35	0,448	0,455	0,479	0,214	0,097	4,178
1600	369	176	76	37	0,432	0,487	0,477	0,206	0,100	4,336
1700	386	183	76	41	0,415	0,539	0,474	0,197	0,106	4,404
1800	400	186	80	42	0,430	0,525	0,465	0,200	0,105	4,500
1900	418	197	79	48	0,401	0,608	0,471	0,189	0,115	4,545
2000	430	202	79	50	0,391	0,633	0,470	0,184	0,116	4,651
2100	444	205	88	46	0,429	0,523	0,462	0,198	0,104	4,730
2200	457	211	90	46	0,427	0,511	0,462	0,197	0,101	4,814
2300	468	215	91	48	0,423	0,527	0,459	0,194	0,103	4,915
2400	485	220	95	52	0,432	0,547	0,454	0,196	0,107	4,948
2500	497	221	103	54	0,466	0,524	0,445	0,207	0,109	5,030
2600	511	227	109	53	0,480	0,486	0,444	0,213	0,104	5,088
2700	526	236	109	54	0,462	0,495	0,449	0,207	0,103	5,133
2800	533	258	109	58	0,458	0,532	0,447	0,205	0,109	5,253

The ratios $\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$ and $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ are shown on the same set of axes.

The proportions $\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ appear together on the same set of axes.

$\bar{r}_N = \frac{N}{\hat{n}_N}$ appears by itself.

The graphs give a clear picture of the trends described in paragraph 5.9.

$\hat{n}_N$ shows a monotonic increase with increasing N and the steadily decreasing slope of the graph can be clearly seen.

Sudden changes in slope can be seen in the case of Sandré, who on being reminded to use as many words as possible, made lists principally between N = 9 000 and N = 10 000 and in this interval there is a marked increase in $\hat{n}_N$ and concomitant increase in the slope of the graph. As soon as the listing stops, the graph resumes its decreasing slope.

$\hat{f}_1(1|N)$ shows a generally increasing trend with decreasing slope.

$\hat{f}_1(2|N)$ shows an increasing trend with decreasing slope.

$\hat{f}_1(3|N)$ shows an increasing trend which falls off very rapidly. In the case of Kerry 1 and Kerry 2 all four graphs i.e. forward, backward, random forward, random backward from each text flatten out from about N = 800 onward.

In the case of Bronwyn 1 and 2 the graphs flatten out at about N = 2 500.

In Sandré's case the graph flattened from N = 6 600 onward but rose between N = 11 800 and N = 13 000.

The ratios $\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ show heavy scatter. In no case can any reliable trend be observed.

The graphs of the younger children appear to wander about aimlessly between ordinates of about 0,4 and 0,7.

In the case of Sandré the limits appear to be about 0,3 and 0,6.

The proportion $\hat{\phi}(1|N) = \dfrac{\hat{f}_1(1|N)}{\hat{n}_N}$ showed a general decline in every case, with a decreasing slope. The larger the text the flatter the graph became.

In the case of Bronwyn the lower limit observed was approximately 0,33 and in the case of Sandré 0,4.

The proportions $\hat{\phi}(2|N) = \dfrac{\hat{f}_1(2|N)}{\hat{n}_N}$ and $\hat{\phi}(3|N) = \dfrac{\hat{f}_1(3|N)}{\hat{n}_N}$ showed only deviations about a mean line in all cases. As the text length increased so the deviations decreased.

The average appearance rate $\bar{r}_N = \dfrac{N}{\hat{n}_N}$. In all cases the graph showed a general upward trend with local inversions. As N increases so the slope of the graph decreases. This is true in all cases, with local deviations. The deviations are slight and do not markedly affect the line.

5.11 The Tables: Type 2, Calculations

The programs of the Graduate School of Business Administration of the University of the Witwatersrand, were used to obtain estimates of the total vocabulary of the children using the equations as detailed in Chapter 4.11.

The results of the various steps are shown in the tables in vertical columns against the size of text, N.

Sandré's calculation tables differ from those of the younger children and are described separately in 5.12.

The tables are numbered 17 to 29 and appear on pages 150 to 173.

Column 1 shows, under the heading, N, the number of hundreds of words of text used. The one exception is Sandré, in which case the number of text words under N, appears at two hundred word intervals.

Column 2. Heading $\bar{r}$.

The average number of appearances of the different words (types) in that length of text. The values are shown to five figures.

Column 3. Heading $\hat{\phi}(1|N)$

The proportion of words appearing once only in a text N words long. The values are quoted to four figures.

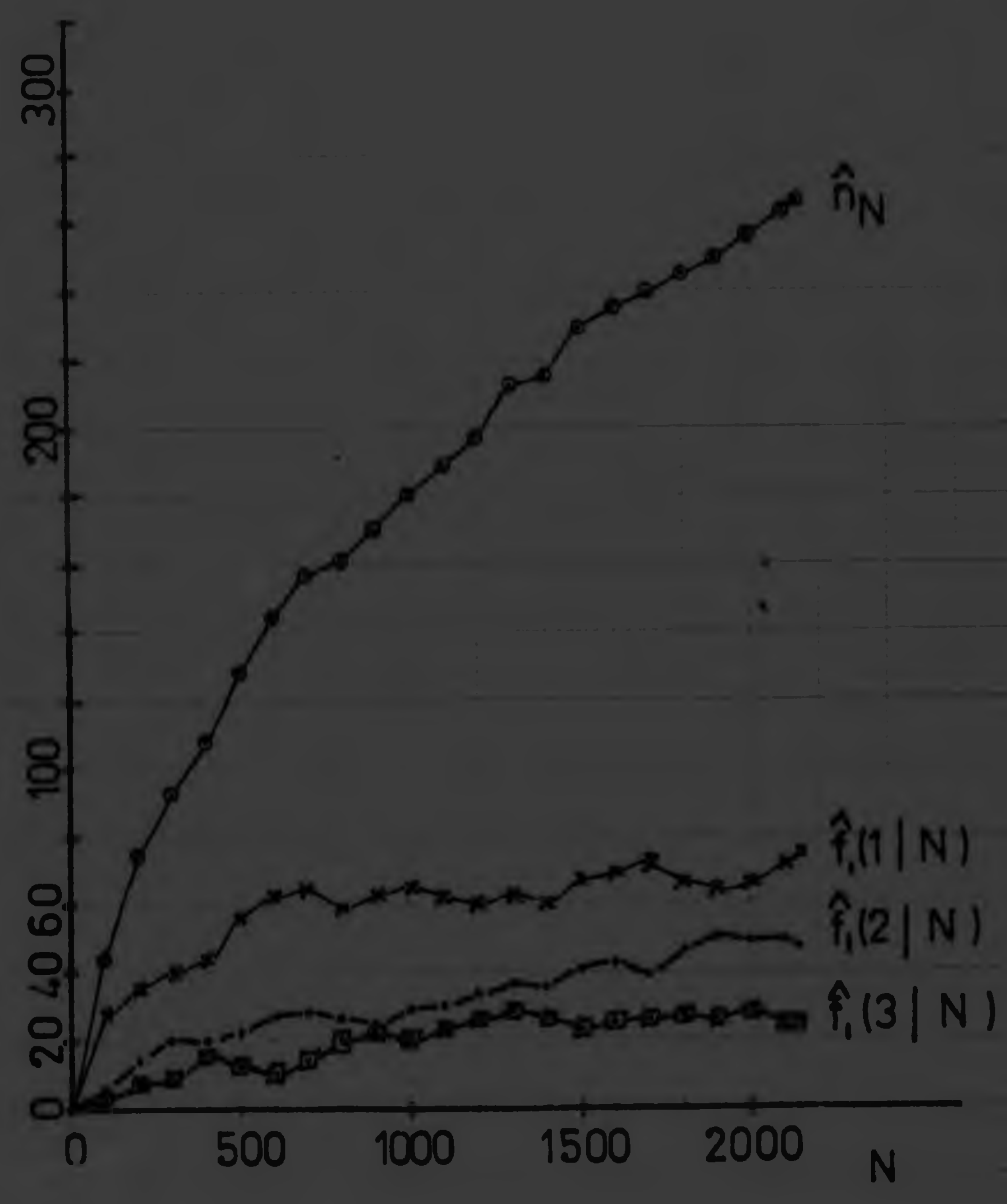


FIGURE 11 KERRY 1 FORWARD
$\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$, $\hat{n}_N$ vs N

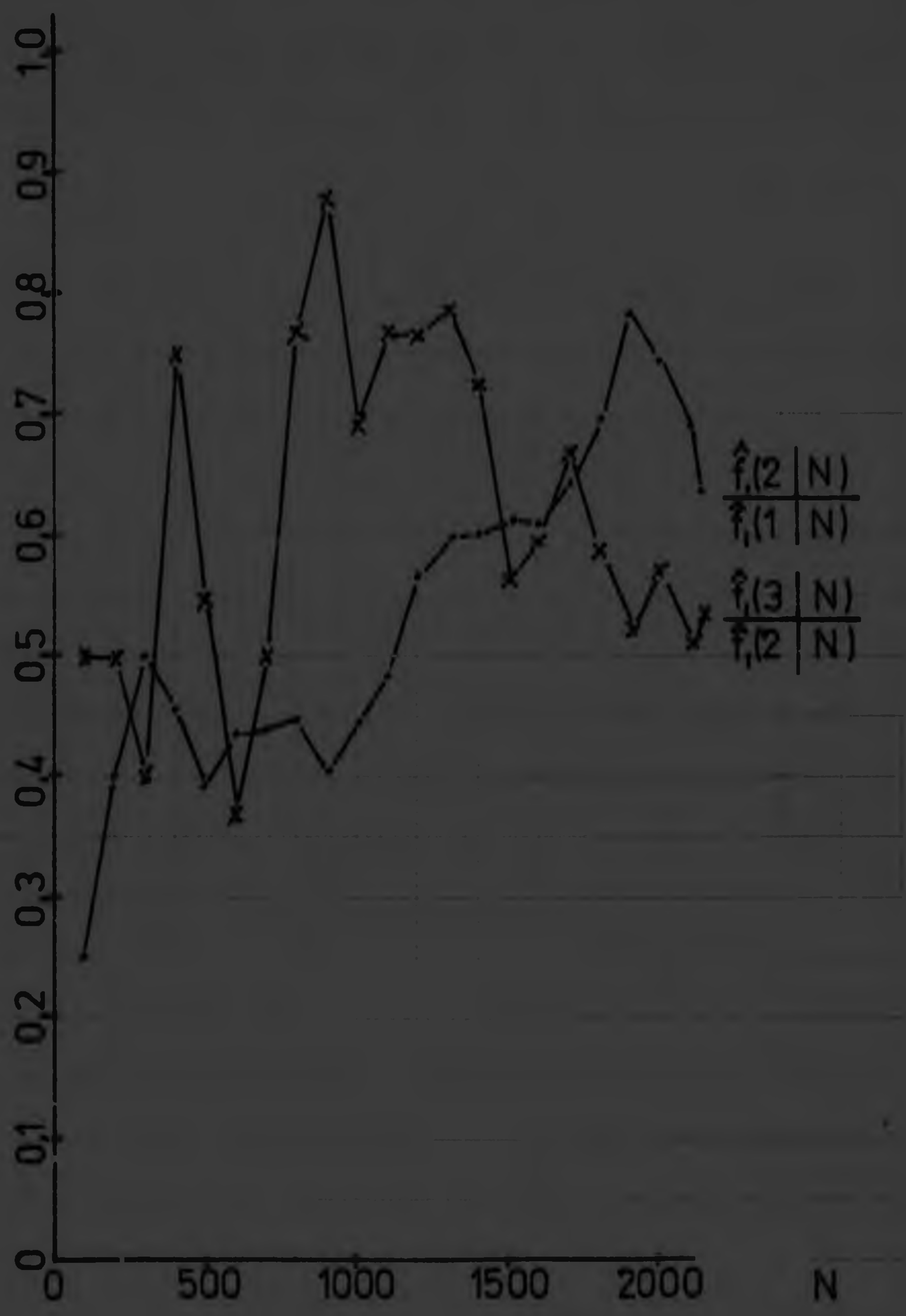


FIGURE 12 KERRY 1 FORWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

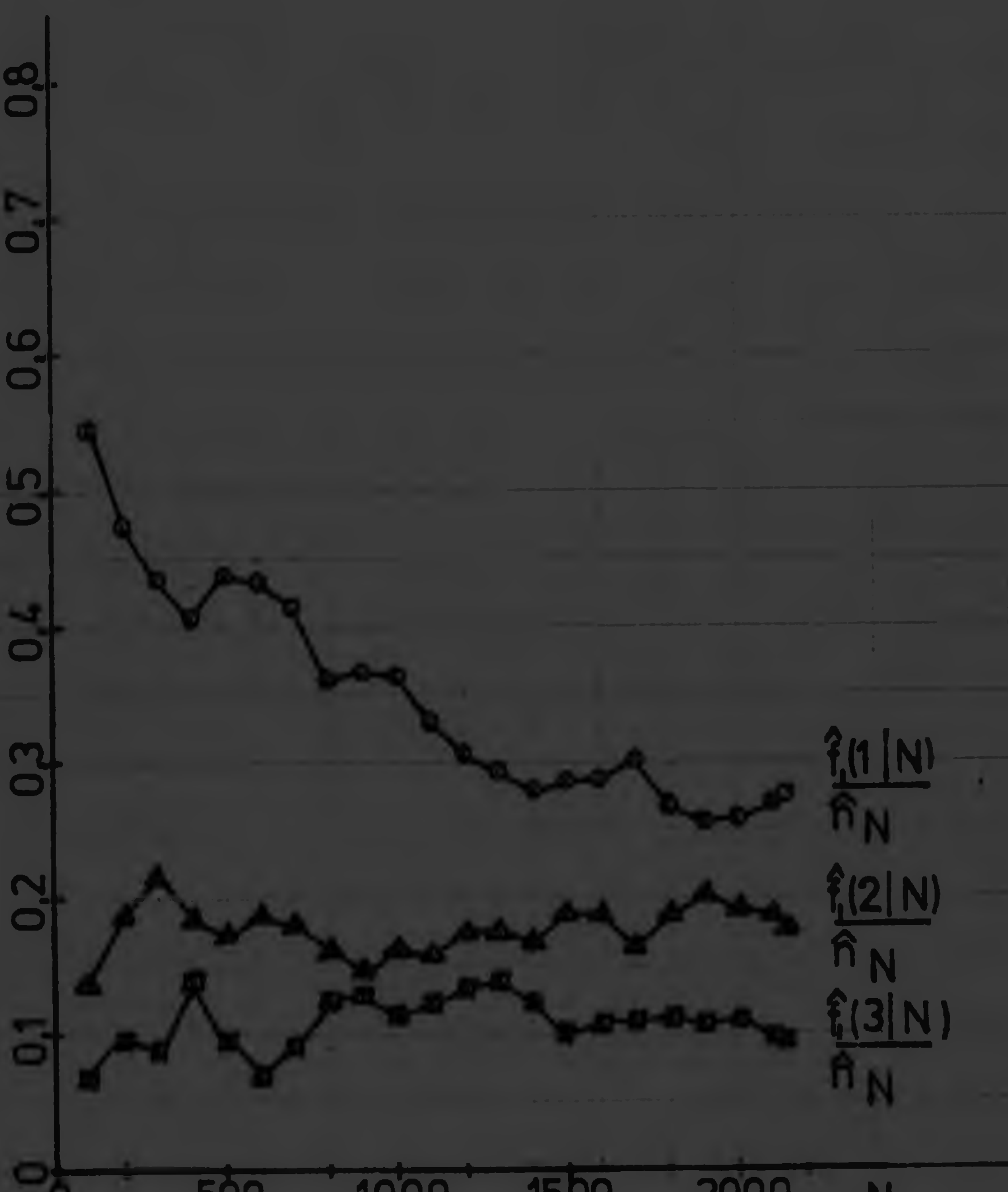


FIGURE 13 KERRY 1 FORWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}, \frac{\hat{f}_1(2|N)}{\hat{n}_N}, \frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

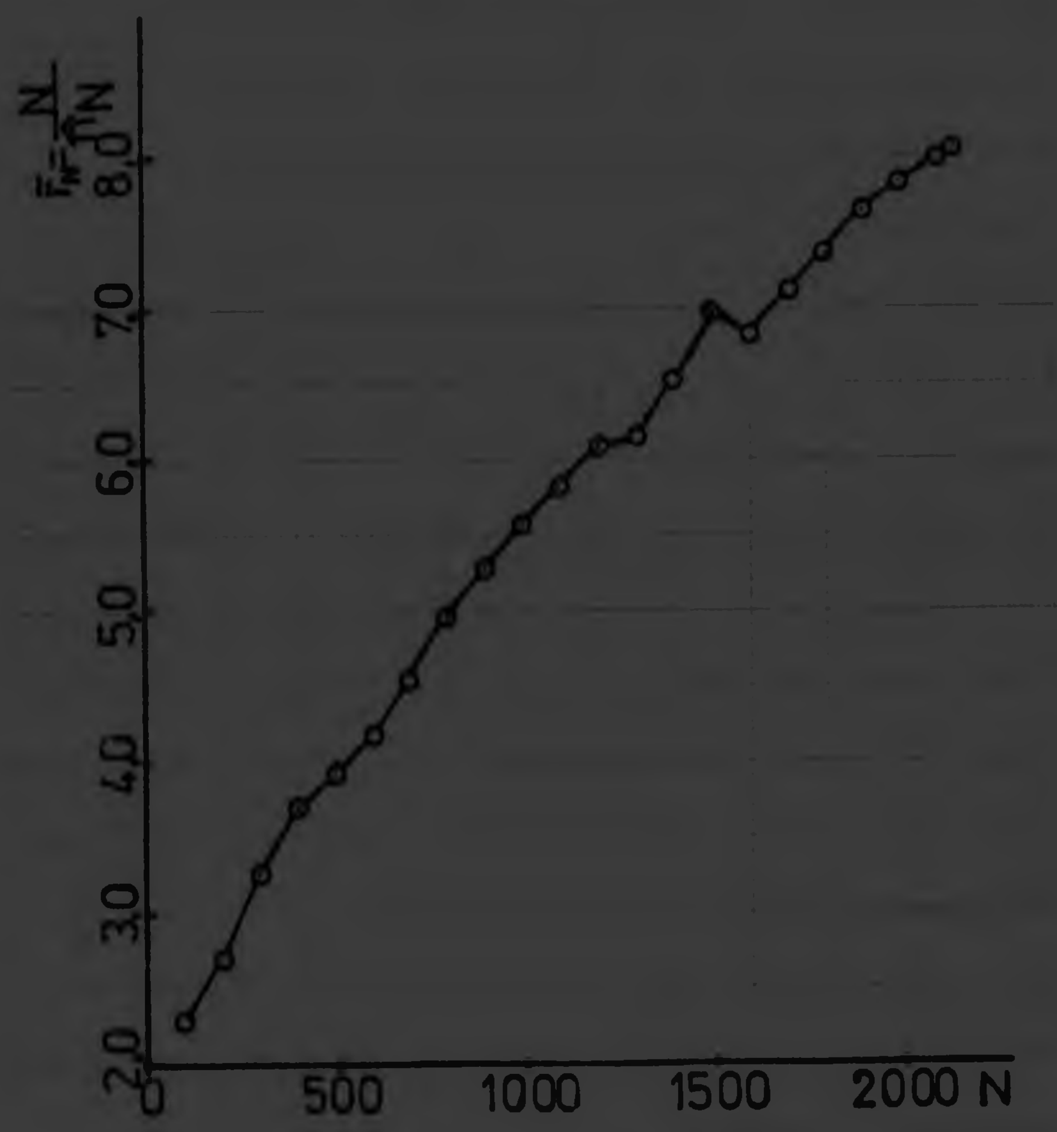


FIGURE 14 KERRY 1 FORWARD $\bar{r}_N = \frac{N}{\bar{n}_N}$ vs N

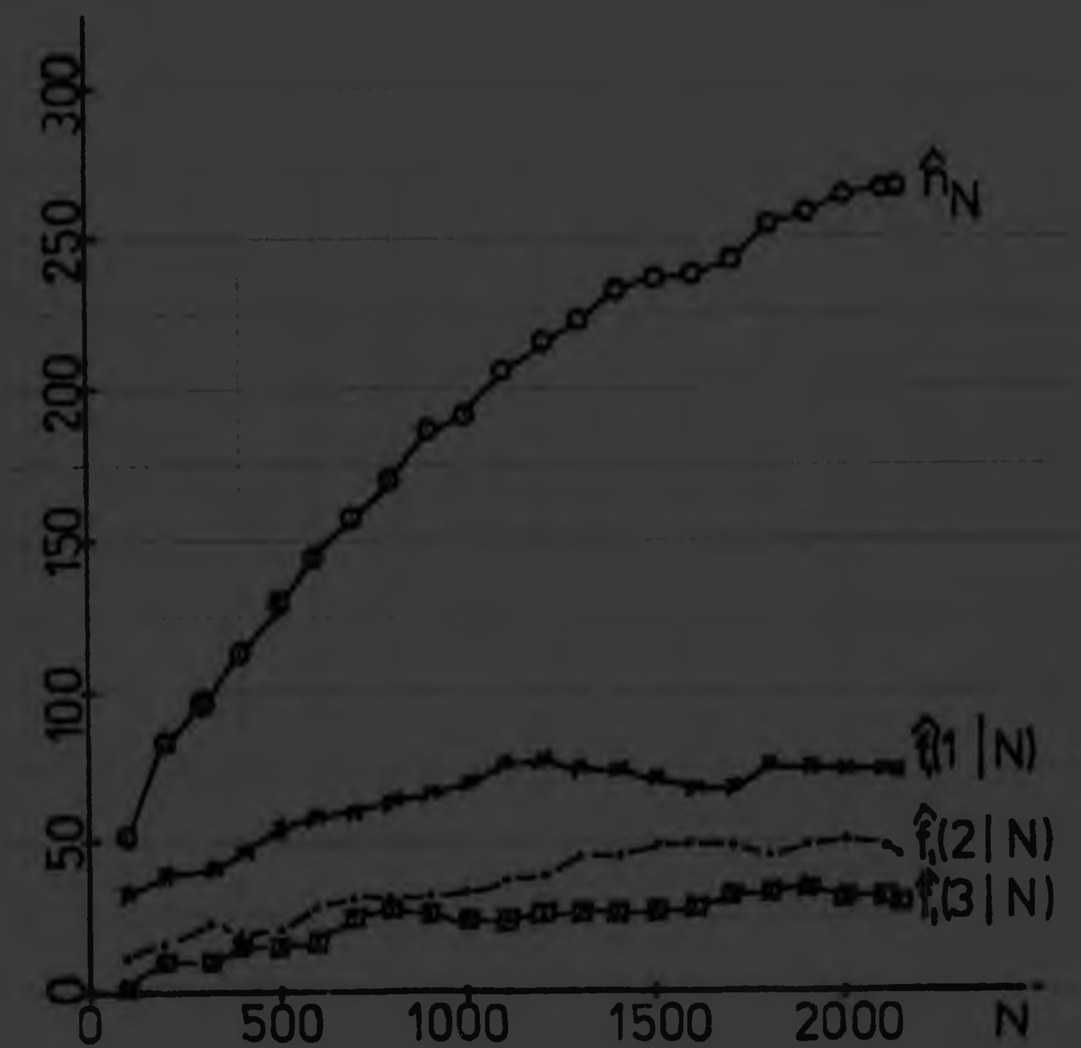


FIGURE 15 KERRY 1 BACKWARD
$\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$, $\hat{n}_N$ vs N

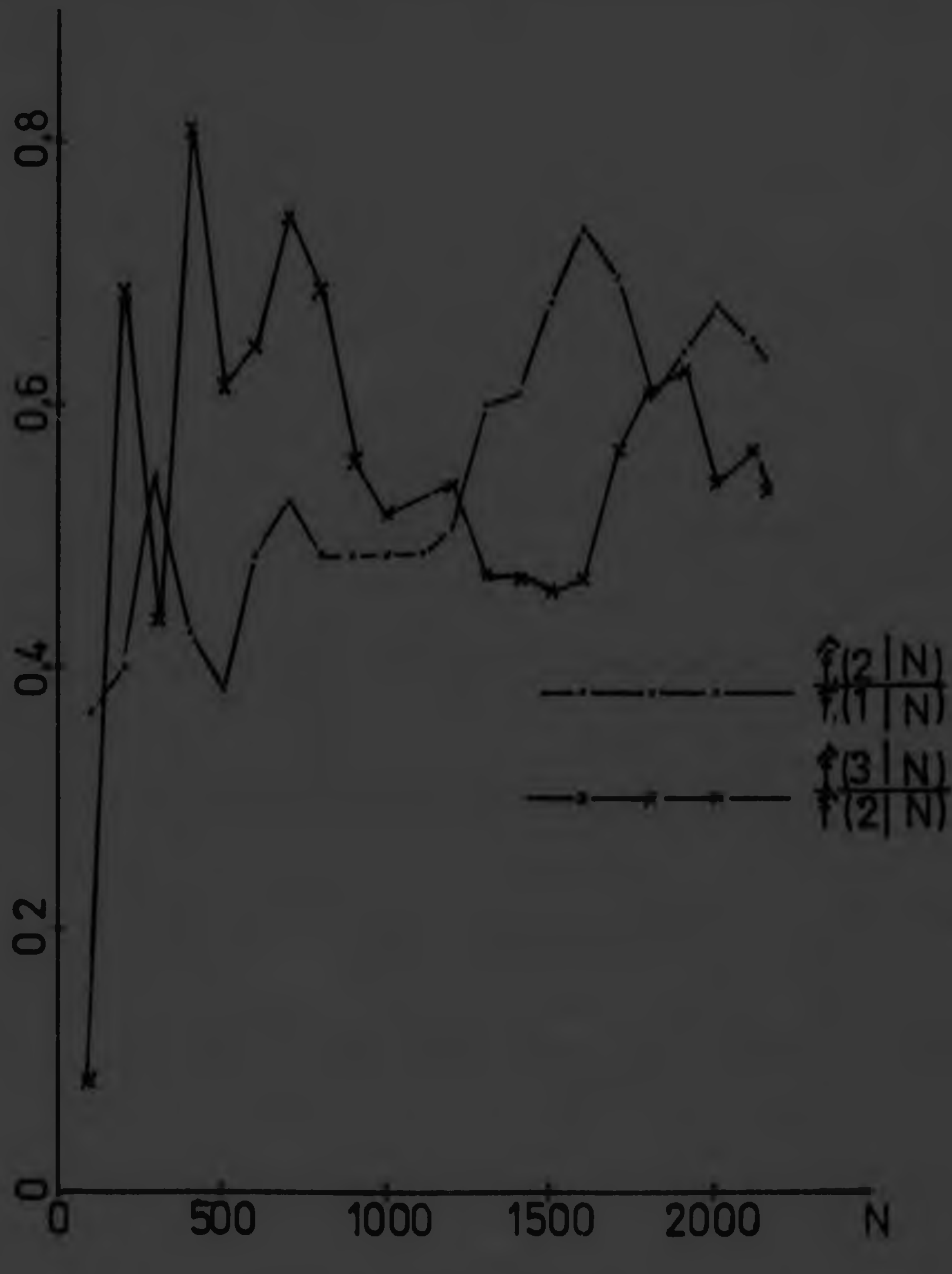


FIGURE 16 KERRY 1 BACKWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}(2|N)}$ vs N

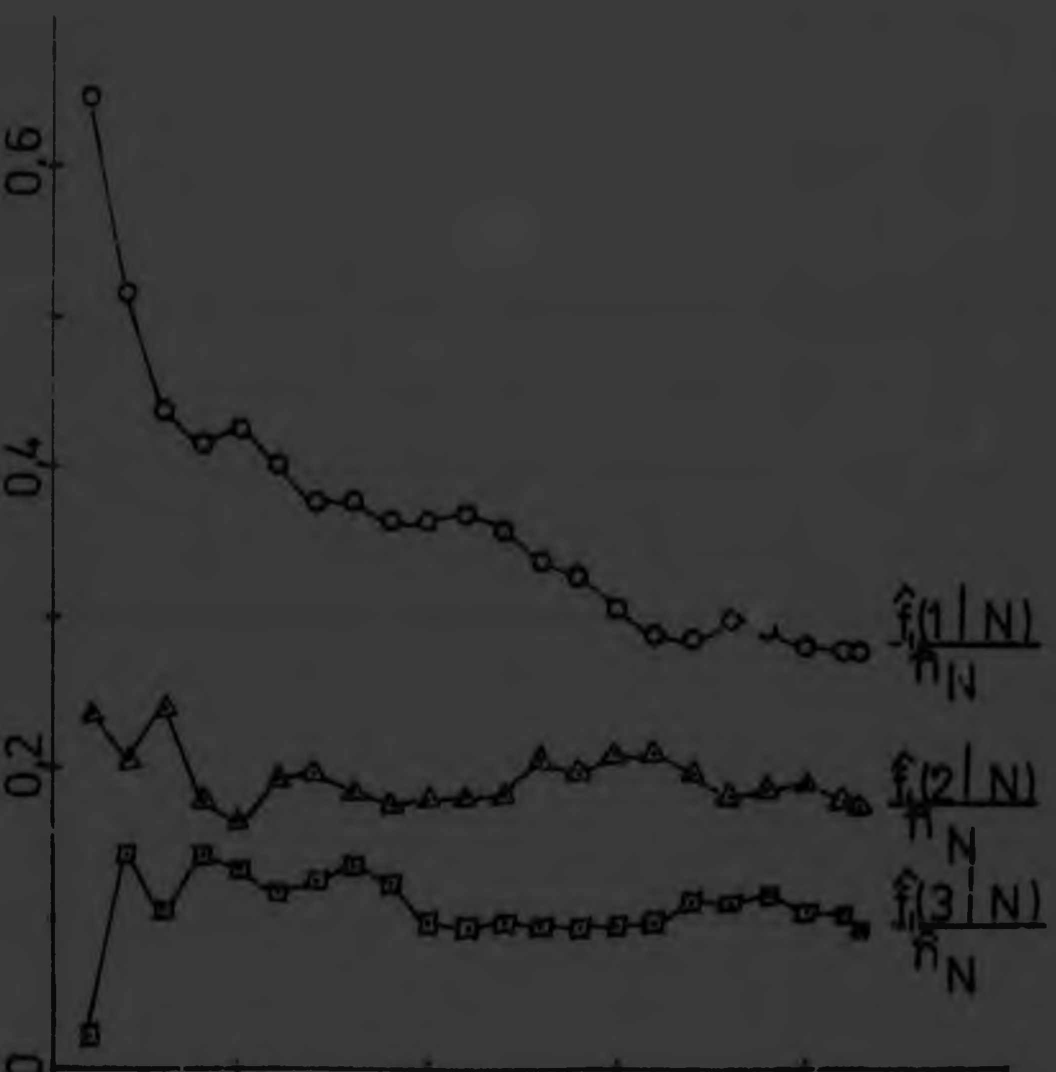


FIGURE 17 KERRY 1 BACKWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

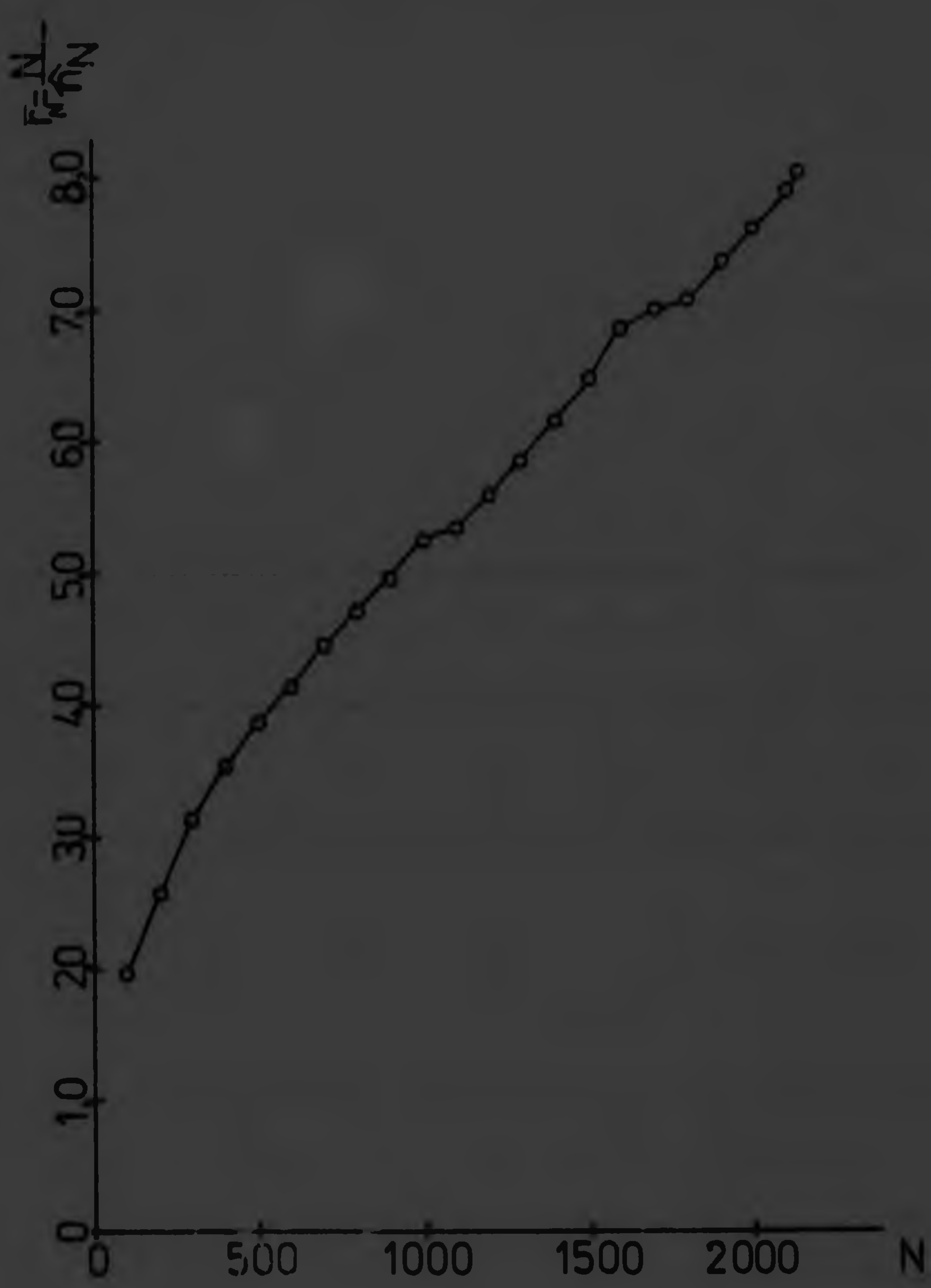


FIGURE 18 KERRY 1 BACKWARD $\bar{r}_N = \frac{N}{\bar{n}_N}$ vs N

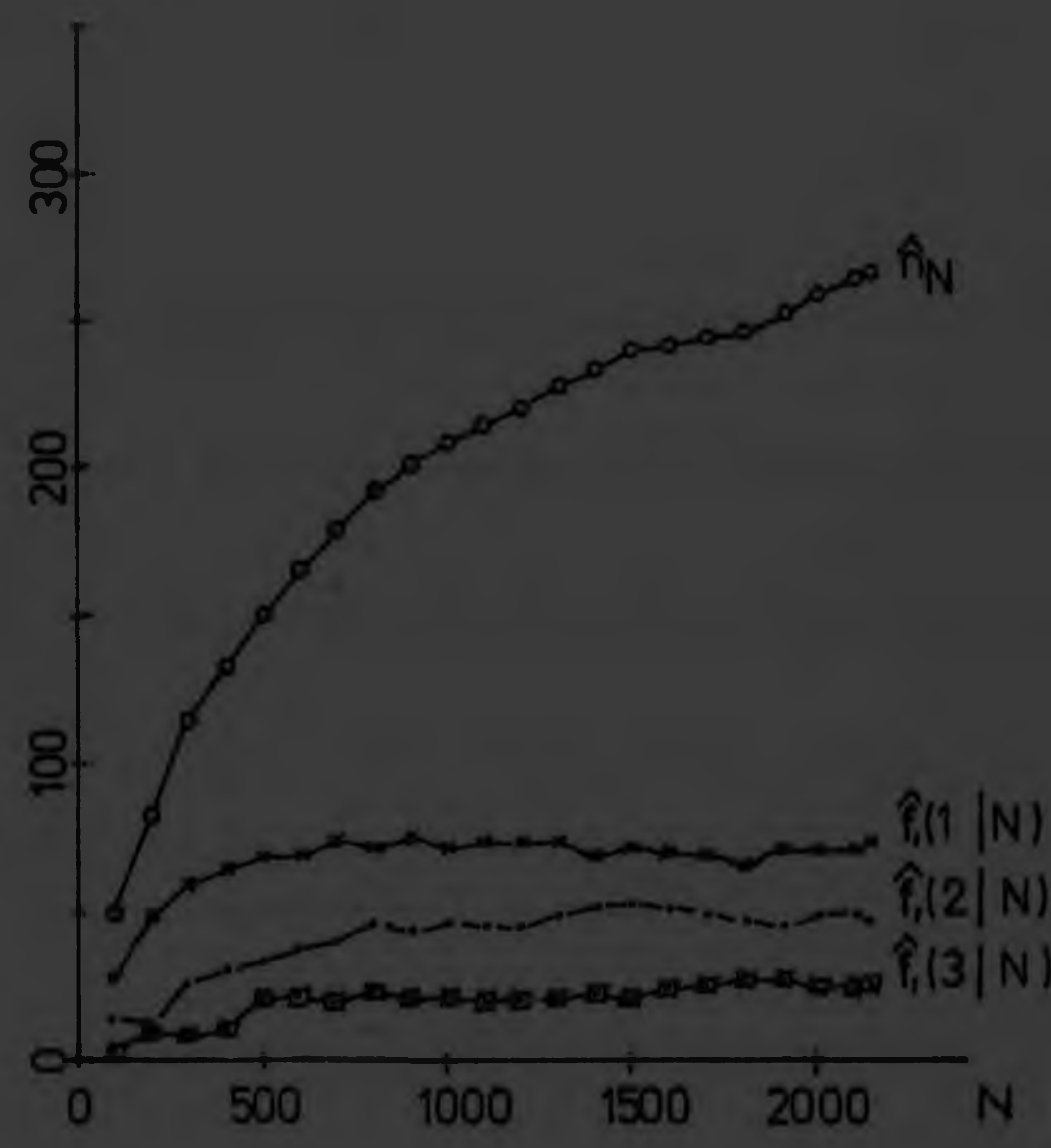


FIGURE 19 KERRY 1 RANDOM FORWARD

$\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$, $\hat{n}_N$ vs (N)

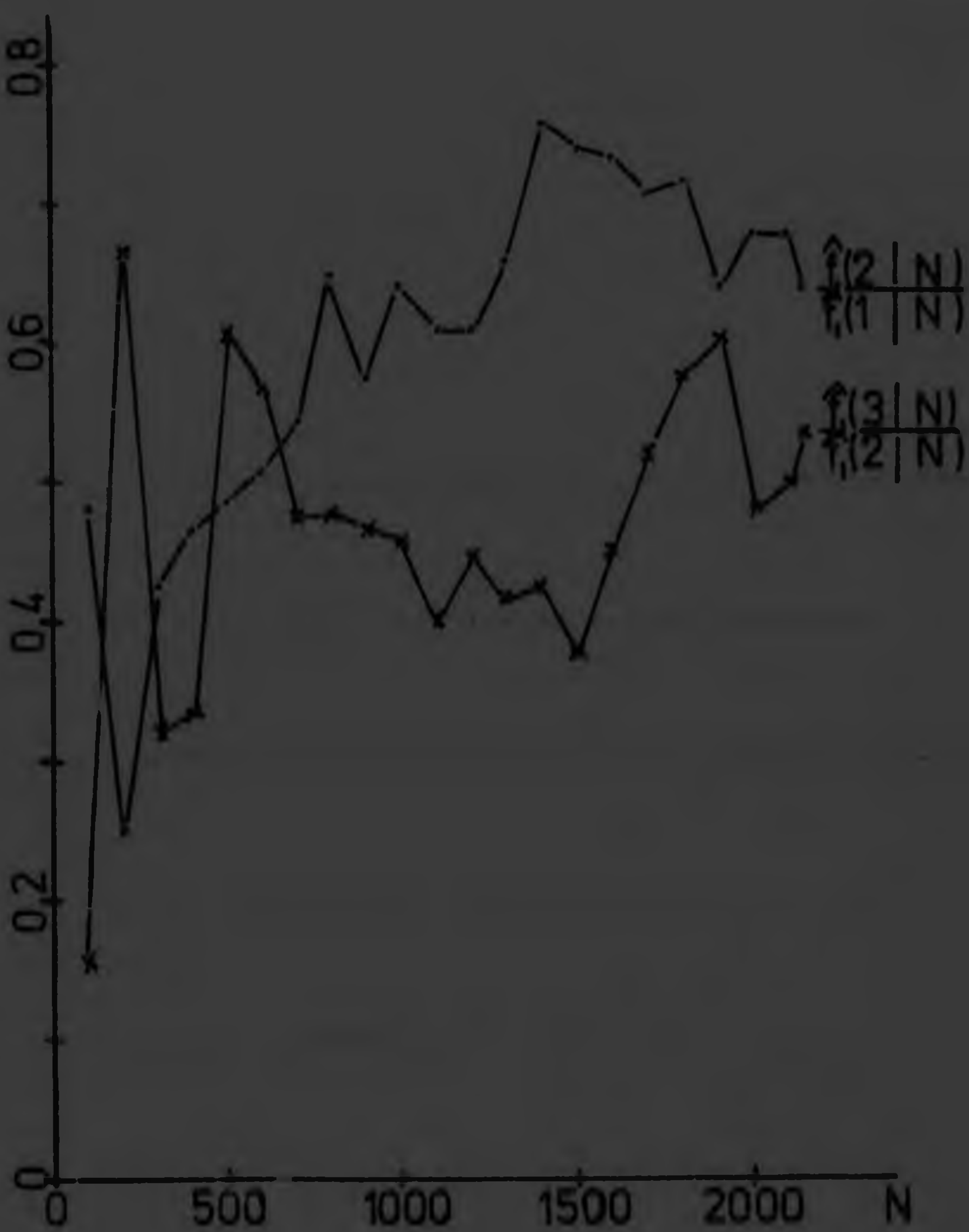


FIGURE 20 KERRY 1 RANDOM FORWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

FIGURE 21 KERRY 1 RANDOM FORWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

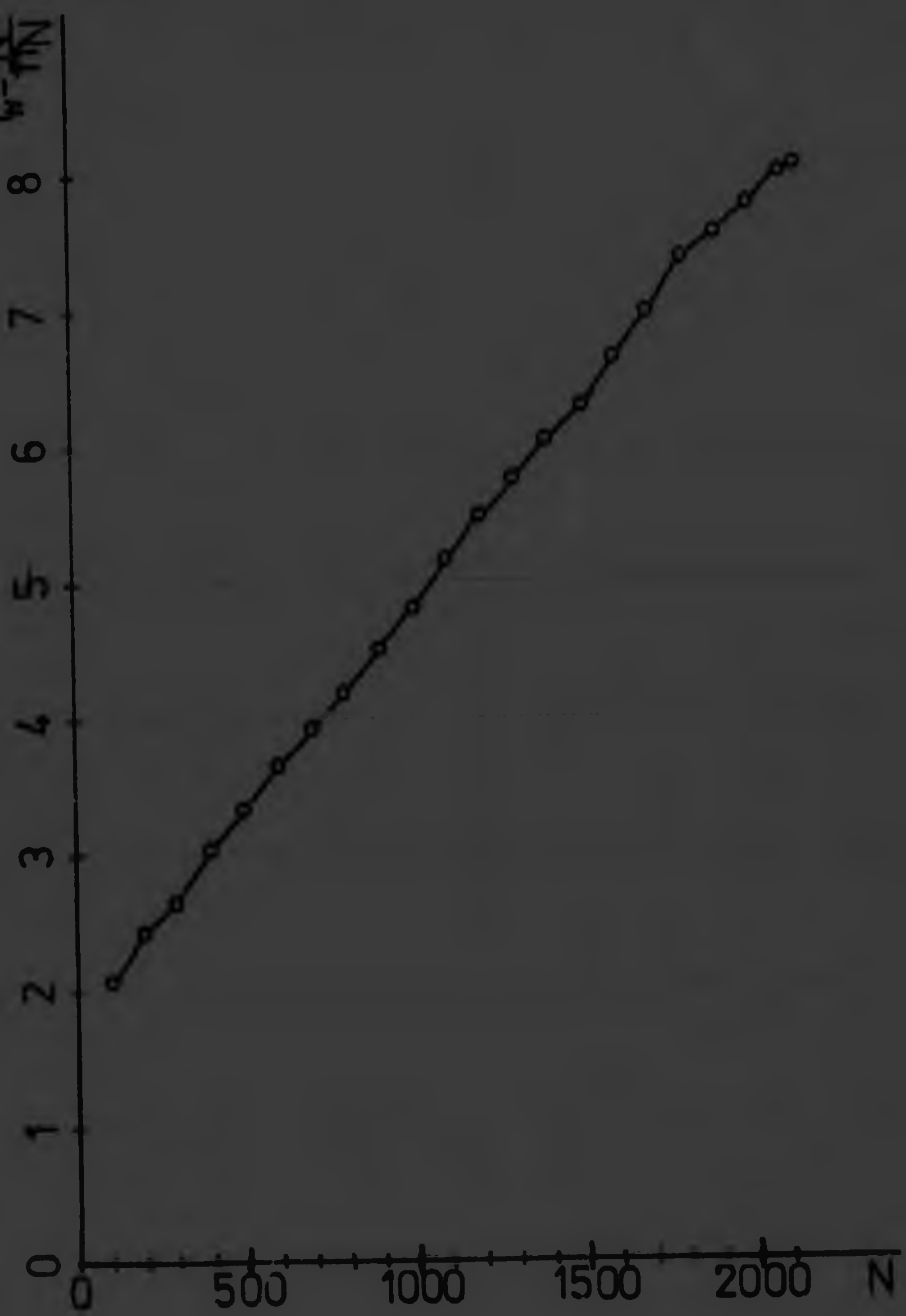


FIGURE 22 KERRY 1 RANDOM FORWARD

$\bar{r} = \frac{N}{\bar{n}_N}$ vs N

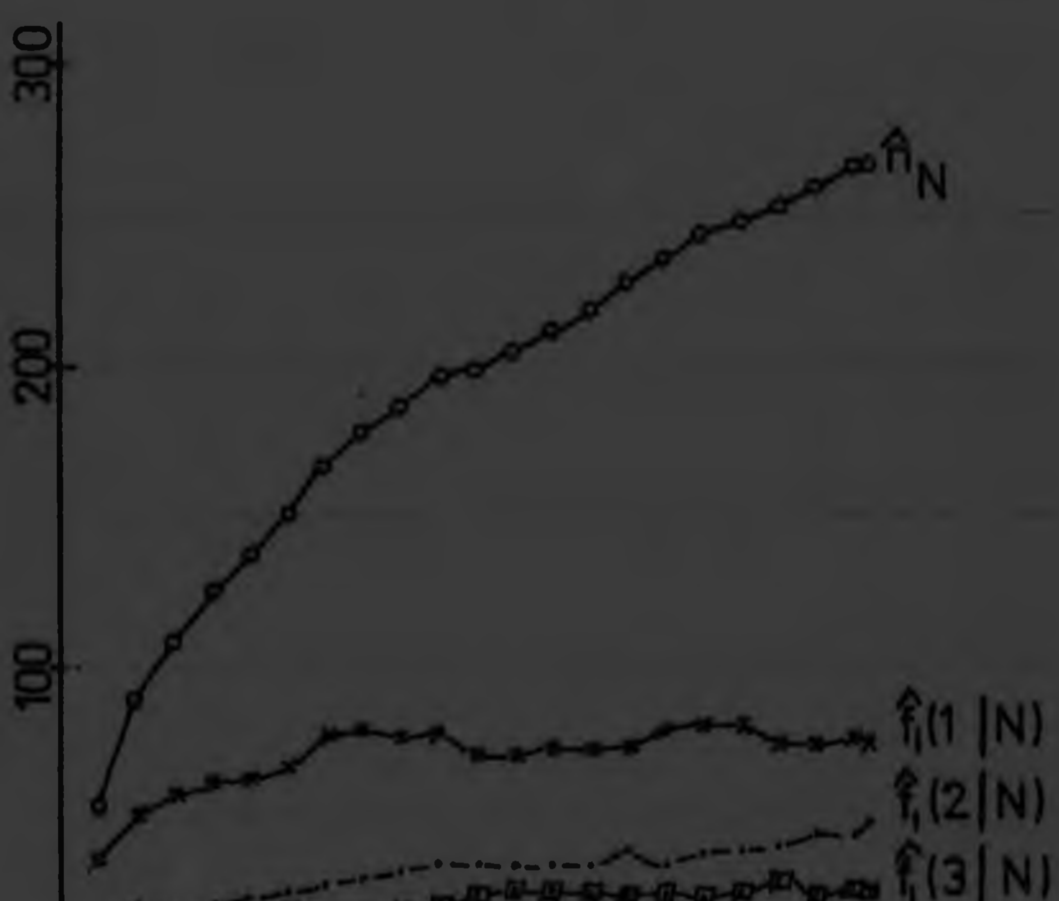


FIGURE 23 KERRY RANDOM BACKWARD

$\bar{f}_1(1|N)$, $\bar{f}_1(2|N)$, $\bar{f}_1(3|N)$, $\hat{n}_N$ vs N

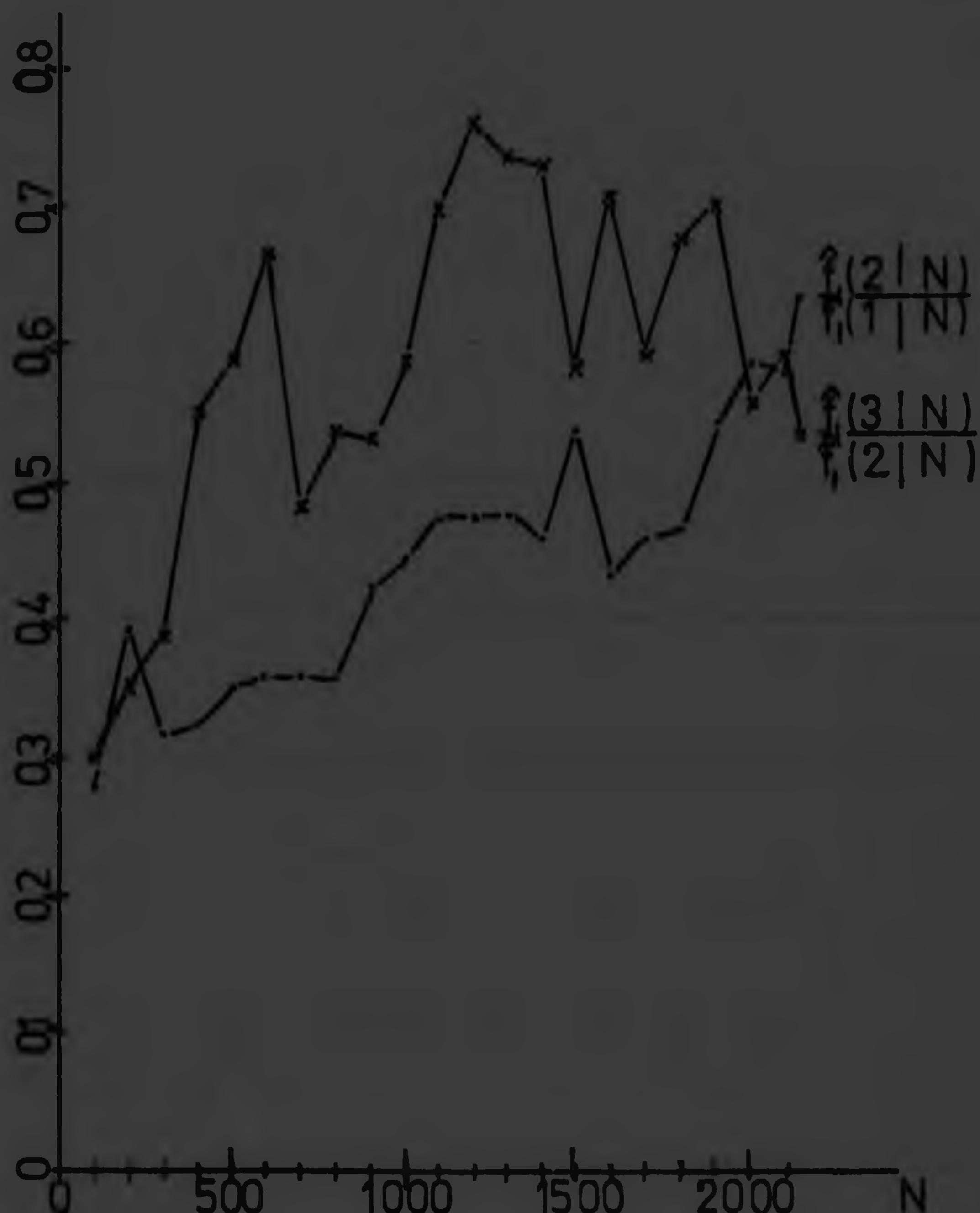


FIGURE 24 KERRY 1 RANDOM BACKWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

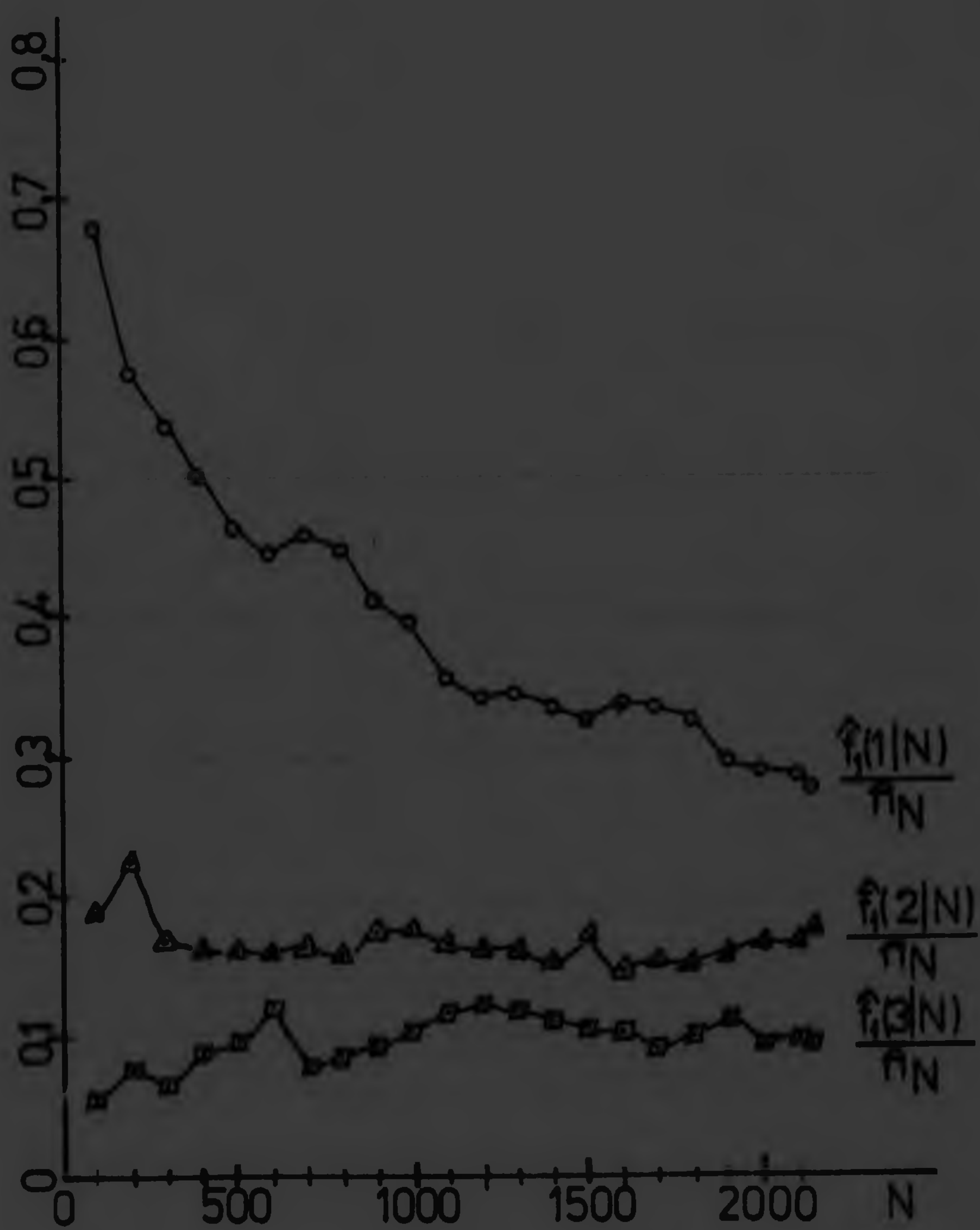


FIGURE 25 KERRY 1 RANDOM BACKWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$

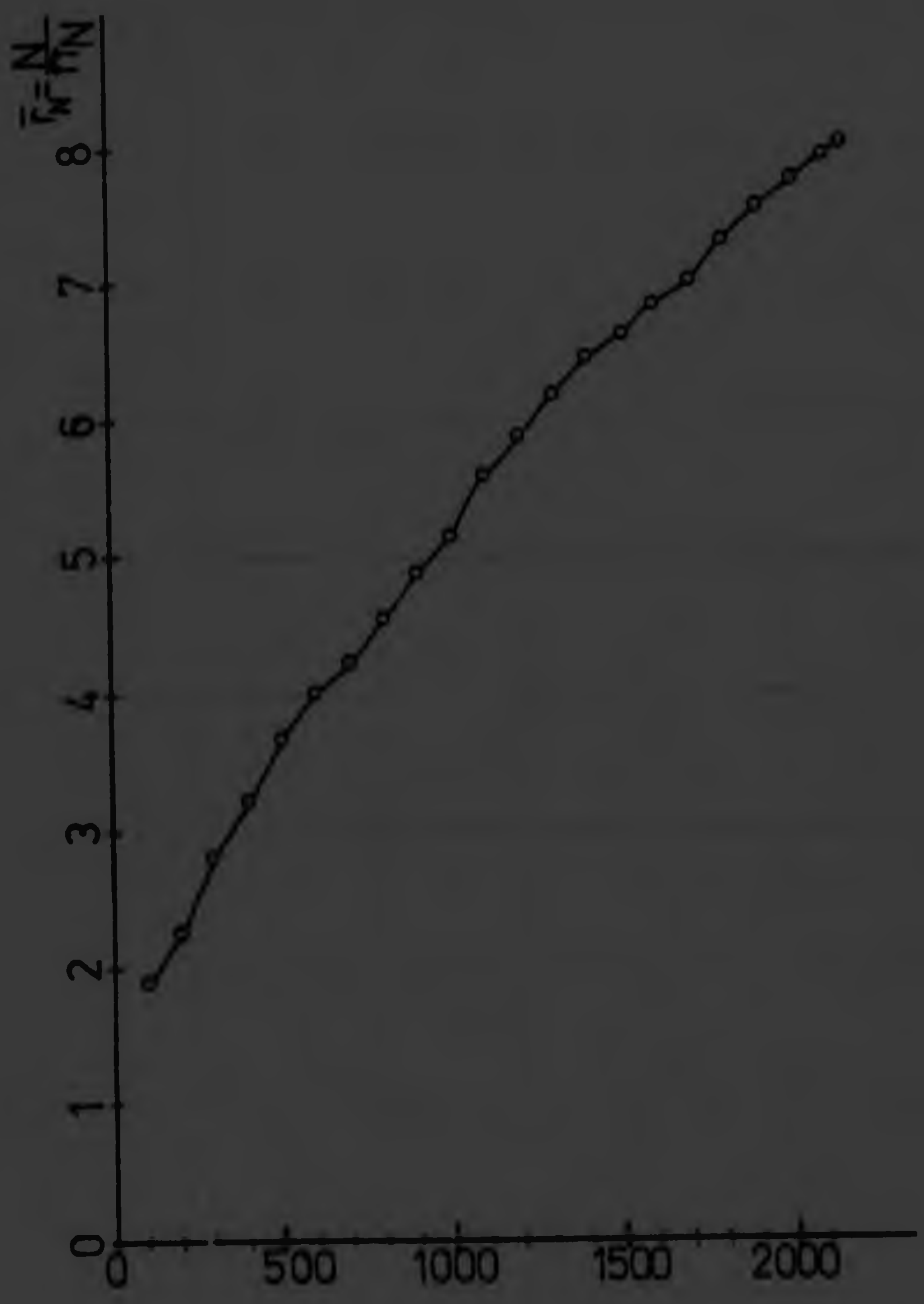


FIGURE 26 KERRY 1 RANDOM BACKWARD

$\bar{r}_N = \frac{N}{\bar{n}_N}$ vs N

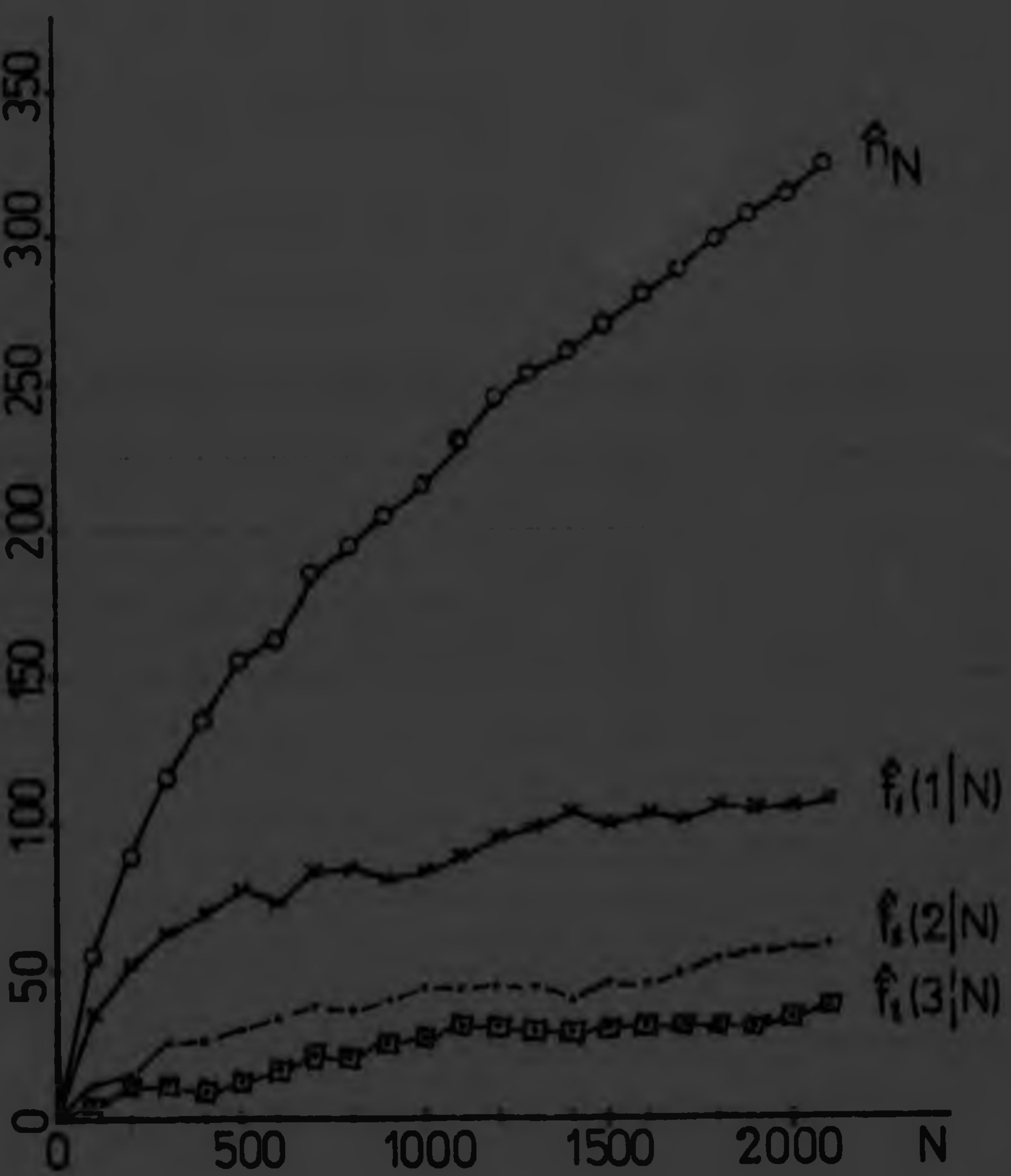


FIGURE 27 KERRY 2 FORWARD

$\bar{f}_1(1|N)$, $\bar{f}_1(2|N)$, $\hat{f}_1(3|N)$, $\hat{n}_N$ vs N

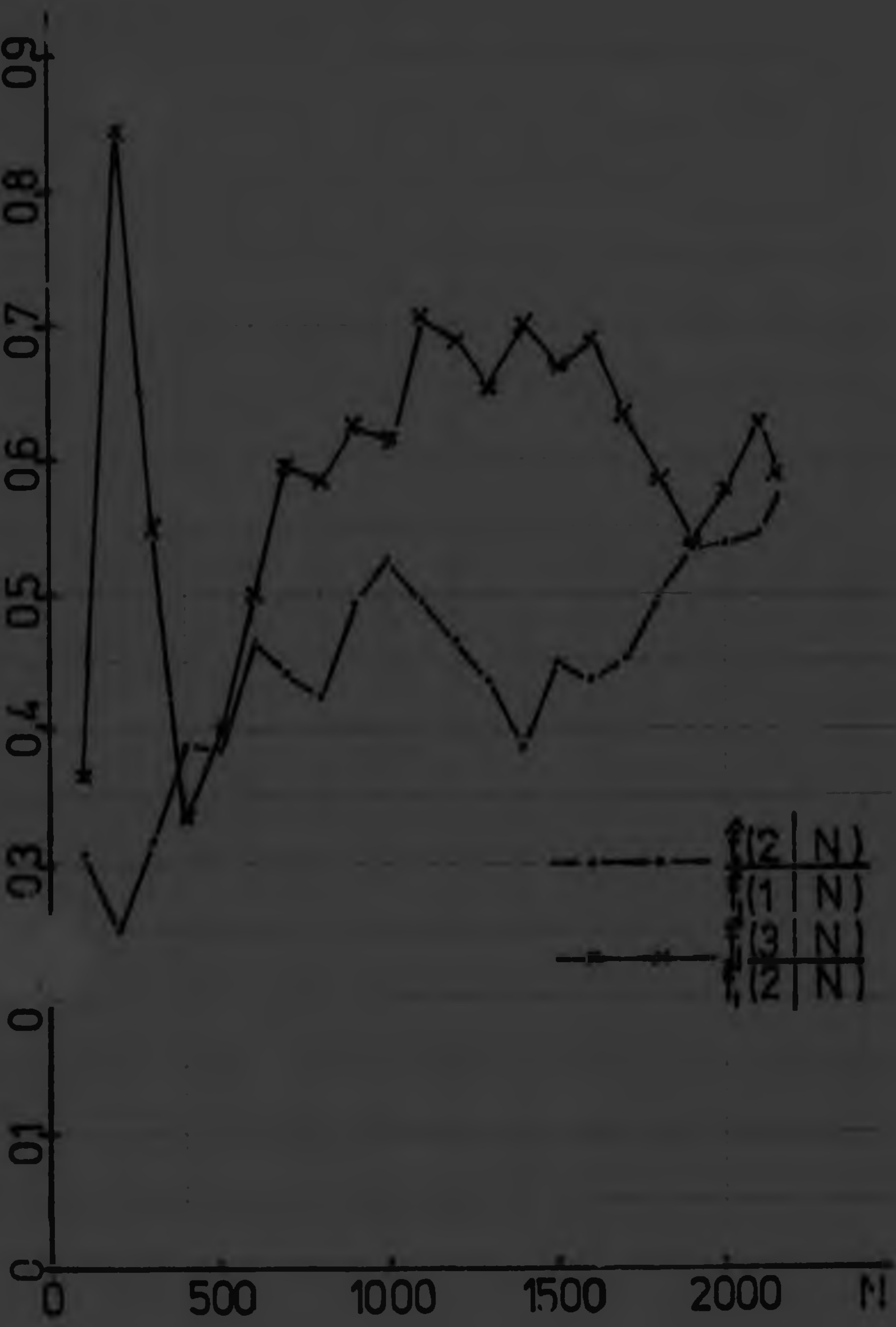


FIGURE 28 KERRY 2 FORWARD

$\dfrac{\hat{\bar{f}}_1(2|N)}{\bar{f}_1(1|N)}, \dfrac{\hat{\bar{f}}_1(3|N)}{\bar{f}_1(2|N)}$ vs N

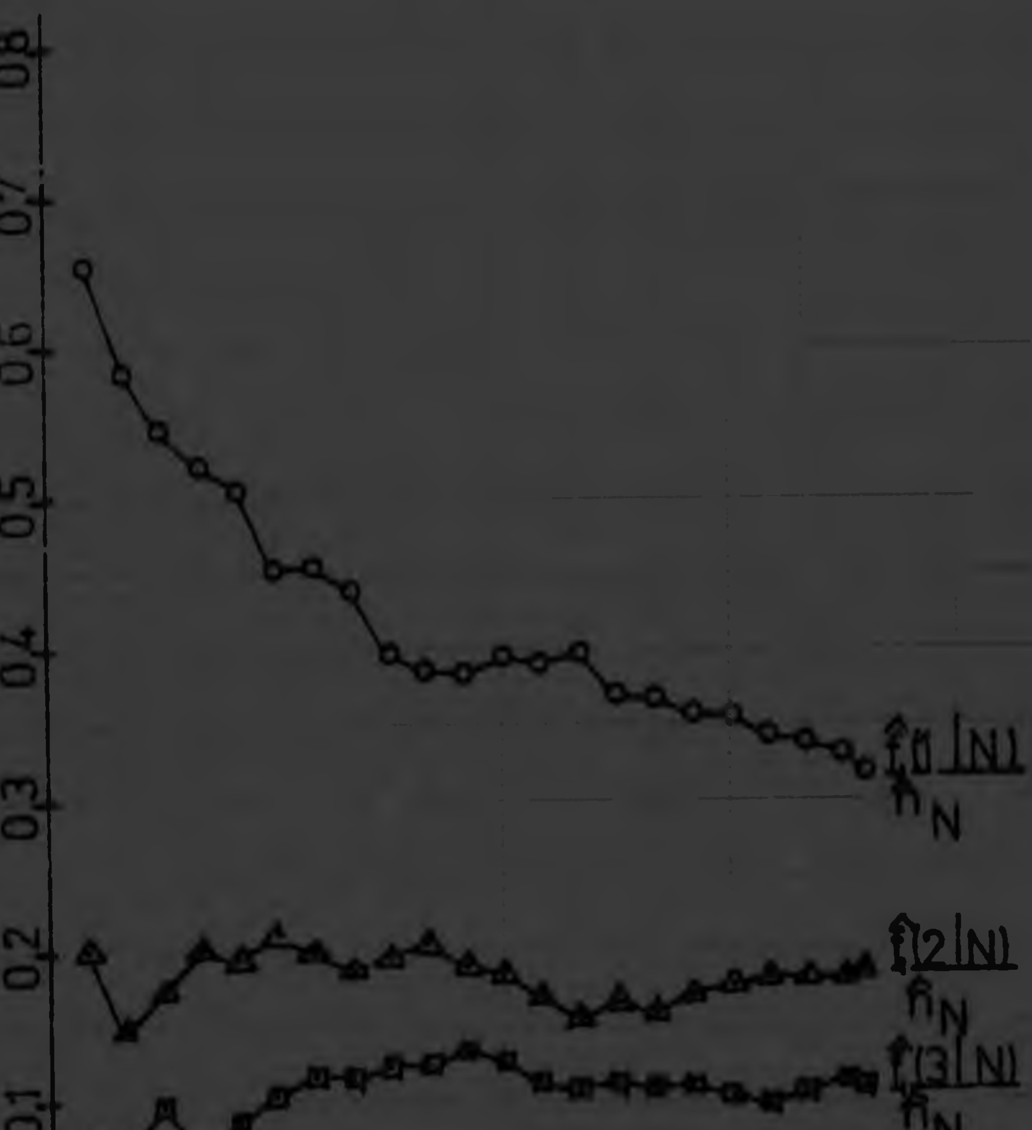


FIGURE 29 KERRY 2 FORWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

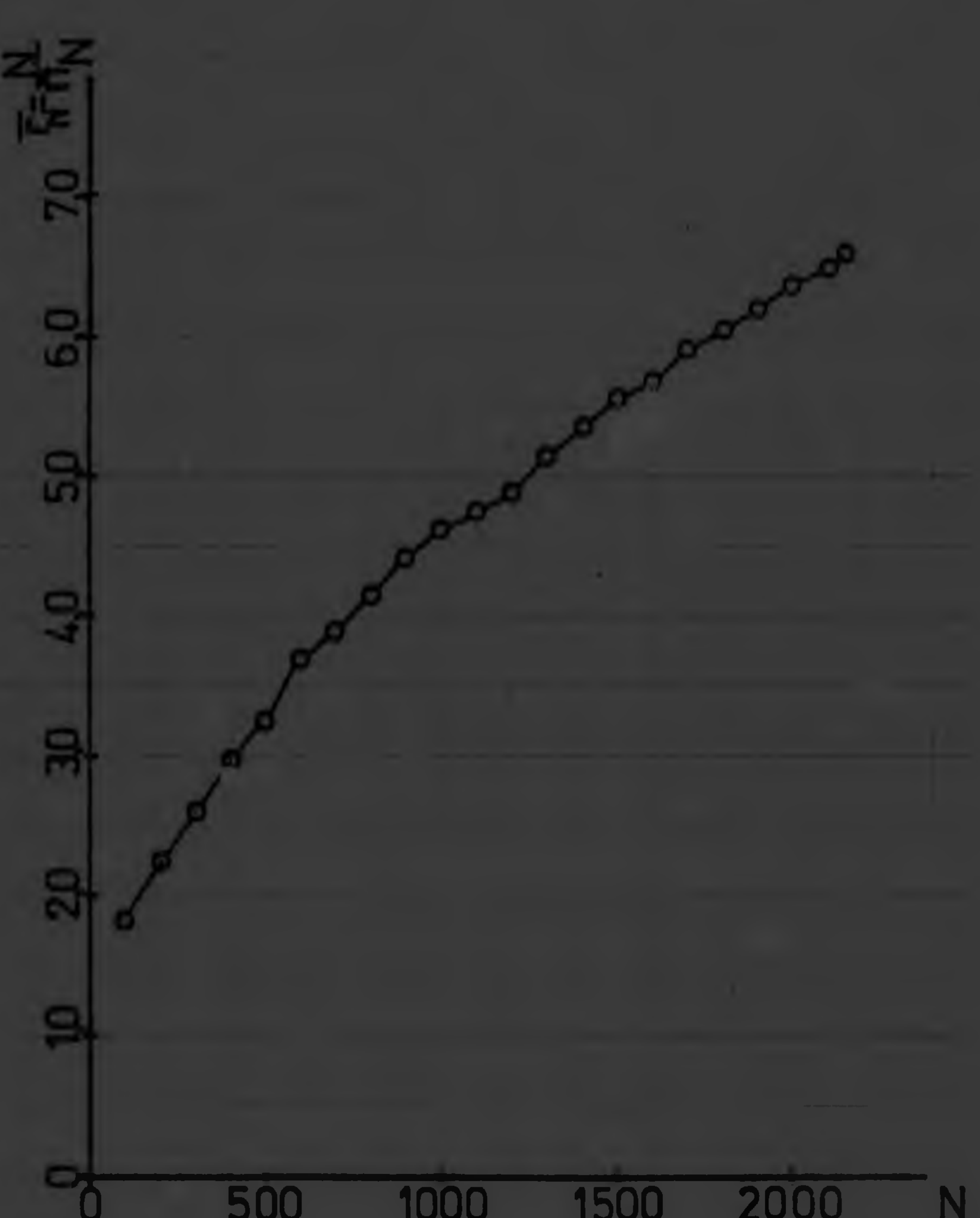


FIGURE 30 KERRY 2 FORWARD $\bar{r}_N = \frac{N}{n_N}$ vs N

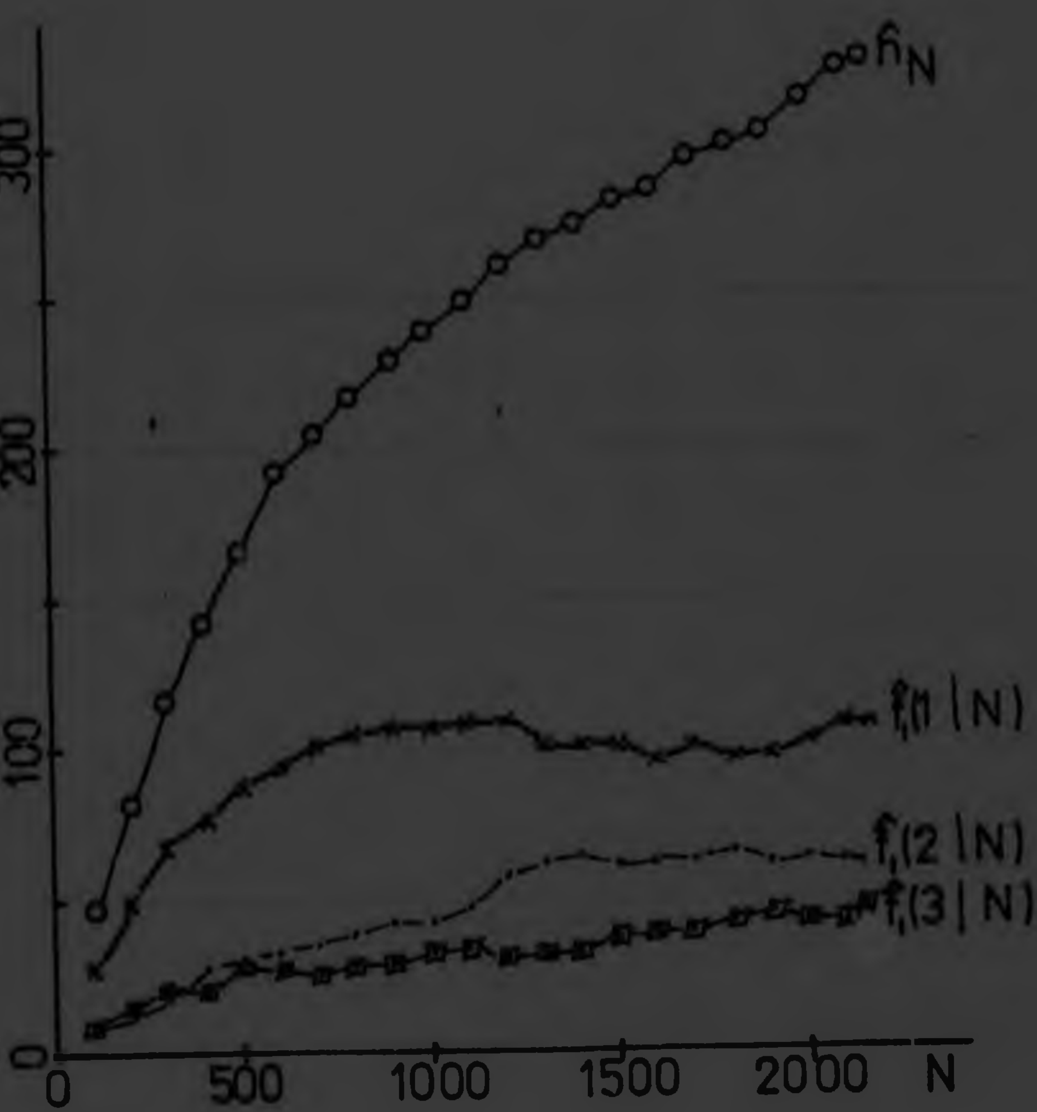


FIGURE 31 KERRY 2 BACKWARD
$\bar{f}_1(1|N)$, $\bar{f}_1(2|N)$, $\bar{f}_1(3|N)$, $\hat{n}_N$ vs N

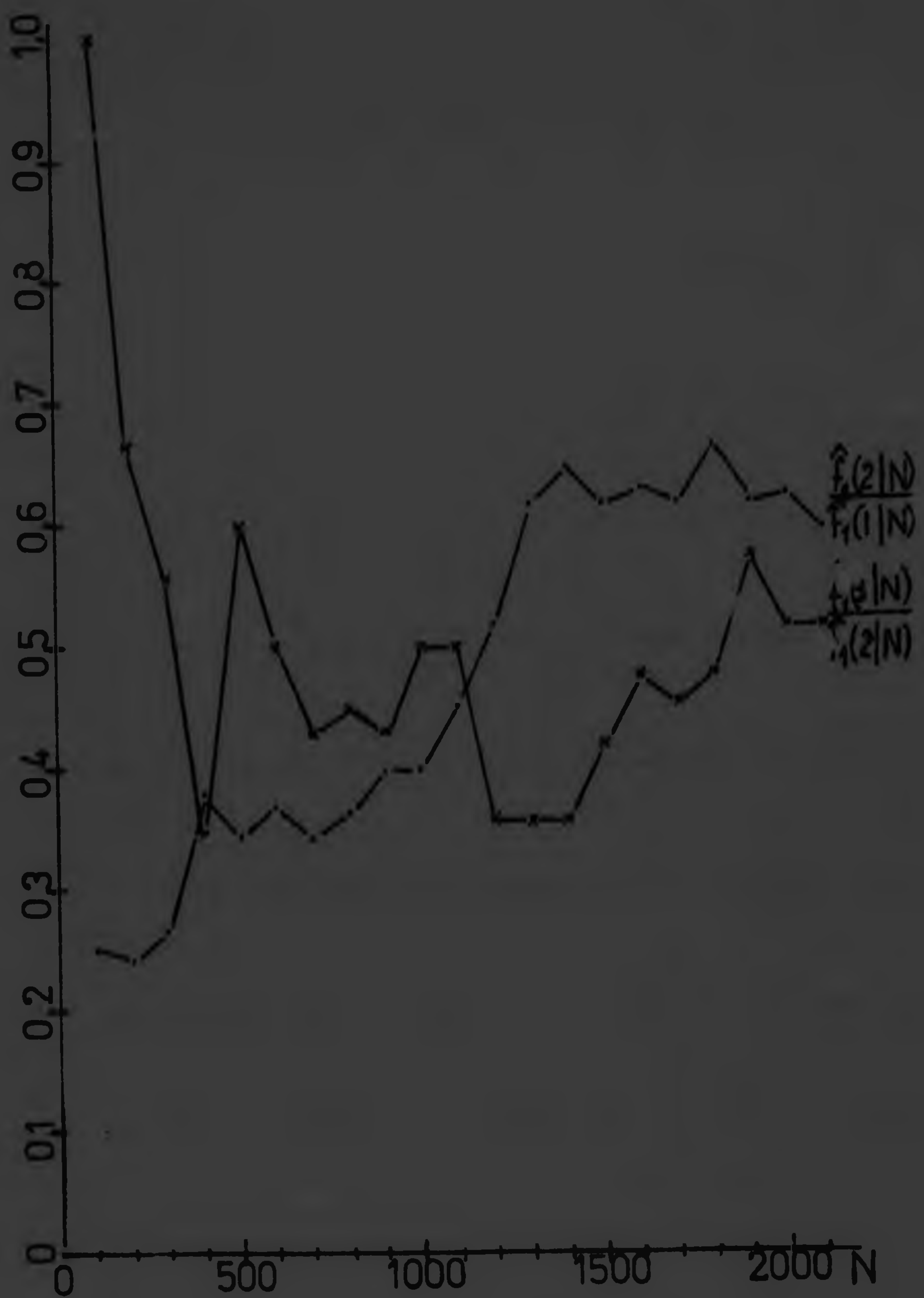


FIGURE 32 KERRY 2 BACKWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$ & $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

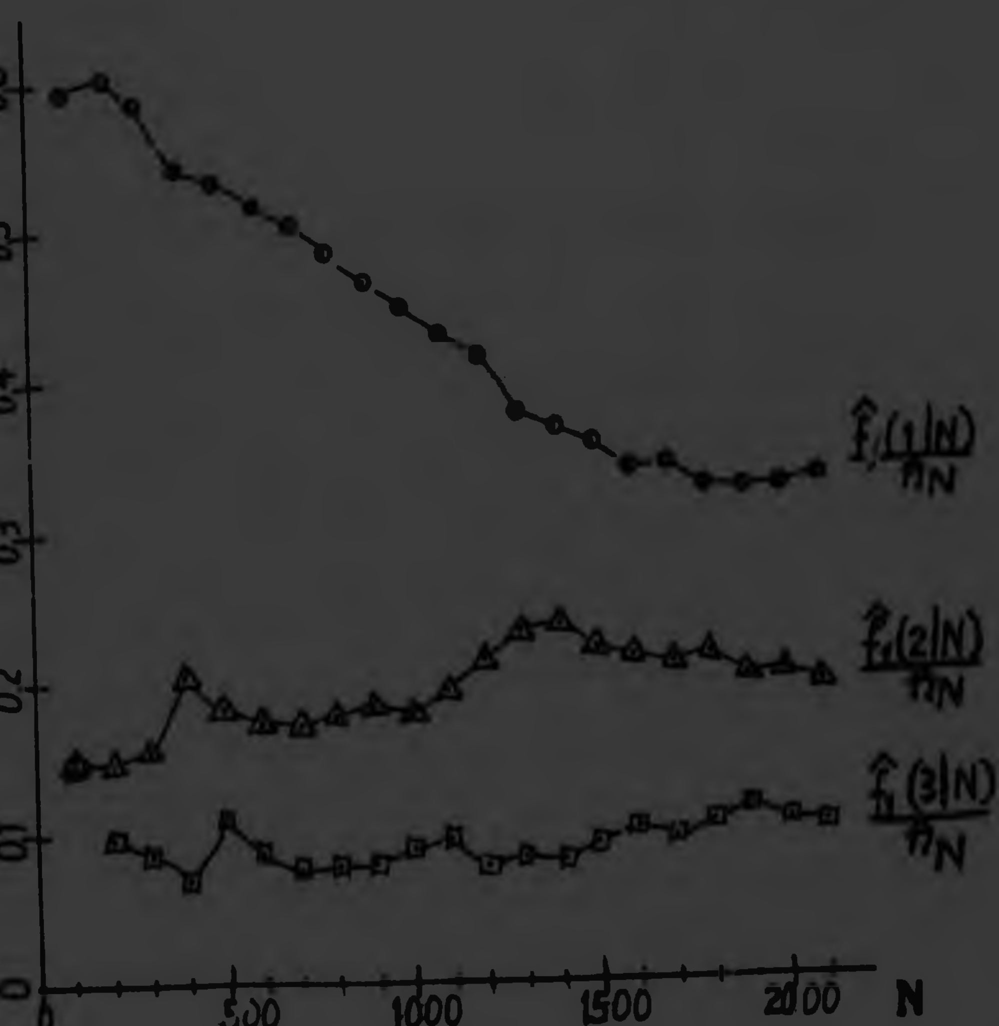


FIGURE 33 KERRY 3 BACKWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(1|N)}{\hat{n}_N}$ vs N

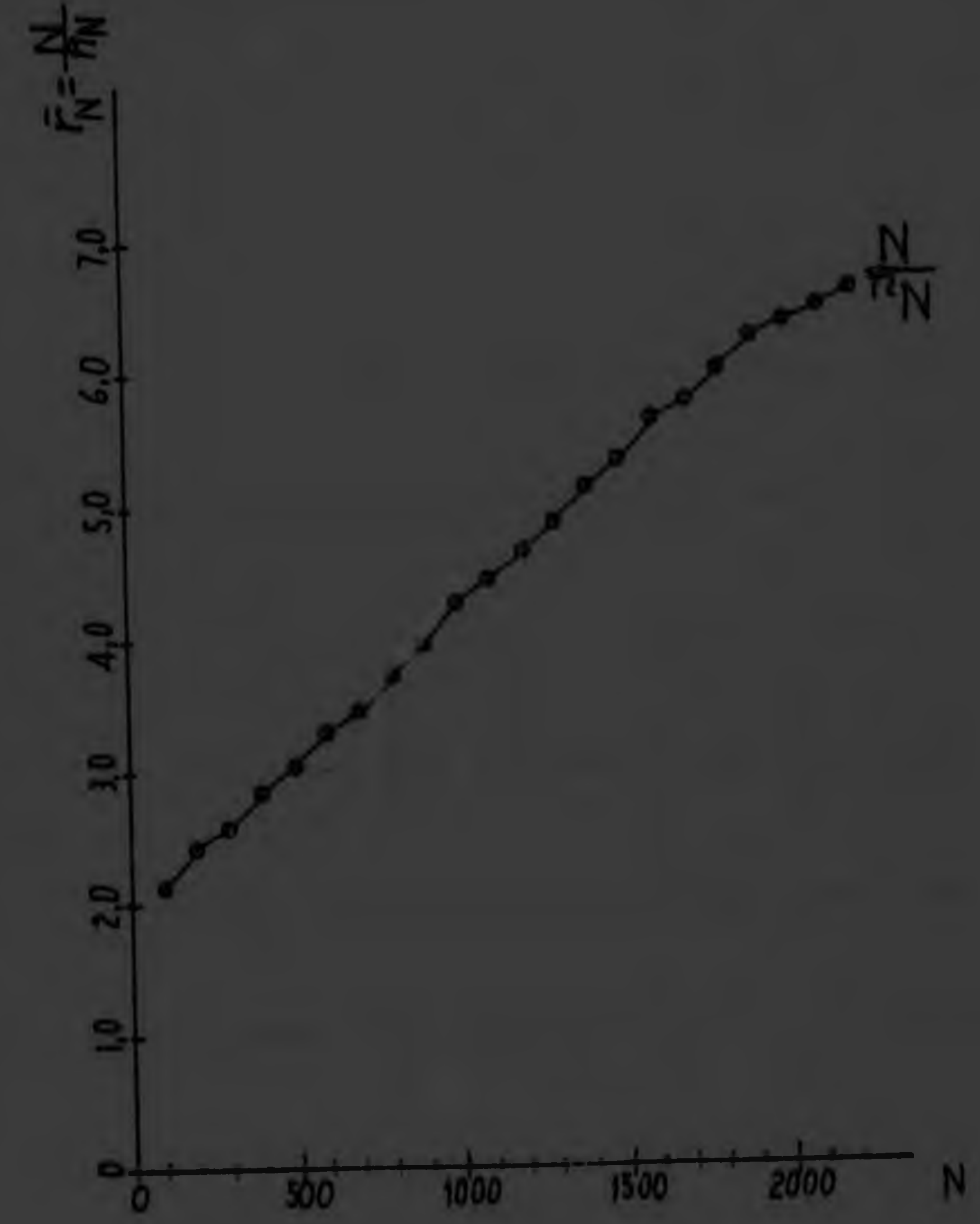


FIGURE 34 KERRY 2 BACKWARD $\bar{r}_N = \frac{N}{n_N}$ vs N

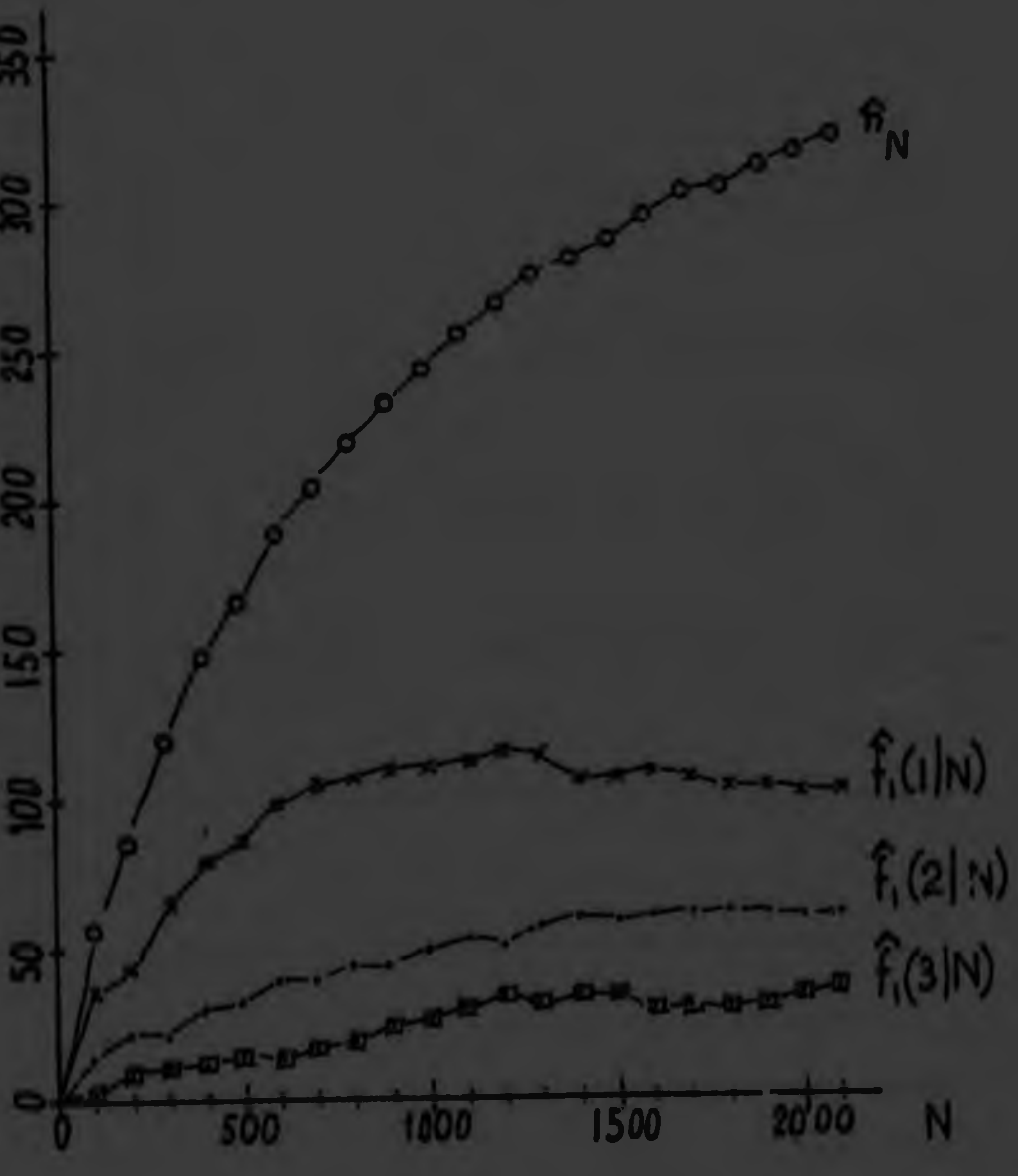


FIGURE 35 KERRY 2 RANDOM FORWARD

$\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

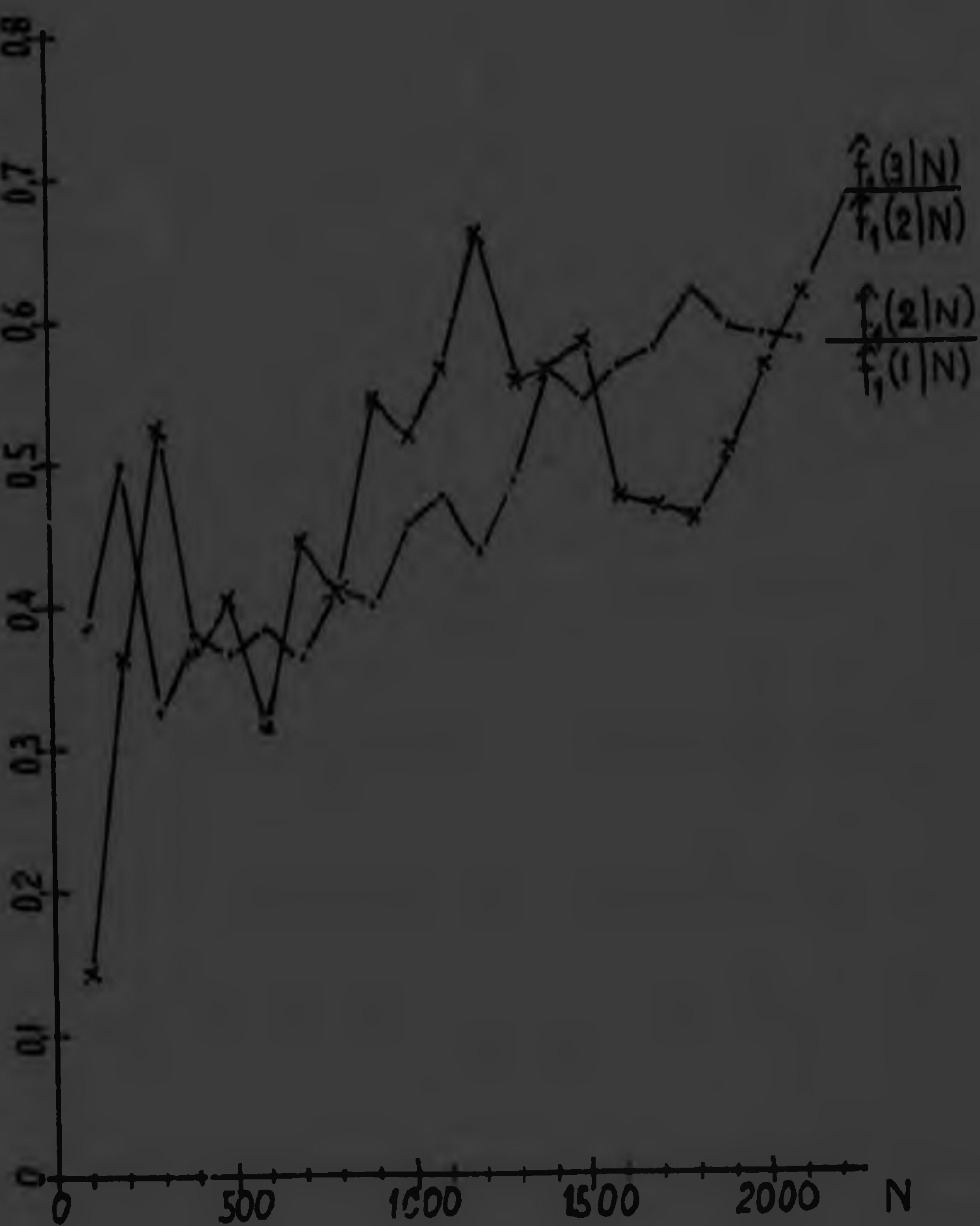


FIGURE 36 KERRY 2 RANDOM FORWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

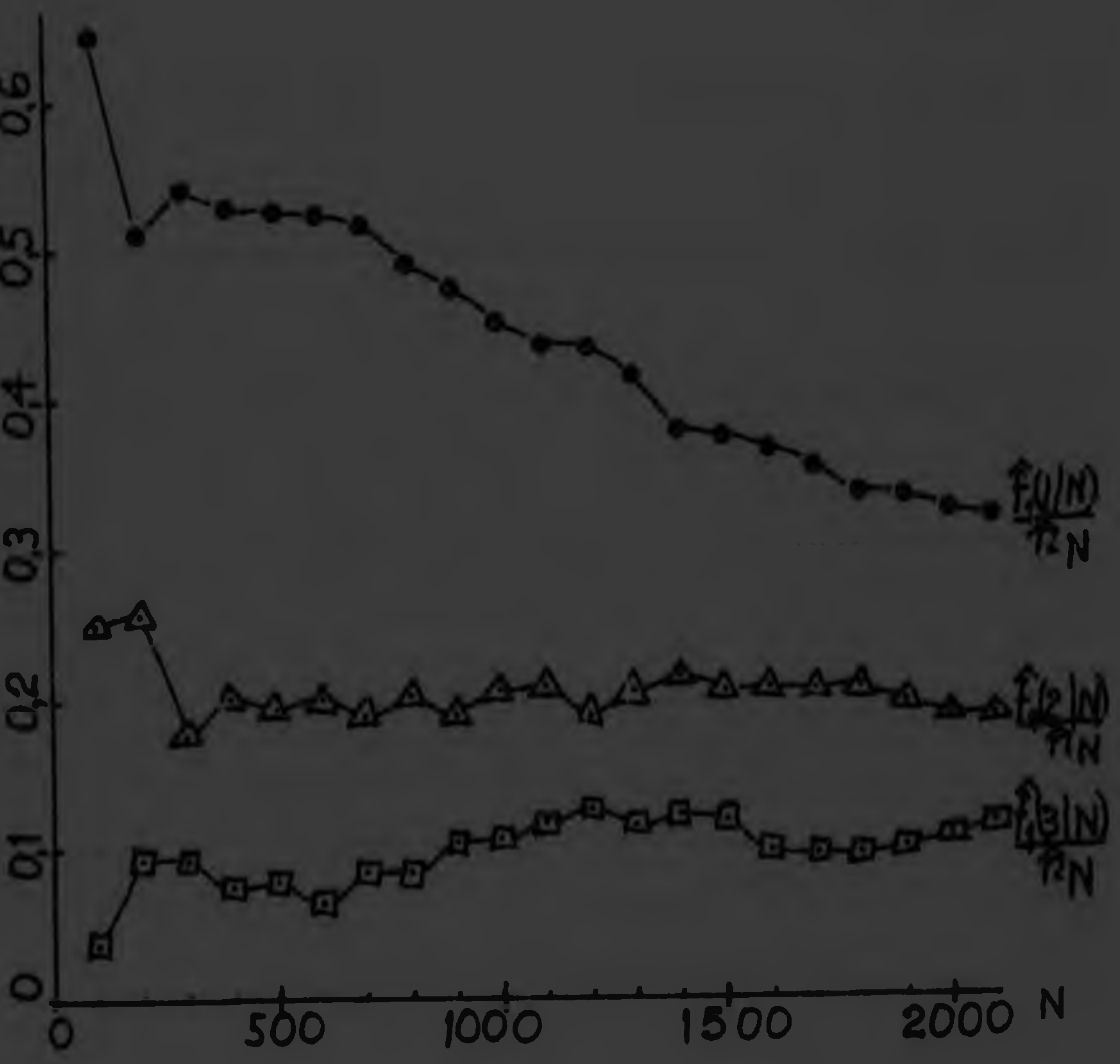


FIGURE 37 KERRY 2 RANDOM FORWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

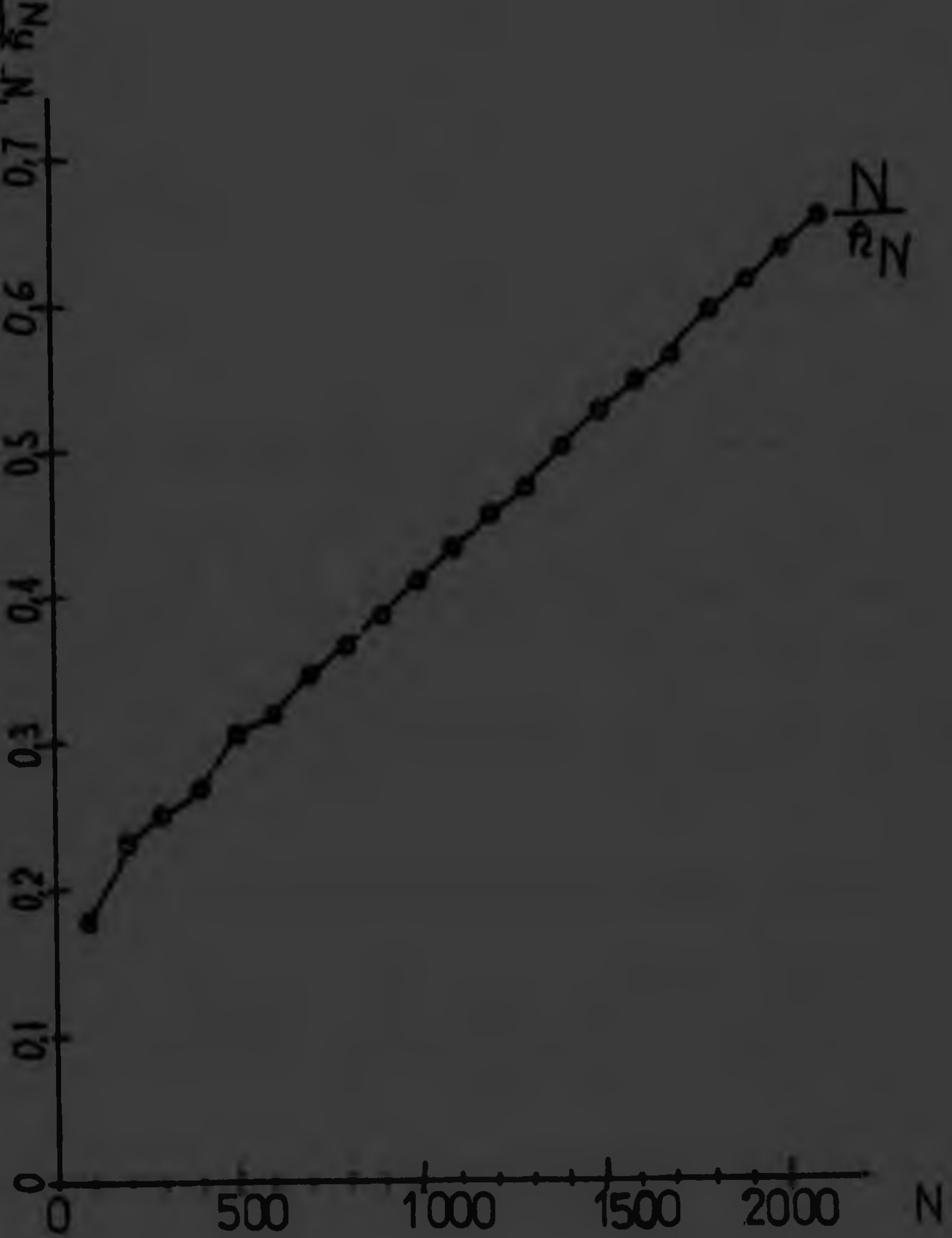


FIGURE 38 KERRY 2 RANDOM FORWARD

$\bar{r}_N = \frac{N}{\hat{n}_N}$ vs N

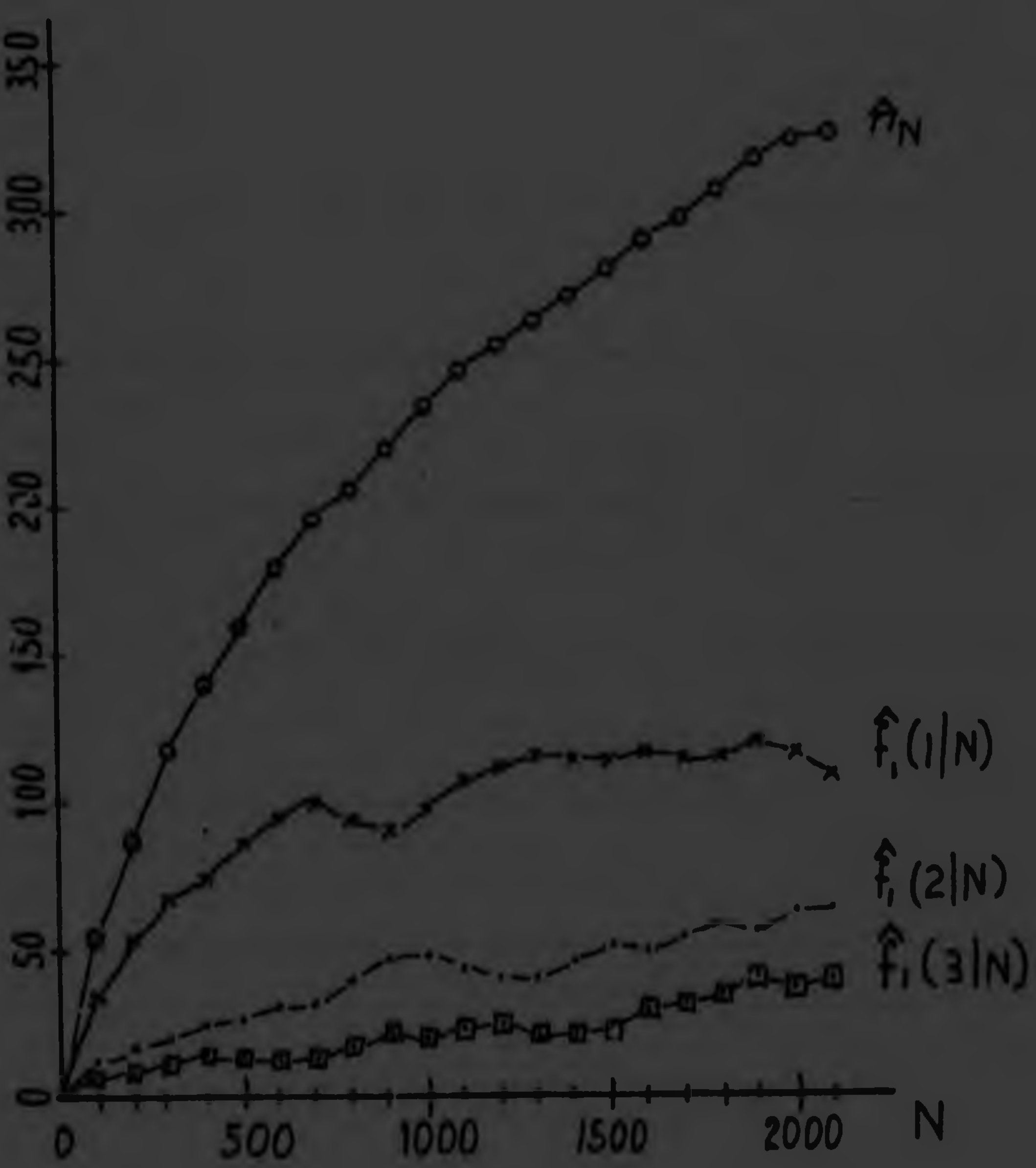


FIGURE 39 KERRY 2 RANDOM BACKWARD
$\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

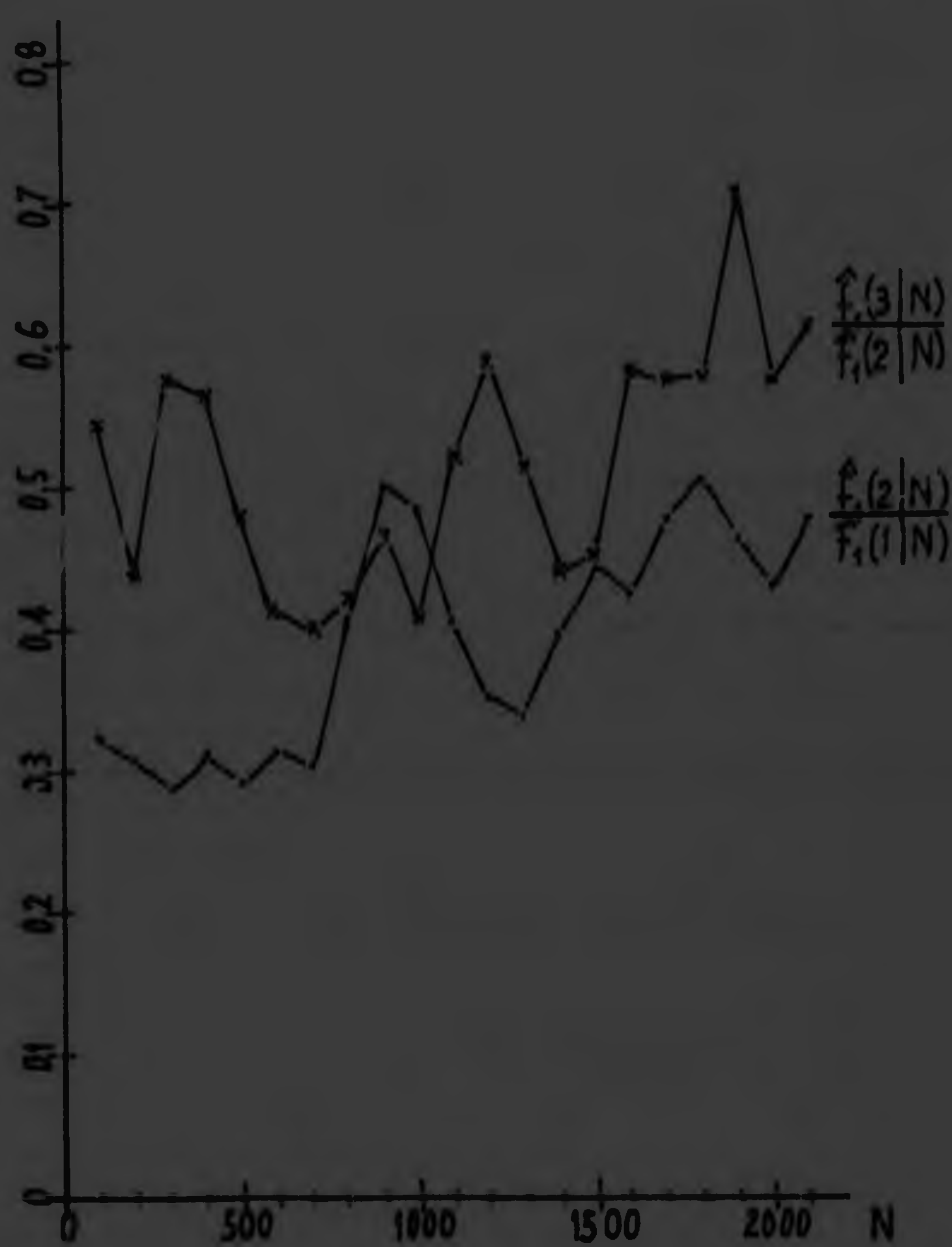


FIGURE 40 KERRY 2 RANDOM BACKWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

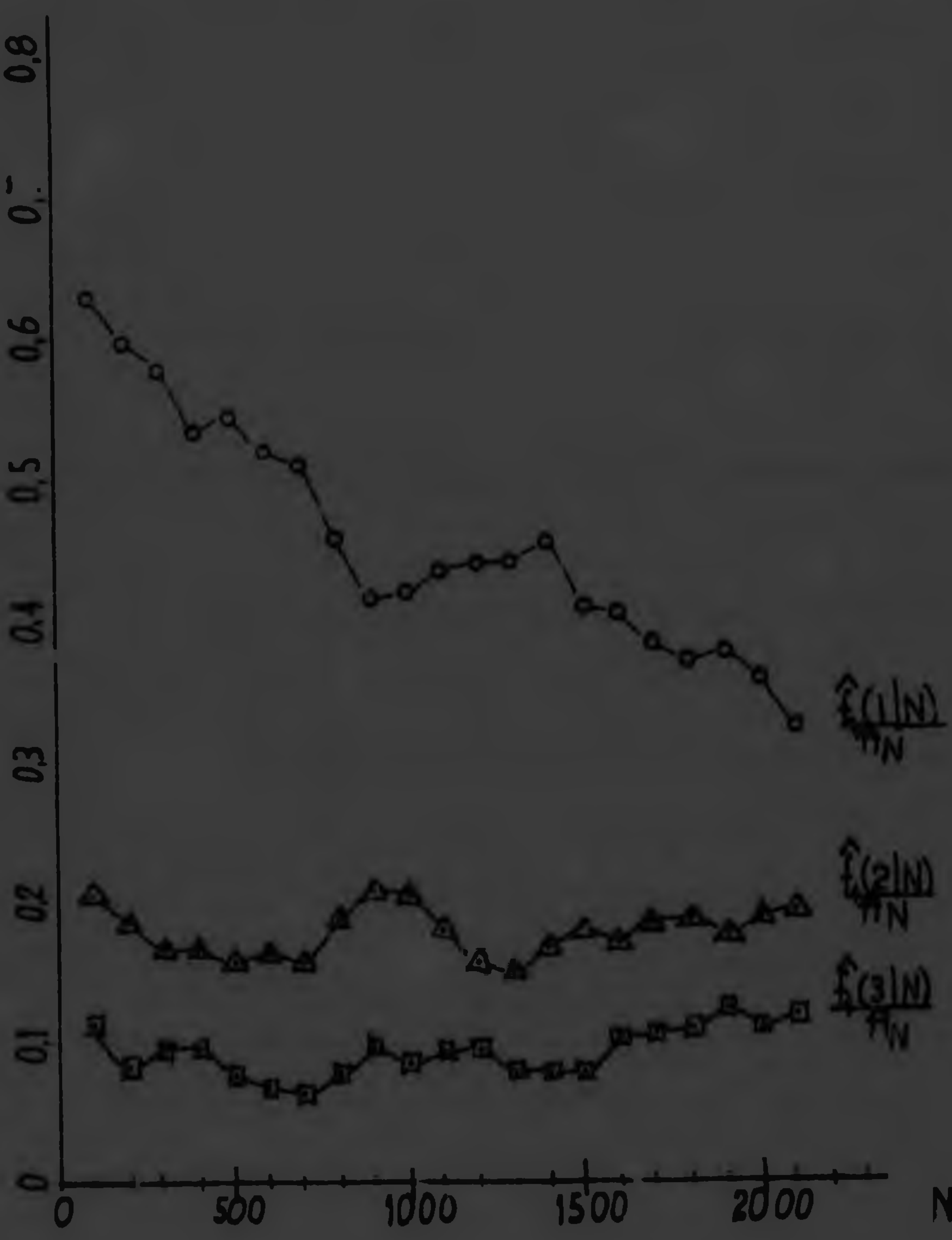


FIGURE 41 FERRY 2 RANDOM BACKWARD

$\frac{\tilde{f}_1(1|N)}{\hat{n}_N}$, $\frac{\tilde{f}_1(2|N)}{\hat{n}_N}$, $\frac{\tilde{f}_1(3|N)}{\hat{n}_N}$ vs N

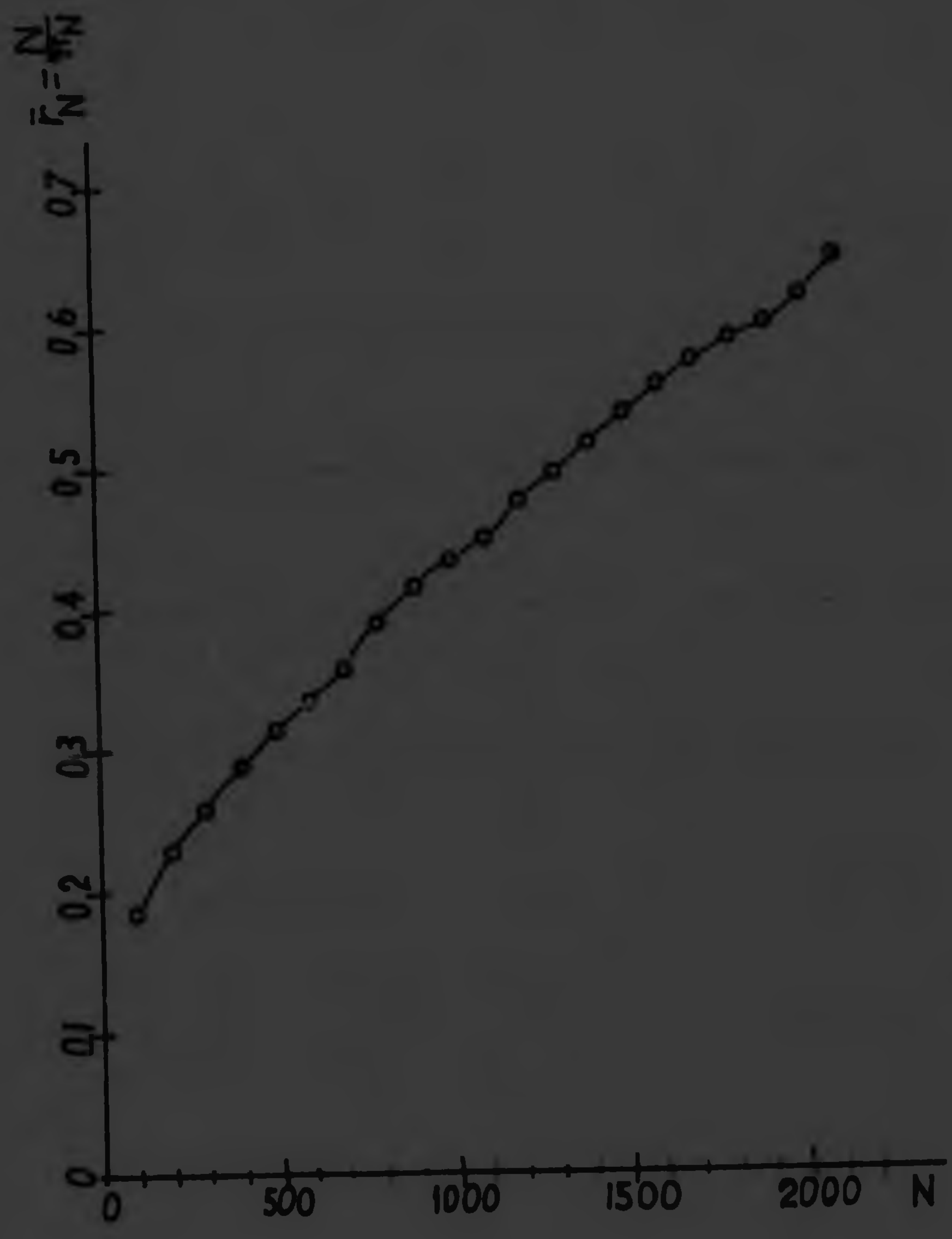


FIGURE 42 KERRY 2 RANDOM BACKWARD

$\bar{r}_N = \frac{N}{n_N}$ vs N

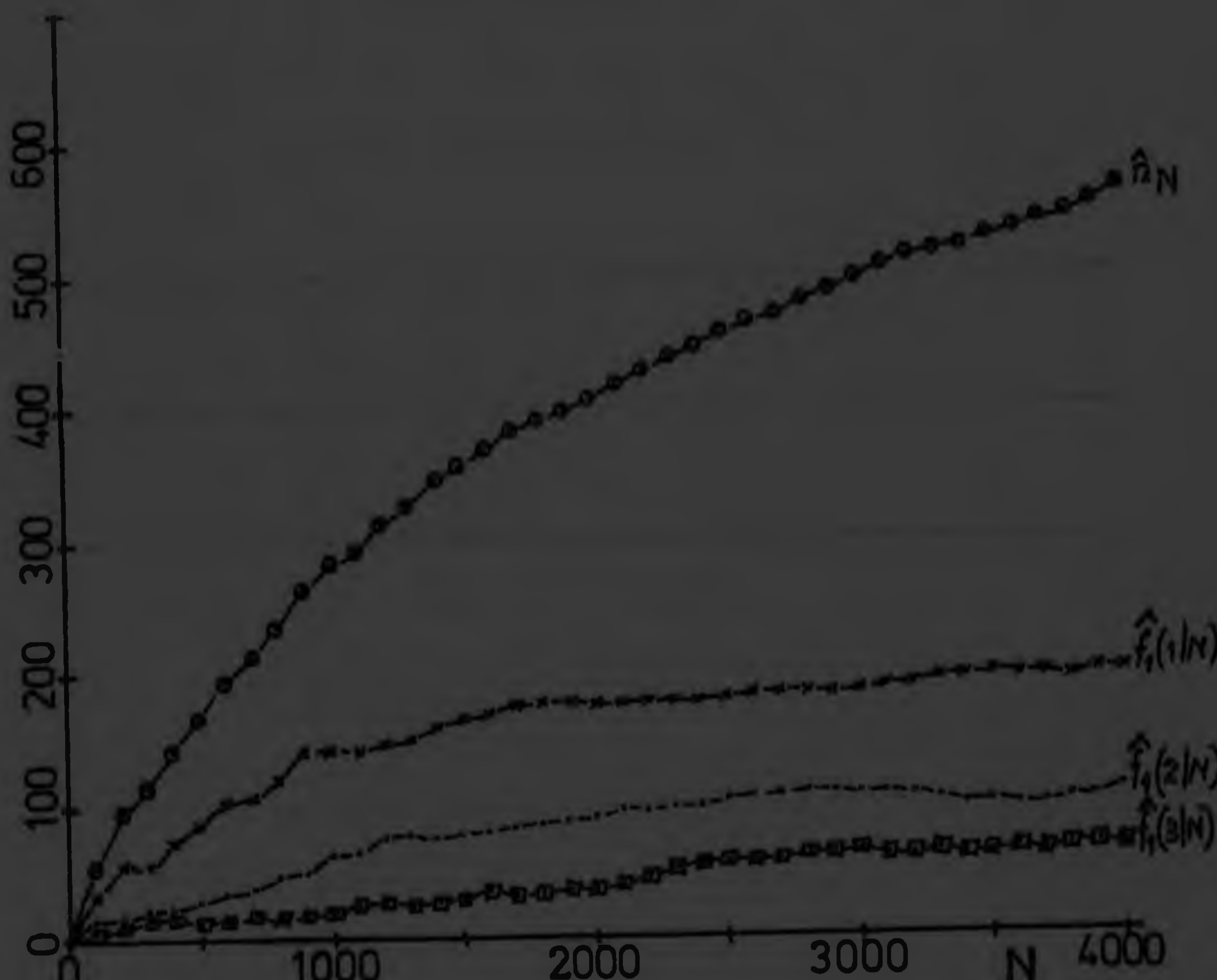


FIGURE 43 BRONWYN 1 FORWARD

$\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

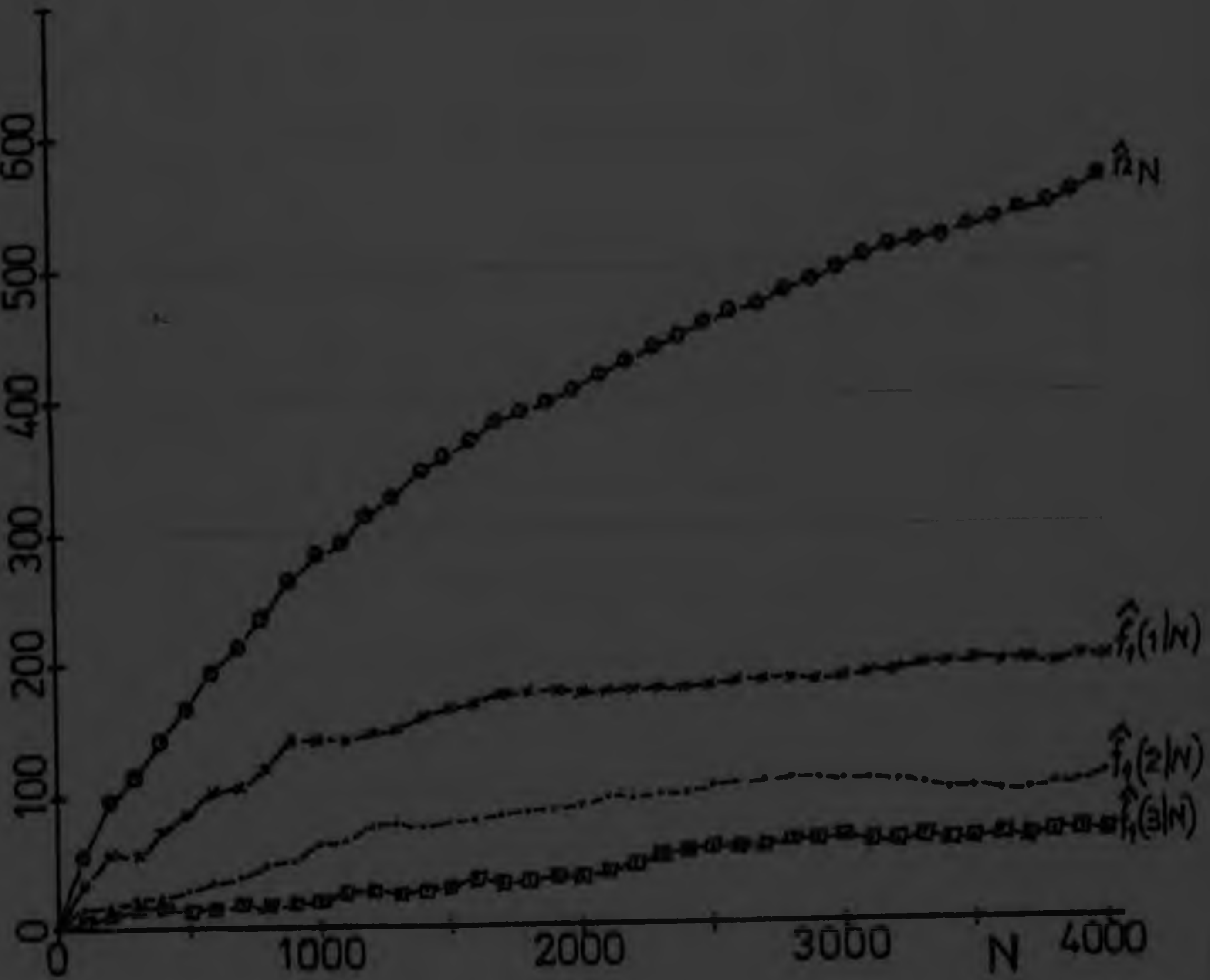


FIGURE 43 BRONWYN 1 FORWARD
$\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

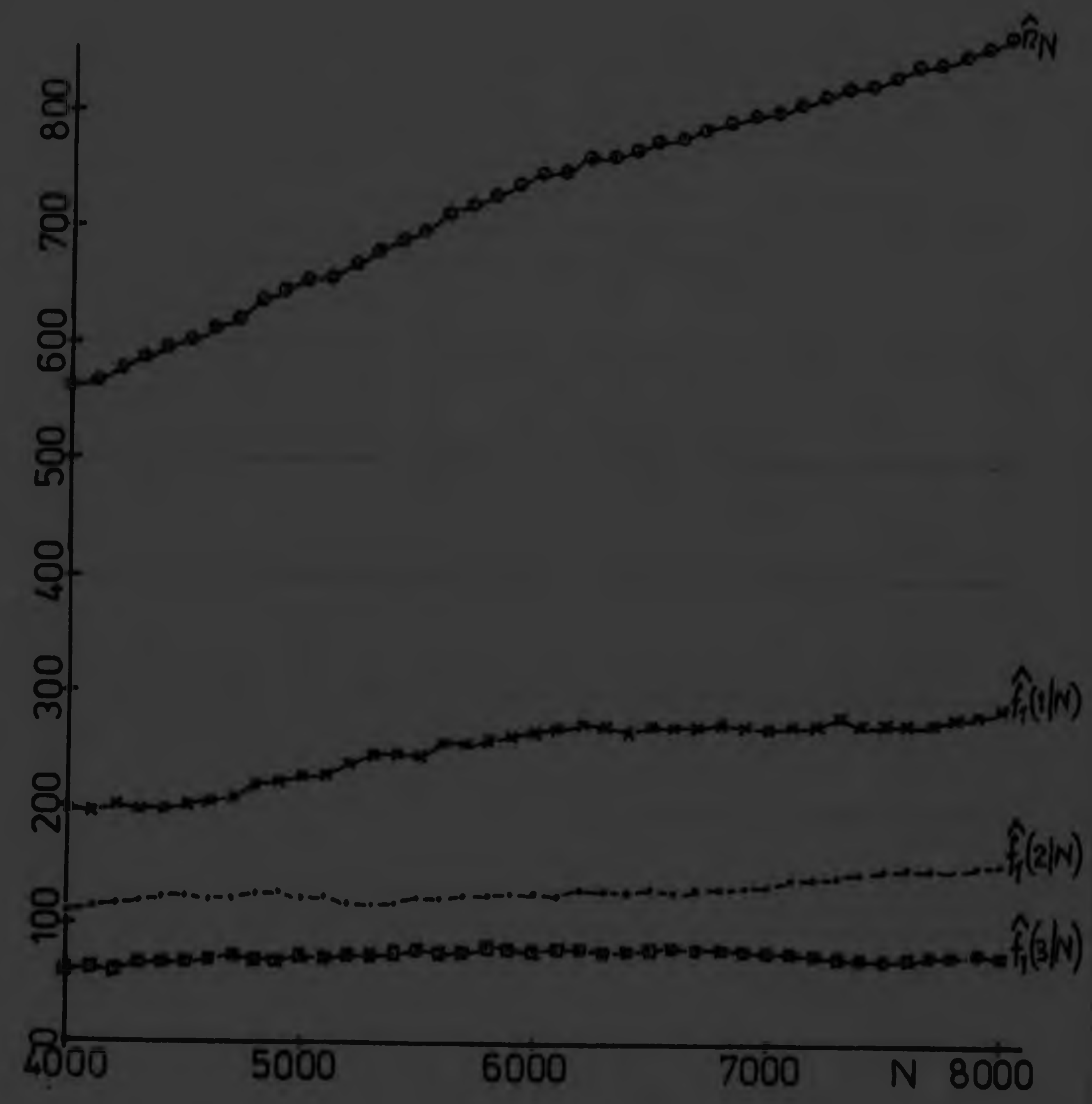


FIGURE 43 contd.

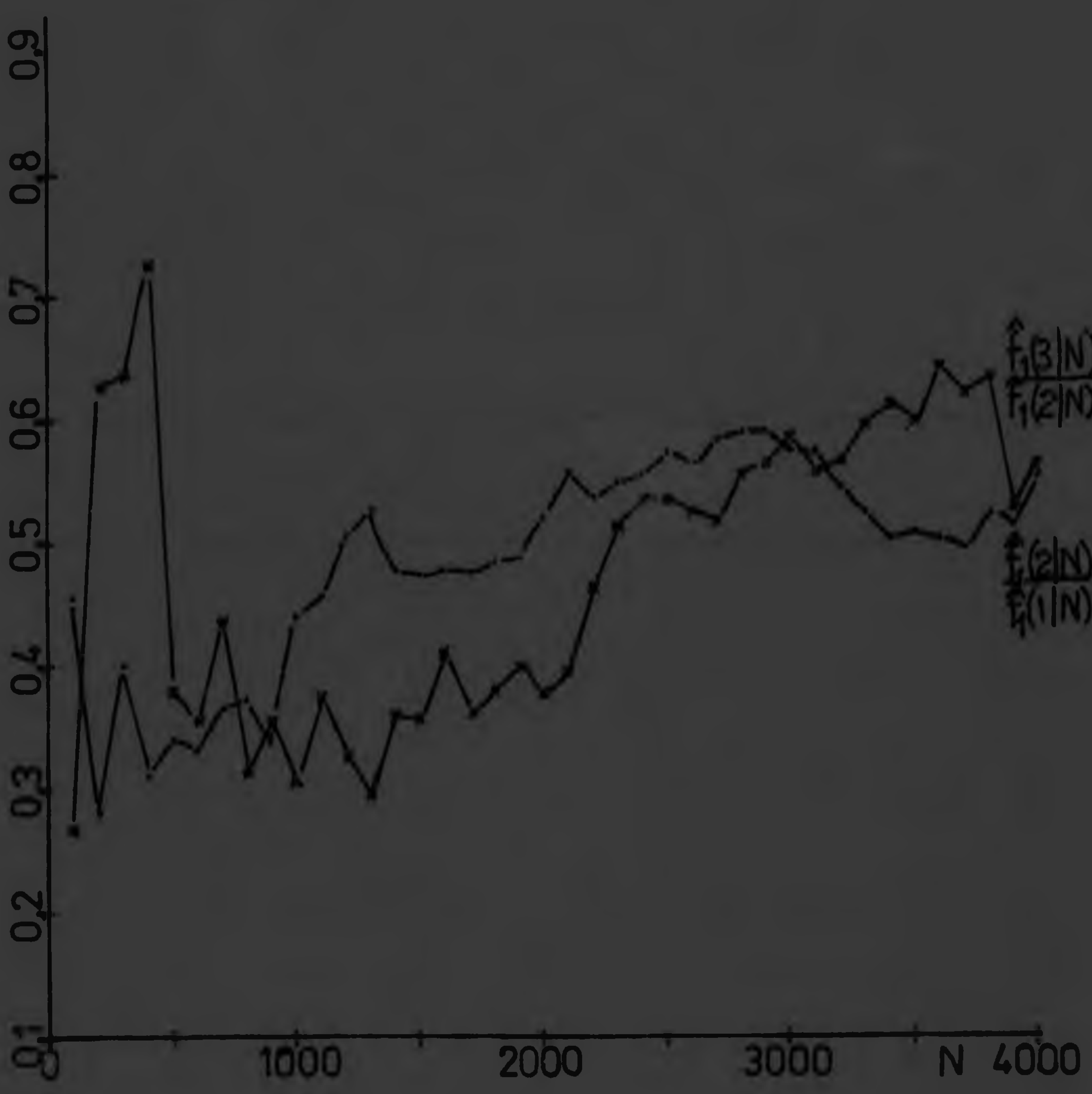


FIGURE 44 BRONWYN 1 FORWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

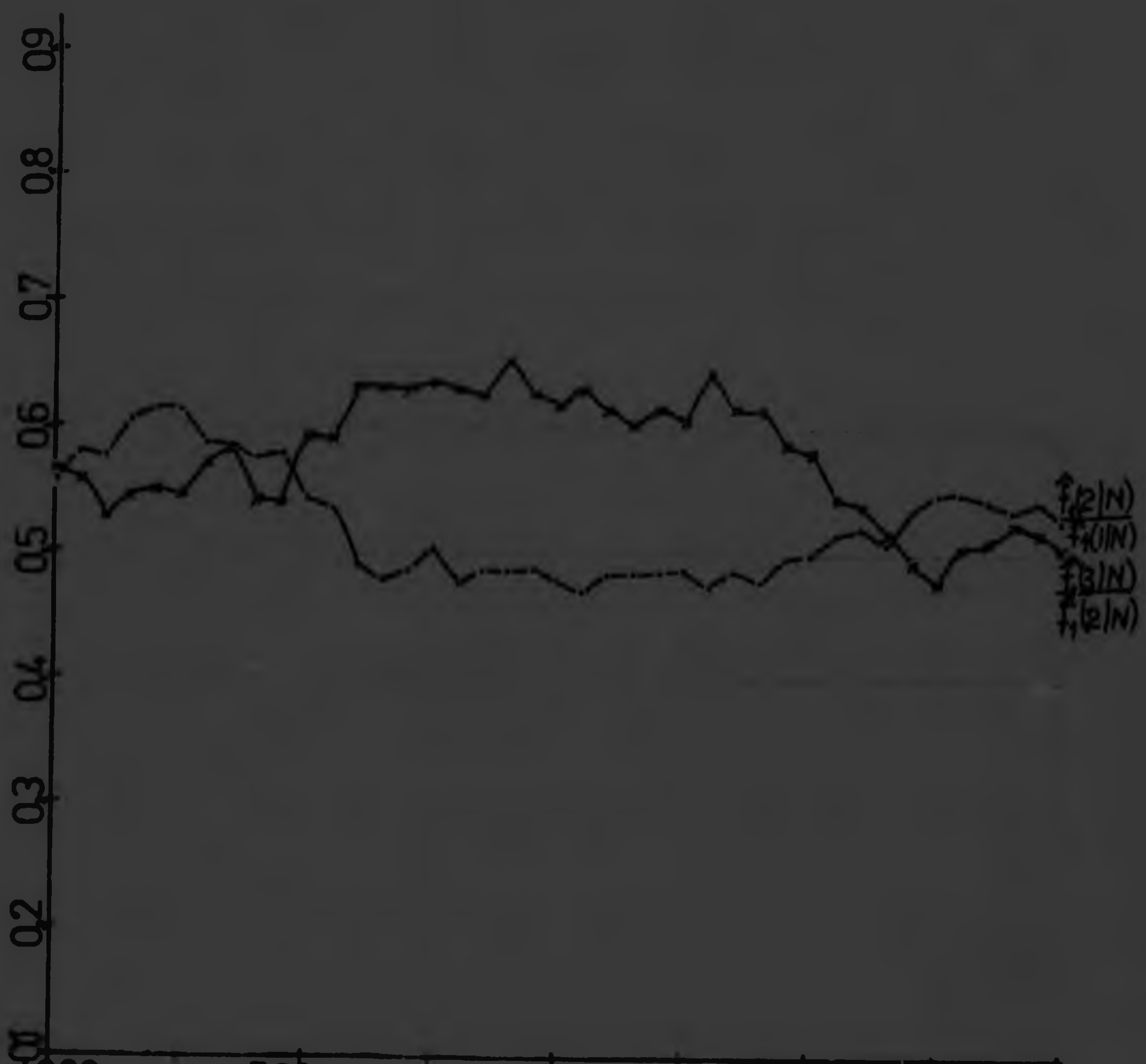


FIGURE 44 contd.

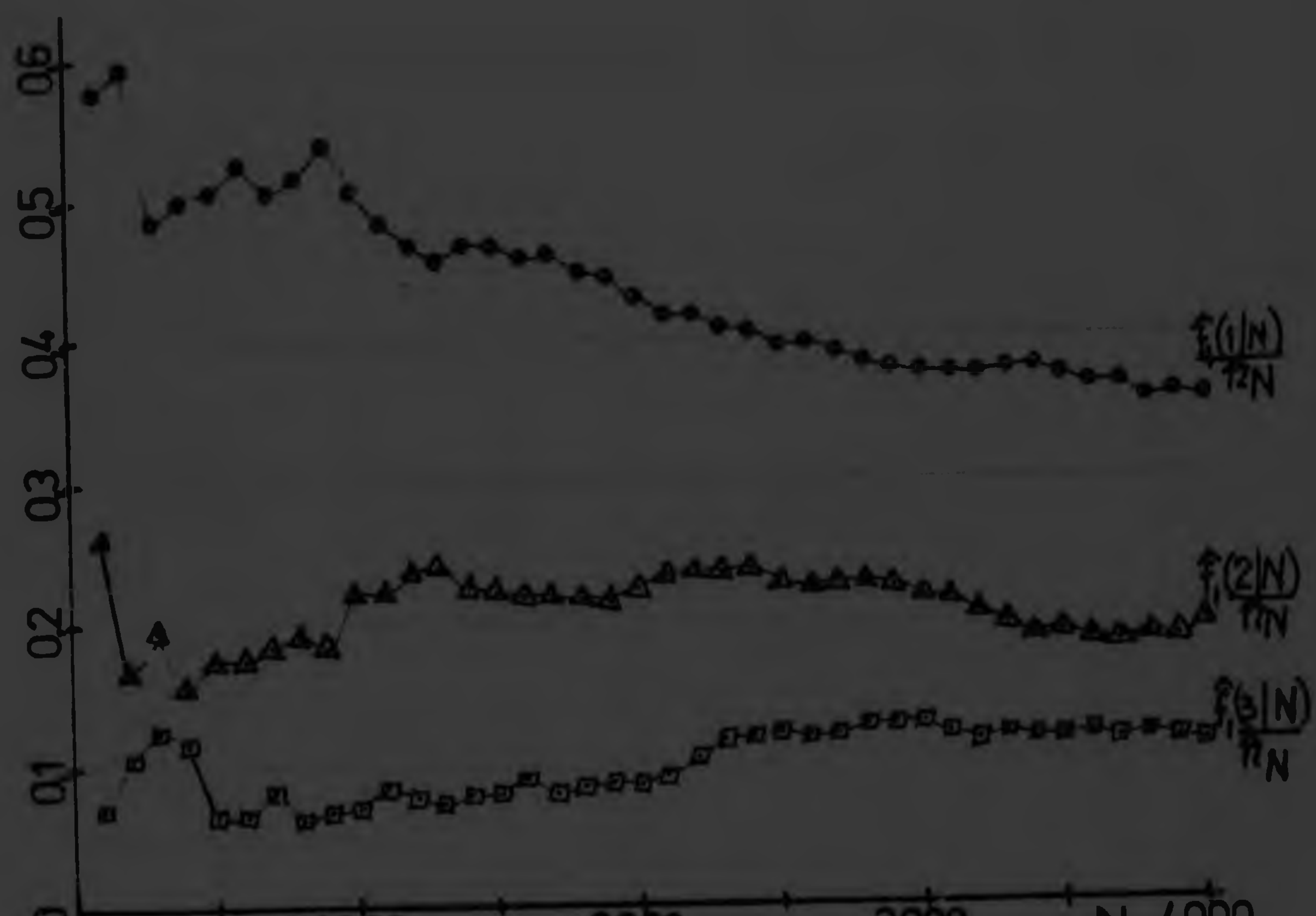


FIGURE 45 BRONWYN 1 FORWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

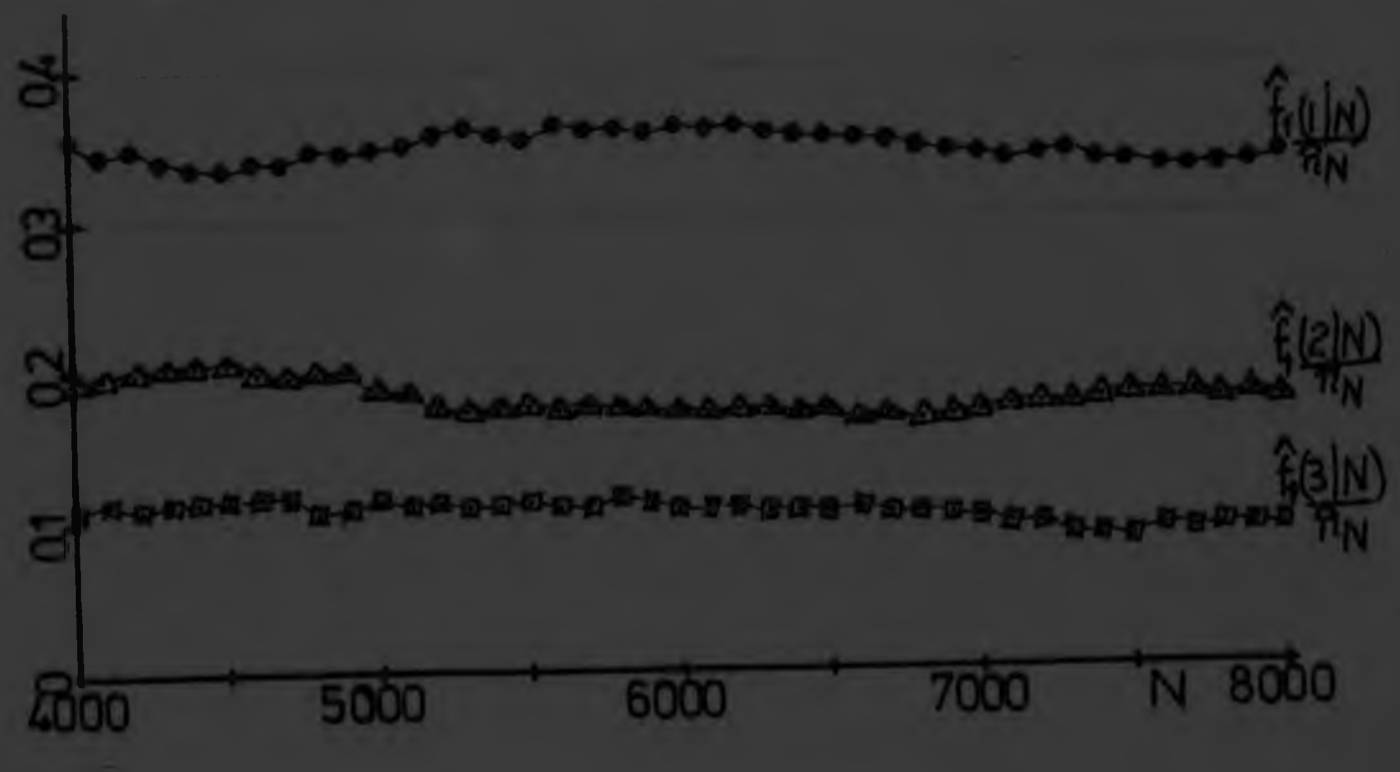


FIGURE 45 contd.

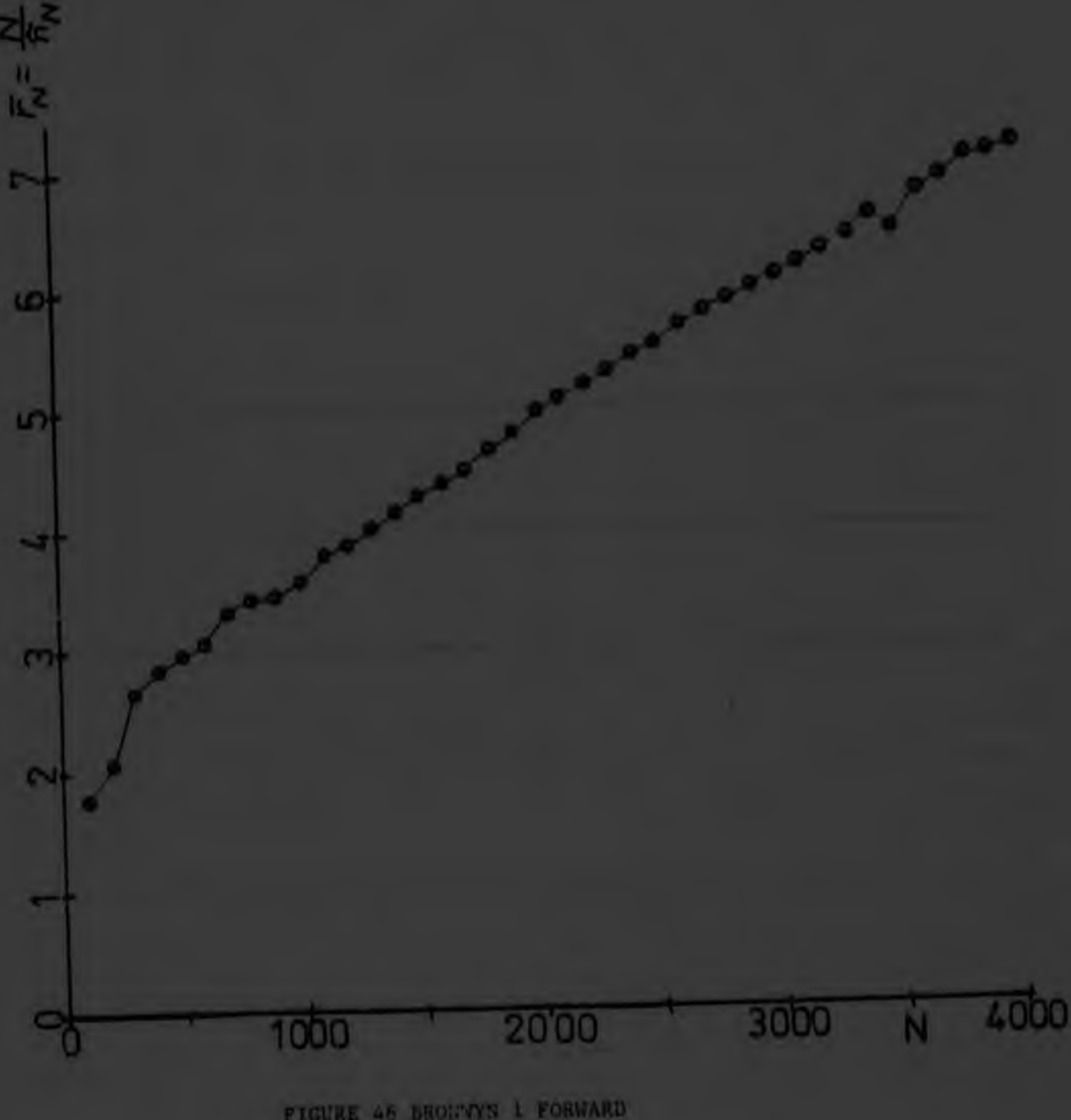


FIGURE 46 BRONWYN 1 FORWARD

$\bar{r}_N = \frac{N}{\bar{n}_N}$ vs N

FIGURE 46 contd.

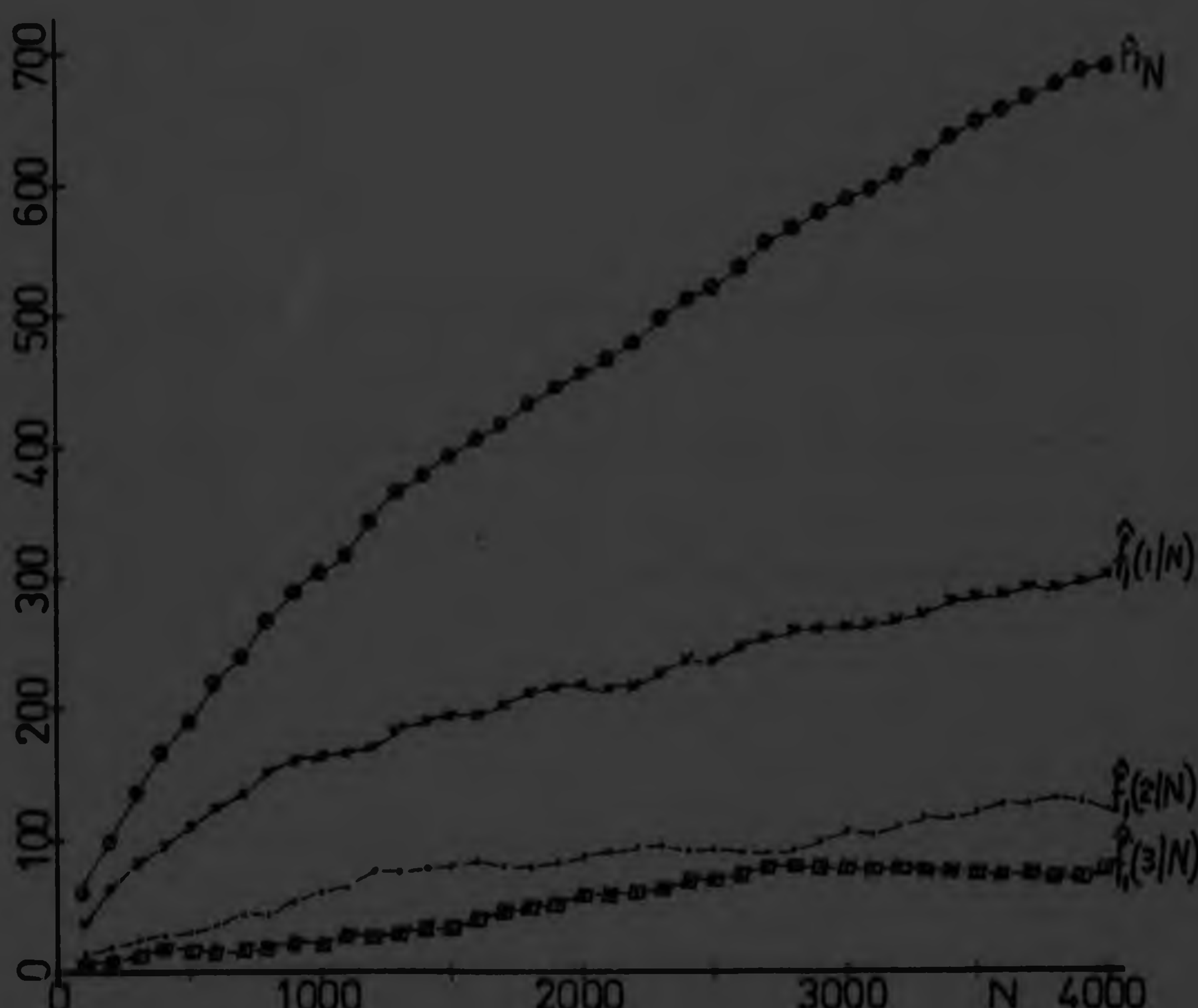


FIGURE 47 BRONWYN 1 BACKWARD

$\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

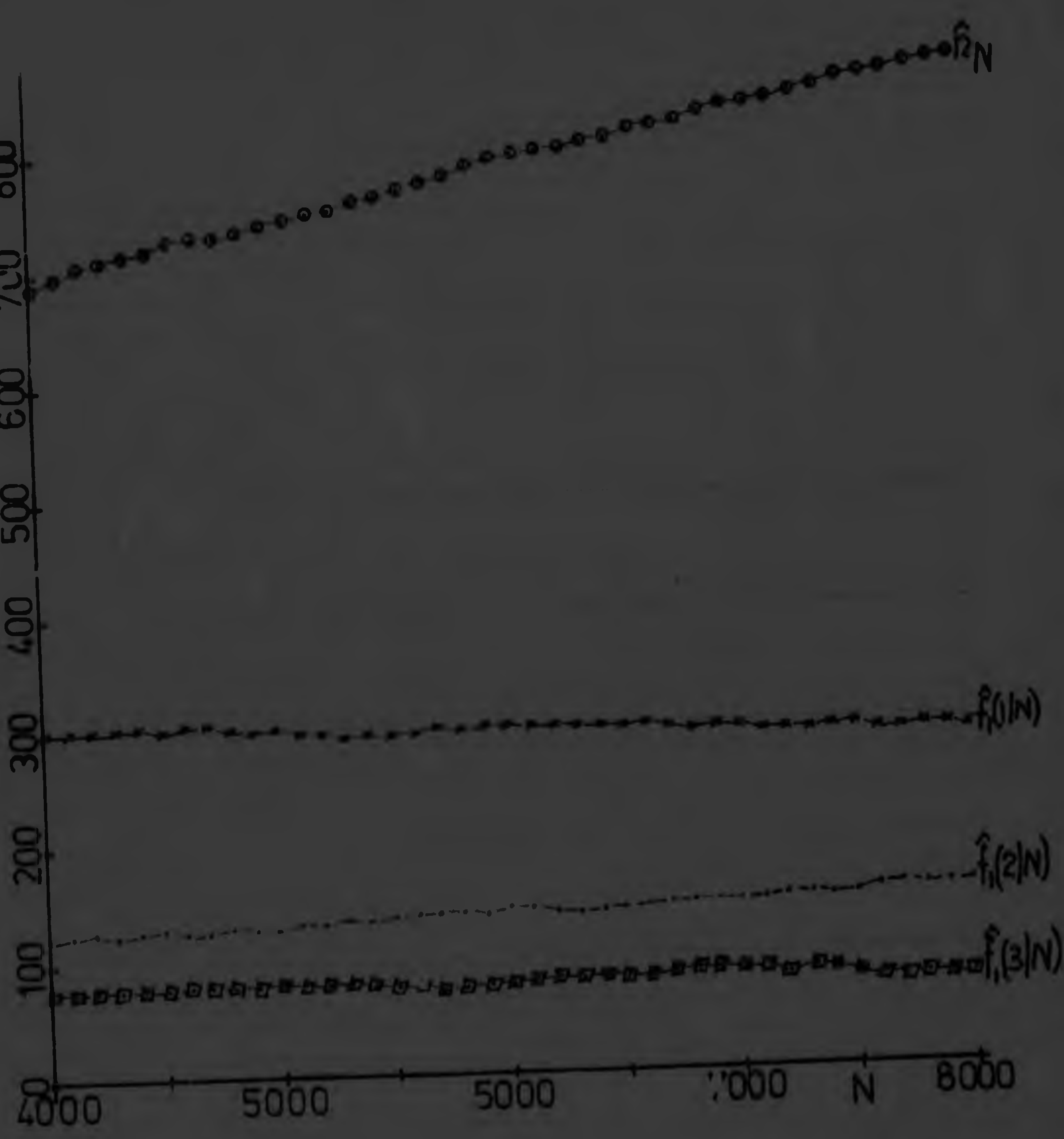


FIGURE 47 contd.

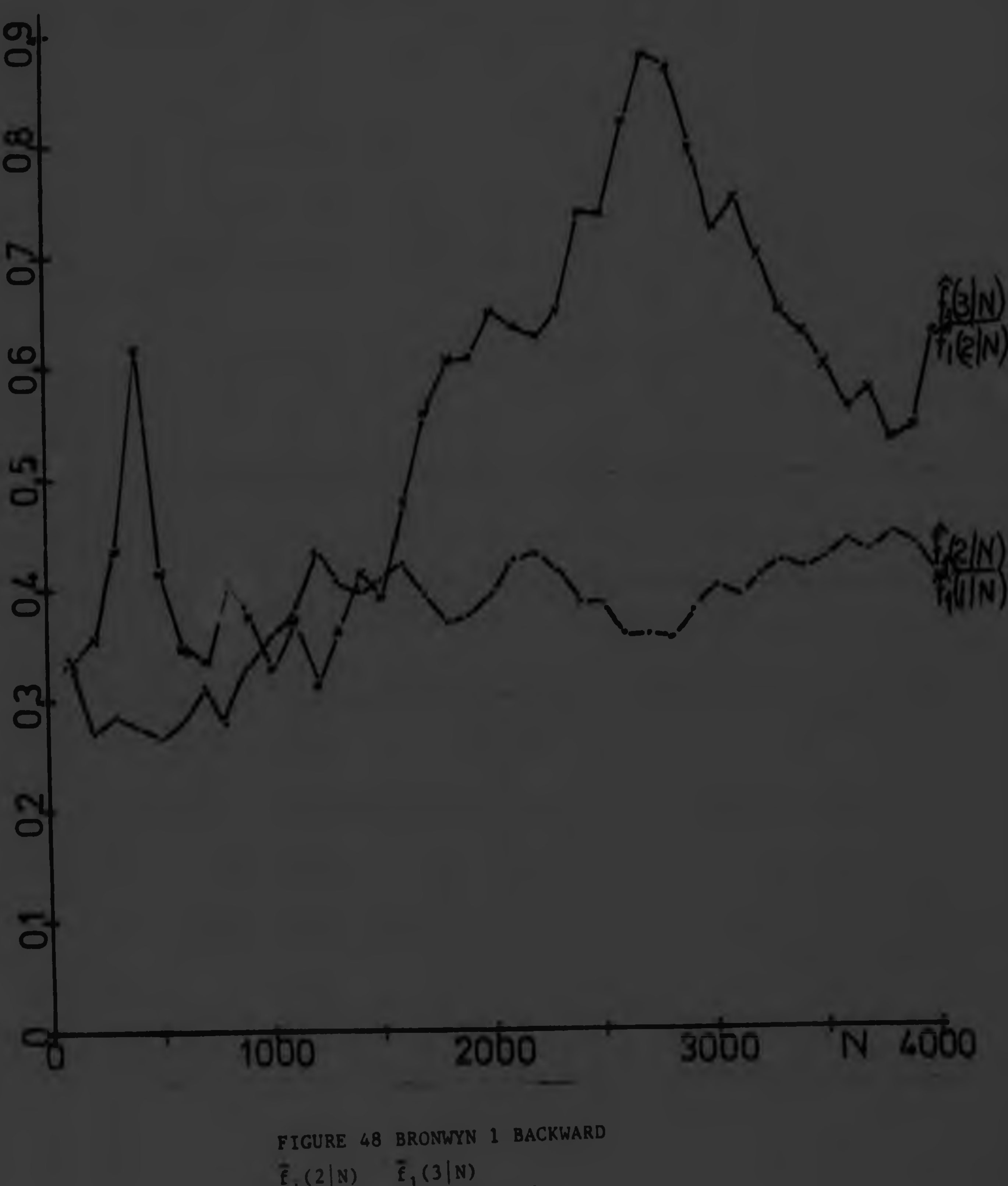


FIGURE 48 BRONWYN 1 BACKWARD

$\frac{\bar{f}_1(2|N)}{\bar{f}_1(1|N)}$, $\frac{\bar{f}_1(3|N)}{\bar{f}_1(2|N)}$ vs N

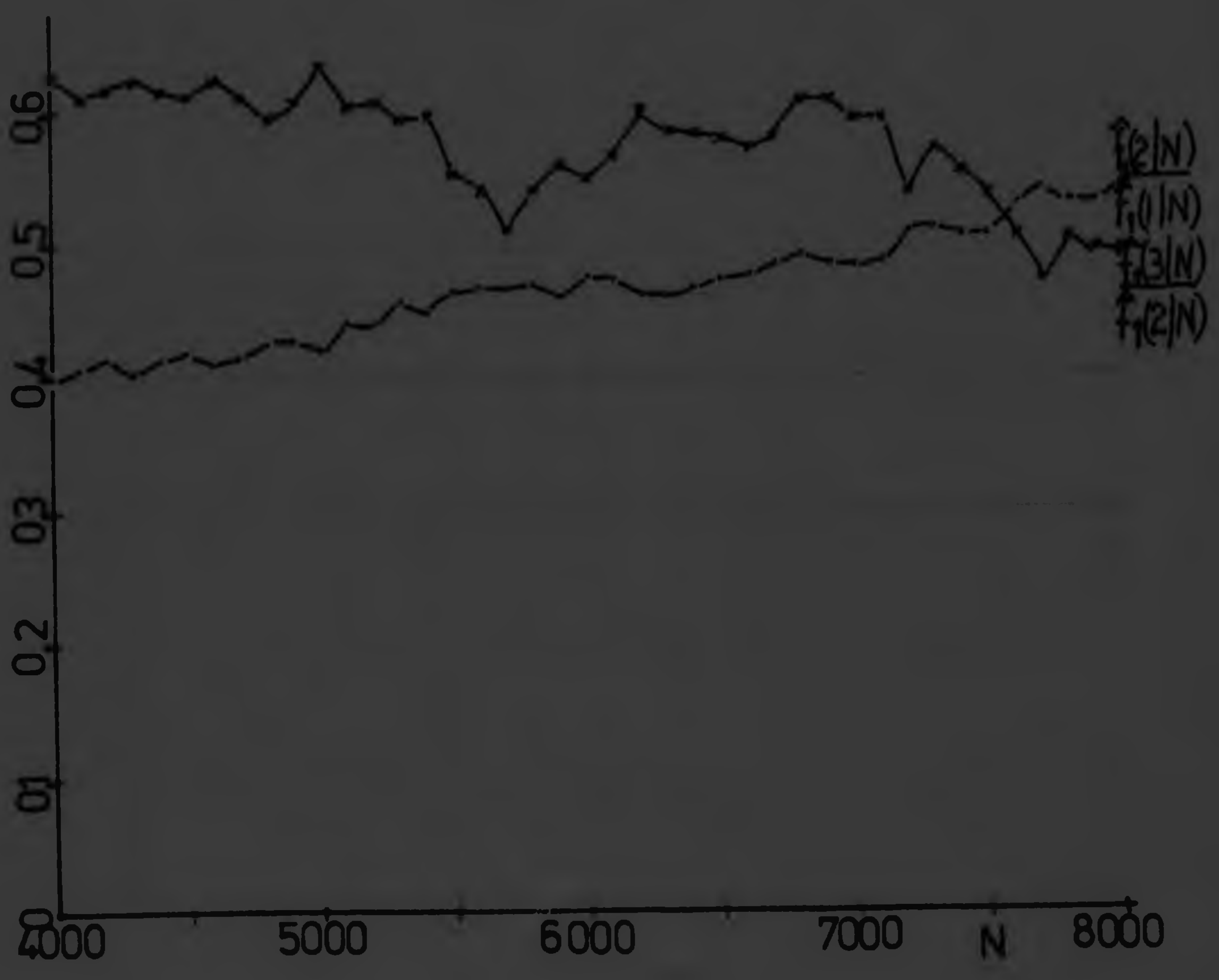


FIGURE 48 contd.

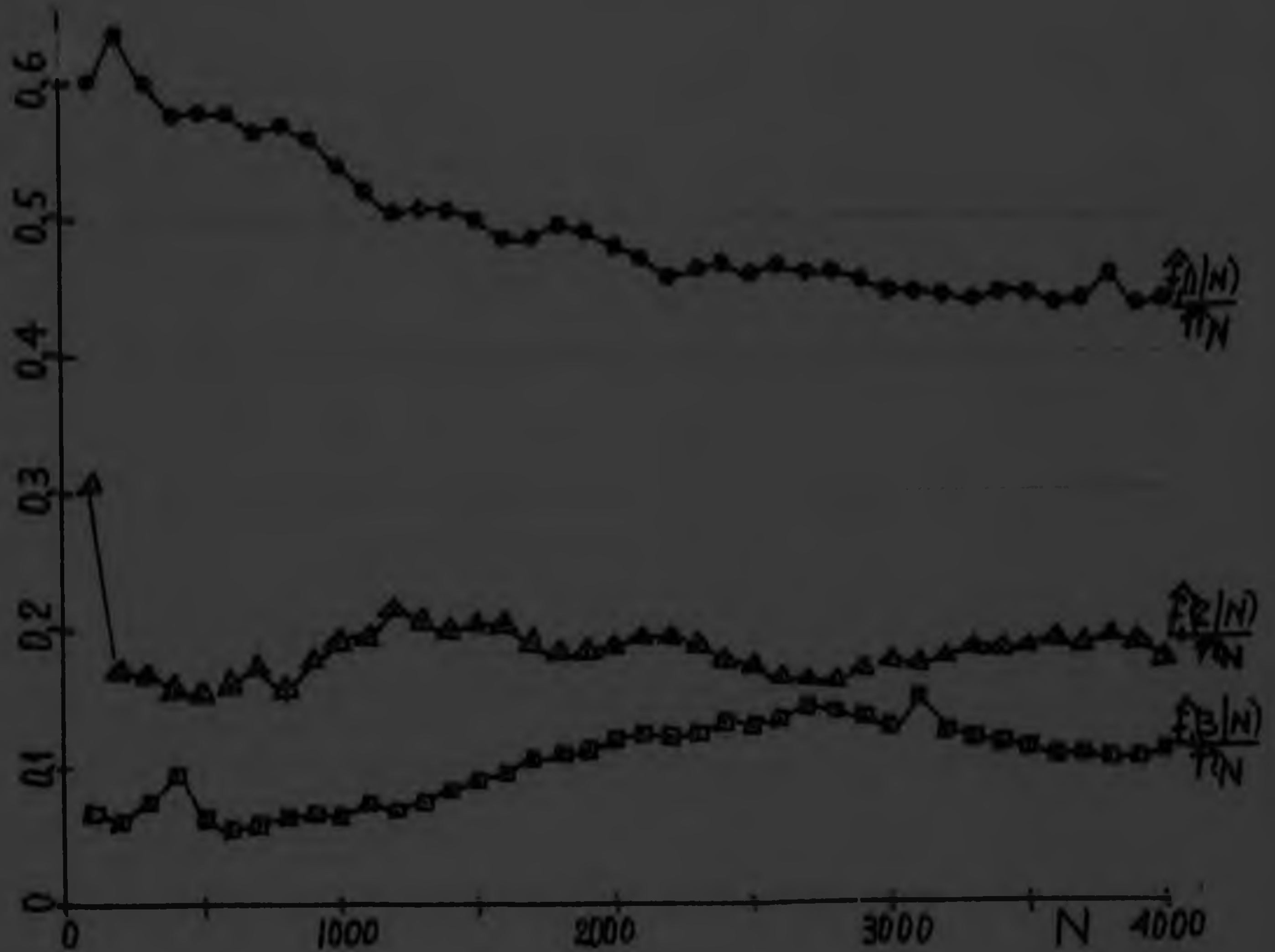


FIGURE 49 BRONWYN 1 BACKWARD

$\frac{\tilde{f}_1(1|N)}{\tilde{n}_N}$, $\frac{\tilde{f}_1(2|N)}{\tilde{n}_N}$, $\frac{\tilde{f}_1(3|N)}{\tilde{n}_N}$ vs N

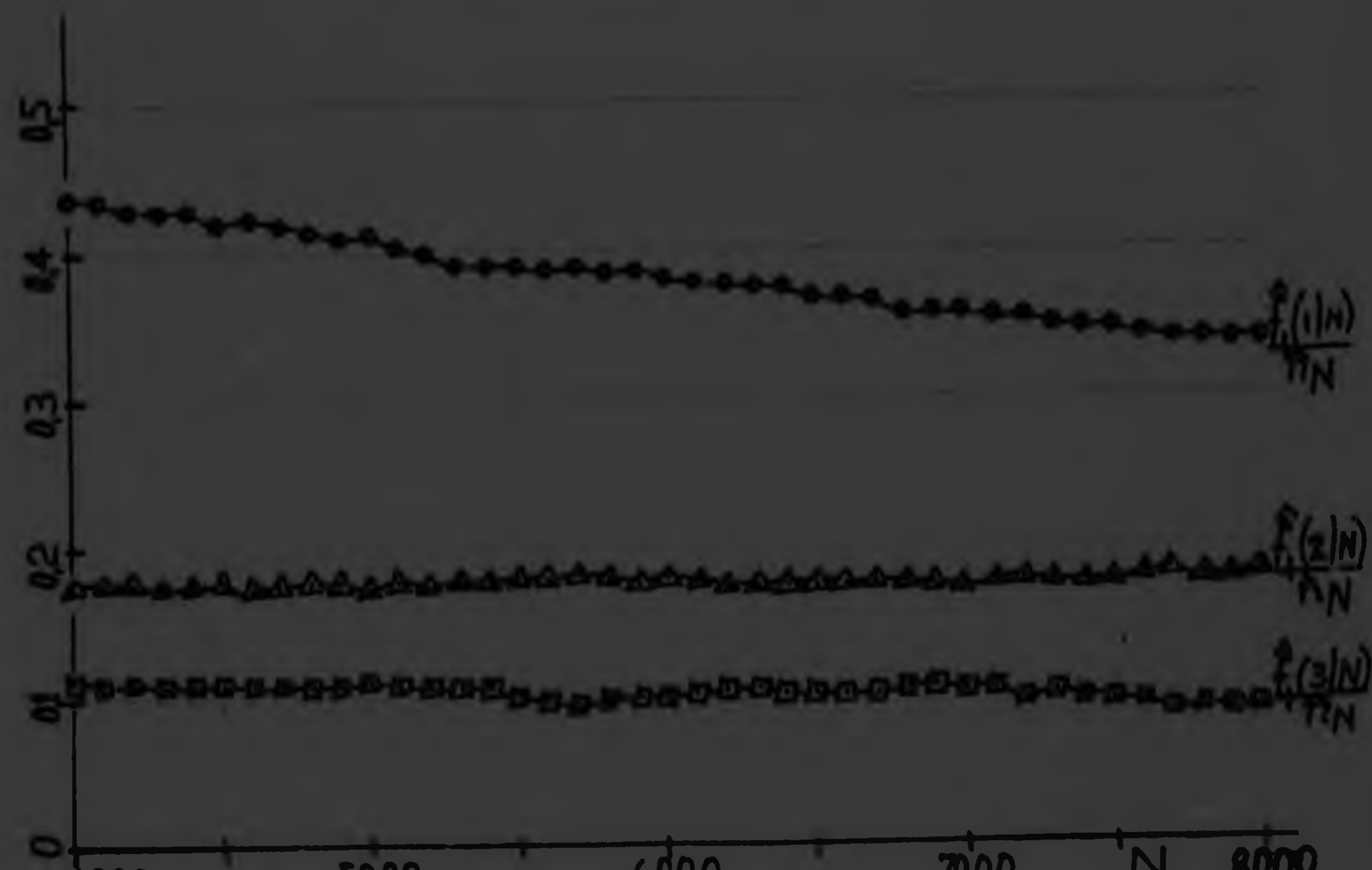


FIGURE 49 contd.

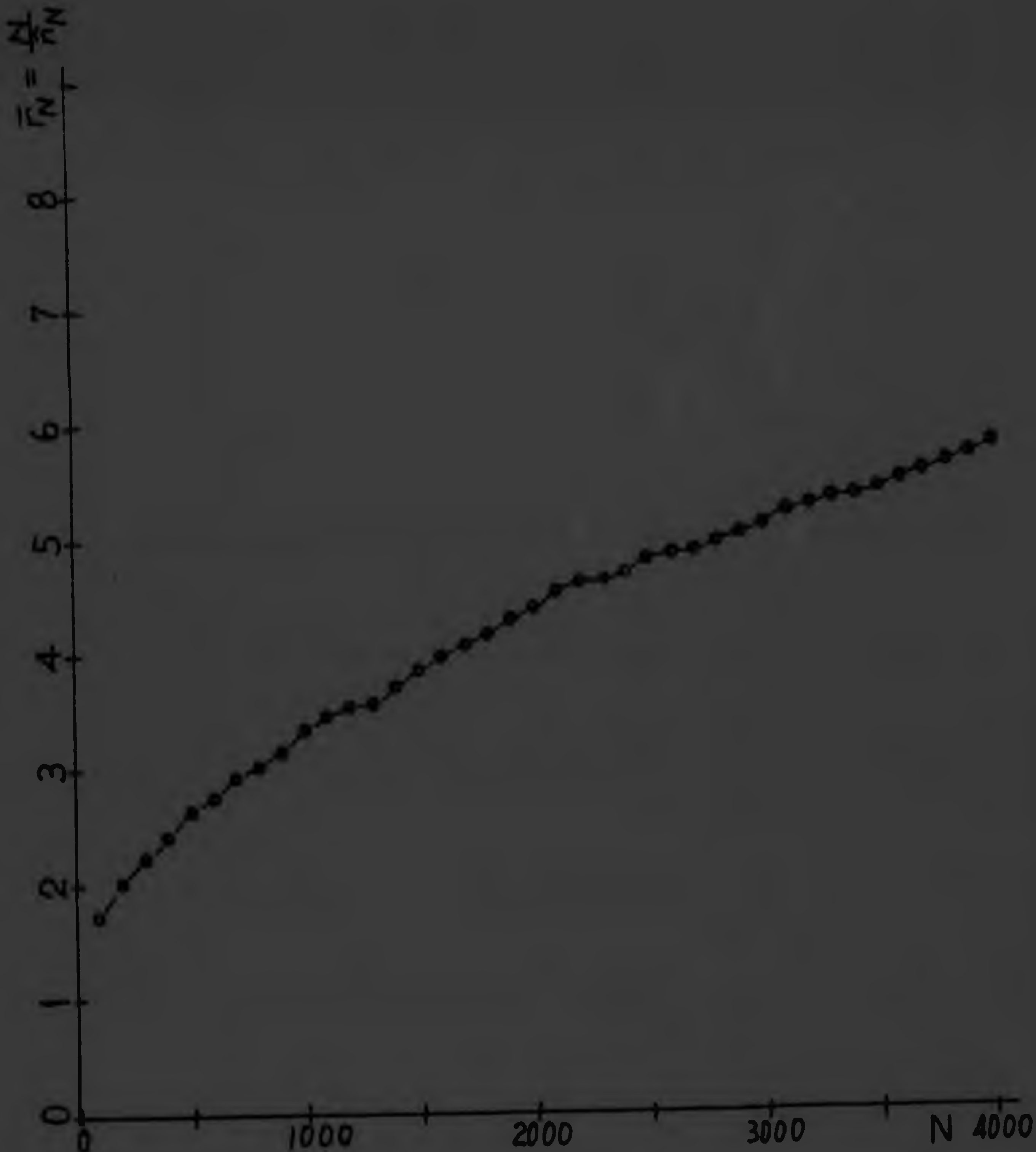


FIGURE 50 BRONWYN 1 BACKWARD

$\bar{r}_N = \frac{N}{n_N}$ vs N

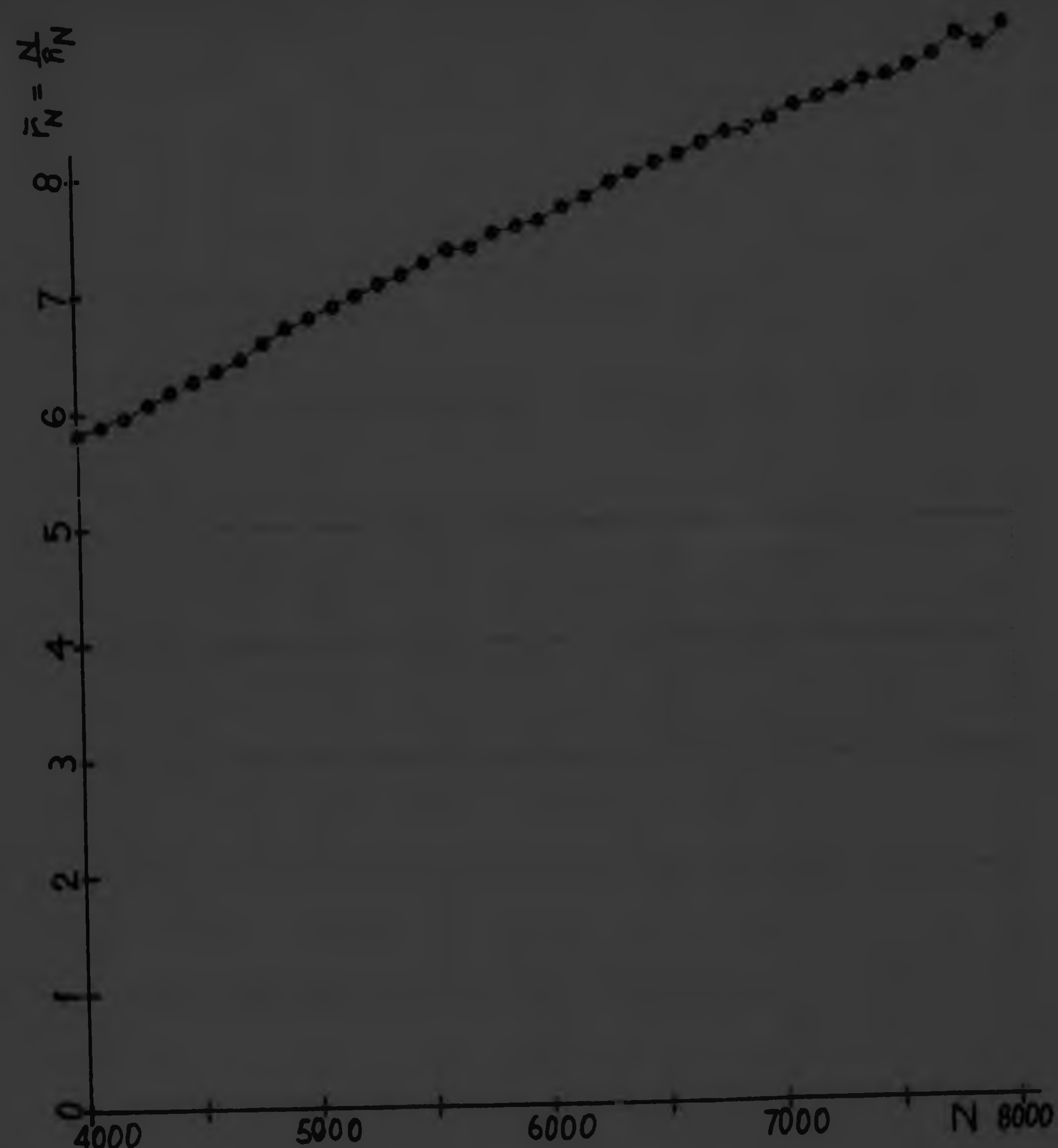


FIGURE 50 contd.

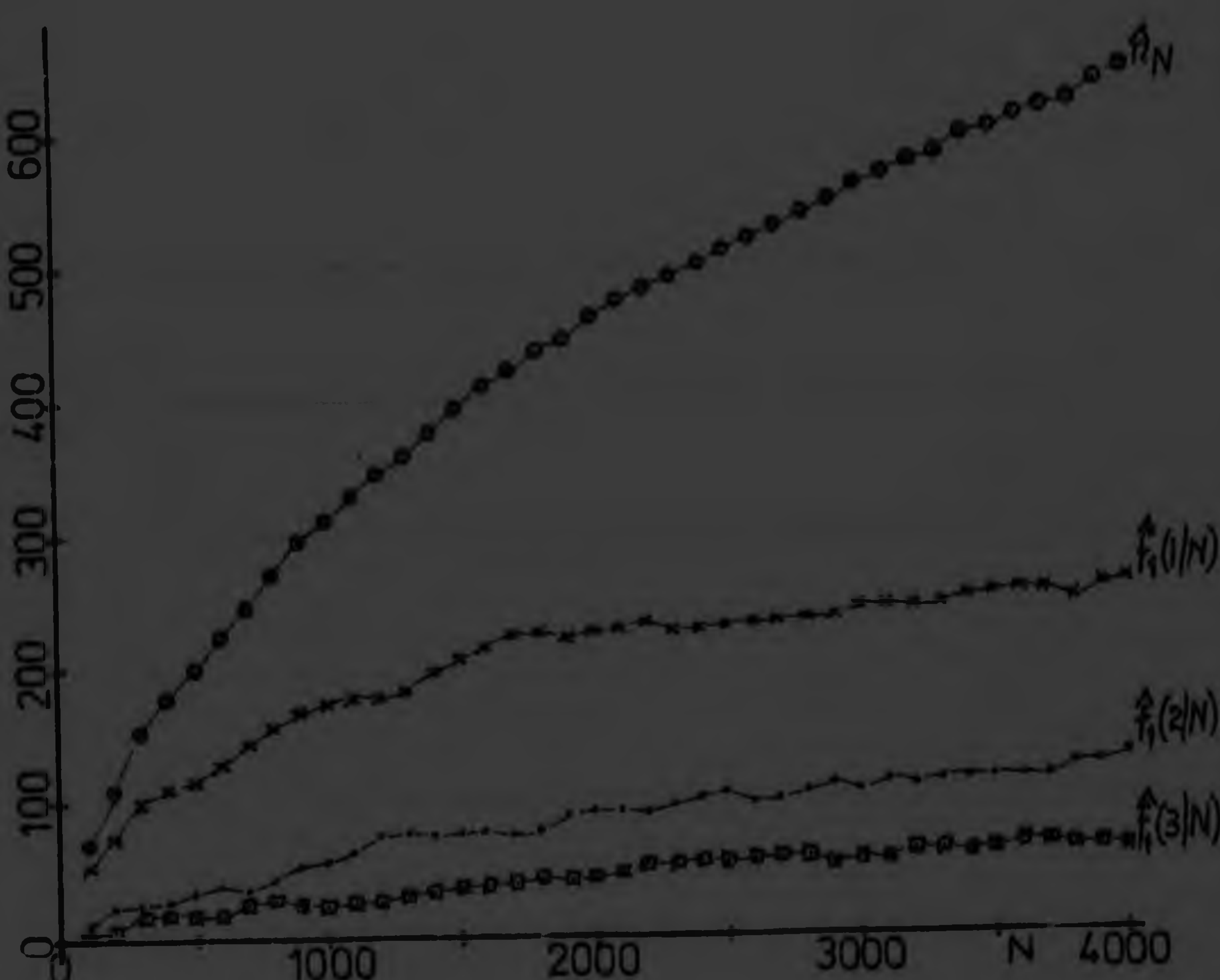


FIGURE 51 BRONWYN 1 RANDOM FORWARD
$\hat{a}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

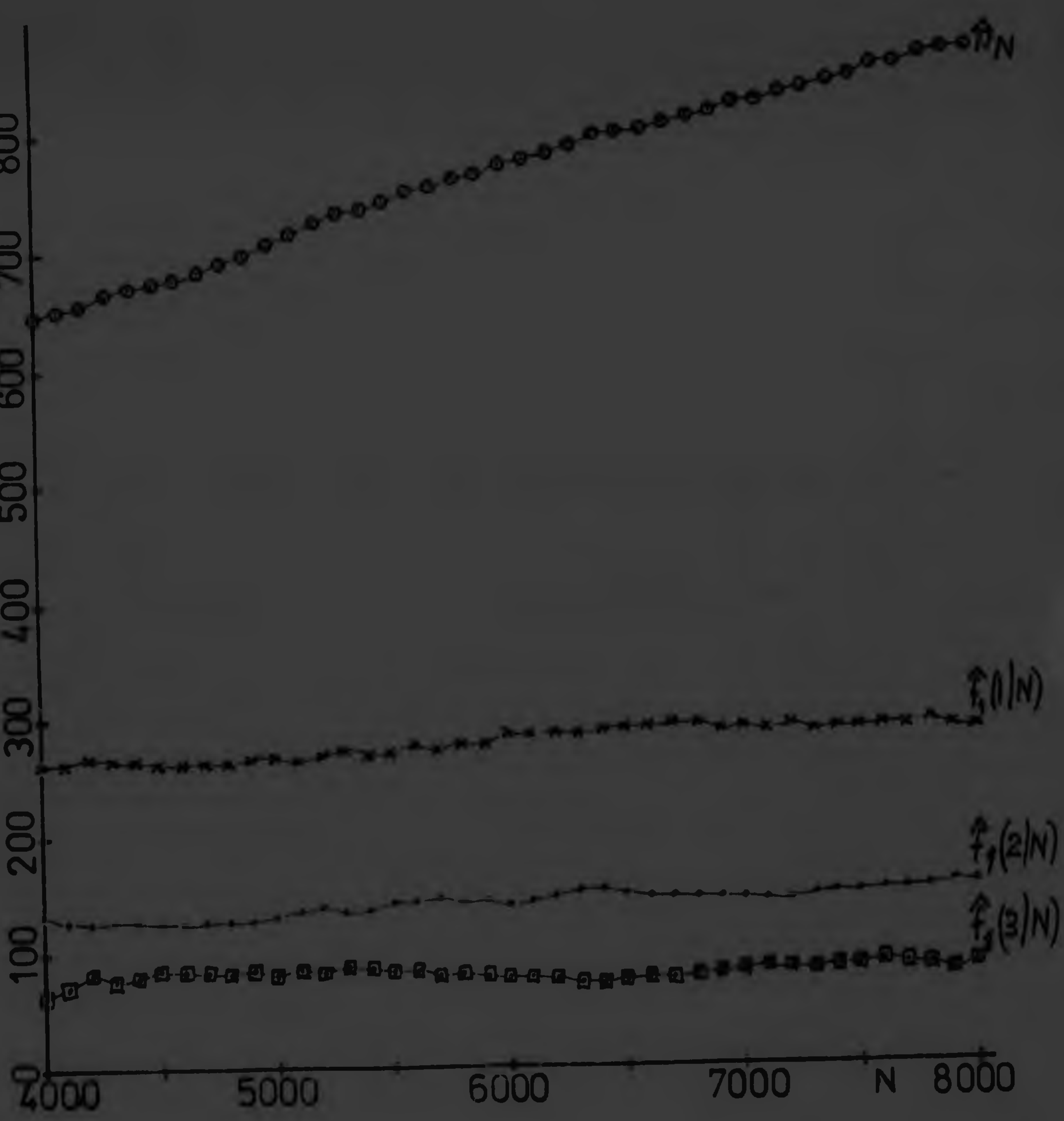


FIGURE 51 contd.

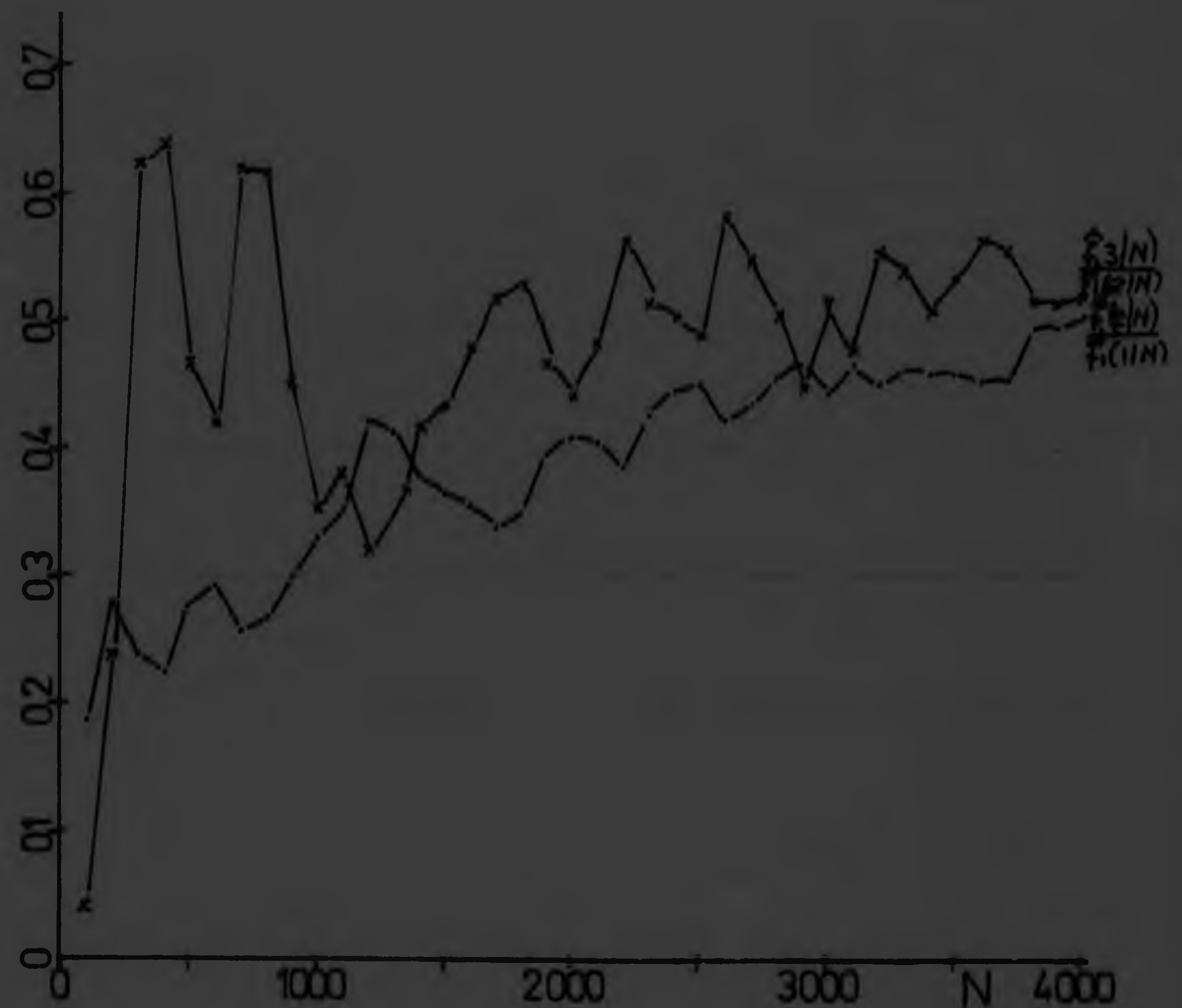


FIGURE 52 BRONWYN 1 RANDOM FORWARD

$$\frac{\bar{f}_1(2|N)}{\bar{f}_1(1|N)}, \frac{\bar{f}_1(3|N)}{\bar{f}_1(2|N)} \text{ vs N}$$

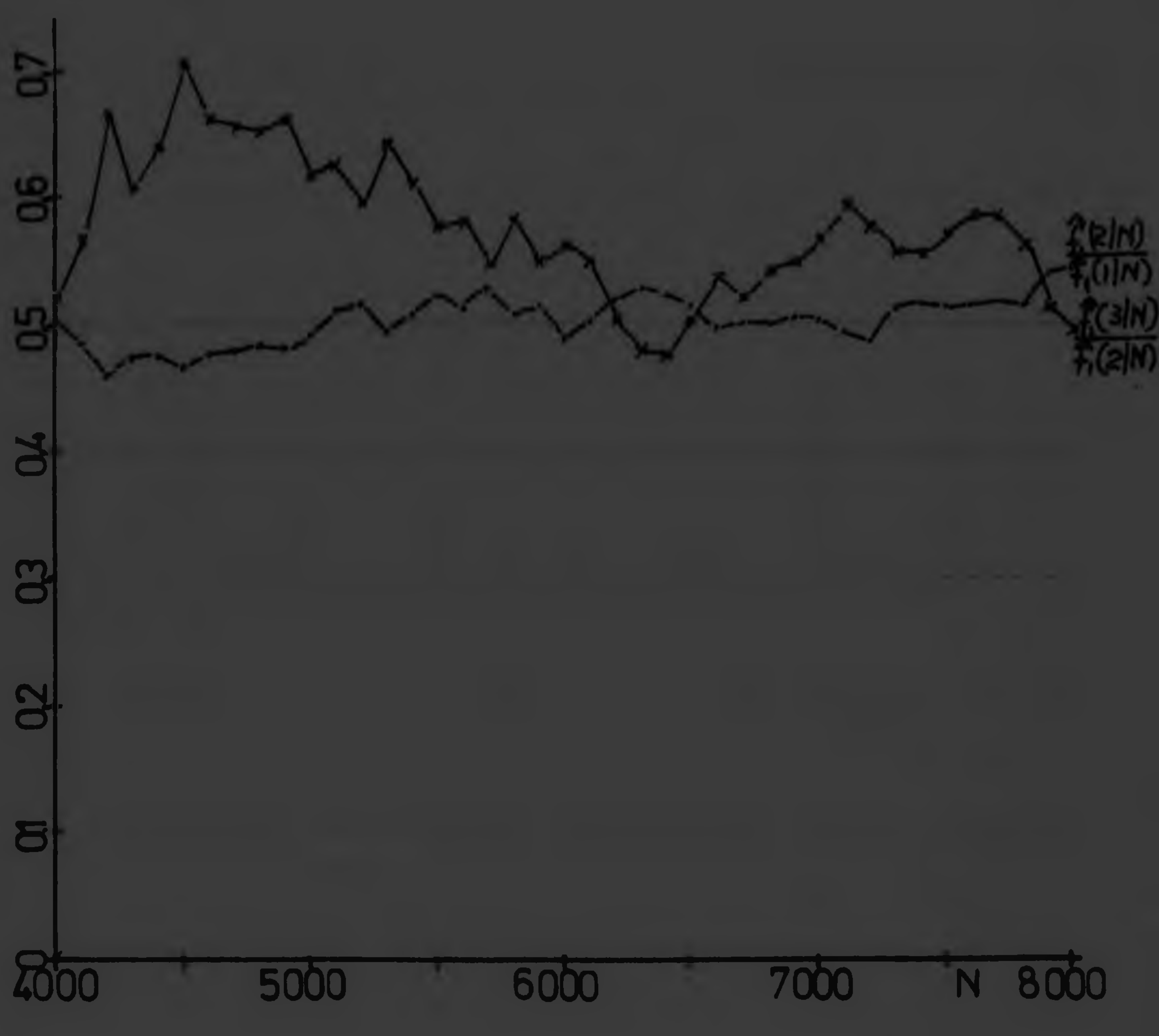
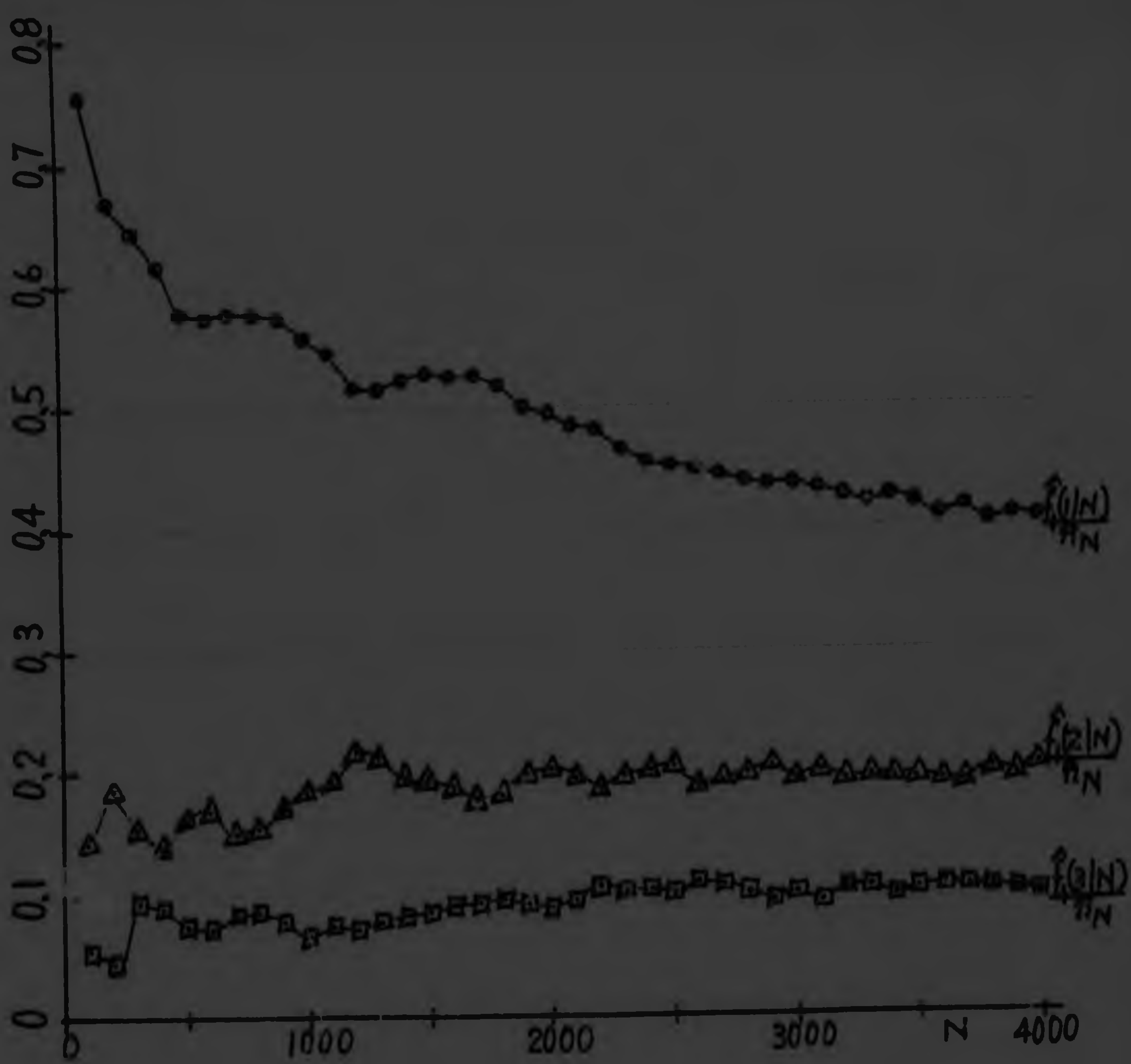


FIGURE 52 contd.



?IGURE 53 BRONWYN 1 RANDOM FORWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$

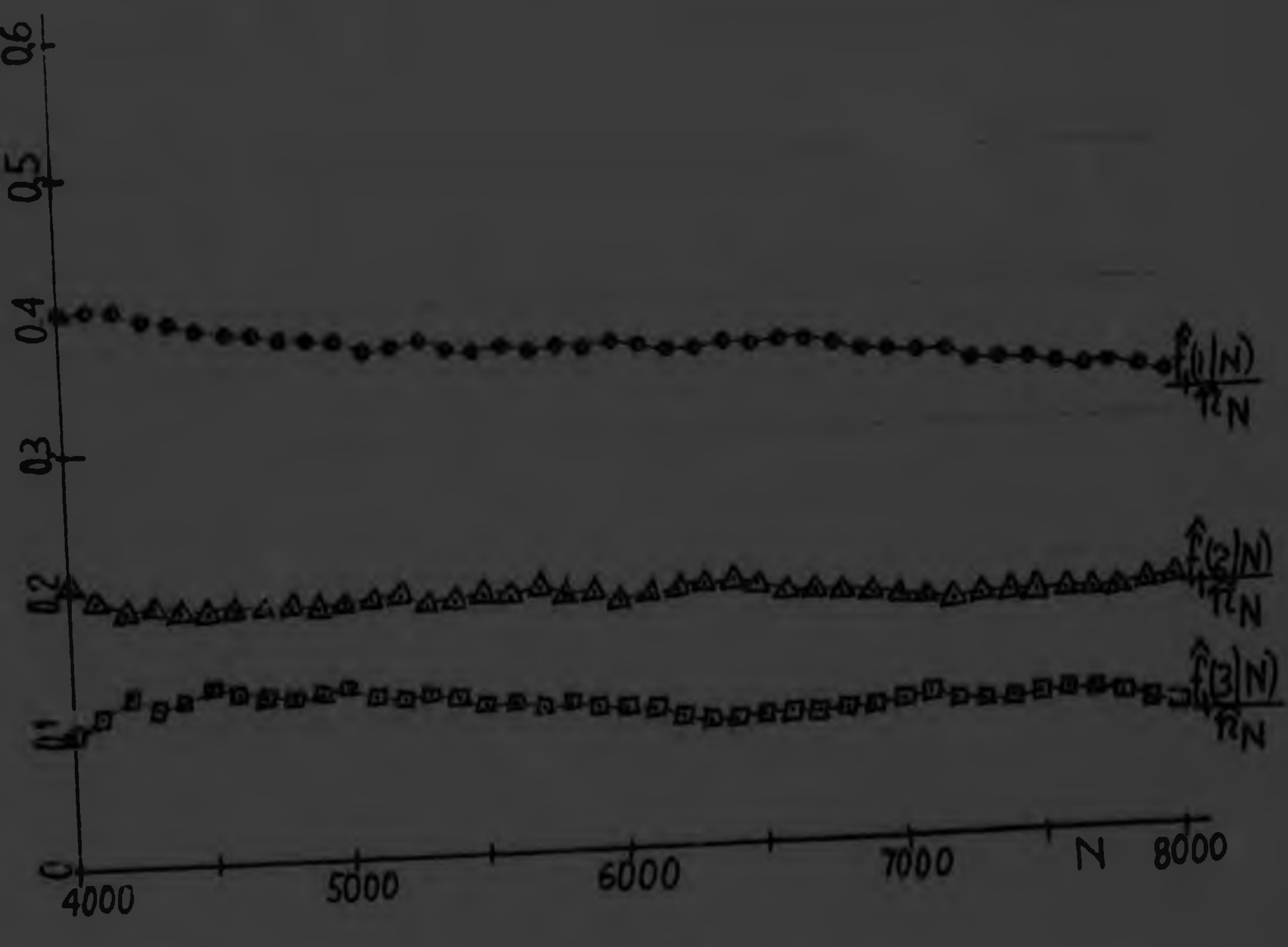


FIGURE 53 contd.

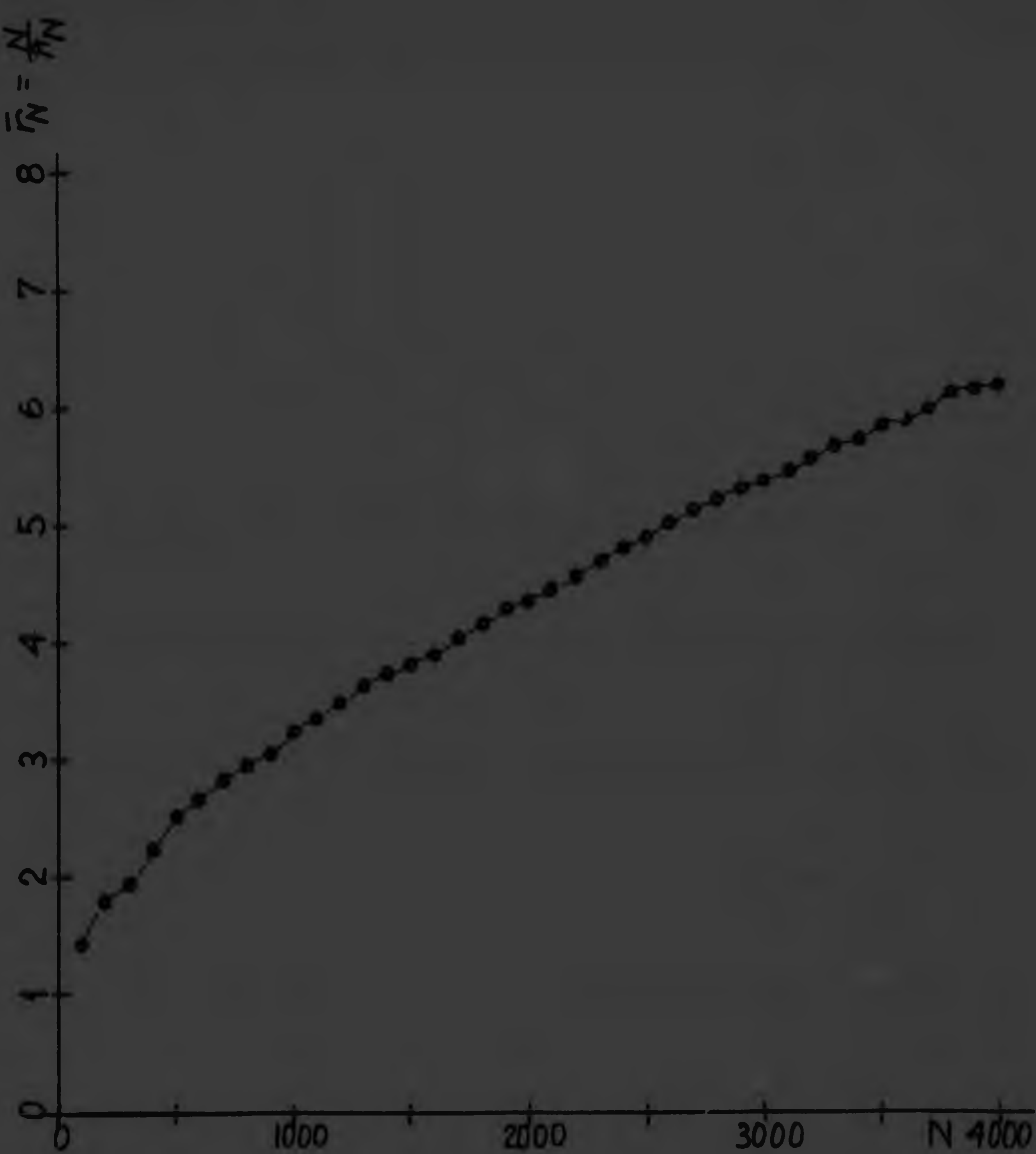


FIGURE 54 BRONWYN 1 RANDOM FORWARD

$\bar{r} = \frac{N}{n_N}$ vs N

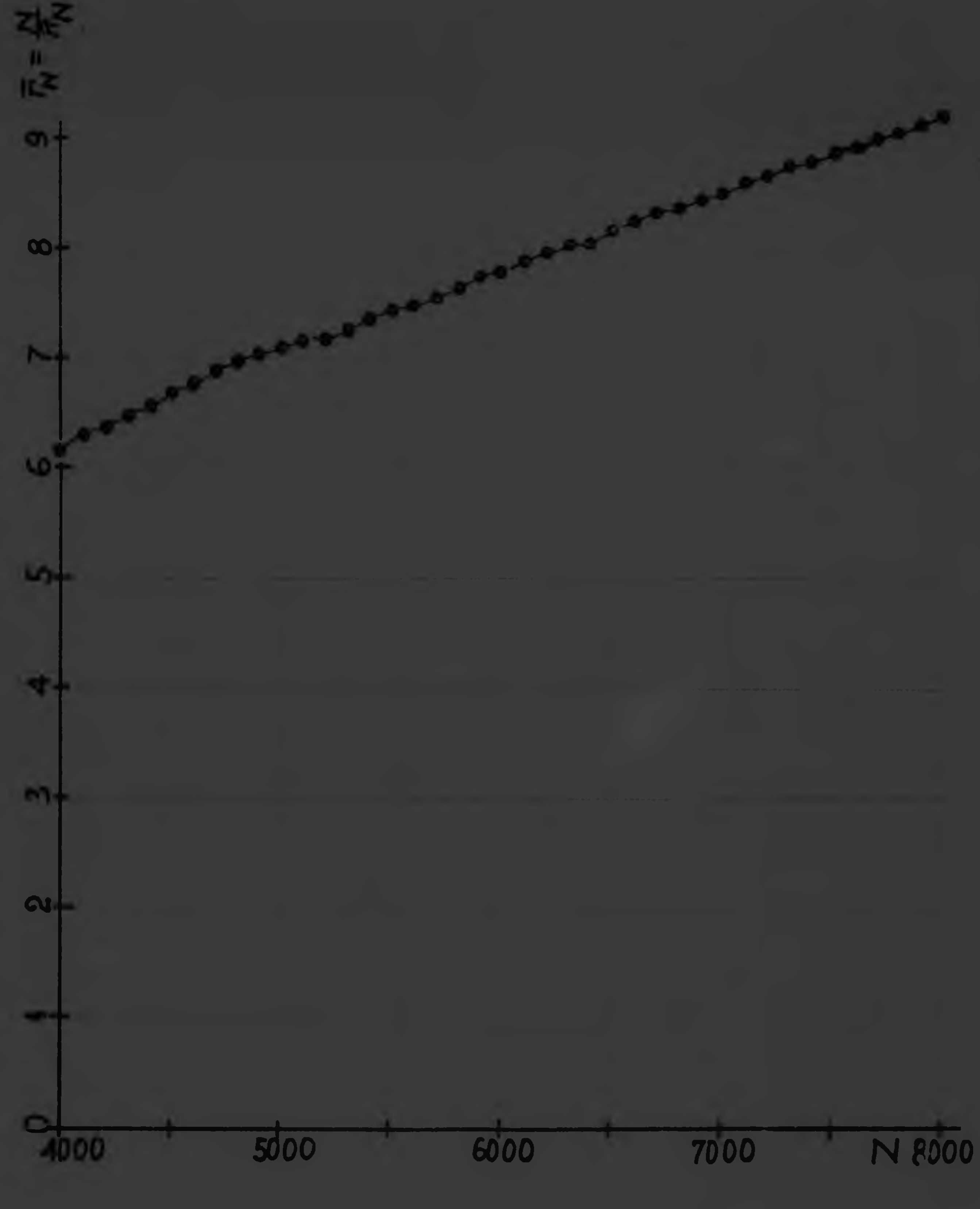


FIGURE 54 contd.

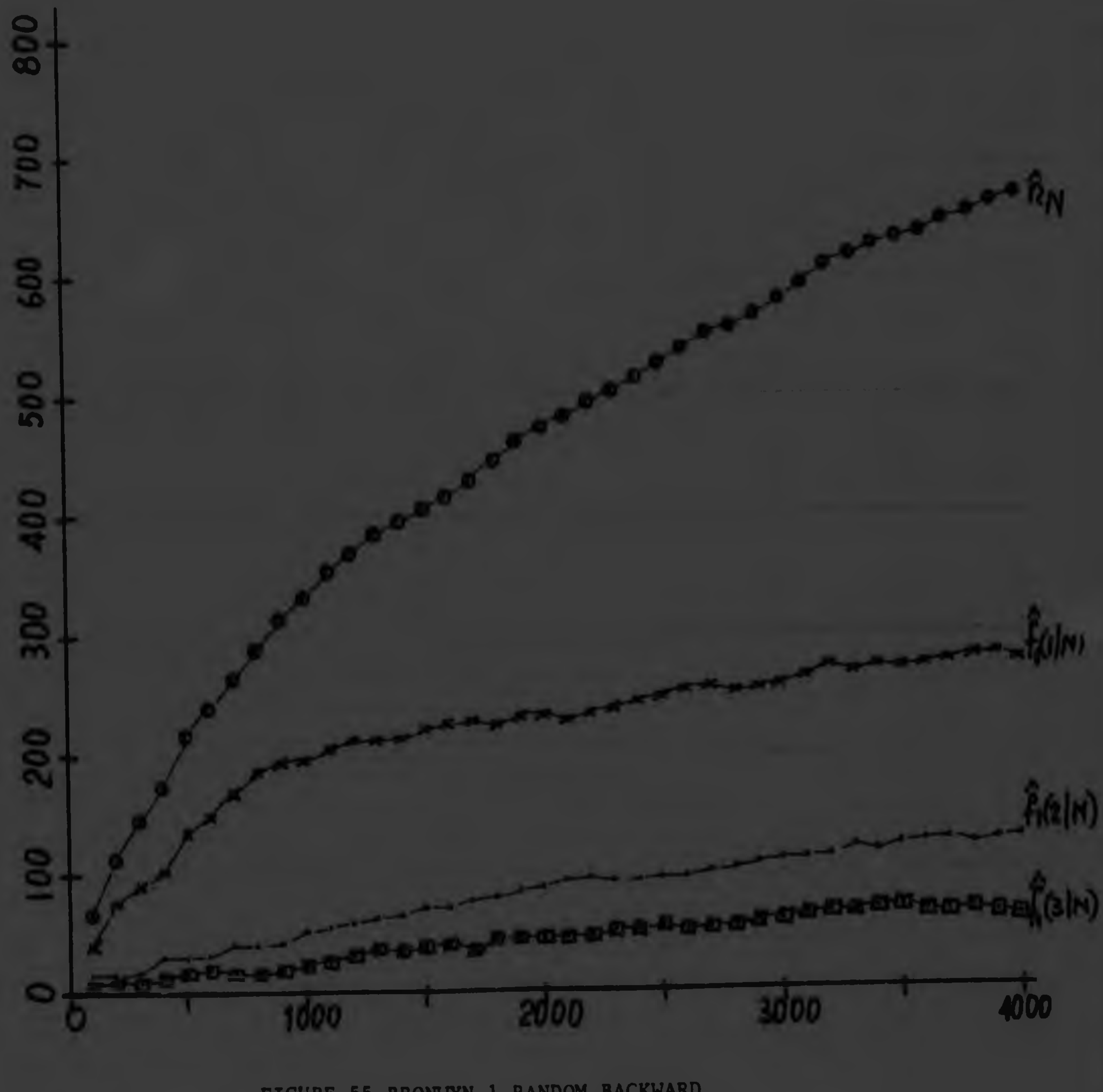


FIGURE 55 BRONWYN 1 RANDOM BACKWARD

$\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

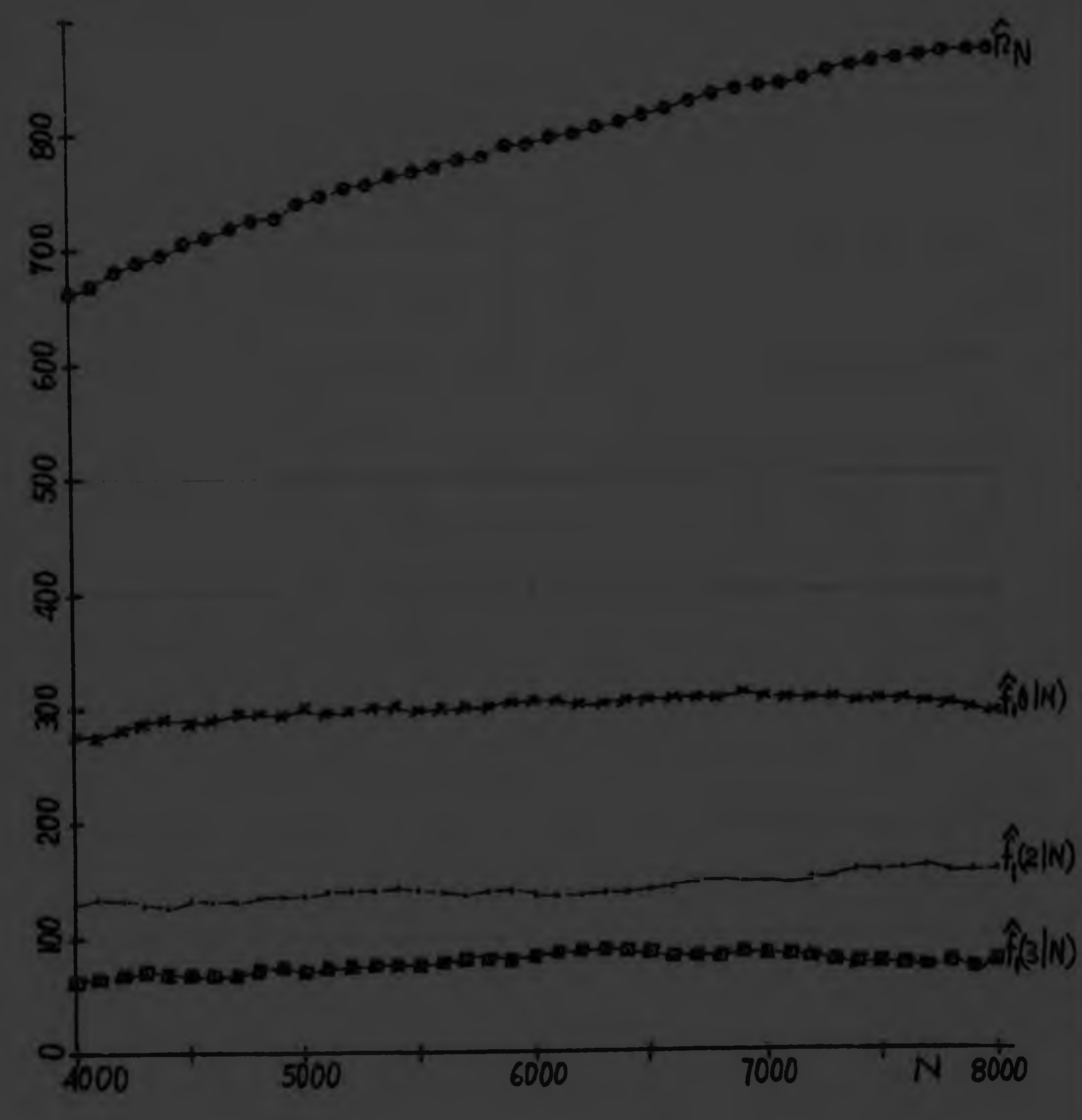


FIGURE 55 contd.

FIGURE 56 BRONWYN 1 RANDOM BACKWARD

$\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

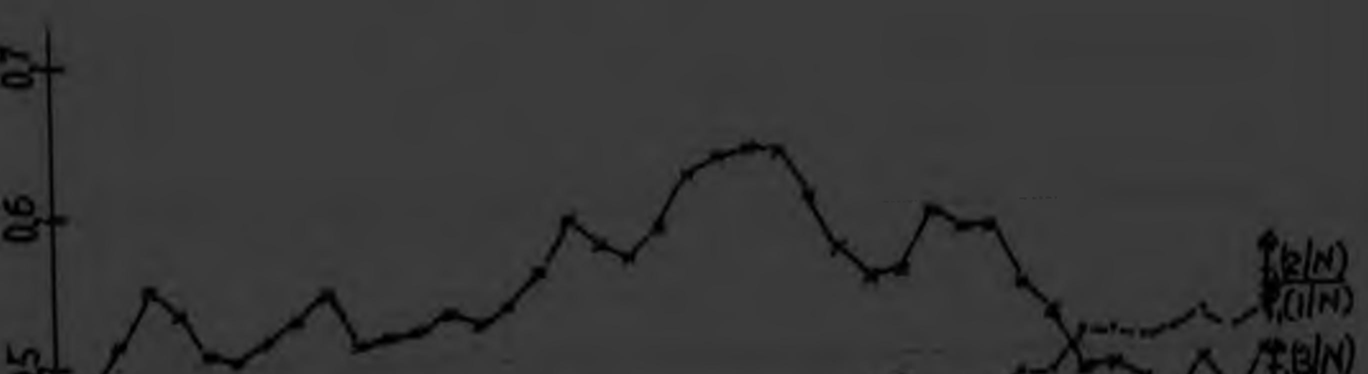


FIGURE 56 contd

FIGURE 57 BRONWYN 1 RANDOM BACKWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

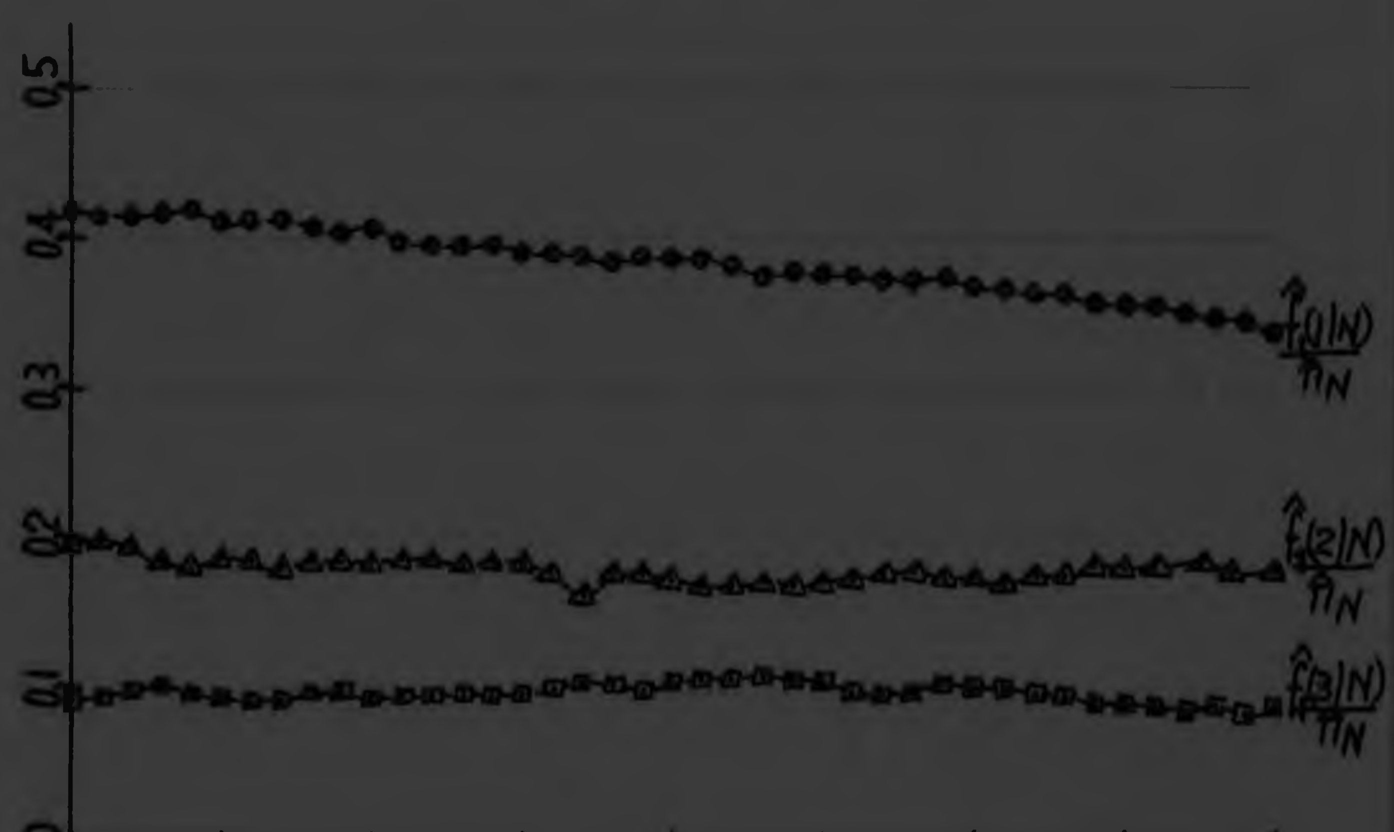


FIGURE 57 contd

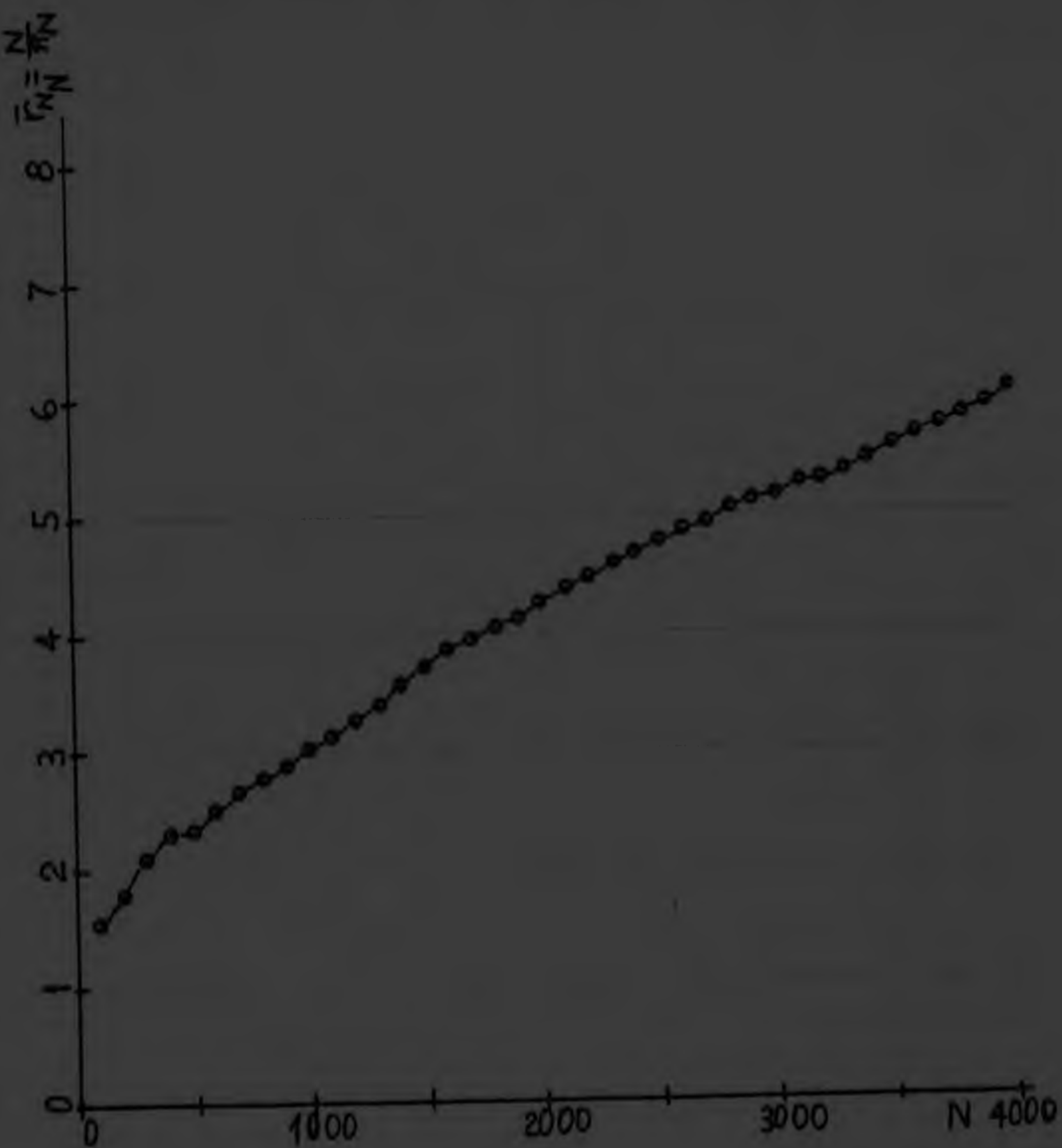


FIGURE 58 BRONWYN 1 RANDOM BACKWARD

$\bar{r}_N = \frac{N}{\bar{n}_N}$ vs N

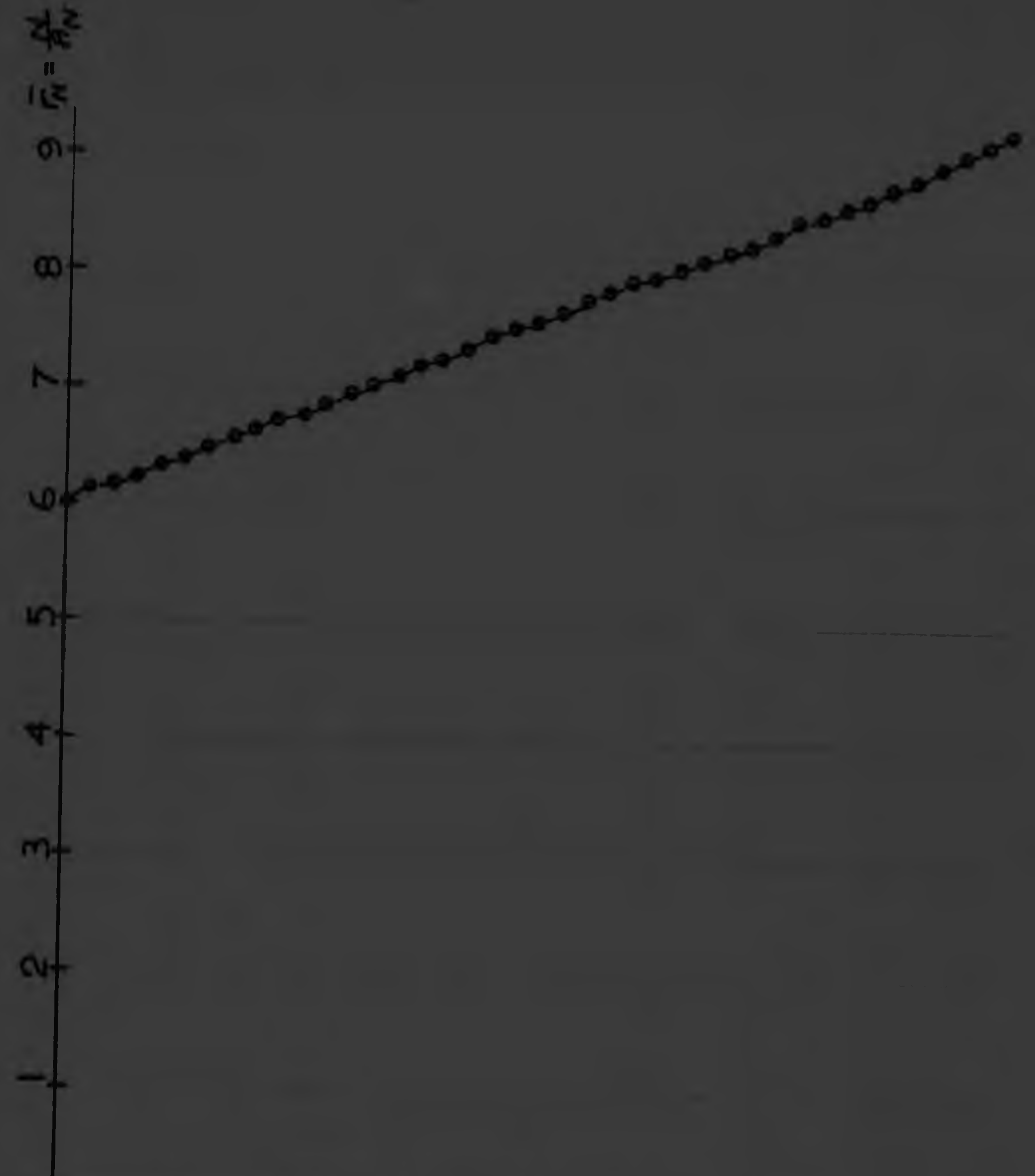


FIGURE 58 contd

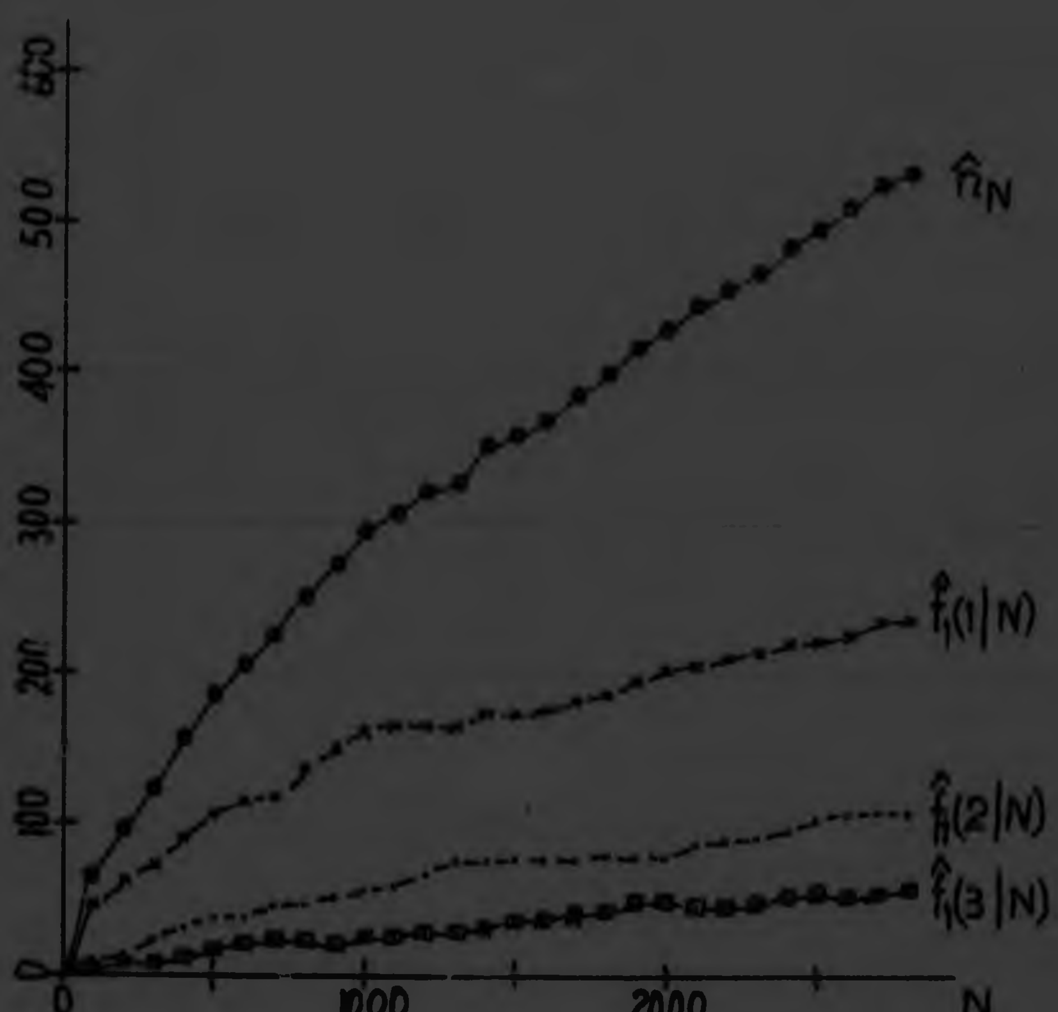


FIGURE 59 BRONWYN 2 FORWARD
$\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

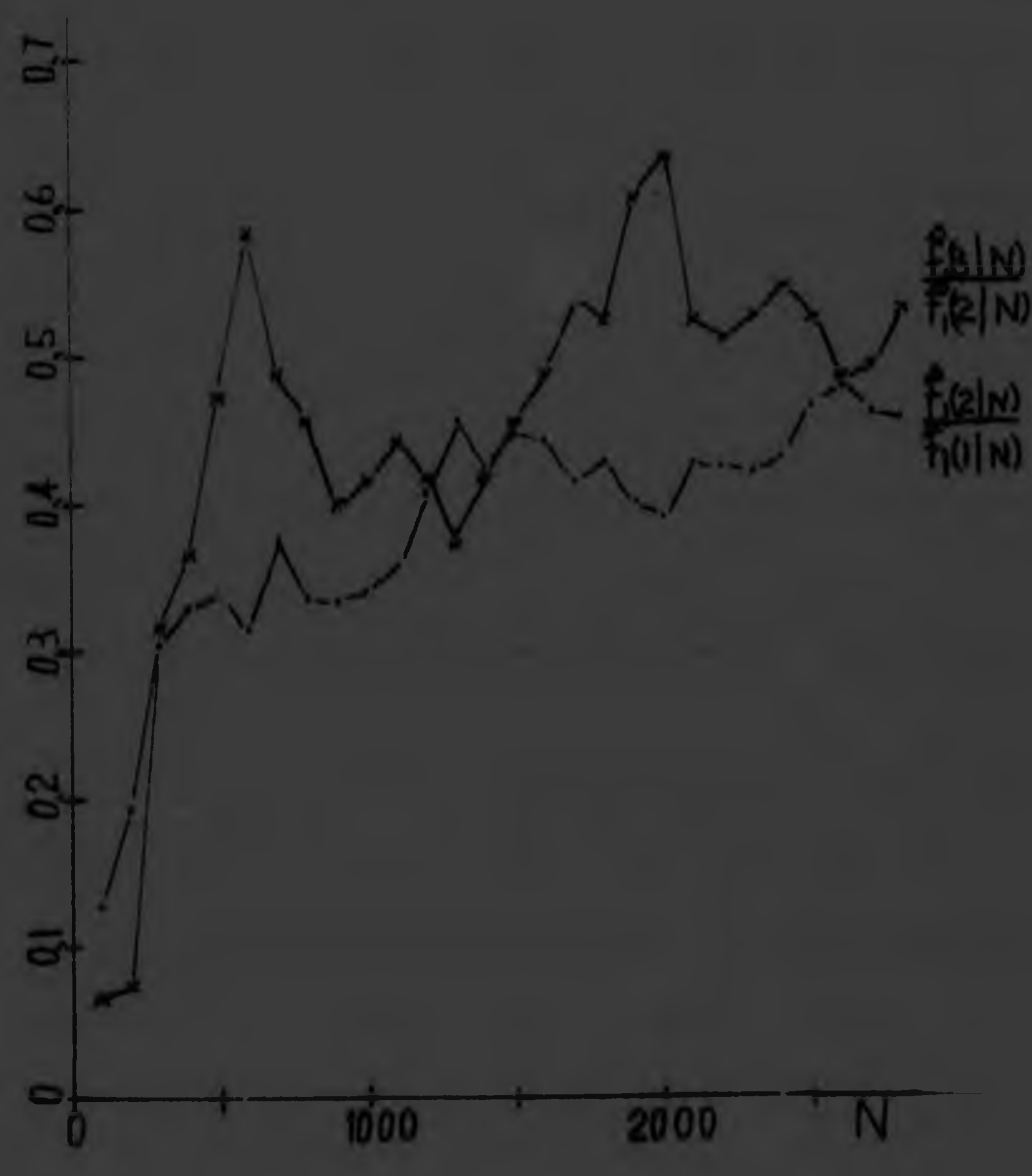


FIGURE 60 BRONWYN 2 FORWARD

$\frac{\hat{f}_1(2|N)}{\hat{n}_{..}}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_{..}}$ vs N

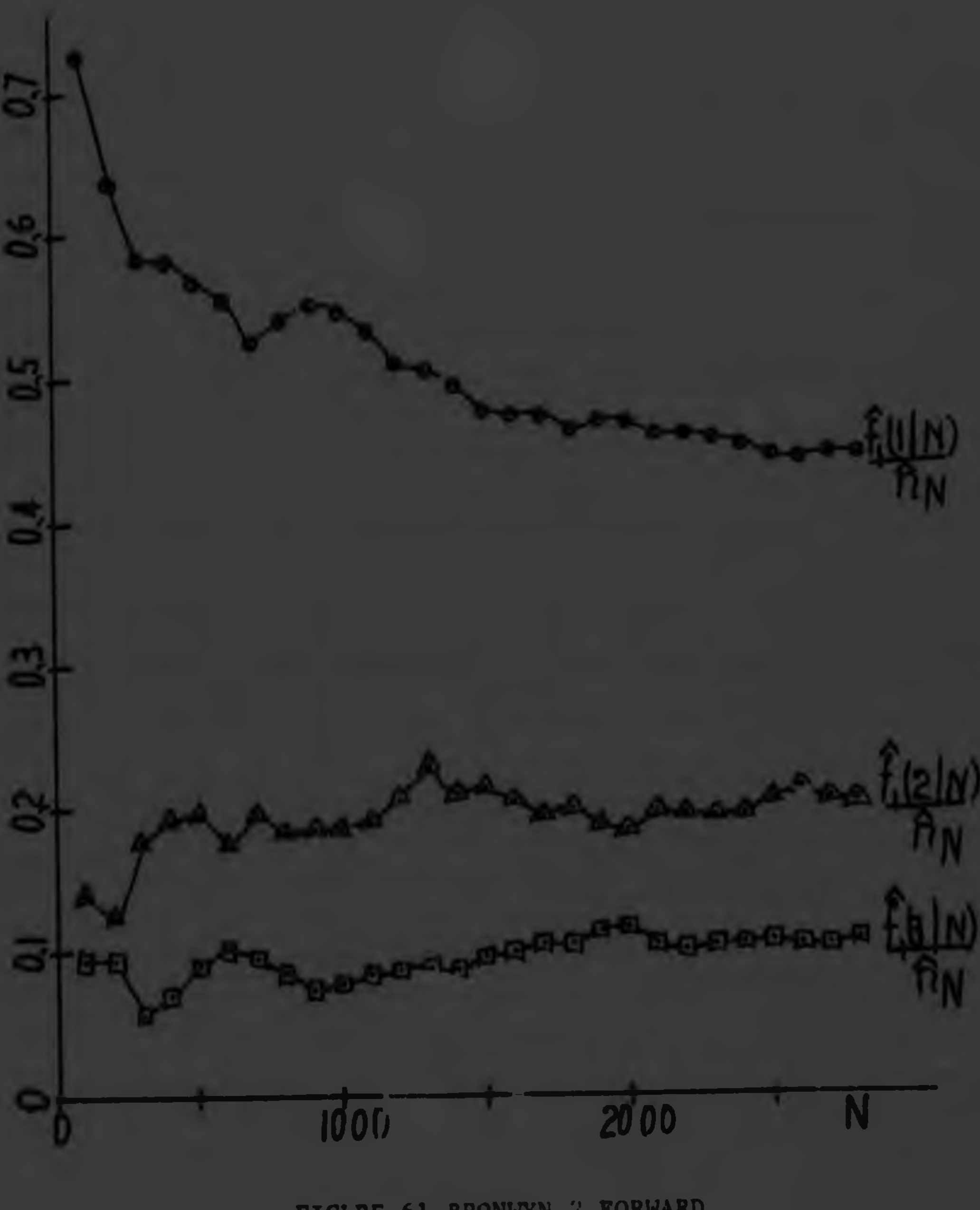


FIGURE 61 BRONWYN 2 FORWARD

$\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

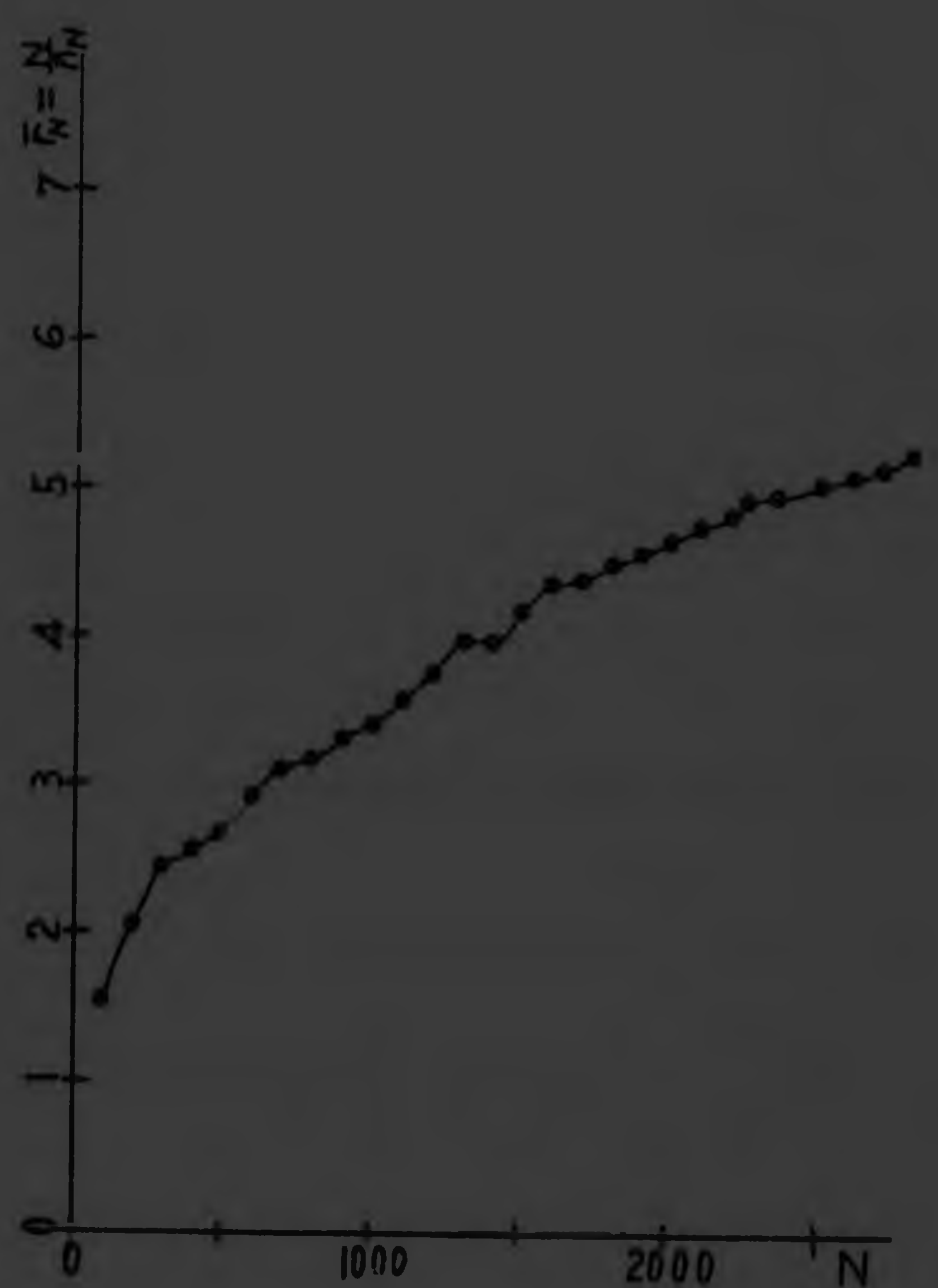


FIGURE 62 BRONWYN 2 FORWARD

$\bar{r}_N = \frac{N}{\bar{n}_N}$ vs N

5.11.1 <u>Column 4.</u> <u>Heading</u> $\hat{g}_N$ i.e. Root $\hat{g}_N$

The heading $\hat{g}_N$ does not mean the "square root of $\hat{g}_N$" but means the value of $\hat{g}_N$ which is the root of equation (73) namely:

$$(1+\hat{g}_N)\ln\hat{g}_N-a\hat{g}_N+b = 0$$

The equation has 3 roots, two less than 1, and one greater than 1, shown as points $\hat{g}_{N1}$, $\hat{g}_{N2}$ and $\hat{g}_{N3}$ in Figure 63 below.

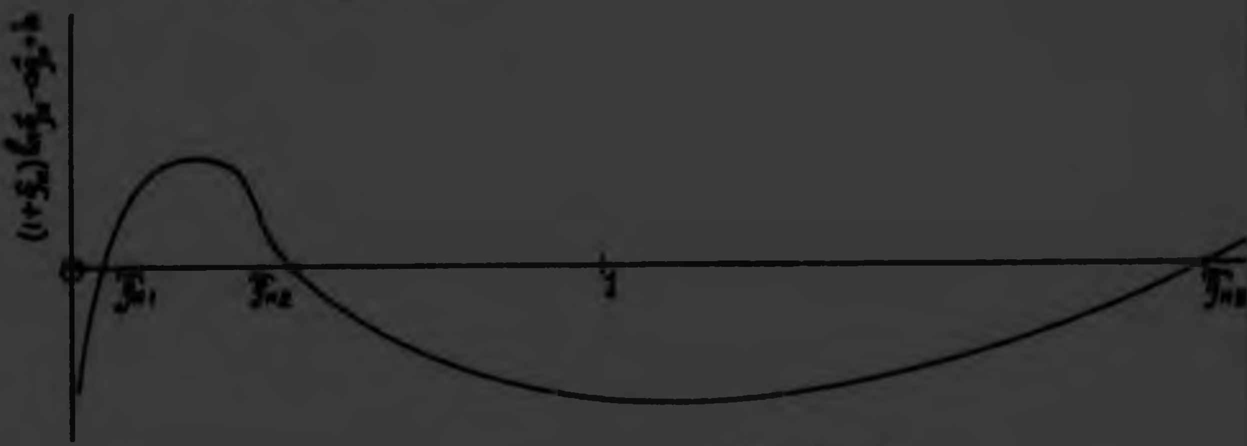


Figure 63 <u>Sketch of the three roots of</u> $(1+\hat{g}_N)\ln\hat{g}_N-a\hat{g}_N+b=0$

The root, $\hat{g}_{N1} = \dfrac{\hat{\Phi}(1|N)}{\bar{r}_N}$

when substituted into equation (73) gives

$$\left(1+\frac{\hat{\Phi}(1|N)}{\bar{r}_N}\right)\ln\frac{\hat{\Phi}(1|N)}{\bar{r}_N} - a\,\frac{\hat{\Phi}(1|N)}{\bar{r}_N} + b \qquad \text{.........} \ (90)$$

but since a and b are defined as

$$a = 2\bar{r}_N - \ln\left[\bar{r}_N/\hat{\Phi}(1|N)\right]$$

$$\text{and } b = 2\hat{\Phi}(1|N) + \ln\left[\bar{r}_N/\hat{\Phi}(1|N)\right],$$

substituting for a and b in (90) makes all the terms cancel, giving 0.

Incidentally substituting

$$\frac{\hat{f}_1(1|N)}{\hat{n}_N} \text{ for } \hat{\Phi}(1|N) \text{ and } \frac{N}{\hat{n}_N} \text{ for } \bar{r}_N \text{ makes } \frac{\hat{\Phi}(1|N)}{\bar{r}_N} = \frac{\hat{f}_1(1|N)}{N}$$

which depends only on Hapax Legomena and the size of the text, N.
But this result is trivial because substituting

$$\frac{\hat{\Phi}(1|N)}{\bar{r}_N} \text{ for } \hat{g}_N \text{ in equation (77), namely } \delta_N=(1-\hat{g}_N)^{-1}\ln\left[\hat{g}_N\,\bar{r}_N/\hat{\Phi}(1|N)\right]$$

$$\text{makes } \delta_N=(1-\hat{g}_N)\ln\left[\frac{\hat{\Phi}(1|N)}{\bar{r}_N}\cdot\frac{\bar{r}_N}{\hat{\Phi}(1|N)}\right] = 0$$

Since equation (44) requires $a>$, $\overline{g}_N = \frac{\hat{\phi}(1|N)}{\hat{r}_N}$ is unacceptable.

The third root $\hat{g}_N$ is greater than 1 and on substituting in equation (7(makes $\hat{a}_N$ negative which is unacceptable by equation (44).

Thus the only acceptable root is the second, which was solved by iteration using the programs of the Graduate School of Business Administration at the University of the Witwatersrand. The results are quoted to eight figures in tables 17 to 29 on pages 150 to 173.

The values of $\hat{g}_N$ are crucial to the entire calculations since all the other estimates with the exception of a and b are calculated from $\hat{g}_N$ via simple equations, namely equations (76), (77), (83), (84) and (89).

The possibility of confusion between roots $\frac{\hat{\phi}(1|N)}{\hat{r}_N}$ and $\hat{g}_N$ was investigated in the case of Kerry 1 forward. The results are shown in table 15 on page 146.

5.11.1.1. Table 15 Kerry 1 $\frac{\hat{\phi}(1|N)}{\hat{r}_N}$, $\hat{g}_N$ vs N (see page 146)

Column 1 shows the text size, N at 100 word intervals up to N = 2100.

Column 2 shows $\frac{\hat{\phi}(1|N)}{\hat{r}_N}$ as calculated from Table 2 on page 48
The values are given to four decimal places.

Column 3 shows $\hat{g}_N$ as in table 17 on page 150.
The values are given to four decimal places.

5.11.1.2. Kerry 1 forward. Graph of $\frac{\hat{\phi}(1|N)}{\hat{r}_N}$ and $\hat{g}_N$ vs N

The results shown in Table 15 are graphed on the same set of axes in Figure 63 on page 147.

It is clear from the graphs that $\hat{g}_N$ is always larger than $\frac{\hat{\phi}(1|N)}{\hat{r}_N}$ and that as the text size increases both (generally) decrease but the ratio

$$\hat{g}_N : \frac{\hat{\phi}(1|N)}{\hat{r}_N} \text{ increases.}$$

5.11.1.3. Kerry 1. The behaviour of $f(\hat{g}_N)$ between $\frac{\hat{\phi}(1|N)}{\hat{r}_N}$ and $\hat{g}_N$ where $f(\hat{g}_N) = (1+\hat{g}_N)\ \ln\hat{g}_N + a\hat{g}_N - b$

An investigation of the behaviour of the graph of $f(\hat{g}_N)$ between $\hat{\phi}(1|N)/\hat{r}_N$ and $\hat{g}_N$ was made, in the case of Kerry 1 forward, for values of N = 100, 200, 500, 1 000, 2 000

Ordinates were calculated at integral and half integral tenths between the two points of intersection on the $\bar{g}_N$ axis and at the turning points of the graph.

The abscissae of the turning points were calculated by iteration using a H.P.25c pocket calculator.

A maximum of twelve iterations was necessary to stabilize all ten figures on the display.

The equation which was solved is:

$$\frac{1}{\bar{g}_N} + 1 + \ln\bar{g}_N - a = 0 \qquad \text{(91)}$$

which is the derivative of the function

$$f(\bar{g}_N) = (1+\bar{g}_N)\ln\bar{g}_N - a\bar{g}_N + b$$

put equal to 0

The results appear in table 16 on page 148.

5.11.1.4. Table 15 Kerry 1 (forward) $f(\bar{g}_N)$ between $\frac{\bar{f}(1|N)}{\bar{r}_N}$ and $\hat{\bar{g}}_N$

The abscissae are quoted to four decimal places and the ordinates to four decimal places.

Under the heading N = 100 etc. appear two columns. The left hand column shows values of $\bar{g}_N$ from $\frac{\bar{f}(1|N)}{\bar{r}_N}$ to $\hat{\bar{g}}_N$, included is the abscissa of the maximum value of $f(\bar{g}_N)$.

The right hand column shows the calculated value of $f(\bar{g}_N)$ corresponding to the value of $\bar{g}_N$ shown in the left hand column.

The table shows that as N increases, both $\frac{\bar{f}(1|N)}{\bar{r}_N}$, and $\hat{\bar{g}}_N$ decrease, but the maximum value of $f(\bar{g}_N)$ increases very rapidly.

5.11.1.5. Kerry 1 (forward) Graphs of $f(\bar{g}_N)$ vs $\bar{g}_N$ for various N

The graphs of $f(\bar{g}_N)$ vs $\bar{g}_N$ for N = 100, 200, 500, 1 000, 2 000 are shown together on the same set of axes in Figure 64 on page 149

As N increases the graph of $f(\bar{g}_N)$ vs $\bar{g}_N$ moves to the left i.e. to lower abscissae and in general the ordinates increase.

This results in the slope at which the line crosses the axis at $\hat{\bar{g}}_N$ becoming ever steeper as N increases. The effect of this increase on the estimate of V is that small variations in $f_1(1|N)$ and n_N which cause the curve to rise or fall as a whole will not have as marked an influence on the point of intersection $\hat{\bar{g}}_N$ as for small N.

An example is the comparison of the graphs for N = 100 and N = 200. There is an obvious deviation from the general trend and the conclusion is that either the graph for N = 100 is too high or the graph for N = 200 is too low. The conclusion is again, that the largest possible text be analysed to minimise vacillations.

5.11.2 The Tables Type 2 (continued)

Column 5. Heading $\hat{\theta}_N$

The solution to equation (76), i.e. $\hat{\theta}_N = 1-\hat{g}_N^2$. The solutions are quoted to eight figures.

Column 6. Heading $\hat{a}_N$.

The solution to equation (77), namely:

$$\hat{a}_N = (1-\hat{g}_N)^{-2}\ \ln\left[\hat{g}_N\hat{\theta}_N/\phi(1|N)\right]$$

The values are quoted to eight figures.

Column 7. Heading $\hat{d}_N$.

The solution to the equation (84)

$$\hat{d}_N = \frac{\hat{\theta}_N}{1-\hat{\theta}_N}$$

The values of $\hat{d}_N$ are quoted to seven figures.

Column 8. Heading $\hat{c}_N$

The solution to the equation (83)

$$\hat{c}_N = \hat{a}_N\sqrt{1-\hat{\theta}_N}$$

The values are quoted to eight figures.

Column 9. Heading $\hat{V}_N$

The estimate of the total vocabulary, i.e. $\hat{V}_N = \dfrac{2}{\hat{c}_N\hat{d}_N}$

The results are quoted to six figures.

5.12 Sandré's Calculation Tables

In the case of Sandré's data the results of the calculations were not shown in as great detail as for the younger children, Kerry and Bronwyn.

Firstly the results are quoted at two hundred word intervals. Secondly $\hat{r}_N$, $\phi(1|N)$ and $\hat{g}_N$ are not shown, and thirdly the number of figures shown in the columns differs from the younger children.

It must be borne in mind that the number of figures shown in the tables does not affect the final result because the computer performs all calculations with all the figures at its disposal and merely rounds off as instructed before printing.

Sandré's data and calculations appear separately in Chapter VI, see 6.4.

TABLE 15 KERRY 1 (FORWARD)

Values of $\frac{\hat{\Phi}(1|N)}{T_N}$ and $\widehat{\hat{\varepsilon}}_N$ with increasing N

N	$\frac{\hat{\Phi}(1\|N)}{T_N}$	$\widehat{\hat{\varepsilon}}_N$
100	0,2400	0,37498
200	0,1750	0,33755
300	0,1333	0,27410
400	0,1100	0,24300
500	0,1120	0,20639
600	0,1033	0,19171
700	0,0914	0,17963
800	0,0725	0,18192
900	0,0669	0,16497
1000	0,0650	0,15633
1100	0,0564	0,16119
1200	0,0500	0,16239
1300	0,0477	0,16680
1400	0,0429	0,16054
1500	0,0447	0,15308
1600	0,0431	0,14528
1700	0,0424	0,13450
1800	0,0367	0,14171
1900	0,0337	0,14029
2000	0,0330	0,13590
2100	0,0338	0,12756

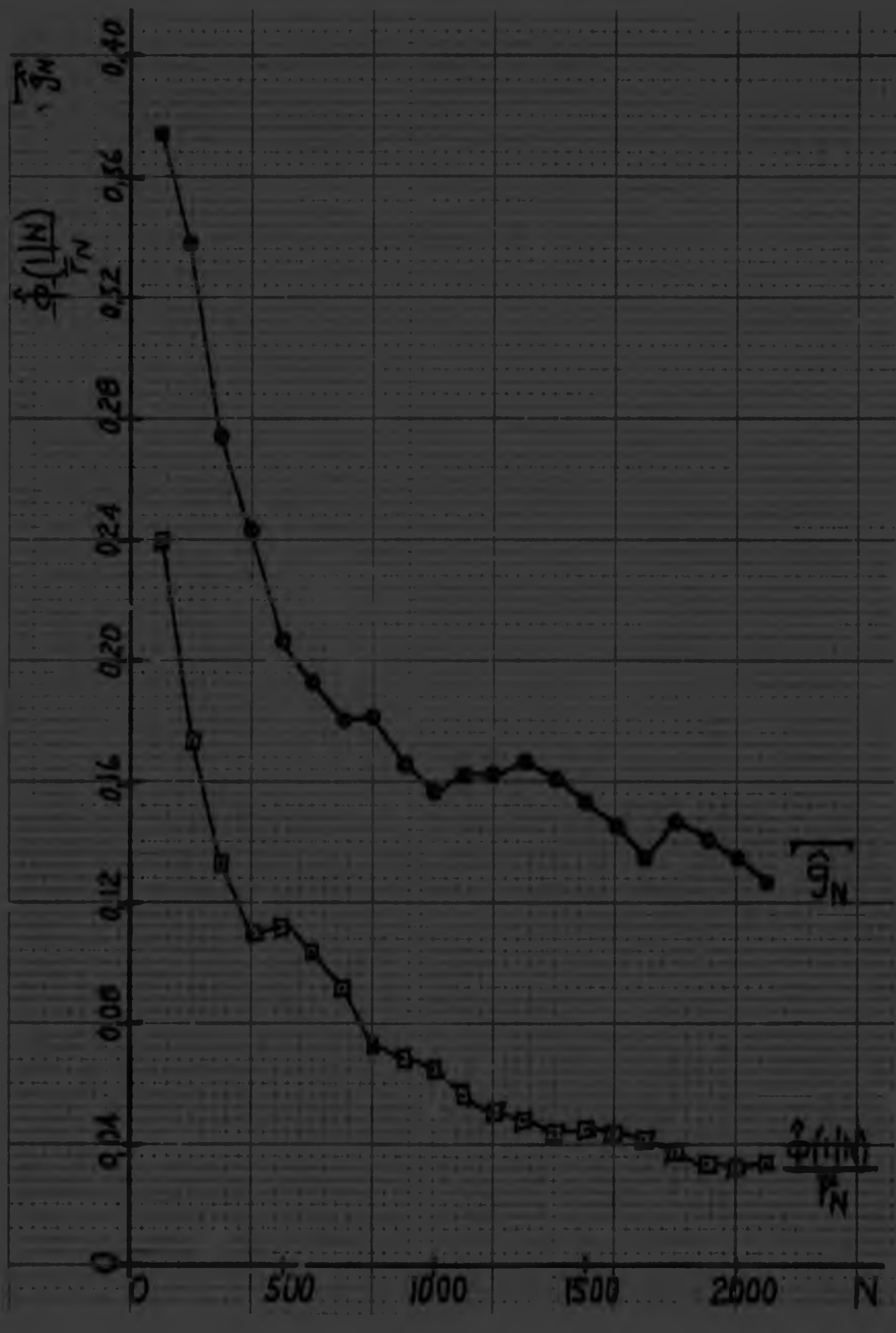


FIGURE 63 KERRY 1 FORWARD

$\frac{\phi(1|N)}{\bar{r}}$ and $\widehat{g}_N$ vs N

TABLE 16 [illegible] 1 (FORWARD): $f(\hat{\varepsilon}_N)$ vs $\hat{\varepsilon}_N$
FOR INCREASING TEXT LENGTH

N = 100		N = 200		N = 500		N = 1 000		N = 2 000	
$\hat{\varepsilon}_N$	$f(\hat{\varepsilon}_N)$	$\hat{\varepsilon}_N$	$f(\hat{\varepsilon}_N)$	$\hat{\varepsilon}_N$	$f(\hat{\varepsilon}_N)$	$\hat{\varepsilon}_N$	$f(\hat{\varepsilon}_N)$	$\hat{\varepsilon}_N$	$f(\hat{\varepsilon}_N)$
0,2400	0,0000	0,1750	0,0000	0,1120	0,0000	0,0650	0,0000	0,0530	0,0000
0,3000	0,0174	0,2000	0,0252	0,1500	0,0386	0,1000	0,0822	0,0725	0,2277
0,3014	0,0174	0,2500	0,0405	0,1535	0,0389	0,1036	0,0845	0,1000	0,1735
0,3500	0,0035	0,3000	0,0251	0,2224	0,0000	0,1503	0,0000	0,1359	0,0000
0,3750	0,0000	0,3376	0,0000						

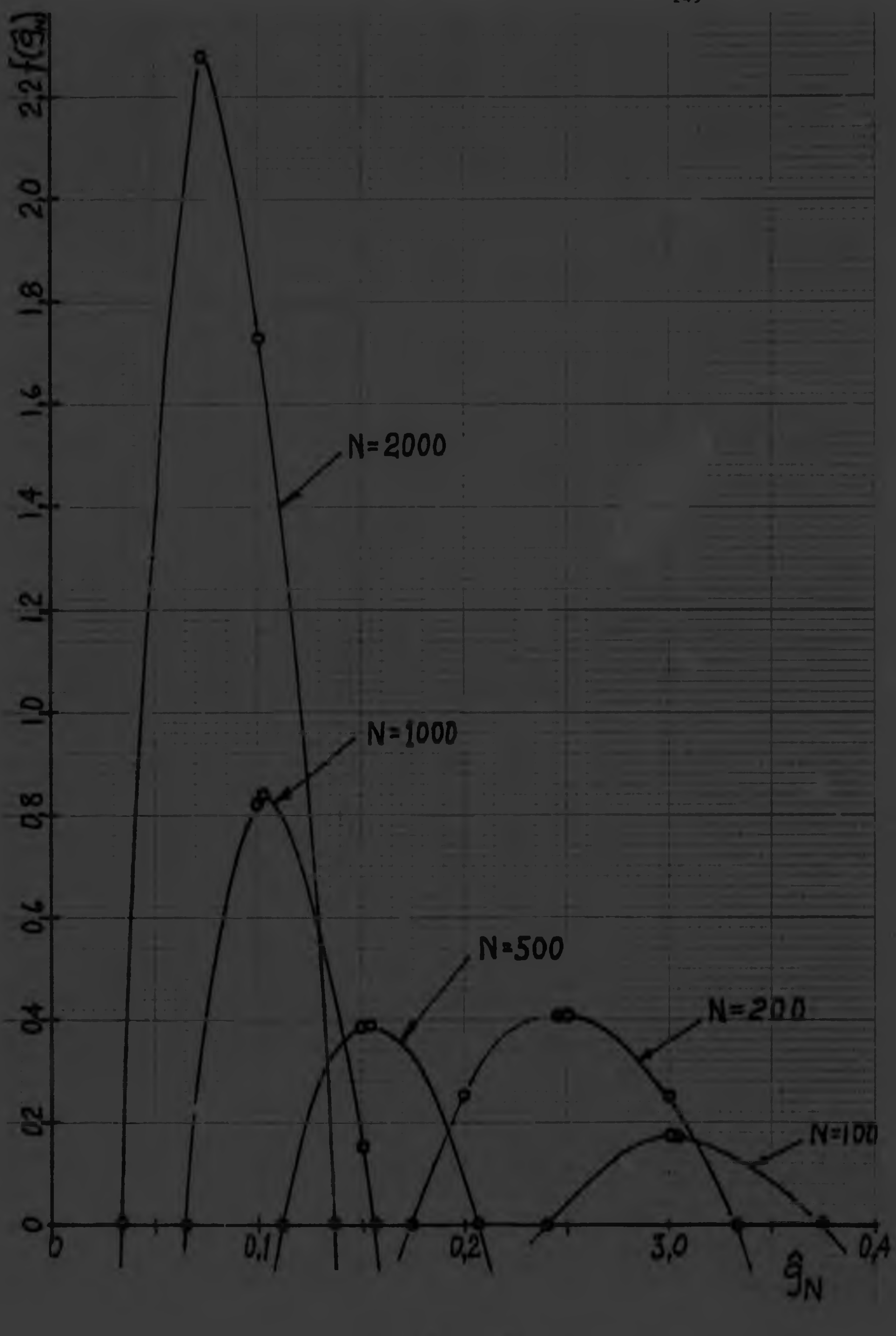


FIGURE 64 KERRY 1 FORWARD $f(\hat{g}_N)$ vs $\hat{g}_N$ FOR INCREASING N

TABLE 17 KERRY (1) FORWARD (CALCULATIONS

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\bar{g}_N$	$\hat{\Omega}_N$	π_N	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	2.2727	.5455	.37497771	.8[illegible]939175	.71378529	.00111958	.30765352	22.258
2	2.7027	.4730	.33755332	.88605779	.99159485	.03888190	.33471602	153.676
3	3.2609	.4348	.27419084	.92481941	.99329311	.04100436	.27235180	179.089
4	3.7037	.4074	.24300158	.94095027	1.04703522	.03983720	.25443107	197.320
5	3.9063	.4375	.20639265	.95740211	.77027243	.04495070	.15897846	279.869
6	4.1667	.4306	.19170737	.96324831	.76446784	.04368276	.14655399	312.408
7	4.5161	.4129	.17963195	.96773237	.82323128	.04284404	.14787859	315.671
8	4.9689	.3602	.18192327	.96690392	1.12473965	.03651884	.20461619	267.654
9	5.2941	.3647	.16497475	.97278339	1.04584789	.03971365	.17253828	291.880
10	5.5556	.3611	.15633231	.97556075	1.04026031	.03991696	.16262609	308.093
11	5.8201	.3280	.16118908	.97401810	1.25282192	.03408030	.20194113	290.604
12	6.0914	.3046	.16238809	.97363013	1.40621948	.03076840	.22835308	284.655
13	6.1321	.2925	.16680276	.97217685	1.50250530	.02687793	.25062191	296.903
14	6.5116	.2791	.16054392	.97422570	1.57313824	.02699881	.25255746	293.309
15	6.5502	.2926	.15307570	.97656786	1.45422649	.02778430	.22260654	323.365
16	6.8085	.2936	.14527994	.97889376	1.42106724	.02898709	.20645237	334.199
17	7.1130	.3013	.13449687	.98191065	1.33489609	.03193009	.17953897	348.876
18	7.3469	.2694	.14170635	.97991931	1.57502747	.02711062	.22319126	330.[illegible]32
19	7.6305	.2570	.14028788	.98031932	1.65959263	.02621647	.23282057	327.668
20	7.8125	.2578	.13589746	.98153192	1.63804531	.02657375	.22260588	338.096
21	7.9848	.2700	.12755758	.9837[illegible]912	1.52180004	.02879024	.19411671	357.867

TABLE 18 KERRY (1) BACKWARD (CA[illegible]ATIONS)

N	$\bar{r}_N$	$\hat{l}(1\|N)$	$\hat{\rho}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.9608	.6471	.36877084	.86400813	.17588848	.06353378	.06486249	485.324
2	2.5641	.5128	.32895356	.89178956	.74158400	.04120629	.24394667	198.963
3	3.1250	.4735	.26268119	.93099862	.74624819	.04497489	.19602531	226.855
4	3.5398	.4159	.25426728	.93534815	1.03523731	.03616866	.26322693	210.072
5	3.8760	.4264	.21517450	.95369995	.85480714	.04119650	.18393260	263.943
6	4.1379	.4000	.21046158	.95569754	.98556525	.03595353	.20744318	268.157
7	4.4586	.3758	.20244330	.95901674	1.09865093	.03342887	.22241437	268.996
8	4.7059	.3765	.18734509	.96490186	1.04700089	.03416443	.19615024	296.710
9	4.9451	.3626	.1816[illegible]734	.96691340	1.11054325	.03247081	.20200461	304.913
10	5.2356	.3665	.16661334	.97224003	1.04052258	.03502312	.17336476	329.393
11	5.3398	.3689	.16125560	.97399664	1.01062107	.03405144	.16296822	360.406
12	5.5814	.3581	.15665823	.97545820	1.05851936	.03312237	.16582566	364.130
13	5.8559	.3378	.15557832	.97579539	1.17493916	.03101111	.18279493	352.816
14	6.2674	.3260	.14727056	.97831142	1.22047424	.03221945	.17973965	345.357
15	6.4935	.3074	.14831465	.97800279	1.34085560	.02964020	.19886833	339.299
16	6.8670	.2876	.14624119	.97861356	1.46459579	.02859912	.21418393	326.506
17	6.9388	.2840	.14585471	.97872645	1.48776054	.02706279	.21699661	340.568
18	7.0864	.2992	.13593119	.98152274	1.35317993	.02951143	.18393916	368.439
19	7.3643	.2907	.13274986	.98237753	1.39846325	.02933987	.18564540	367.187
20	7.6046	.2814	.13103306	.98283035	1.45508852	.02862116	.19066459	366.499
21	7.8947	.2782	.12[illegible]812	.98405135	1.46091175	.02938154	.18449551	368.952

TABLE 19 KERRY (I) RANDOM FORWARD (CALCULATIONS)

N	$\bar{\hat{v}}$	$\hat{\phi}(1\|N)$	$\bar{\hat{g}}_N$	$\hat{\theta}_N$	$\hat{a}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	2.0408	.5510	.4662722[illegible]	.78259021	1.02369690	.03599611	.47732133	116.403
2	2.4390	.5854	.29087883	.91538954	.27103734	.05409436	.07883900	468.961
3	2.6316	.5175	.30941695	.90426117	.65636092	.03148362	.20308912	312.794
4	3.0303	.4924	.26173651	.93149406	.64571458	.03399318	.16[illegible]00700	348.124
5	3.3333	.4467	.25573134	.93460149	.86823773	.02858174	.22203553	315.151
6	3.6364	.4182	.24228412	.94129843	.98340935	.02672554	.23826438	314.083
7	3.9106	.4134	.22018099	.95152038	.94089085	.02803880	.20716614	344.311
8	4.1885	.3717	.22412950	.94976598	1.19411564	.02363354	.26763642	316.196
9	4.4776	.3731	.20276803	.95888513	1.11548710	.02591345	.22618502	341.225
10	4.8077	.3462	.19811451	.96075064	1.26208305	.02447813	.25003695	326.774
11	5.1643	.3474	.17894357	.96797925	1.19158936	.02748160	.21322703	341.307
12	5.4794	.3379	.16975051	.97118479	1.21960163	.02808659	.20702785	343.955
13	5.7269	.3216	.16767418	.97188538	1.31425381	.02659130	.22036630	341.307
14	6.0345	.2974	.16790271	.97180873	1.[illegible]314262	.02462285	.24734432	328.390
15	6.2762	.3012	.15774608	.97511619	1.[illegible]34752	.02612453	.22287107	343.501
16	6.6390	.2904	.15142757	.97706974	1.46340847	.02663156	.22160012	338.894
17	6.9959	.2798	.14614528	.97864157	1.51764584	.02695291	.22179669	334.556
18	7.3469	.2694	.14170635	.97991931	1.57502747	.02711062	.22319126	330.532
19	7.5397	.2817	.13231748	.98249209	1.45737934	.02953529	.19283664	351.155
20	7.7519	.2752	.13029313	.98302370	1.49501419	.02895284	.19478995	354.627
21	7.9848	.2700	.12755758	.98372912	1.52180004	.02879024	.19411671	357.867

TABLE 20 KERRY (1) RANDOM BACKWARD (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\hat{\bar{g}}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.8868	.6792	.36183095	.86907840	.00805940	.06638157	.00291614	10331.750
2	2.2222	.5730	.35961336	.87067825	.51943749	.03366325	.18679661	318.957
3	2.8037	.5327	.26563346	.92943889	.45630920	.04390707	.12121093	375.798
4	3.2258	.5000	.23214632	.94610810	.52606422	.04388918	.12212384	373.140
5	3.6765	.4632	.21043456	.95571733	.64969444	.04316451	.13671803	338.905
6	4.0000	.4467	.19461977	.96212316	.68967921	.04233559	.13422513	351.958
7	4.2169	.4578	.17555571	.96918023	.58296686	.04492387	.10234308	435.005
8	4.5198	.4463	.16445339	.97295511	.61050177	.04496945	.10039896	442.979
9	4.8649	.4108	.16327739	.97334051	.78806216	.04056679	.12867266	383.154
10	5.1282	.3949	.15870422	.97481298	.85958362	.03870300	.13641948	378.799
11	5.5838	.3553	.15776968	.97510874	1.07814026	.03561342	.17009777	330.155
12	5.8824	.3431	.15242141	.97676772	1.13333607	.03503631	.17274457	330.451
13	6.1905	.3429	.14281523	.97960383	1.10492802	.03694525	.15780038	343.055
14	6.4516	.3318	.13967848	.98048997	1.16135406	.03589692	.16221595	343.462
15	6.6372	.3230	.13806564	.98093790	1.20979786	.03430676	.16703135	349.022
16	6.8376	.3376	.12782425	.98366100	1.09064102	.03762704	.13941014	381.273
17	6.9959	.3333	.12569106	.98420179	1.10957050	.03664608	.13946289	391.331
18	7.2874	.3239	.12259245	.98497111	1.15633392	.03641031	.14175767	387.489
19	7.5697	.2948	.12681490	.98391801	1.35204220	.03220074	.17145884	362.247
20	7.7519	.2868	.12599289	.98412579	1.40202141	.03099765	.17664462	365.259
21	7.9245	.2868	.12263638	.98496038	1.39098167	.03118622	.17058462	375.948

TABLE 21 KERRY (2) FORWARD (CALCULATIONS)

N	$\bar{F}_M$	$\bar{\phi}(1\|N)$	$\tilde{g}_N$	$\bar{b}_N$	$\bar{a}_N$	$\hat{d}_N$	$\bar{c}_N$	$\hat{v}_N$
1	1.8182	.6545	.42643446	.81815368	.29540318	.04499149	.12597007	352.884
2	2.2472	.5843	.33969802	.88460529	[illegible]485454	.03832955	.13752824	379.406
3	2.6087	.5478	.28769714	.91723037	.44201618	.03693909	.12716675	425.765
4	2.9851	.5224	.24676192	.93910861	.45619476	.03855671	.11257142	460.789
5	3.2468	.5065	.22586042	.94898713	.47803003	.03720581	.10796797	497.880
6	3.7037	.4506	.21535277	.95362324	.72775280	.03427088	.15672344	372.367
7	3.7838	.4541	.20673144	.95726216	.68555743	.03199783	.14172614	441.021
8	4.1451	.4404	.18815529	.96459764	.70396531	.03404836	.13245463	443.342
9	4.4118	.3971	.19366241	.96249491	.95023590	.02851451	.18402481	381.143
10	4.6083	.3871	.18730664	.96491623	.98674721	.02750321	.18482423	393.448
11	4.74[illegible]	.3836	.18182796	.96693861	.98975855	.02658798	.17996562	417.980
12	4.8780	.3943	.16994578	.97111845	.89524543	.02802015	.15214312	469.145
13	5.1383	.3913	.15980452	.97446257	.88217419	.02935248	.14097524	483.329
14	5.3435	.3969	.14957047	.97762871	.82307041	.03121439	.12310690	520.467
15	5.5556	.3704	.15246665	.97675395	.97597194	.02801205	.14880306	479.815
16	5.6940	.3665	.14921725	.97773427	.98828638	.02744504	.14746910	494.157
17	5.9028	.3542	.14722151	.97832584	1.05244541	.02655171	.15494251	486.146
18	6.0201	.3545	.14343131	.97942752	1.03932190	.02644924	.14907104	507.252
19	6.1889	.3420	.14321238	.97949022	1.11144733	.02513539	.15917289	499.891
20	6.3492	.3365	.14071989	.98019797	1.13641739	.02474994	.15991622	505.316
21	6.4815	.3333	.13829380	.98087484	1.14805794	.02442245	.15876919	515.79[illegible]

TABLE 22 KERRY (2) BACKWARD (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\bar{S}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	2.1277	.5957	.36571425	.86625314	.42119241	.06476808	.15403599	200.469
2	2.4096	.6024	.28325176	.91976845	.17422324	.05731967	.04934903	707.046
3	2.[illegible]862	.5862	.26229078	.93120360	.19788498	.04511882	.05190338	854.037
4	2.8169	.5423	.25667721	.93411684	.38695967	.03544597	.09932369	568.081
5	3.0303	.5335	.23397285	.94525671	.37176394	.03453416	.08698261	665.807
6	3.2787	.5191	.21522659	.95367754	.39124876	.03431300	.08420706	692.186
7	3.4314	.5049	.20867628	.95645422	.44152778	.03137763	.09213632	691.797
8	3.7037	.4861	.19598335	.96159053	.49867970	.03129406	.09773284	653.924
9	3.9474	.4649	.18905205	.96425933	.58358413	.02997704	.11032772	604.723
10	4.2373	.4454	.18007809	.96757191	.65657926	.02983748	.11823541	566.918
11	4.4177	.4297	.17711627	.96862984	.72833812	.02807039	.12900048	552.319
12	4.6154	.4154	.17320406	.97000039	.79176301	.02694481	.13713646	541.255
13	4.8689	.3783	.17790961	.96834821	1.00774288	.02353368	.17928702	474.014
14	5.1095	.3686	.17115486	.97070605	1.04236317	.02366911	.17840534	473.631
15	5.3191	.3582	.16689003	.97214776	1.08934784	.02326918	.18180108	472.773
16	5.6140	.3404	.16323096	.97335571	1.18348885	.02283219	.19318175	453.437
17	5.7627	.3424	.15685499	.97539651	1.15134144	.02332039	.18059355	474.889
18	6.0000	.3267	.15548295	.97582507	1.24242783	.02242510	.19317621	461.681
19	6.2500	.3257	.14791977	.97811979	1.22438526	.02352811	.18111056	469.353
20	6.3492	.3270	.14444768	.97913492	1.20540142	.02346349	.17411721	489.549
21	6.4615	.3323	.13921636	.98061883	1.15691280	.02409358	.16106105	515.393

TABLE 23 KERRY (2) RANDOM FORWARD (CALCULATIONS)

N	$\bar{\bar{r}}_N$	$\hat{\phi}(1 \mid N)$	$\hat{\sigma}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.7857	.6429	.46458691	.78415900	.47621071	.03633041	.22124124	248.825
2	2.3256	.5116	.39960092	.84031916	.99417347	.02631247	.39727253	191.329
3	2.5000	.5417	.31590033	.90020704	.55109608	.03006916	.17409134	382.060
4	2.6846	.5302	.28790975	.91710800	.52932090	.02765973	.15239656	474.468
5	3.0303	.5273	.23776984	.94346553	.40958035	.03337666	.09738576	615.307
6	3.1746	.5238	.22326791	.95[illegible]15144	.38937960	.03176793	.08693594	724.172
7	3.4483	.5172	.20065641	.9597[illegible]706	.36410481	.03405249	.07305986	803.900
8	3.6697	.4908	.19620605	.96150321	.47678655	.03122025	.09354836	684.790
9	3.8793	.4741	.18920273	.96420234	.53904343	.02992755	.10198843	655.251
10	4.0984	.4508	.18602049	.96539640	.64550662	.02789873	.12007737	597.014
11	4.3137	.4353	.18037146	.96746618	.70858711	.02703387	.12780881	578.843
12	4.5283	.4340	.16931391	.97133285	.68504322	.02823595	.11598718	610.686
13	4.7273	.4145	.16806149	.97175539	.78204167	.02646537	.13143092	574.982
14	5.0000	.3786	.17147845	.97059518	.98659450	.02357717	.16917956	501.407
15	5.2448	.3741	.16283214	.97348571	.98598295	.02447701	.16054964	508.935
16	5.4422	.3673	.15794784	.97505254	1.00975800	.02442764	.15948886	513.355
17	5.6291	.3543	.15653569	.97549659	1.08017445	.02341804	.16908574	505.094
18	5.9211	.3355	.15433770	.97617990	1.18491936	.02276743	.18287754	480.348
19	6.1290	.3323	.14892334	.97782189	1.18723011	.02320501	.17680603	487.474
20	6.3492	.3238	.14573878	.97876024	1.22915268	.02304076	.17913502	484.566
21	6.5625	.3219	.14048290	.98026460	1.22419453	.02365256	.17197818	491.676

TABLE 24 KERRY (2) RANDOM BACKWARD (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1 \mid N)$	$\hat{g}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{\ell}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.8519	.6296	.44503003	.[illegible]9194831	.48519474	.04049188	.21592617	228.748
2	2.2989	.5977	.31257755	.90229529	.26794797	.04617461	.08375448	517.153
3	2.5862	.5776	.26856381	.92787349	.25210094	.04288178	.10927278	688.867
4	2.8777	.5324	.25471264	.93512148	.42900437	.03603356	.06369460	507.938
5	3.1447	.5409	.21657217	.95309651	.29410088	.04064076	.07711810	772.626
6	3.3520	.5196	.20794040	.95676082	.37066666	.03687864	.0759354[illegible]	703.232
7	3.5897	.5077	.19392323	.96239382	.39157528	.03655913	.13126475	720.426
8	3.9024	.4537	.19791275	.96083057	.66324621	.03066264	.18233746	496.904
9	4.1284	.4128	.20383078	.95845506	.89455390	.02563240	.15546602	427.922
10	4.3103	.4181	.18900752	.96427619	.82253933	.02699254	.12006968	476.597
11	4.4898	.4327	.17187375	.97045946	.69859314	.02986527	.09885949	557.738
12	4.7431	.4387	.15716380	.97529960	.62902302	.03290432	.08940387	614.835
13	4.9618	.4389	.14799714	.97809690	.60409294	.03435051	.10180515	651.239
14	5.1661	.4207	.14701021	.97838801	.69250435	.03233616	.11168694	607.536
15	5.3763	.4050	.14530534	.97888637	.76863676	.03090851	.10959923	579.361
16	5.5363	.4014	.14120442	.98006135	.77617556	.03072117	.12741762	593.998
17	5.7239	.3805	.14281917	.97960269	.89216107	.02825065	.13514137	555.613
18	5.8824	.3693	.14202750	.97982824	.95151711	.02698570	.12307376	548.413
19	5.9937	.3754	.13646597	.98137707	.90186530	.02773539	.13975787	585.909
20	6.1920	.3560	.13782197	.98100513	1.01404762	.02582290	.16323674	554.177
21	6.4815	.3302	.13947457	.98054689	1.17037106	.02400270		510.448

TABLE 25 BRONWYN (1) FORWARD (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\Psi}(1\|K)$	$\hat{g}_N$	$\hat{\theta}_N$	$\hat{a}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.7544	.5789	.63507140	.59668434	1.79413986	.01479448	1.13940620	118.646
2	2.0833	.5938	.38449335	.65216486	.48632550	.02882147	.18698889	371.106
3	2.6549	.4867	.33426964	.88826382	.90232730	.02649885	.30162054	250.232
4	2.8169	.5000	.28923821	.91634130	.68697894	.02738333	.19870043	367.574
5	2.9762	.5060	.25943124	.93269545	.57064462	.02771568	.14804298	487.435
6	3.0928	.5258	.23121846	.94653803	.40000695	.02950815	.09248894	732.821
7	3.3019	.5047	.22119617	.95107228	.47455251	.02776903	.10496914	686.132
8	3.4043	.5149	.20571250	.95768237	.38719636	.02828852	.07965106	887.622
9	3.4221	.5399	.19150281	.96332669	.23966920	.02918644	.04589731	1493.007
10	3.5714	.5036	.19743812	.96101820	.41940671	.02465301	.08260683	979.702
11	3.7931	.4828	.19101030	.96351510	.50174642	.02400782	.09583867	869.233
12	3.8462	.4679	.19465148	.96211082	.58364946	.02116063	.11360818	831.939
13	4.0000	.4554	.19020444	.96382231	.63376176	.02049334	.12054420	809.601
14	4.1056	.4663	.17825317	.96822584	.54849845	.02176580	.09777147	939.817
15	4.2493	.4674	.16950506	.97126806	.52071440	.02253632	.08826363	1005.461
16	4.3597	.4578	.16771096	.97187304	.56255972	.02159568	.09434736	981.597
17	4.6322	.4593	.15397477	.97629178	.52021617	.02422322	.08010012	1030.778
18	4.6272	.4473	.15893215	.97474062	.59115225	.02143848	.09395295	992.946
19	4.7980	.4419	.15352738	.97642940	.60367423	.02180303	.09268034	989.750
20	4.9628	.4293	.15159750	.97701824	.66129237	.02125638	.10025012	938.546
21	5.0481	.4159	.15341586	.97646362	.73438507	.01975591	.11266613	898.544
22	5.1765	.4165	.14820665	.97805479	.71713334	.02023935	.10628384	929.750
23	5.2874	.4069	.14777052	.97816390	.76554394	.01947640	.11312473	907.745
24	5.4054	.4032	.14496905	.97898400	.77714789	.01940950	.11266226	914.613
25	5.4945	.3956	.14474726	.97904825	.81652194	.01869149	.11818916	905.333
26	5.6277	.3961	.14016950	.98035252	.80118762	.01919121	.11230201	927.983

2/......

TABLE 25 [illegible] (1) FORWARD (CONT.) (CALCULATIONS)

N	$\bar{t}_N$	$\hat{\phi}(1\|N)$	$\hat{g}_N$	$\hat{\theta}_N$	$\bar{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
27	5.7692	.3889	.13832933	.98086500	.85424902	.01898528	.11540103	912.858
28	5.8455	.3820	.13848490	.98082197	.87173456	.01826536	.12072188	907.018
29	5.9548	.3758	.13745880	.98110509	.90252459	.01790493	.12405986	900.380
30	6.0484	.3750	.13504553	.98176271	.90002358	.01794424	.12154406	917.004
31	6.1386	.3723	.13347453	.98218459	.19031003	.01778425	.12150300	925.567
32	6.2622	.3718	.13033456	.98301291	.90411770	.01808383	.11783767	938.544
33	6.3953	.3760	.12557352	.98423129	.86784303	.01891417	.10897503	970.299
34	6.5511	.3757	.12196499	.98512459	.85939127	.01947795	.10481542	979.625
35	6.6414	.3719	.12109226	.98533672	.87750959	.01919930	.10625941	980.343
36	6.7669	.3647	.12054384	.98546922	.91531664	.01883873	.11033559	962.196
37	6.8646	.3655	.11818951	.98603129	.90427685	.01907799	.10687578	980.886
38	7.0111	.3542	.11859185	.98593599	.96813250	.01844828	.11481255	944.243
39	7.0652	.3587	.11611104	.98651826	.93591058	.01876267	.10866934	980.906
40	7.1048	.3535	.11688983	.98633677	.96717131	.01804729	.11305245	980.252
41	7.2310	.3439	.11731583	.98623705	1.02289772	.01747776	.12000179	953.578
42	7.2539	.3489	.11533745	.98669732	.98864317	.01766020	.11402738	993.174
43	7.3254	.3390	.11697555	.98631674	1.05015373	.01676327	.12284213	971.234
44	7.4074	.3350	.11662900	.98639768	1.07242203	.01648112	.12507540	970.222
45	7.4875	.3344	.11529374	.98670739	1.07197094	.01649551	.12559130	981.015
46	7.5163	.3382	.11360556	.98709381	1.04487133	.01662657	.11870295	1013.366
47	7.5806	.3371	.11276412	.98728430	1.04877377	.01651979	.11826378	1023.701
48	7.5472	.3443	.11124176	.98762530	1.00293636	.01662710	.11156821	1078.136
49	7.6087	.3432	.11047184	.98779601	1.00698185	.01651846	.11124295	1088.398
50	7.6570	.3476	.10837859	.98825413	.97597247	.01682728	.10577428	1123.663
51	7.7626	.3486	.10632932	.98869413	.96449624	.01714700	.10255283	1137.351
52	7.7728	.3572	.10583517	.98921829	.90958595	.01764419	.09444684	1200.165

TABLE 25 BROWN'S (1) FORWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\bar{\bar{g}}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\bar{d}_N$	$\hat{c}_N$	$\bar{v}_N$
53	7.7941	.3603	.10267901	.98945707	.88935536	.01770762	.09131783	1236.843
54	7.8374	.3570	.10286367	.98941910	.90797277	.01731667	.09339714	1236.609
55	7.8910	.3515	.10347843	.98929226	.94026224	.01679827	.09728622	1223.809
56	7.8762	.3615	.10109186	.98978049	.87841273	.01729502	[illegible]	1302.254
57	7.9387	.3579	.10106397	.98978609	.89800525	.01700100	.09075582	1296.226
58	7.9780	.3570	.10054749	[illegible]	.89821571	.01688175	.09031320	1311.780
59	8.0163	.3573	.10005212	.98998958	.89848715	.01676203	.08989549	1327.288
60	8.0537	.3597	.09889323	.99022019	.88213933	.01687525	.08723730	1358.556
61	8.1442	.3605	.09739178	.99051487	.87364072	.01740470	.08508521	1373.0490
62	8.1365	.3598	.09772390	.99045008	.87945765	.01672790	.08594382	1391.1560
63	8.2677	.3556	.09686744	.99061674	.89897454	.01675758	.08708113	1370.5440
64	8.3660	.3516	.09650040	.99068773	.92000955	.01662268	.08878100	1355.2160
65	8.3979	.3514	.09611601	.99076176	.92004323	.01649934	.08843064	1370.7590
66	8.4833	.3496	.09541166	.99089664	.92855749	.01649236	.08854729	1369.5330
67	8.5350	.3478	.09516728	.99094319	.93739253	.01633047	.08920902	1372.8491
68	8.5859	.3485	.09433329	.99110126	.93115574	.01637875	.08783871	1390.1580
69	8.6466	.3421	.09510225	.99095559	.96918720	.01587906	.09217167	1366.4913
70	8.7391	.3396	.09452522	.99106503	.98170024	.01584567	.09279519	1360.1700
71	8.7980	.3383	.09409672	.99114585	.98786962	.01576639	.09295505	1364.6587
72	8.8452	.3378	.09362727	.99123394	.98977452	.01570505	.09263235	1374.7620
73	8.8700	.3414	.09245765	.99145162	.96565211	.01588783	.08928168	1409.9501
74	8.9480	.3349	.09306097	.99133968	1.00431538	.01546880	.09346235	1383.3666
75	9.0056	.3325	.09296465	.99135762	1.01782513	.01529452	.09462142	1381.9900
76	9.0476	.3298	.09308589	.99133503	1.03376389	.01505359	.09622866	1380.6588
77	9.1232	.3294	.09227866	.99148470	1.03375340	.01512152	.09539306	1386.4937
78	9.1549	.3310	.09152251	.99162364	1.02232075	.01517738	.09356529	1408.3767
79	9.1967	.3306	.09113097	.99169517	1.02350521	.01511541	.09327286	1418.5822
80	9.2166	.3353	.08980113	.99193579	.99272549	.01537559	.08914763	1459.1120

TABLE 26 BRONWYN (1) BACKWARD (CALCULATIONS)

N	$\bar{t}_N$	$\hat{\phi}(1\|N)$	$\widehat{K}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.7241	.6207	.55821389	.68839729	.99278522	.02209215	.55418640	163.356
2	2.0202	.6364	.35823703	.87166625	.20032859	.03396091	.07176507	820.611
3	2.2222	.6000	.33105141	.89040500	.30471778	.02708168	.10087723	732.084
4	2.4242	.5758	.30228794	.90862203	.34558731	.02485890	.10446680	770.140
5	2.6312	.5789	.25994492	.93242866	.22536081	.02759835	.05858138	1237.050
6	2.7523	.5780	.24202198	.94142538	.18719590	.02678708	.04530551	1647.987
7	2.9289	.5607	.22949642	.94733143	.23533612	.02569522	.05400877	1441.164
8	3.0075	.5677	.21628976	.95321876	.17370468	.02547012	.03757053	2090.027
9	3.1359	.5575	.20836741	.95658308	.20054299	.02448056	.04178659	1955.113
10	3.3113	.5397	.20105916	.95957524	.26275563	.02373732	.05282941	1594.860
11	3.4700	.5205	.19714636	.96113336	.34042358	.02248092	.06711322	1325.586
12	3.5191	.5015	.20283705	.95885718	.44282514	.01942132	.08982128	1146.495
13	3.5714	.5082	.19511199	.96193135	.39217371	.01943719	.07651770	1344.729
14	3.7135	.5040	.18646210	.96523190	.39042515	.01983000	.07279944	1385.413
15	3.8462	.4974	.18041152	.96745169	.40621704	.01981571	.07328616	1377.204
16	3.9604	.4827	.17973977	.96769363	.47357994	.01872103	.08512110	1255.056
17	4.0767	.4844	.17190248	.97014957	.44596583	.01931792	.07666254	1350.475
18	4.1763	.4919	.16320854	.97336298	.38979667	.02030097	.06361812	1548.575
19	4.2986	.4887	.15819520	.97497433	.39245373	.02050473	.06208422	1571.067
20	4.3860	.4759	.15893024	.97474122	.45379698	.01929510	.07212198	1437.194
21	4.5295	.4591	.15873921	.97480190	.53325331	.01842168	.08464813	1282.577
22	4.6025	.4540	.15738535	.97522986	.55446327	.01789600	.08726436	1280.670
23	4.6371	.4597	.15360522	.97640544	.51737118	.01799247	.07947087	1398.722

2/.......

TABLE 26 BROWN'S (1) BACKWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\phi(1\|N)$	$\bar{R}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
24	4.6967	.4578	.15174150	.97697455	.52170097	.01767926	.07917571	1428.809
25	4.8170	.4566	.14720106	.97833186	.51641931	.01806029	.07597268	1457.632
26	4.8507	.4627	.14367509	.97935748	.47832066	.01824757	.06872212	1594.867
27	4.8913	.4583	.14361050	.97935879	.49917209	.01757288	.07171625	1586.972
28	4.9645	.4592	.14060497	.98023027	.48729682	.01770800	.06851625	1648.417
29	5.0347	.4514	.14074779	.98019010	.52183869	.01706201	.07386976	1586.840
30	5.1195	.4454	.13979518	.98045737	.55133587	.01672340	.07707393	1551.665
31	5.2189	.4444	.13670814	.98131090	.54837734	.01693779	.07496756	1575.0800
32	5.2805	.4406	.13593304	.98152226	.56482291	.01659974	.07677794	1569.2480
33	5.3312	.4378	.13521194	.98171777	.57662332	.01627210	.07796621	1576.4540
34	5.3628	.4432	.13241506	.98246628	.54335374	.01648028	.07194811	1686.7315
35	5.4180	.4412	.13134062	.98274964	.55029851	.01627709	.07227647	1700.0270
36	5.4962	.4351	.13092335	.98285913	.57888114	.01592780	.07518886	1656.7960
37	5.5639	.4376	.12811255	.98358721	.55959350	.01619677	.07169080	1722.4134
38	5.6464	.4324	.12737107	.98377663	.58303225	.01595776	.07426137	1683.5938
39	5.7101	.4334	.12527329	.98430663	.57283202	.01608246	.07175054	1732.9823
40	5.8055	.4369	.12164785	.98520184	.54674113	.01664399	.06650972	1806.7080
41	5.8993	.4345	.11993605	.98561537	.55404603	.01671107	.06645000	1800.9846
42	5.9744	.4296	.11948299	.98572385	.57670933	.01643975	.06990684	1765.5190
43	6.0820	.4286	.11714858	.98627627	.57569361	.01671315	.06744146	1774.3718
44	6.1711	.4250	.11607468	.98652673	.59057182	.01664114	.06855023	1753.2240
45	6.2762	.4184	.11553174	.98665273	.62169665	.01642669	.07182556	1695.1230
46	6.3361	.4201	.11371219	.98706555	.60865194	.01659496	.06921107	1741.3185

TABLE 28 [illegible] (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\hat{\hat{g}}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	[illegible]	[illegible]	$\hat{v}_N$
47	6.[illegible]	.4190	.111[illegible]	.97758485	.6[illegible]5044	.01692481	.067[illegible]	1745.[illegible]
48	6.5[illegible]	.4135	.11076175	.98708661	.63700569	.01699027	.06[illegible]	1[illegible]
49	6.[illegible]	.4090	.10906470	.98910491	.65153409	.01695277	.07[illegible]	1599.3154
50	6.[illegible]	.4078	.10797055	.98840700	.65531337	.01705131	.07[illegible]	1662.3150
51	6.[illegible]	.4036	.10778800	.98839609	.68958624	.016[illegible]	.07434440	1612.[illegible]
52	6.9612	.3976	.10713768	.98852158	.70449806	.01688151	.07[illegible]	1599.[illegible]
53	7.0781	.3899	.10705975	.98[illegible]3821	.74585625	.01627290	.07[illegible]	1543.2970
54	7.1925	.3834	.10581315	.98[illegible]343	.74515452	.016[illegible]5454	.07[illegible]	1555.1956
55	7.[illegible]	.[illegible]	.10504282	.98899605	.75875446	.01629625	.07[illegible]	1559.6848
56	7.3236	.38[illegible]	.10591464	.98920178	.76157511	.01655456	.07[illegible]	1544.6751
57	7.3834	.3873	.10256900	.98952043	.76408092	.01655556	.07[illegible]	[illegible]583.[illegible]
58	7.4[illegible]	.3819	.10208529	.98957[illegible]	.77225790	.01637181	.07[illegible]	1549.5984
59	7.5[illegible]	.3827	.10129702	.98975114	.77570985	.016[illegible]	.07[illegible]	1575.7[illegible]
60	7.5949	.3797	.10088303	.98982971	.78047198	.01621953	.07[illegible]	1560.5[illegible]
61	7.6925	.3770	.10000026	.98999995	.79298952	.01622943	.07[illegible]	1559.1957
62	7.[illegible]	.3756	.09886780	.99022576	.79672054	.016[illegible]	.07[illegible]	1555.[illegible]
63	7.9046	.3739	.09756112	.99048185	.80719811	.016[illegible]	.07[illegible]	1547.11[illegible]
64	7.9800	.3716	.097[illegible]373	.99058640	[illegible]1295121	.01[illegible]	.07[illegible]7548	1542.160[illegible]
65	8.0745	.3677	.09561877	.99066484	.83266741	.016[illegible]	.08045115	1522.07[illegible]
66	8.11[illegible]	.3678	.09597999	.99078706	.83039725	.01629582	.07970150	1539.8790
67	8.2209	.3652	.09566122	.99084896	.85425702	.01616077	.08171916	1514.4106
68	8.3130	.3570	.09590113	.99087300	.88862193	.01584[illegible]	.08521968	1481.3545
69	8.3434	.3579	.09527022	.990[illegible]364	.88195937	.01582205	.08402401	1504.3457
70	8.4236	.357[illegible]	.094[illegible]1982	.99119383	.88206941	.01591541	.08519654	1510.4530

4/........

TABLE 26. BROGWYN (1) BACKWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\tilde{g}_N$	$\hat{\theta}_N$	$\hat{u}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
71	8.5337	.3522	.0941[illegible]	.99114025	.91312961	[illegible]	[illegible]	1481.7055
72	8.5124	.3505	.09461341	.99101043	.92143792	.01557110	.08730455	1495.9640
73	8.6801	.3460	.09573075	.99121457	.94349227	.01545546	.08843011	146[illegible].3468
74	8.7574	.3467	.09259551	.99142655	[illegible]	.0158[illegible]	[illegible]	147[illegible].1784
75	8.7925	.3458	.09237152	.99146754	.9497[illegible]	.01549325	[illegible]	1465.5050
76	8.8889	.3404	.09246856	.99144959	.97137135	.01525703	.08982110	1459.4253
77	8.9639	.3370	.09222424	.99149472	.98054097	.01513947	.09098363	1451.9614
78	9.0382	.3384	.09114867	.99169195	.97849318	.01530321	.08923906	1464.6580
79	9.1224	.3372	.09044409	.99181992	.98376322	.01534787	.08897525	1464.5708
80	9.2060	.3337	.09029722	.99184644	1.00331497	.01520573	.09059632	1451.8200

TABLE 27 BRONWYN (1) RANDOM FORWARD (CALCULATIONS)

N	$\bar{x}_N$	$\hat{\phi}(1\|N)$	$\bar{\xi}_N$	$\hat{\theta}_N$	$\hat{\sigma}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.4286	.7571	.56819189	.67715800	.16131729	.02097490	.09165913	1040.290
2	1.7857	.6696	.42268026	.82134145	.20741558	.02230636	.09767045	992.445
3	1.9355	.6452	.38151506	.85444570	.21820551	.01956762	.08324879	1227.761
4	2.2346	.6145	.31374628	.90153193	.19235778	.02258844	.06036113	1447.594
5	2.5126	.5779	.28199255	.92048025	.28323690	.02315096	.08003986	1079.330
6	2.6786	.5759	.25441623	.93527240	.27577089	.02492227	.05743976	1445.839
7	2.8340	.5789	.23042631	.94690377	.15656751	.02547675	.03607725	2175.968
8	2.9520	.5793	.21631194	.95373886	.12244463	.02553889	.02644949	2960.811
9	3.0405	.5743	.20914614	.95625794	.12885171	.02429033	.02694882	3055.322
10	3.2258	.5531	.19949145	.96021914	.11764449	.02413772	.03543143	2338.542
11	3.3435	.5541	.19113940	.96346575	.11639059	.02397417	.03371518	2474.349
12	3.4884	.5174	.19710933	.96111637	.35474211	.02059814	.06995130	1588.053
13	3.6212	.5125	.18916273	.96421748	.35779297	.02072810	.06768107	1425.614
14	3.7234	.5213	.17784166	.96837234	.29094744	.02186995	.05174256	1767.400
15	3.7975	.5266	.17073089	.97085100	.25081134	.02220434	.04282121	2103.455
16	3.8929	.5255	.16541052	.97263938	.24551323	.02221849	.04027962	2234.800
17	4.0284	.5261	.15761268	.97515827	.22319603	.02509110	.07517850	2462.114
18	4.1284	.5183	.15537506	.97585863	.25239271	.02245704	.03921549	2271.014
19	4.2793	.4977	.15566891	.97576720	.34527093	.02119283	.05374793	1755.817
20	4.3384	.4924	.15482688	.97602868	.36740506	.02035626	.05688411	1727.024
21	4.4397	.4841	.15321147	.97652626	.40165114	.01980991	.06153754	1640.618
22	4.5549	.4803	.14946771	.97765946	.410181[illegible]	.01989169	.06130878	1639.970
23	4.6939	.4633	.14979482	.97756153	.490651[illegible]	.01894189	.07549700	1436.604

2/.........

TABLE 27 BREWYN [1] HANSEN FORWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\widehat{\ell}_N$	$\hat{\beta}_N$	$\hat{\alpha}_N$	$\bar{d}_N$	$\hat{c}_N$	$\hat{v}_N$
24	4.8000	.4540	.14884716	.97784454	.53273350	.01838984	.07929581	1371.519
25	4.8924	.4501	.14658338	.97851536	.54581445	.01821622	.08000720	1372.280
26	5.0096	.4470	.14521941	.97948825	.55227047	.01836637	.07909566	1376.747
27	5.1040	.4423	.14143300	.97999674	.57055706	.01814514	.08069545	1365.906
28	5.2045	.4368	.13981104	.98045290	.59328765	.01791374	.08294809	1345.976
29	5.3016	.4351	.13709956	.98120373	.59466357	.01800071	.08152872	1362.792
30	5.3763	.4373	.13392723	.98206353	.57579100	.01825078	.07711599	1421.070
31	7.7720	.3692	.10072947	.96985362	.83581823	.01767431	.08419132	1461.065
32	5.4482	.4306	.13388169	.98207575	.60853285	.01767448	.08147120	1388.940
33	5.5749	.4251	.13180929	.98262632	.63039792	.01752514	.08309221	1361.835
34	5.6799	.4200	.13037372	.98300272	.65209997	.01772393	.08501661	1342.349
35	5.7239	.4242	.12776351	.98367649	.62440151	.01769376	.07977569	1414.488
36	5.8333	.4200	.12604007	.98410887	.64087510	.01755872	.08078867	1399.137
37	5.8920	.4190	.12479430	.98442644	.64258111	.01751846	.08019030	1420.416
38	5.9871	.4159	.12326139	.98480666	.65409338	.01685219	.08062434	1416.018
39	6.1192	.4026	.12399840	.98462445	.72346455	.01703663	.08970827	1322.945
40	6.1321	.4088	.12176776	.98517263	.68595350	.01662524	.08552691	1405.466
41	6.1824	.4049	.12171549	.98518538	.70563966	.01678973	.08588707	1400.665
42	6.2980	.4025	.11966157	.98568112	.71245778	.01696826	.08525372	1397.246
43	6.3830	.4027	.11763352	.98616236	.70607781	.01663221	.08305836	1419.086
44	6.4662	.3970	.11742932	.98621041	.73477191	.01652818	.08628356	1393.645
45	6.5574	.3934	.11646527	.98643589	.75080186	.01637694	.08744216	1383.835
46	6.6766	.3872	.11570477	.98661244	.78108341	.01637371	.09037489	1351.291
47	6.7747	.3844	.11446798	.98689711	.79252446	.01636199	.09071851	1346.439
48	6.8713	.3816	.11329955	.98716325	.80412173	.01624023	.09110647	1341.667

3/.......

TABLE 27 [illegible] (1) [illegible] FORWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\widehat{g}_N$	$\bar{\pi}_N$	$\bar{\mu}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
49	6.9565	.3783	.11254233	.98733425	.81955224	.01616306	.09223413	1335.198
50	7.01[illegible]	.3777	.11166483	.98753101	.[illegible]260	.01585291	.09160322	1350.815
51	7.0621	.3743	.11161923	.98754120	.83835423	.01549680	.09[illegible]	1348.202
52	7.1429	.3683	.11177999	.98750526	.87112778	.01546551	.09737450	1325.386
53	7.1724	.3697	.11082393	.98771811	.86089969	.01568190	.09540802	1355.441
54	7.2404	.3716	.10905502	.98811138	.84574270	.01559891	.09221548	1383.017
55	7.3569	.3666	.10831624	.98826760	.[illegible]	.01547421	.09431487	1359.426
56	7.4224	.3644	.10776508	.98858669	.88116217	.01570856	.09495842	1361.093
57	7.4667	.3680	.10601890	.98875999	.85682702	.01553352	.09083980	1401.578
58	7.5597	.3634	.10567909	.98883[illegible]4	.[illegible]1747	.01581477	.09308392	1383.201
59	7.6316	.3658	.10384858	.98921549	.86273789	.01582[illegible]	.08959395	1411.524
60	7.7326	.3630	.10295397	.98940051	.87545568	.01625955	.09013146	1402.550
61	7.8[illegible]	.3647	.10050887	.98989803	.85944045	.01606405	.08638108	1441.31
62	7.9385	.3624	.09991711	.99001658	.87027133	.01599453	.08695489	1438.03
63	7.9949	.3604	.09958214	.99008346	.88022608	.01584787	.08765447	1439.74
64	8.0402	.3631	.09824067	.99034882	.86186600	.01603347	.08467007	1473.23
65	8.1454	.3622	.09695727	.99059933	.86322200	.01621161	.08369541	1474.02
66	8.2397	.3633	.09537917	.99090284	.85294855	.01650370	.08135337	1489.61
67	8.3230	.3627	.09439415	.99100976	.90108976	.01660156	.08056432	1495.33
68	8.3744	.3596	.09444678	.99107981	.87043285	.01633898	.08220947	1488.96
69	8.4352	.3533	.09515190	.99094617	.90685999	.01586240	.08628911	1461.18
70	8.4951	.3519	.09469777	.99103236	.91332203	.01578744	.08648938	1464.72
71	8.5956	.3499	.09387672	.99118721	.92215431	.01584106	.08656853	1458.43

4/.......

TABLE 27 BRONWYN (1) RANDOM FORWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\hat{g}_N$	$\hat{\theta}_N$	$\bar{a}_N$	a_N	$\hat{c}_N$	$\hat{v}_N$
72	8.6538	.3498	.09315813	.99132156	.92070591	.01586520	.08577120	1469.77
73	8.7425	.3437	.09349388	.99125892	.95567173	.01553456	.08934927	1440.92
74	8.7886	.3420	.09332377	.99129069	.96475351	.01538103	.09003431	1444.23
75	8.8652	.3416	.09247285	.99144882	.96460789	.01545905	.08919972	1450.39
76	8.8993	.3396	.09253150	.99143797	.97006969	.01523614	.09031689	1453.40
77	8.9953	.3376	.09184712	.99156415	.98547822	.01526515	.09051299	1447.50
78	9.0382	.3395	.09089136	.99173380	.97199702	.01539072	.08834589	1470.91
79	9.1224	.3337	.09126556	.99167061	1.00008444	.01507047	.09182078	1445.31
80	9.2060	.3326	.09055477	.99179989	1.01036265	.01511171	.09149277	1445.87

TABLE 28 BRONWYN (1) RANDOM BACKWARD (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\bar{\hat{g}}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.5385	.6308	.81367223	.72813746	3.84146500	.00488400	3.14874172	130.052
2	1.7857	.6875	.39792490	.84165579	.05482989	.02657678	.02181818	3449.127
3	2.0979	.6364	.33201349	.88976705	.13515860	.02690567	.04487447	1656.483
4	2.3121	.5896	.31671262	.89969313	.31715149	.02242352	.00044581	887.962
5	2.3148	.6296	.28185099	.92056006	.04959400	.02317626	.01397811	6173.59
6	2.5000	.6208	.25287235	.93605560	.02431506	.02439764	.00614860	13332.316
7	2.6616	.6084	.23583132	.94458362	.04084282	.02425760	.00963201	8559.832
8	2.7778	.6146	.22129089	.95103037	.00021185	.02427603	.00004688	57334.000
9	2.8662	.6210	.21566586	.95305592	.00001461	.02255771	.00000317	10000.000
10	3.0211	.5952	.20034200	.95986313	.02094460	.02391475	.00419608	19930.613
11	3.1539	.5869	.19366932	.96243283	.04163965	.02332831	.00836431	10631.125
12	3.2520	.5745	.18920267	.96420240	.08459371	.02244570	.01600534	5567.137
13	3.3943	.5535	.18727684	.96492743	.17031091	.02116332	.03189694	2962.766
14	3.5533	.5431	.18002605	.96759063	.19962066	.02132520	.03593691	2609.734
15	3.7129	.5225	.17810583	.96827835	.28707916	.02034947	.05113044	1922.195
16	3.8554	.5181	.17073041	.97085118	.28867787	.02081869	.04928604	1949.370
17	3.9627	.5082	.16844159	.97162747	.32786477	.02014433	.05522600	1797.767
18	4.0449	.5056	.16480887	.97283804	.33105850	.01989789	.05456135	1842.204
19	4.1215	.5033	.16155374	.97390044	.33379354	.01963938	.05392550	1888.461
20	4.2463	.4926	.15929186	.97462612	.37716824	.01920531	.06007978	1733.326
21	4.3659	.4761	.15983838	.97445172	.45510149	.01816266	.07274258	1513.777
22	4.4534	.4777	.15501308	.97597098	.43573916	.01846198	.06754518	1603.826
23	4.5817	.4761	.14984971	.97754508	.43059909	.01892769	.06452507	1637.585
24	4.6784	.4737	.14663345	.97849864	.43394893	.01896195	.06363136	1657.585

TABLE 28 BRONWYE (1) RANDOM BACKWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\bar{\phi}(1\|N)$	$\hat{g}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
25	4.7529	.4715	.14439946	.97914881	.43877757	.01878355	.06335920	1680.516
26	4.8327	.4703	.14170414	.97991997	.43781585	.01876951	.06204024	1717.527
27	4.9091	.4636	.14113599	.98008066	.46761760	.01822315	.06602579	1662.237
28	5.0542	.4531	.13953739	.98052937	.51418650	.017985[illegible]1	.07174808	1549.876
29	5.1146	.4462	.13975132	.98046958	.54774320	.01731110	.07654774	1509.291
30	5.1724	.4466	.13760740	.98106426	.54044658	.01727007	.07436931	1557.192
31	5.2365	.4459	.13561350	.98160899	.53857800	.01721755	.07301128	1590.982
32	5.2718	.4481	.13372338	.98211807	.52307546	.01716324	.06994736	1665.936
33	5.3659	.4374	.13422674	.98198318	.57606947	.01651627	.07732385	1566.048
34	5.4487	.4375	.13159555	.98268265	.56889755	.01668986	.07486427	1600.669
35	5.5644	.4308	.13026673	.98305062	.59826225	.01655133	.07793355	1550.500
36	5.6515	.4286	.12842619	.98350674	.60136726	.01656412	.07761651	1555.632
37	5.7453	.4301	.12532192	.98429447	.51907342	.01693834	.07382363	1599.425
38	5.8372	.4301	.12285012	.98490787	.58219431	.01717359	.07159626	1626.590
39	5.9091	.4273	.12181813	.98516035	.59188113	.01702232	.07234544	1624.051
40	6.0241	.4187	.12147057	.98524495	.63547665	.01669333	.07719159	1552.092
41	6.1286	.4081	.12206721	.98509961	.69028687	.01612498	.08426130	1471.982
42	6.1584	.4150	.11927015	.98577464	.64824522	.01649929	.07731622	1567.811
43	6.2319	.4174	.11683404	.98634982	.62999505	.01680445	.07360476	1616.962
44	6.3218	.4167	.11498499	.98677850	.62873304	.01696236	.07229465	1630.939
45	6.4011	.4097	.11520255	.98672843	.66425912	.01652201	.07652420	1581.861
46	6.4785	.4099	.11345595	.98712778	.65872651	.01667102	.07473630	1605.228
47	6.5460	.4123	.11136341	.98759824	.64132065	.01694335	.07141948	1652.774
48	6.6298	.4061	.11133015	.98760563	.67232460	.01660039	.074849[illegible]	1609.611

TABLE 28 BRONWYN (1) RANDOM BACKWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\hat{R}_N$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\bar{c}_N$	$\hat{v}_N$
49	6.7308	.4038	.10991538	.98791862	.68025494	.01668817	.07477039	1602.844
50	6.7659	.4060	.10863054	.98819941	.66582316	.01674837	.07232869	1651.005
51	6.8548	.3948	.10932283	.98791701	.72617048	.01603157	.07982254	1562.890
52	6.9333	.3960	.10808831	.98831695	.71517146	.01626810	.07730150	1590.396
53	7.0106	.3968	.10643357	.98867190	.70673102	.01646719	.07521981	1614.650
54	7.0959	.3943	.10556507	.98885602	.71775579	.01643233	.07576984	1606.329
55	7.1990	.3901	.10480297	.98901635	.73684400	.01637173	.07722330	1581.929
56	7.2727	.3896	.10365272	.98925614	.73638040	.01644223	.07632768	1593.629
57	7.3643	.3876	.10261244	.98947072	.74396950	.01648654	.07634038	1589.082
58	7.4454	.3851	.10190362	.98961568	.75506324	.01643087	.07694352	1581.967
59	7.5064	.3868	.10048717	.98990238	.74248779	.01661581	.07461023	1613.282
60	7.5949	.3861	.09926057	.99014735	.74286598	.01674927	.07373720	1619.376
61	7.6730	.3682	.10254508	.98948455	.84615242	.01542593	.08676857	1494.256
62	7.7694	.3797	.09815431	.99036574	.77325022	.01658005	.07589775	1589.339
63	7.8456	.3773	.09760267	.99047375	.78437364	.01650366	.07655680	1582.938
64	7.9208	.3775	.09645426	.99069661	.78024828	.01663870	.07525808	1597.200
65	7.9853	.3771	.09562612	.99085569	.78013891	.01667041	.07460141	1608.191
66	8.0586	.3761	.09483242	.99100685	.78327596	.0166963[illegible]	.07427973	1612.644
67	8.1114	.3729	.09485501	.99100256	.80021328	.01643921	.07590407	1602.822
68	8.1928	.3699	.09444708	.99107981	81504053	.01633899	.07697795	1590.155
69	8.2437	.3716	.09336150	.99128366	.80308706	.01648217	.07497722	1618.403
70	8.3532	.3675	.09286684	.99137580	.82356954	.01642183	.07648200	1592.388

4/........

TABLE 28 BRONWYN (1) RANDOM BACKWARD (CONT.) (CALCULATIONS)

N	$\bar{r}_N$	$\bar{\phi}(1\|N)$	$\hat{\bar{g}}_N$	$\bar{\theta}_N$	$\hat{\alpha}_N$	$\hat{a}_N$	$\hat{c}_N$	Q_N
71	8.4524	.3667	.09176826	.99157864	.82486892	.01658389	.07569653	1593.188
72	8.5207	.3621	.09196740	.99154204	.85020238	.01628219	.07619062	1570.950
73	8.5882	.3624	.09105086	.99170977	.84609610	.01638683	.07703757	1584.280
74	8.6651	.3560	.09156507	.99161589	.88220114	.01598287	.08077854	1549.100
75	8.7413	.3555	.09074414	.99176556	.88267481	.01605881	.08009720	1554.890
76	8.8269	.3542	.09000909	.99189842	.88775849	.01610955	.07990599	1553.701
77	8.9327	.3492	.08989376	.99191916	.91495007	.01594148	.08224797	1525.375
78	9.0173	.3468	.08945227	.99199831	.92702635	.01589406	.08292443	1517.442
79	9.1224	.3430	.08910900	.99205959	.94727314	.01581494	.08441049	1498.185
80	9.2060	.3372	.08948600	.99199229	.98105905	.01548498	.08779007	1471.208

TABLE 29 BRONWYN (7) FORWARD (CALCULATIONS)

N	$\bar{r}_N$	$\hat{\phi}(1\|N)$	$\widehat{g_N}$	$\hat{\theta}_N$	$\hat{\alpha}_N$	$\hat{d}_N$	$\hat{c}_N$	$\hat{v}_N$
1	1.5625	.7188	.50158262	.78841487	.17349261	.02974798	.08702081	772.590
2	2.0619	.6352	.34071678	.88591209	.14327878	.03807082	.04881747	1076.124
3	2.4590	.5854	.29087502	.91539174	.27101821	.03606395	.07883239	703.480
4	2.5641	.5833	.26824981	.92804205	.22524637	.03224252	.06042228	1026.606
5	2.6882	.5699	.25707269	.93391365	.25947291	.02926344	.05670338	1060.858
6	2.9268	.5561	.23256105	.94591540	.26336151	.02914926	.06124760	1120.246
7	3.0575	.5265	.23026478	.94697815	.39430279	.02551451	.09079397	863.348
8	3.1875	.5418	.21170712	.95518011	.27812522	.02663940	.05894458	1273.684
9	3.3088	.5515	.19534433	.96184063	.19723541	.02800654	.03852879	1855.468
10	3.4014	.5403	.19299954	.96275121	.24132496	.02584651	.04657558	1661.383
11	3.5851	.5342	.18163106	.96693748	.24265784	.02698703	.04412272	1704.896
12	3.7383	.5109	.18150610	.96705556	.34666818	.02446178	.06292230	1299.384
13	3.9677	.5061	.16785598	.97182441	.33598250	.02653209	.05639663	1336.612
14	3.9886	.4957	.17221391	.97034240	.39407045	.02337012	.06786436	1261.035
15	4.1783	.4791	.16840816	.97163874	.46222788	.02283958	.07784283	1124.924
16	4.3360	.4770	.16096771	.97408944	.45366424	.02349645	.07302517	1165.615
17	4.4041	.4741	.15876228	.97479457	.46184909	.02274941	.07332414	1198.983
18	4.5000	.4650	.15778196	.97510487	.50254661	.02170028	.07929271	1159.130
19	4.5455	.4713	.15324700	.97651541	.46140164	.02188481	.07070827	1292.460
20	4.6512	.4698	.14919484	.97774094	.45847201	.02190277	.06840158	1331.302
21	4.7297	.4617	.14889163	.97783130	.49601752	.02100413	.07385278	1289.313
22	4.8140	.4617	.14547086	.97883826	.48750591	.02102505	.07091779	1341.336
23	4.9145	.4594	.14239901	.97972256	.49078870	.02100692	.06988770	1362.282
24	4.9485	.4536	.14316308	.97950435	.52034724	.01991285	.07449448	1348.256
25	5.0302	.4447	.14327556	.97947216	.56357563	.01908573	.08074647	1297.770
26	5.0881	.4442	.14133543	.98002434	.56106299	.01886958	.07929736	1336.613
27	5.1331	.4487	.13818330	.98090541	.53136092	.01902625	.07342505	1431.635
28	5.2553	.4465	.13485706	.98181361	.5335902C	.01928078	.07195830	1441.532

5.13 Calculation trends

Columns 2 and 3 have been described already in Data Tables (see 5.8 columns 11 & 8) and the trends have been discussed in the Data Trends (see 5.9 $\bar{r}_N$ and $\hat{\phi}(1|N)$. They are repeated here only for completeness.

$\hat{g}_N$ shows a general downward trend for all samples. Although small local inversions of the trend appear, they are only slight, and localised to isolated points.

$\bar{\theta}_N$ shows a general rising trend with only very rare and very small inversions of the trend. The inversions are so small that they usually appear in the fourth figure and very rarely in the third figure quoted.

$\bar{\alpha}_N$ shows an upward trend in the small samples. (Kerry 1 and 2 and Bronwyn 2). The trend is not at all smooth and many inversions appear. In the case of Bronwyn 1 forward the upward trend appears to peter out after about N = 7 000. In the three other Bronwyn 1 cases, namely Bronwyn 1 backward, Bronwyn 1 random forward and Bronwyn 1 random backward, the upward trend continues throughout. The difference is only slight and is probably not significant. In Sandré's case the trend appears to be an increase up to N = 500 followed by a decline up to N = 1 000. Thereafter $\bar{\alpha}_N$ appears to level out. The fluctuations are rather large and the apparent trends may not be significant.

$\bar{d}_N$ showed a general downward trend with many and large local inversions. This downward trend was observed in all cases and since

$$\hat{V}_N = \frac{2}{\bar{c}_N \bar{d}_N}$$

the trend is highly significant and will be discussed in much greater detail in 7.7.1 to 7.7.4.

$\bar{c}_N$, although exhibiting large fluctuations, appeared to be distributed about a constant mean. The deviations from the mean decreased with increasing N. Again since

$$\hat{V}_N = \frac{2}{\bar{c}_N \bar{d}_N}$$

the behaviour of $\bar{c}_N$ is very important and will be discussed in more detail in 7.7.4.

$\hat{V}_N$ shows a general upward trend in the case of the forward natural data. But in all other cases no trend is immediately obvious apart from large fluctuations. The case of Sandré, because of the listing, will be discussed later under 6.4.4.4. to 6.5.1.

5.14 The Graphs type 2 (Calculations)

Graphs of $\bar{c}_N$, $\hat{d}_N$ and $\hat{V}_N$ are shown in figures 65 to 77 on pages 177 to 189. The graphs of all three parameters, $\bar{c}_N$, $\hat{d}_N$ and $\hat{V}_N$ are shown on the same page so that their trends can be easily compared.

Kerry 1 and Kerry 2 (all four variations of each) and Bronwyn 2 were plotted at 100 word intervals, but to enable Bronwyn 1 to be shown on one page the graphs were plotted at intervals of 400 words.

Some of the fine structure of the variations and fluctuations may have been lost but the overall trends are clear.

Sandré was plotted at 200 word intervals.

$\bar{c}_N$, after some wild fluctuations, appears to settle down to deviating less and less from a horizontal average line. This trend is visible in all cases whether reversed, randomised, or both.

The only exception was Sandré but as has already been mentioned in 5.10.1 Sandré made lists of words which dramatically affected Hapax Legomena, $\Phi(1|N)$, and in turn this affected $\bar{F}_N$ on which the estimates of $\hat{V}_N$ were calculated.

The effect of listing is discussed in more detail in 6.4.6. to 6.5.1.

$\hat{d}_N$ showed a continuous downward trend for all samples whether forward, backward, natural or randomised.

$\hat{V}_N$ being calculated by $\hat{V}_N = \frac{2}{c_N d_N}$ showed a steady increase with increasing N in the cases of natural i.e. forward data, only, namely

Kerry 1 forward,

Kerry 2 forward,

Bronwyn 1 forward,

Bronwyn 2 forward.

All the other cases showed no clear trend at all only some rather wild fluctuations, namely:

- Kerry 1 random forward
- Kerry 1 random backward
- Kerry 2 random backward
- Bronwyn 1 backward
- Bronwyn 1 random forward
- Bronwyn 1 random backward

The only exception appears to be in Kerry 2,

Kerry 2 shows a slight downward trend,

Kerry 2 random forward, after an initial rise from N = 200 to N = 700 shows a steady decline thereafter.

FIGURE 65 KERRY 1 FORWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ *vs* N

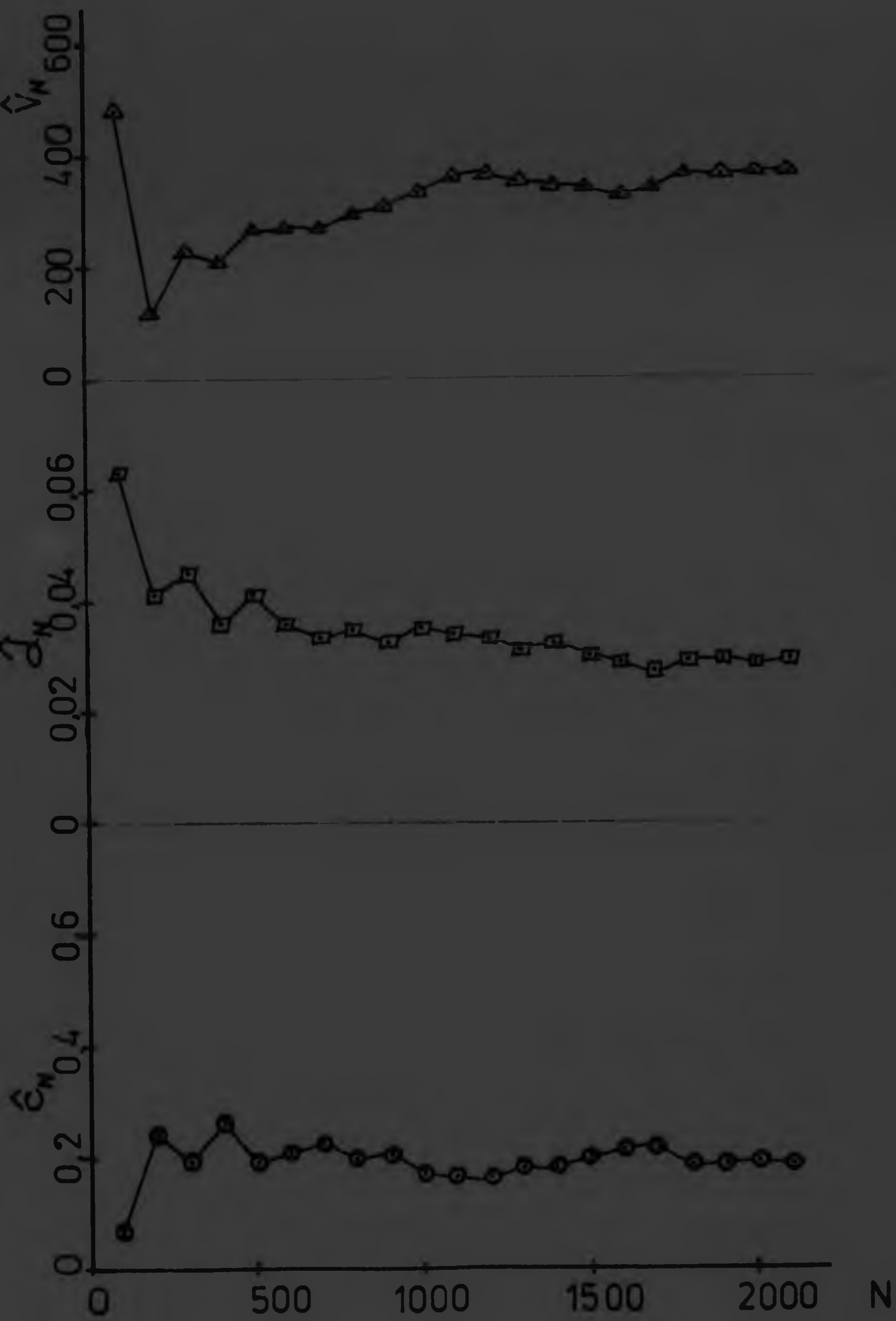


FIGURE 66 KERRY 1 BACKWARD $\hat{c}_N$, $\hat{J}_N$ AND $\hat{V}_N$ vs N

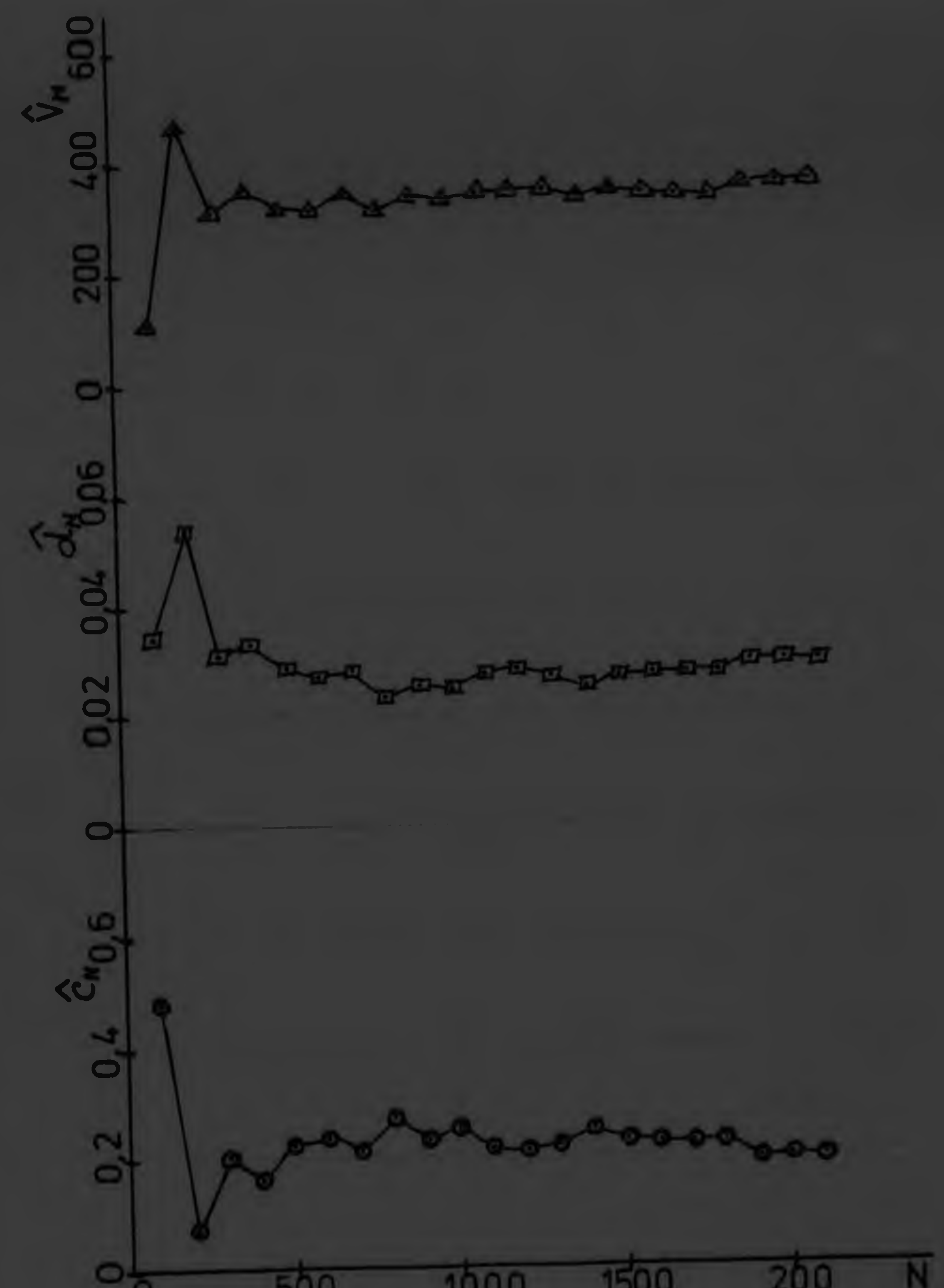


FIGURE 67 KERRY 1 RANDOM FORWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ vs N

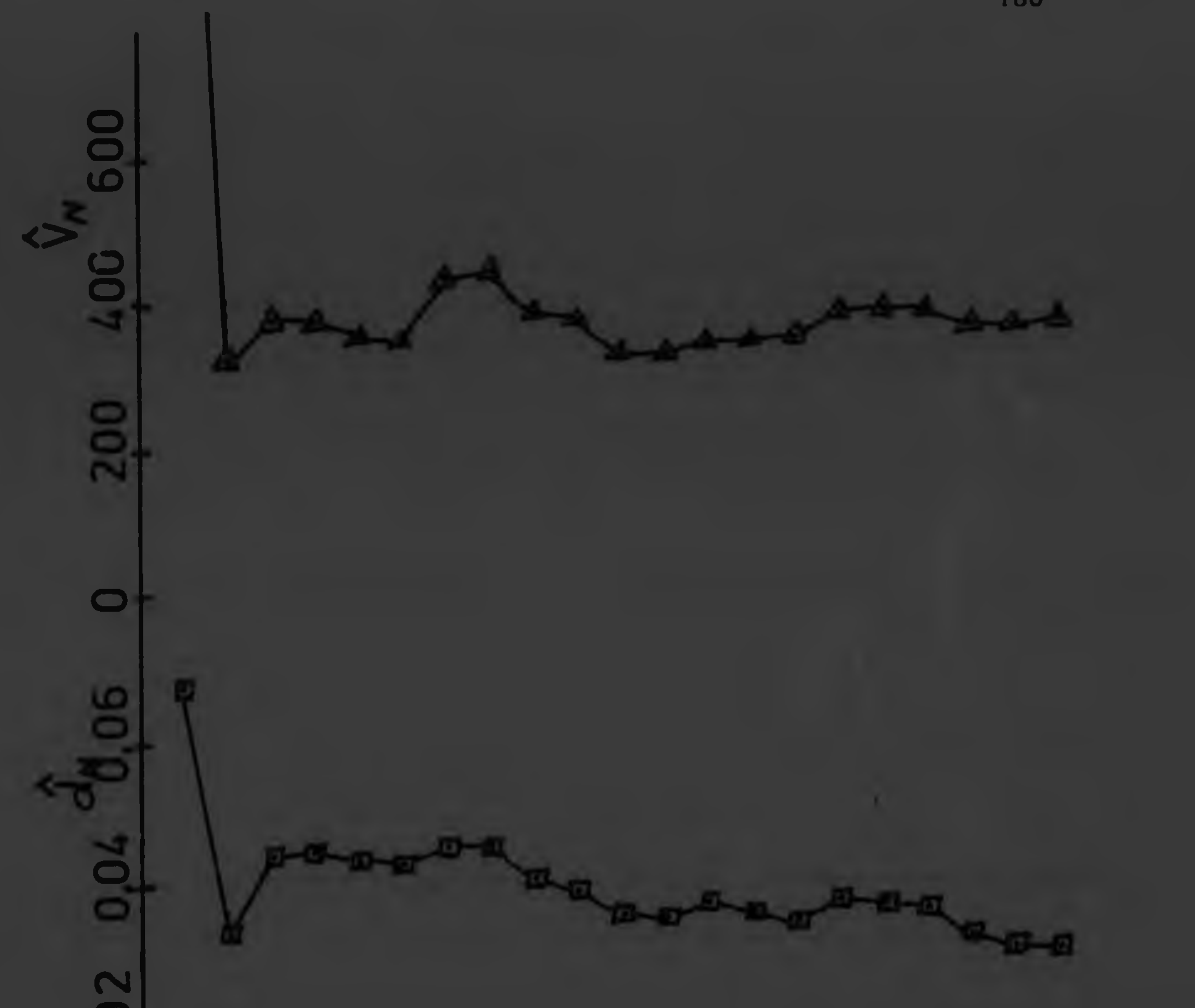


FIGURE 68 KERRY 1 RANDOM BACKWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{v}_N$ vs N

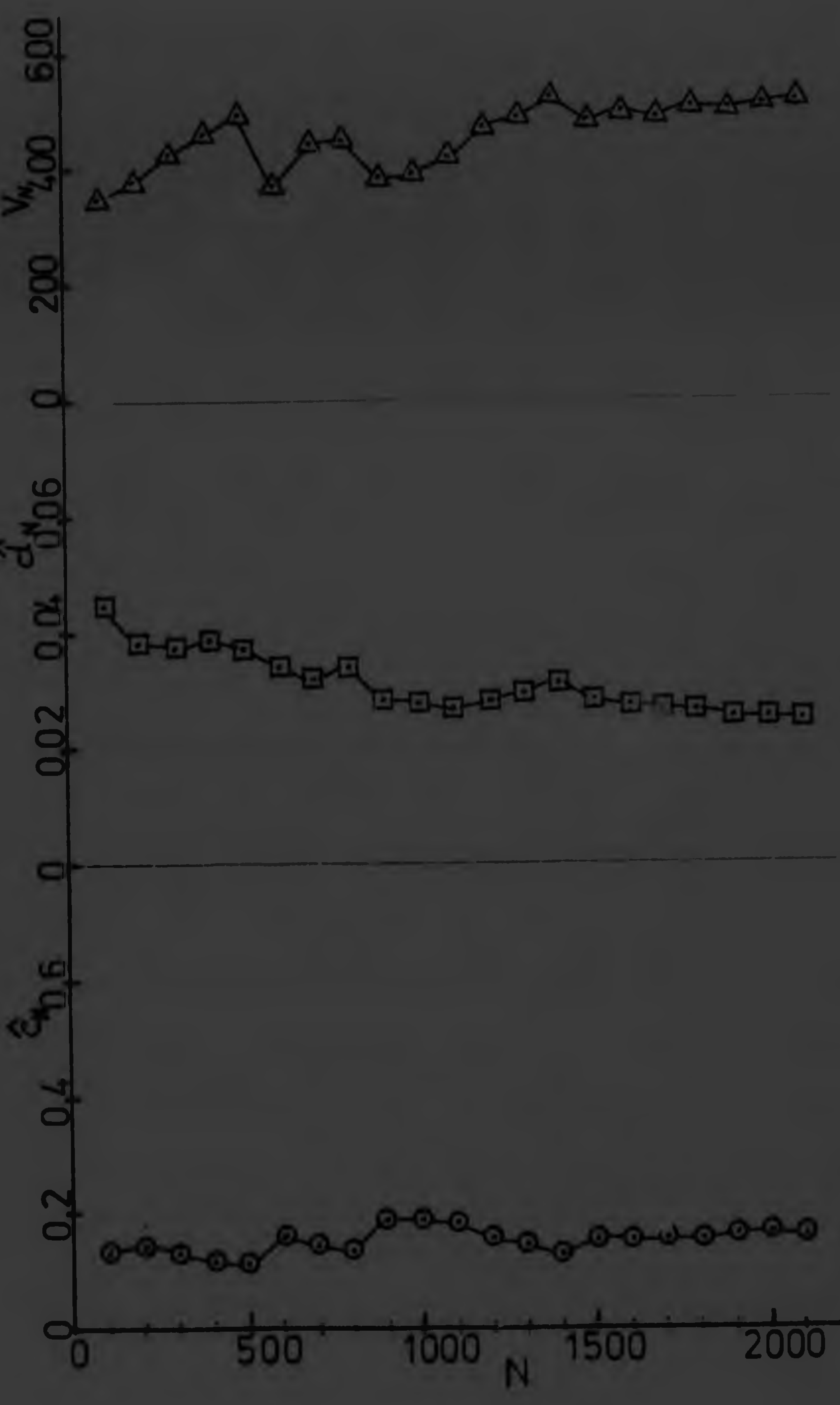


FIGURE 69 KERRY 2 FORWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ *vs N*

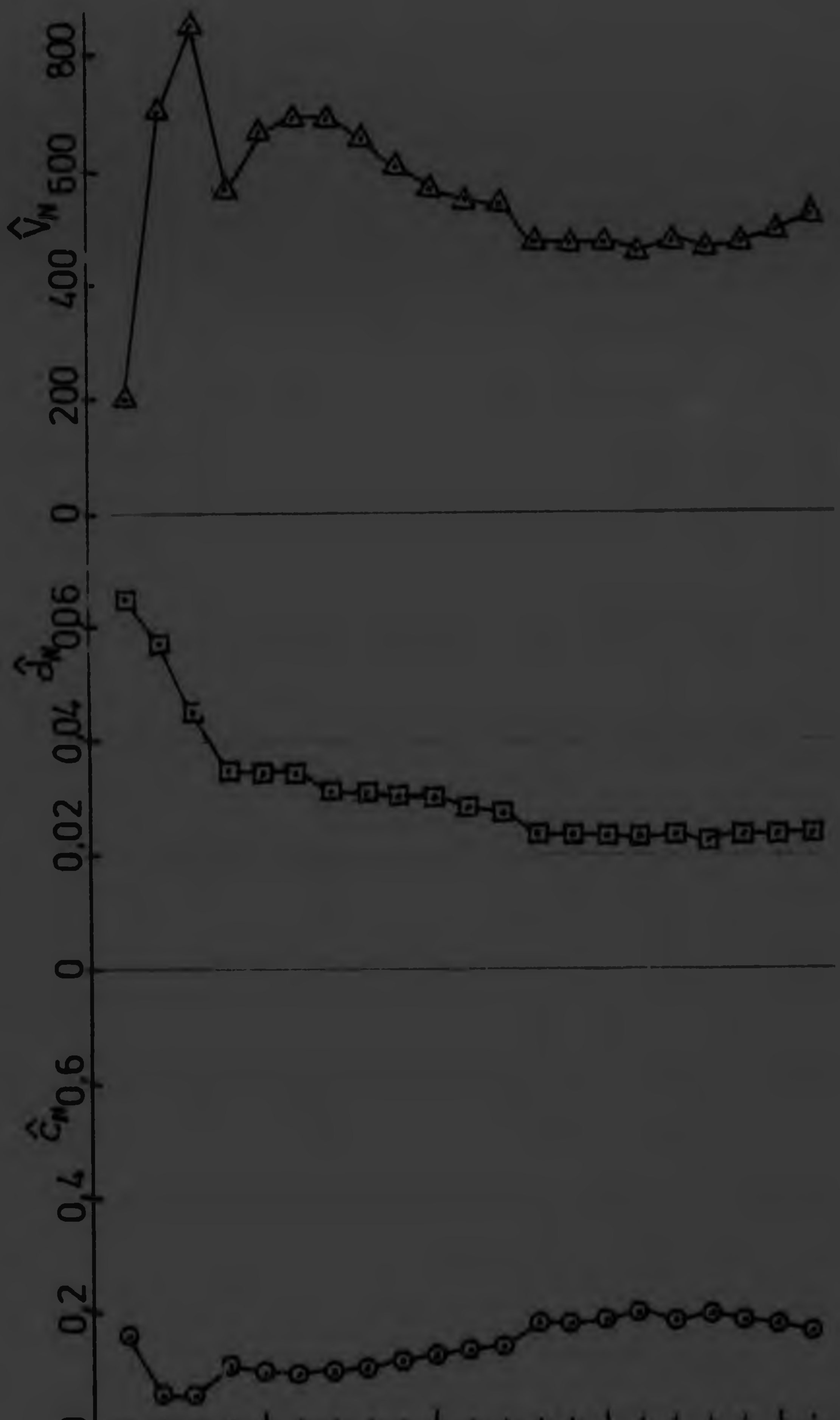


FIGURE 70 KERRY 2 BACKWARD $\hat{c}_N$, $\bar{d}_N$ AND $\hat{V}_N$ vs N

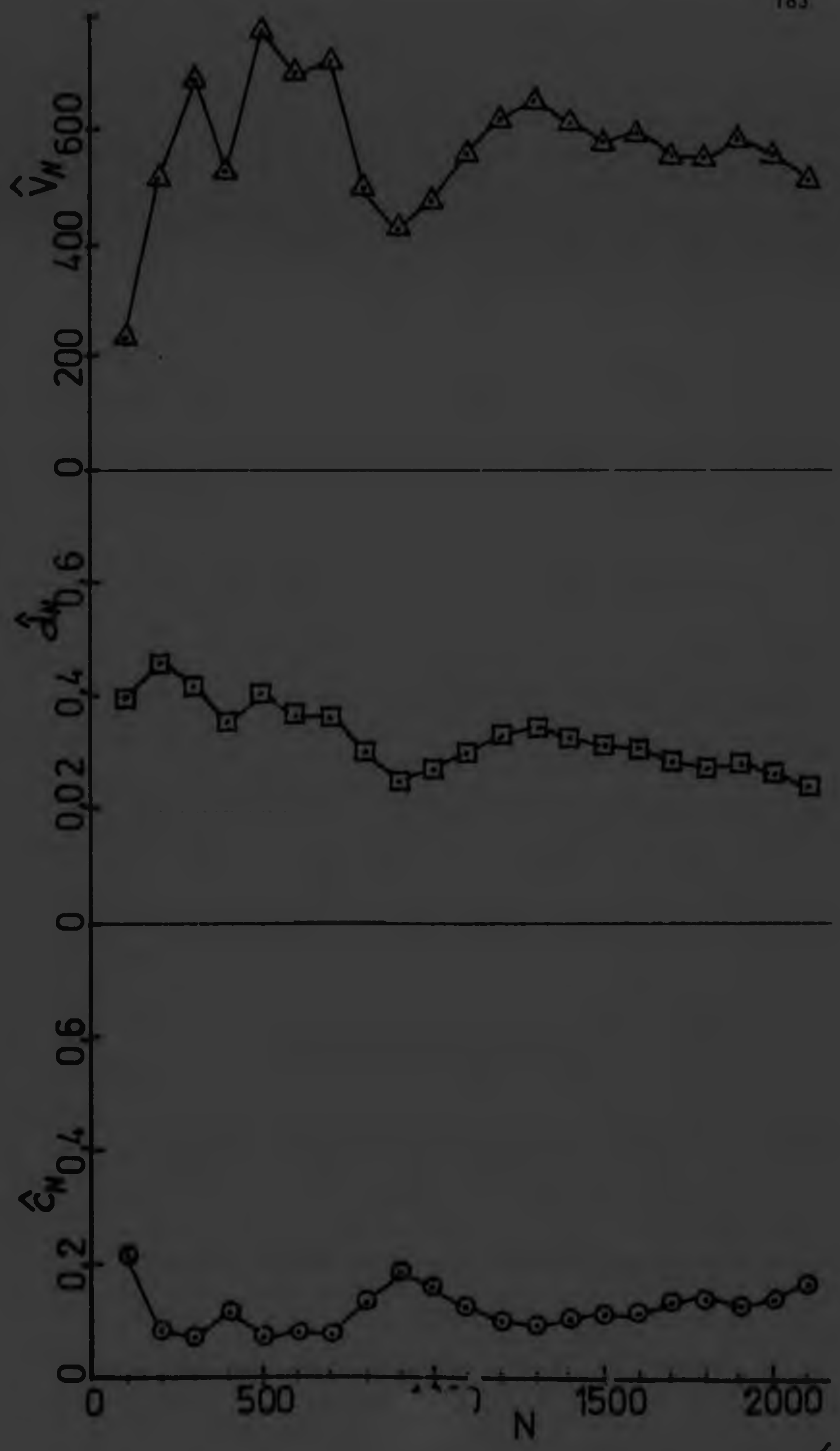


FIGURE 71 KERRY 2 RANDOM BACKWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ *vs* *N*

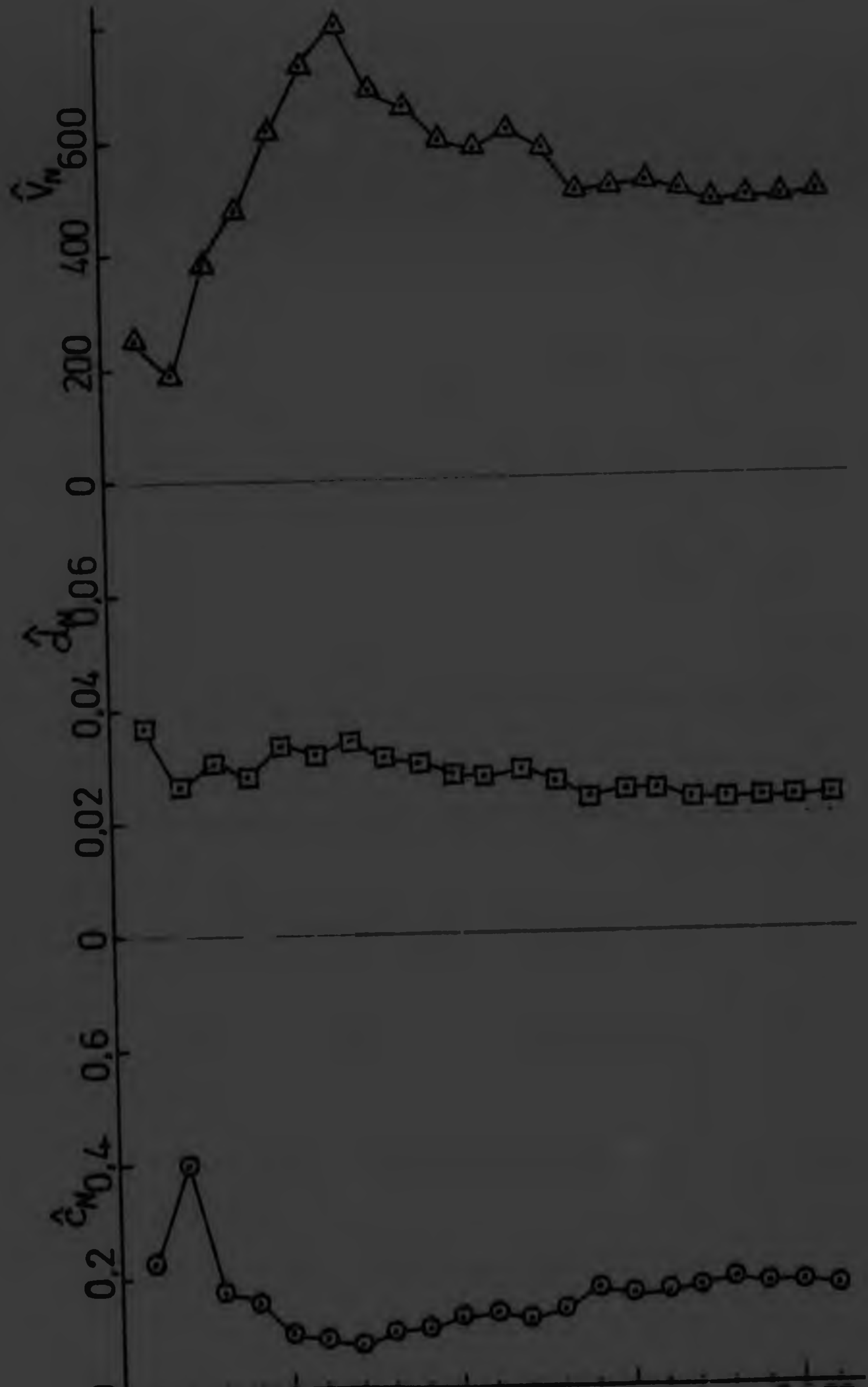


FIGURE 72 KERRY 2 RANDOM FORWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ *vs* *N*

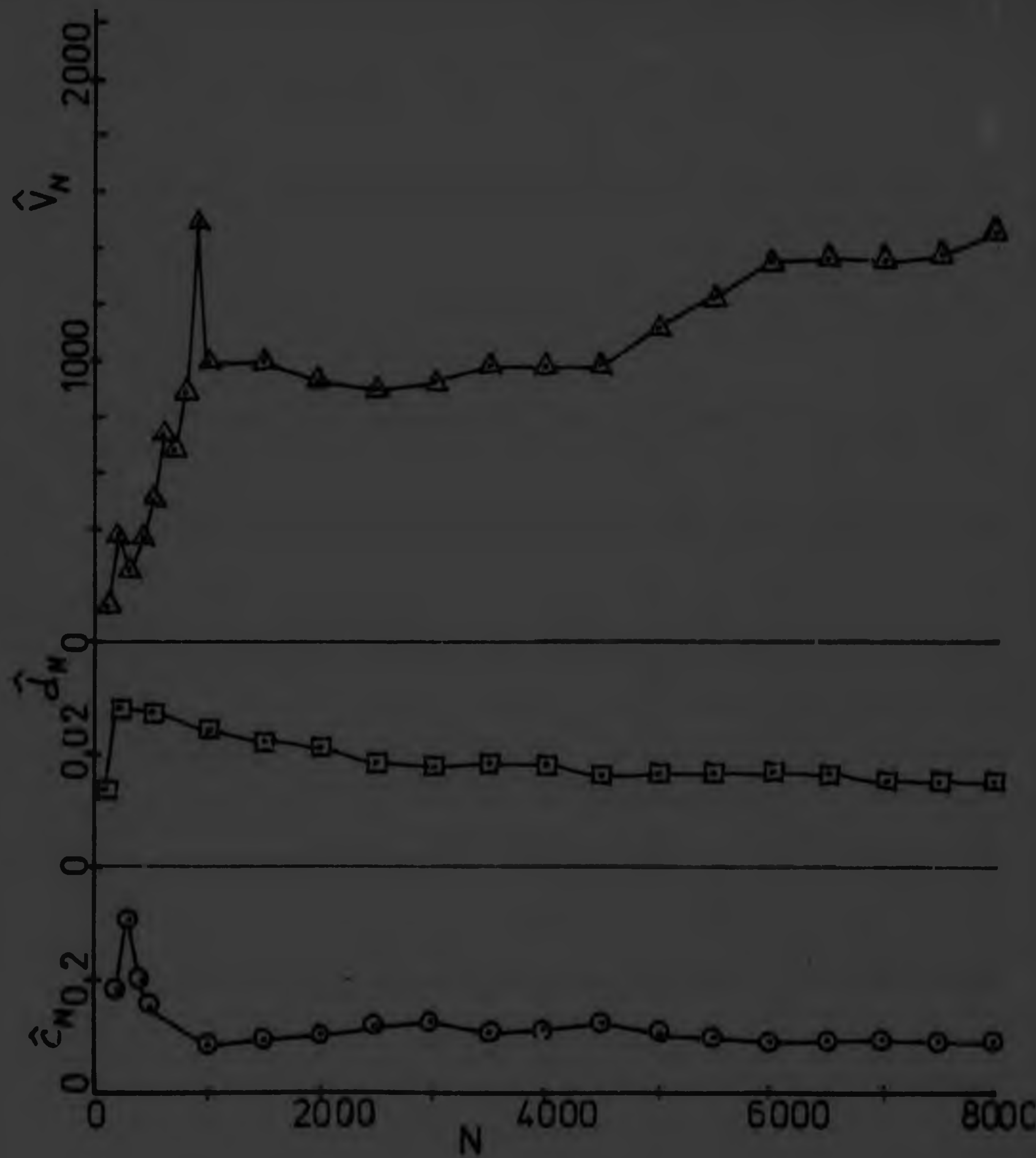


FIGURE 73 BRONWYN 1 FORWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{v}_N$ *vs* *N*

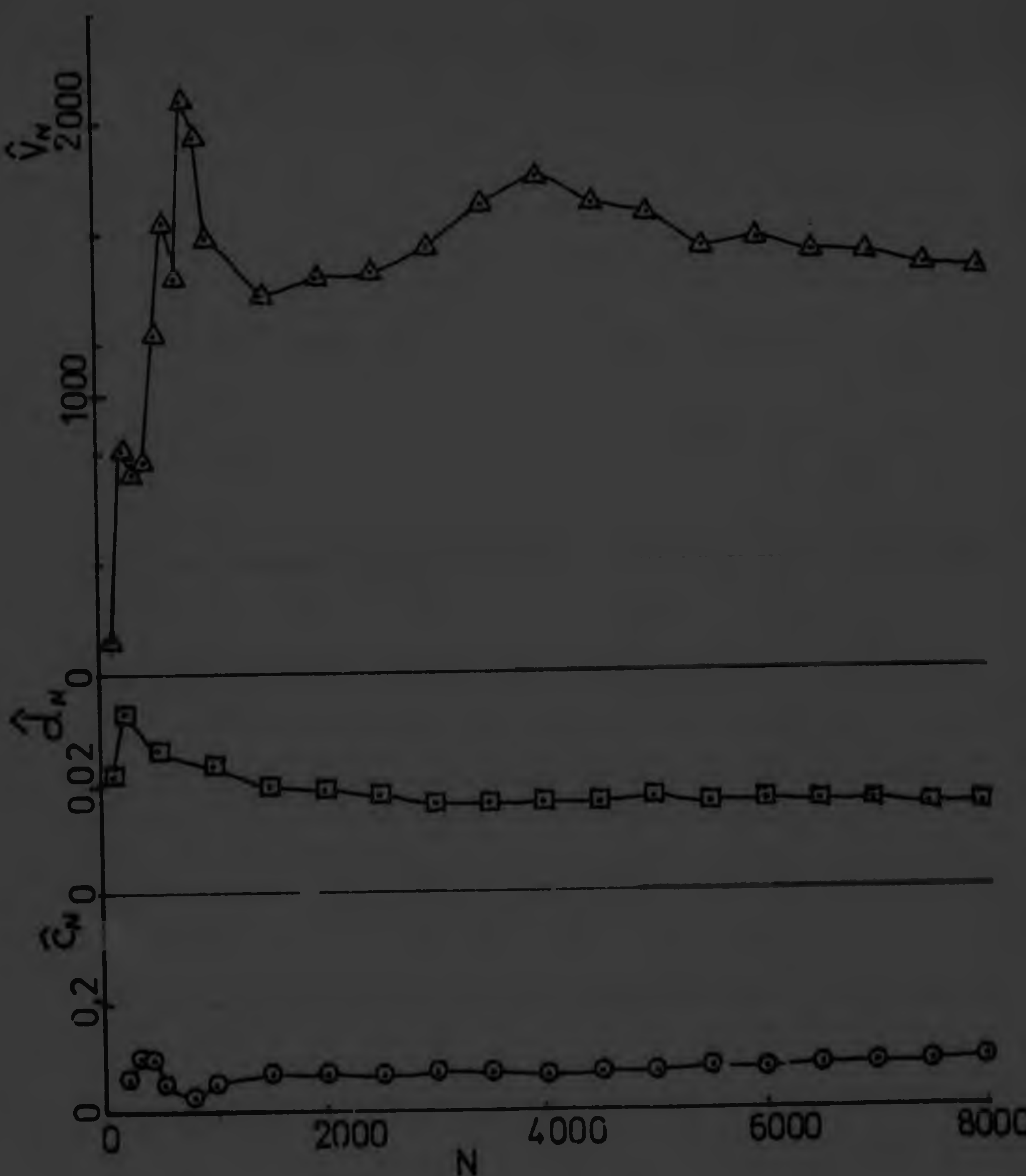


FIGURE 74 BRONWYN 1 BACKWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{v}_N$ *vs N*

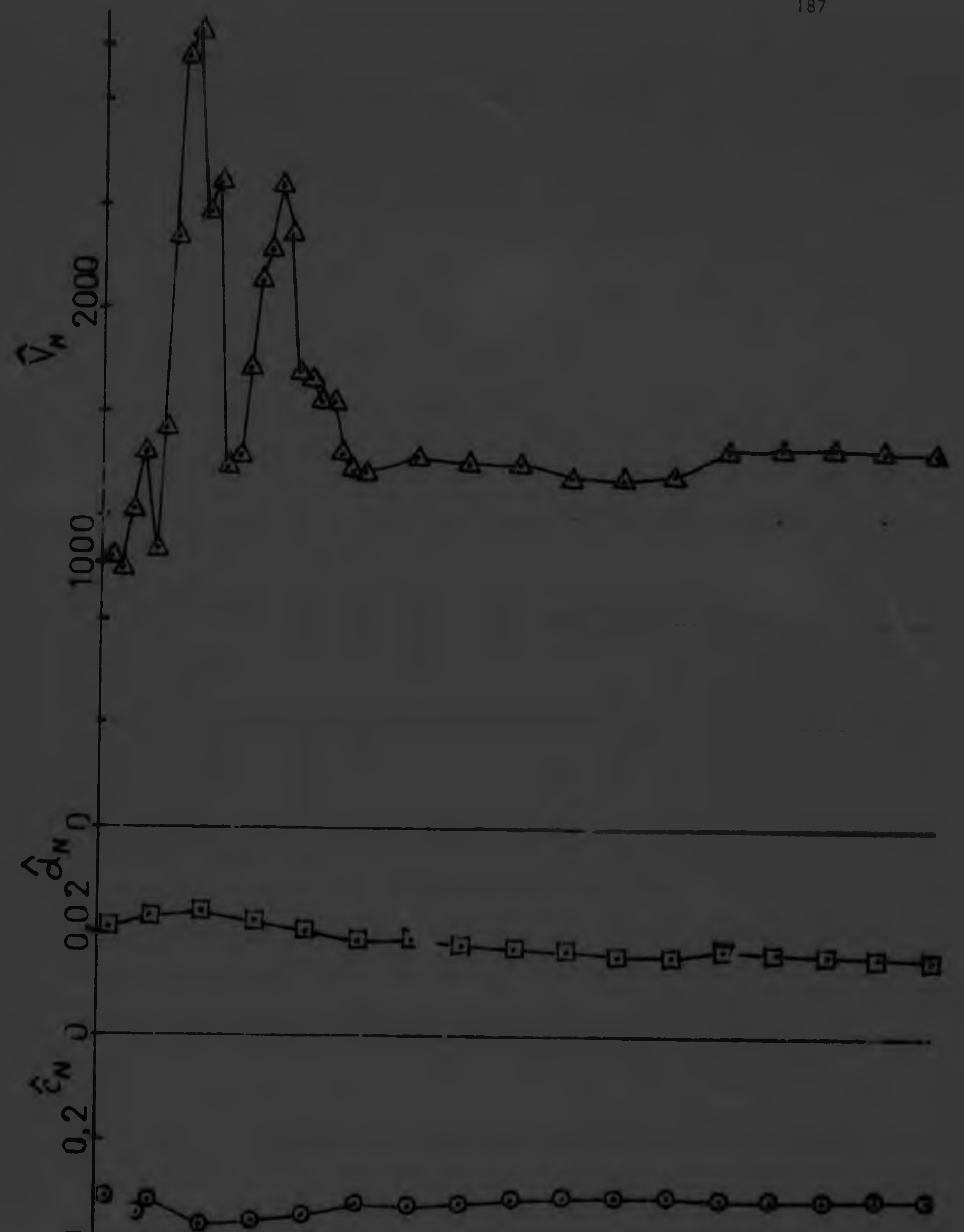


FIGURE 75 BRONWYN 1 RANDOM FORWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ vs N

FIGURE 76 BRONWYN 1 RANDOM BACKWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ vs N

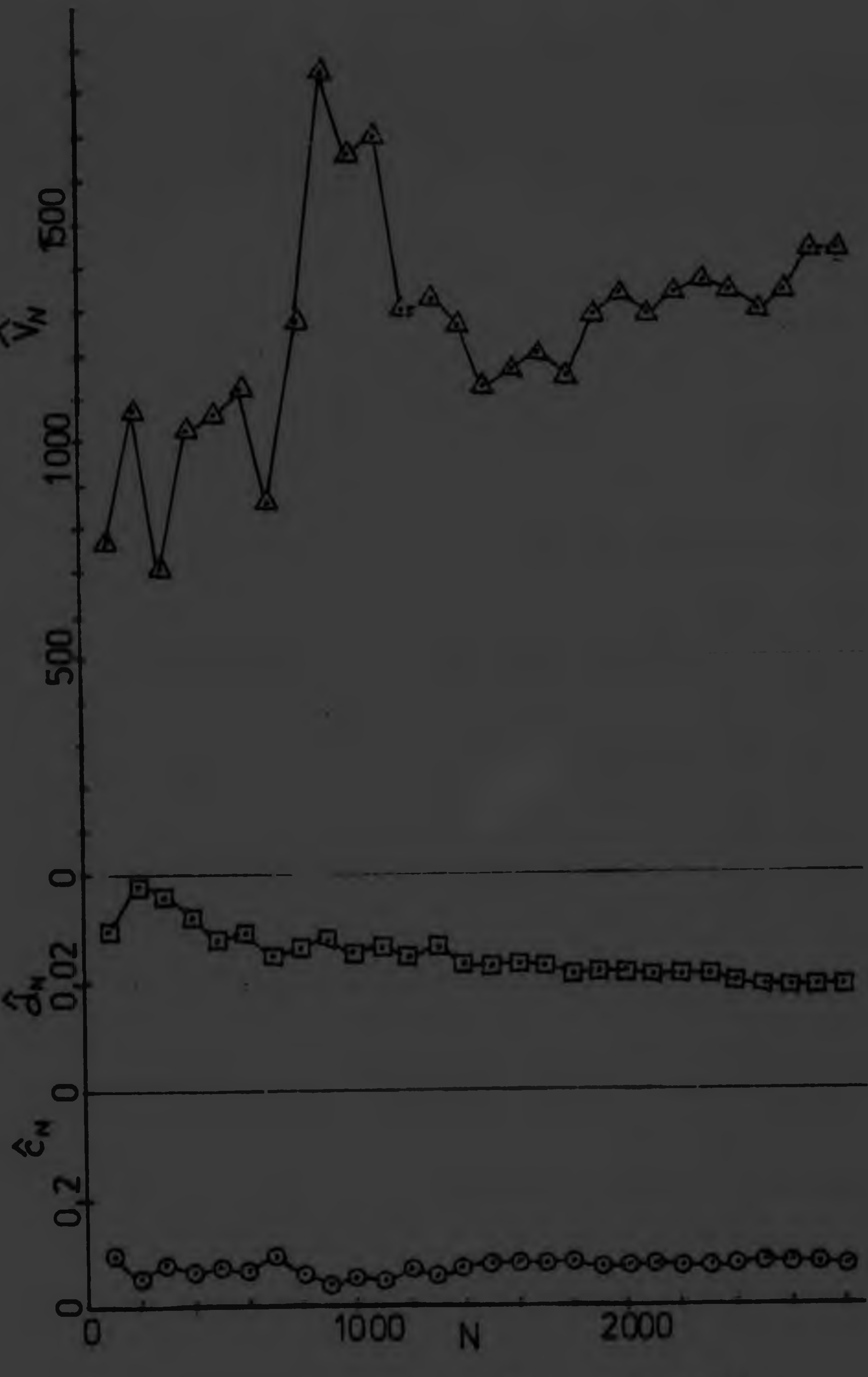


FIGURE 77 BRONWYN 2 FORWARD $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ vs N

CHAPTER VI Results

6.1 Hysteresis

As has been mentioned the data were reversed to determine whether the asymptote was "running away from you".

The data were reversed in the cases of Kerry 1, Kerry 2, Bronwyn 1 and also randomised and reversed.

Graphs were prepared for all the cases and are described and discussed one by one.

The graphs show $\bar{n}_N$ vs N in the forward and backward data cases, together on the same axes. The data are taken from tables 2 to 13 pages 48 to 70

6.1.1 Kerry 1 forward/backward Hysteresis

Figure 78 on page 199 shows the Kerry 1 forward/backward hysteresis loop.

The forward part is distinctly higher than the backward especially tween N = 300 and N = 1 600 with maximum separation at N = 1 200 where the difference is 18.

6.1.2 Kerry 1 random forward/backward hysteresis

Figure 79 on page 200 shows the Kerry 1 random forward/backward hysteresis loop.

As can be readily seen a clear separation between the two lines is clearly visible especially between N = 700 and N = 1 400 with a maximum difference of 17 at N = 1 300.

If the "running asymptote because of learning during the test period", contention was correct there should be no observable separation in the random case and if separation did occur, it would be ascribed to nothing other than chance, because "random forward" differs from "random backward" only in respect of the order in which the letters of the words were read.

Since the separation seen in the random case is as large as in the natural case, no significance can be attached to the natural forward/backward hysteresis.

6.1.3 Kerry 2 forward/backward hysteresis

The hysteresis loop Kerry 2 forward/backward is shown in figure 80 on page 201.

There is a clearly observable separation from N = 400 to N = 1 700 with a maximum separation of 24 at N = 900.

6.1.4 Kerry 2 random forward/backward hysteresis

The hysteresis loop Kerry 2 random forward/backward is shown in figure 81 on page 202.

There is a clear separation between the two curves from N = 400 to $N = 1\ 700$, thereafter the two lines cross and separate in the opposite direction. The maximum difference between the ordinates is 14 and occurs at $N = 900$.

Again, the hysteresis in this case can only be due to chance and is not markedly different from the natural data, forward/backward hysteresis.

6.1.5 Bronwyn 1 forward/backward hysteresis

The graph of Bronwyn 1 forward/backward hysteresis appears in Figure 82 on page 203.

The forward curve is markedly higher than the backward for its entire length. A maximum separation of 128 occurs at $N = 4\ 100$.

6.1.6 Bronwyn 1 random forward/backward hysteresis

The graph of Bronwyn 1 random forward/backward hysteresis appears in Figure 83 on page 204.

Hysteresis does occur but it is much less marked than in the natural data case. A maximum separation of 47 occurred at $N = 4\ 600$.

6.1.7 Hysteresis abandoned

Because $\bar{n}_N$ vs N is cumulative a sudden influx of new words due to a change of topic for example would raise the value of $\bar{n}_N$ above the trend and the effect would be long lasting as has already been pointed out in 3.7.3. The converse is equally true in that if there was a portion of text during which, for whatever reason, few or no new words were introduced, the growth of $\bar{n}_N$ would be stunted and the curve of $\bar{n}_N$ vs N would flatten and take a long time to catch up with the trend. A sudden influx or stunting of growth in the random case could equally well be due to chance.

If a sudden influx occurred at, say, a quarter of the way along the text and the words appeared only once in the complete text, then on reversing the data the sudden influx would occur at three quarters of the way along the text and thereby create an hysteresis. Thus, there would tend to be an element of rotational symmetry about the hysteresis loop.

This is clearly visible in the case of Bronwyn 1 (natural unrandomised data) where in the forward direction a more rapid increase in $\bar{n}_N$ takes place from about $N = 2\ 000$ to $N = 4\ 000$, while in the backward direction a more rapid increase in $\bar{n}_N$ takes place between about $N = 4\ 000$ to $N = 6\ 000$.

For this reason hysteresis was abandoned for the remaining data, namely Bronwyn 2 and Sandré - in which cases the natural, forward analysis was made.

6.2 Kerry The estimate of V, $\hat{V}_N$

The natural forward data gave an estimate of $\hat{V}_N$ which rose steadily with increasing text size N. This could easily be interpreted as evidence for vocabulary growth during the test period. The two samples of text were recorded contiguously (Kerry 1 from 15th February to 24th February and Kerry 2 from 25th February to 22nd March) and a graph of the estimates at the beginning and the end of the recording periods was plotted against the date. See Figure 91 on page 222. Although only four points were plotted, the estimate at the beginning of the second recording is very nearly equal to the estimate at the end of the first recording. It is tempting to claim this as evidence for continuous growth during the period 11 February to 22nd March but there are two pitfalls.

FIRST: The estimate plotted against 24th February was not calculated on data collected on 24th February but was calculated on data accumulated from 11th February to 24th February.

SECOND: The evidence supplied by Bronwyn.

6.3 Bronwyn

Bronwyn 1 was recorded from 11th February to 2nd March and Bronwyn 2 from 10th March to 17th March.

The estimate at the beginning of the second recording is considerably lower than the estimate at the end of the first recording and there are 10 days clear in between the two recordings. See Figure 92 on page 223.

It is also significant that the two estimates at the ends of the recordings are almost equal.

6.3.1 Bias

The only conclusion which can be drawn from this evidence is that the method of estimating $\hat{V}_N$ is at fault and applies a downward bias to $\hat{V}_N$ especially with small text. The estimate of $\hat{V}_N$ appears to approach the true value of V asymptotically.

An asymptotic treatment of $\hat{V}_N$ vs N appears in 7.8 on page 228.

6.4 Sandré

Since the text was so large (14 000 words) it was analysed at 200 word intervals. The data were analysed and can be seen in table 31 on pages 216 and 217.

Only the natural data were used, no reversal or randomisation of text was made.

6.4.1 Sandré's Data Table - Table 30 on pages 205, 206 and 207

Column 1 shows the size of text analysed, i.e. the number of tokens, N.

Column 2 shows the number of types appearing in that text, $\bar{n}_N$, which increases continuously through the entire text but shows a rapid rise from N = 9 000 to N = 10 000. This occurred because Sandré was reminded to use as many different words as possible. This she did by the very obvious process of lists. A table of the numbers of words in the lists is shown in table 32 on page 219.

Column 3, under the heading of $\bar{f}_1(1|N)$ shows Hapax Legomena, corresponding to N. The trend is upward with some local inversions and a dramatic increase from N = 9 000 onward.

Column 4 under the heading $\bar{f}_1(2|N)$ shows Hapax Dislegomena, which exhibits a continuous upward trend with some local inversions.

Column 5 under the heading $\bar{f}_1(3|N)$ shows Hapax Trilegomena, which also shows a steady rise but there are many local inversions of the trend and no sudden rises. It appears as though the words which appeared once because of the listing may have appeared twice rarely and almost never thrice.

Column 6 under the heading $\dfrac{\bar{f}_1(2|N)}{\bar{f}_1(1|N)}$ shows the ratio of Hapax Dislegomena to Hapax Legomena, to three figures. The ratio rises for the first 2000 words of text, thereafter it appears to remain constant but with many fluctuations and no decrease where the listing occurs.

Column 7 under the heading $\dfrac{\bar{f}_1(3|N)}{\bar{f}_1(2|N)}$ shows the ratio of Hapax Trilegomena to Hapax Dislegomena, to three figures. The ratio rises to a maximum of 0,623 at N = 7 600 and thereafter appears to drop slightly and then rise again towards the end of the text.

Column 8 under the heading $\dfrac{\bar{f}_1(1|N)}{\bar{n}_N}$ shows the proportion of Hapax Legomena, to three figures. This ratio is also known, more correctly as $\phi(1|N)$. It shows a dropping trend at first then appears to level off.

Column 9 under the heading $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$ shows the proportion of Hapax Dislegomena. The ratio appears to remain constant but fluctuates.

Column 10 under the heading $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ shows the proportion of Hapax Trilegomena to three figures. The ratio rises for the first 3 000 text words and thereafter fluctuates about a mean without any apparent trend either way.

Column 11 under the heading $\frac{\hat{N}}{\hat{n}_N}$ shows the average number of times the words appear. This ratio is also known as $\bar{r}_{..}$. It shows an upward trend with few and small inversions of the trend.

6.4.2 Sandré's Data Graphs

As in the other cases the graphs of $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ and $\hat{n}_N$ are shown on the same axes, because they are all natural numbers.

$\hat{n}_N$ vs N appears in Figure 84 on pages 208 and 209. As can be clearly seen the trend is generally upward with a few local inversions and a sharp rise from N = 9 000 to N = 10 000.

$\hat{f}_1(1|N)$ vs N appears in Figure 84 on pages 208 and 209. It shows a general upward trend and the sharp rise between N = 9 000 and N = 10 000 can be clearly seen.

$\hat{f}_1(2|N)$ vs N appears in Figure 84 on pages 208 and 209. It shows a general upward trend, not as steep as $\hat{f}_1(1|N)$.

$\hat{f}_1(3|N)$ vs N appears in Figure 84 on pages 208 and 209. It shows a gentle upward slope with some small inversions.

The ratios $\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$ and $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ are shown together in Figure 85 on pages 210 and 211. $\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$ remains remarkably steady during its entire length but remains between the limits 0,3 and 0,4 from N = 1 800 onward. In contrast $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ rises to a maximum of just greater than 0,6 at N = 7 800, declines thereafter to a minimum at about 11 000 and rises again. The trends are general with many inversions. The local fluctuations decline with increasing N, but are larger than those observed in $\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$.

The proportions
of axes.

The proportions $\frac{\hat{f}_1(1\ N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2\ N)}{\hat{n}_N}$ and $\frac{\hat{f}_1(3\ N)}{\hat{n}_N}$ are graphed on the same set of axes.

The proportion of Hapax Legomena $\frac{\hat{f}_1(1|N)}{\hat{n}_N}$ otherwise known as $\hat{\phi}(\_|N)$, appears in Figure 86 on pages 212 and 213. The downward trend is general with some fluctuation initially. There is a slight rise in the neighbourhood of N = 10 000 where the listing takes place. Apart from this local rise the graph appears to be quite flat generally, from N = 7 000 onward.

The proportion of Hapax Dislegomena $\frac{\hat{f}_1(1|N)}{\hat{n}_N}$ otherwise denoted as $\hat{\phi}(2|N)$, appears in Figure 86 on pages 212 and 213. The graph remains practically level throughout its complete length except for local fluctuations which show a decreasing trend along the length of the graph.

The proportion of Hapax Trilegomena appears in Figure 86 on pages 212 and 213. There is no trend, apart from slight local fluctuations and a gentle overall undulation.

The average appearance rate $\bar{f}_N = \frac{N}{\hat{n}_N}$ appears in Figure 87 on pages 214 and 215. The trend is upward except for a few local inversions and downward slope between N = 9 000 and N = 10 000 where the listing occurs.

6.4.3 Sandré's Calculation Tables

The calculations made from the data appear in Table 31 on pages 216 and 217. In contrast with the younger children $\bar{f}_N$, $\phi(1|N)$ $\hat{g}_N$ are not shown.

Column 1 shows the size of text analysed and increases in steps of 200.

Column 2 shows to five figures, the values of $\hat{\theta}_N$, which is the solution to equation 76 on page 37. The value of $\hat{\theta}_N$ rises gradually throughout the text with only one small inversion of the trend at N = 9 200.

Column 3 shows, to five figures, the values of $\hat{\alpha}_N$, which is the solution to equation 76 on page 37. The trend appears to be a gradual general rise to a maximum value of 0,36924 at N = 6 600, thereafter the value appears to drop. The trend appears to be only general, with many local inversions.

Column 4 shows, to five figures, the values of $\hat{d}_N$. The trend is generally downward with many local inversions. As has already been pointed out, it was expected that $\hat{d}_N$ would deviate on either side of a fixed value d and that as the size of text increased the deviations would decrease in size. The downward trend therefore presents a problem because since $\hat{V}_N = \frac{2}{\hat{c}_N \hat{d}_N}$ any trend in $\hat{d}_N$ must affect the estimate of $\hat{V}_N$. $\hat{d}_N$ is the solution of equation 84 on page 38.

Column 5 shows to five figures, the values of $\hat{c}_N$ which is the solution to equation 83 on page 38. The trend is downward but appears to level out from N = 9 800 onwards. There are many local deviations from the trend.

Column 6 shows to five figures, the values of $\hat{V}_N$ which are calculated by

$$\hat{V}_N = \frac{2}{\hat{c}_N \hat{d}_N}$$

For values of N = 1 800 words the estimate is unacceptably large. From then on the value of $\hat{V}_N$ drops, generally to a minimum of 4 778, at N = 6 600, thereafter there is a general rise until N = 9 600 whereupon $\hat{V}_N$ rises dramatically from 991 to 13 115 at N = 10 000. Thereafter the value of $\hat{V}_N$ remains in the vicinity of twelve and a half thousand, with many fluctuations.

6.4.4 Sandré's Calculation Graphs

$\hat{d}_N$, $\hat{c}_N$ and $\hat{V}_N$ were graphed on the same page, but different axes, thereby they can be easily compared.

6.4.4.1 The Graph of $\hat{d}_N$ is shown in Figure 88 on page 218. As can be seen the trend is generally downward with fluctuations on either side of the trend. The downward drift in $\hat{d}_N$ appears to be more pronounced in Sandré's case than in the case of the younger children.

6.4.4.2 The Graph of $\hat{c}_N$ is shown in Figure 88 on page 218. The graph follows a very erratic course but appears to rise to a maximum in the vicinity of N = 5 000 and then sinks to a minimum somewhere in the vicinity of N = 11 000. The curve is very erratic indeed, the value varying between 0,02 and 0,05.

6.4.4.3 The Graph of $\hat{V}_N$ appears in Figure 88 on page 218. The curve appears to start on one level and to proceed to a higher level, but with many local deviations and fluctuations on the way.

6.4.4.4 Sandré: Comparison of graphs of $\hat{c}_N$ and $\hat{V}_N$ (Fig. 88 page 218). Inspection of the two curves $\hat{c}_N$ and $\hat{V}_N$ reveals the strikingly well related trends of $\hat{c}_N$ and $\hat{V}_N$, when one increases the other decreases.

One is practically a mirror image of the other. This is not surprising since the two are connected via the equation

$$\hat{V}_N = \frac{2}{\hat{c}_N \hat{d}_N}$$

Thus, mathematically it is necessary that provided $\hat{d}_N$ remains constant, $\hat{V}_N$ is inversely proportional to $\hat{c}_N$.

$\hat{d}_N$ is not constant as can be seen by the graph. The trend is gradually and continuously downward, and shows no large sudden changes.

The value of $\hat{d}_N$ varies between 0,0074 and 0,010, as compared with the sharp increases and decreases in $\hat{c}_N$. Thus changes in $\hat{d}_N$ cannot have a very marked, i.e. obvious effect on $\hat{V}_N$.

6.4.5 $\hat{V}_N$: Dependence on Hapax Legomena

Since the estimates of $\hat{V}_N$ were based on $\bar{f}_1(1|N)$ and $\bar{r}_N$, any sudden change in $\bar{f}_1(1|N)$ can be expected to have an effect on $\bar{r}_N$ and also on $\hat{V}_N$.

Inspection of the data tables shows that, in Sandré's case, $\bar{f}_1(1|N)$ made up almost half the types, so that any change in $\bar{f}_1(1|N)$ would have a marked effect on $\bar{r}_N$.

The lowest ratio $\bar{f}_1(1|N)$: $\bar{r}_N$ was recorded at the end of Bronwyn 1 where the ratio was 0,108.

To illustrate the effect of variations in $\bar{f}_1(1|N)$ and $\bar{r}_N$ and $\hat{V}_N$, which variations were due to the listing, a table was compiled showing the effects of the listing. The table has already been mentioned in 5.14 but is discussed in more detail here.

6.4.6 Table of numbers in Sandré's lists - Table 32 page 219

Column 1 shows 200 word intervals from N = 6 401 onwards. Listing began in the interval N = 6 401 to N = 6 600 and continued intermittently up to the interval N = 12 201 to N = 12 400. For completeness all 200 word intervals from N = 6 401 are accounted for whether listing took place or not.

Column 2 shows the topic or topics under which the listed words were classified.

Column 3 shows, under the heading "No. of words in List", the number of words listed, without any conjunctions or prepositions in between.

Column 4 shows, under the heading, "increase in $\hat{f}_1(1|N)$, $\Delta\hat{f}_1(1|N)$" the increase in the value of Hapax Legomena. In some cases $\Delta\hat{f}_1(1|N) <$ No. of words in list. This is not surprising because $\Delta\hat{f}_1(1|N)$ is the difference between the number of first appearances in the interval and the number of words already used once which are repeated during that interval. Thus, although many words may be listed, if they or other words used once already, are repeated, the increase in $\hat{f}_1(1|N)$, $\Delta\hat{f}_1(1|N)$ may be less than the number of words in the list. During three of the intervals, where no listing occurred, $\hat{f}_1(1|N)$ actually decreased.

Column 5, under the heading "Increase in $\bar{r}_N$, $\Delta \frac{N}{\bar{n}_N}$" shows the change in the average number of times the words occur, to two figures, occurring during the interval. As can be seen, the value is negative in six instances.

Column 6, under the heading "Increase in $\hat{V}_N$, $\Delta\hat{v}_N$", shows by how much the estimate of $\hat{V}_{..}$ alters. In 19 instances an increase was recorded and in 14 a decrease.

6.5 Graph and Correlation of $\Delta\hat{r}_N$ with $\Delta\hat{f}_1(1|N)$ and $\Delta\hat{v}_N$ with $\Delta\hat{f}_1(1|N)$

Figure 89 on page 220 shows the scattergram of $\bar{r}_N$ vs $\hat{f}_1(1|N)$. The scattergram shows good grouping about a line with negative slope. A product/moment correlation coefficient of r = $^-$0,87 was obtained.

Figure 90 on page 221 shows the scattergram of $\Delta\hat{v}_N$ vs $\Delta\hat{f}_1(1|N)$. The grouping is good and a product/moment correlation coefficient r = 0,89 was obtained.

The scattergrams and correlation coefficients show quite clearly the heavy dependence of the estimates on Hapax Legomena.

6.5.1 No Lists $\hat{V}_{..}$ decreases

Conveniently, between N = 8 000 and N = 9 000 Sandré made no lists and the estimate of $\hat{V}_N$ shows a steady, almost linear decrease. Between N = 10 000 and N = 10 400 no lists were made, and although a small increase in $\hat{f}_1(1|N)$ appeared, $\hat{V}_N$ shows a steady decrease. From N = 11 800 onward only one list of 17 words appears and the graph shows a continuous downward trend except between N = 12 200 and N = 12 400 where the list of 17 words appears.

The effect of listing on $\hat{V}_{..}$ though dramatic when it occurs, is thus seen to be temporary and with large enough N should disappear.

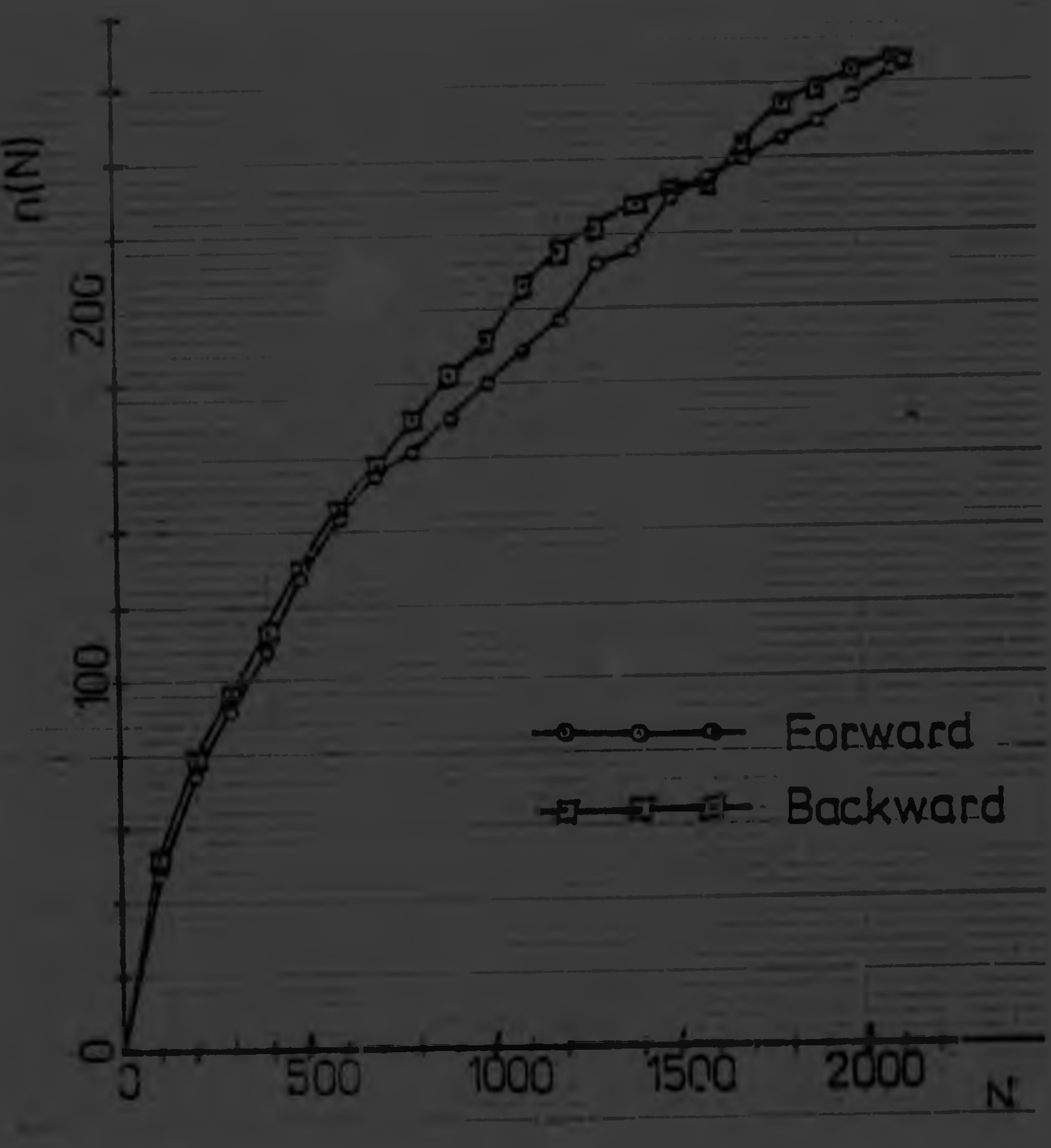


FIGURE 78 KERRY 1 n_N vs N FORWARD/BACKWARD HYSTERESIS

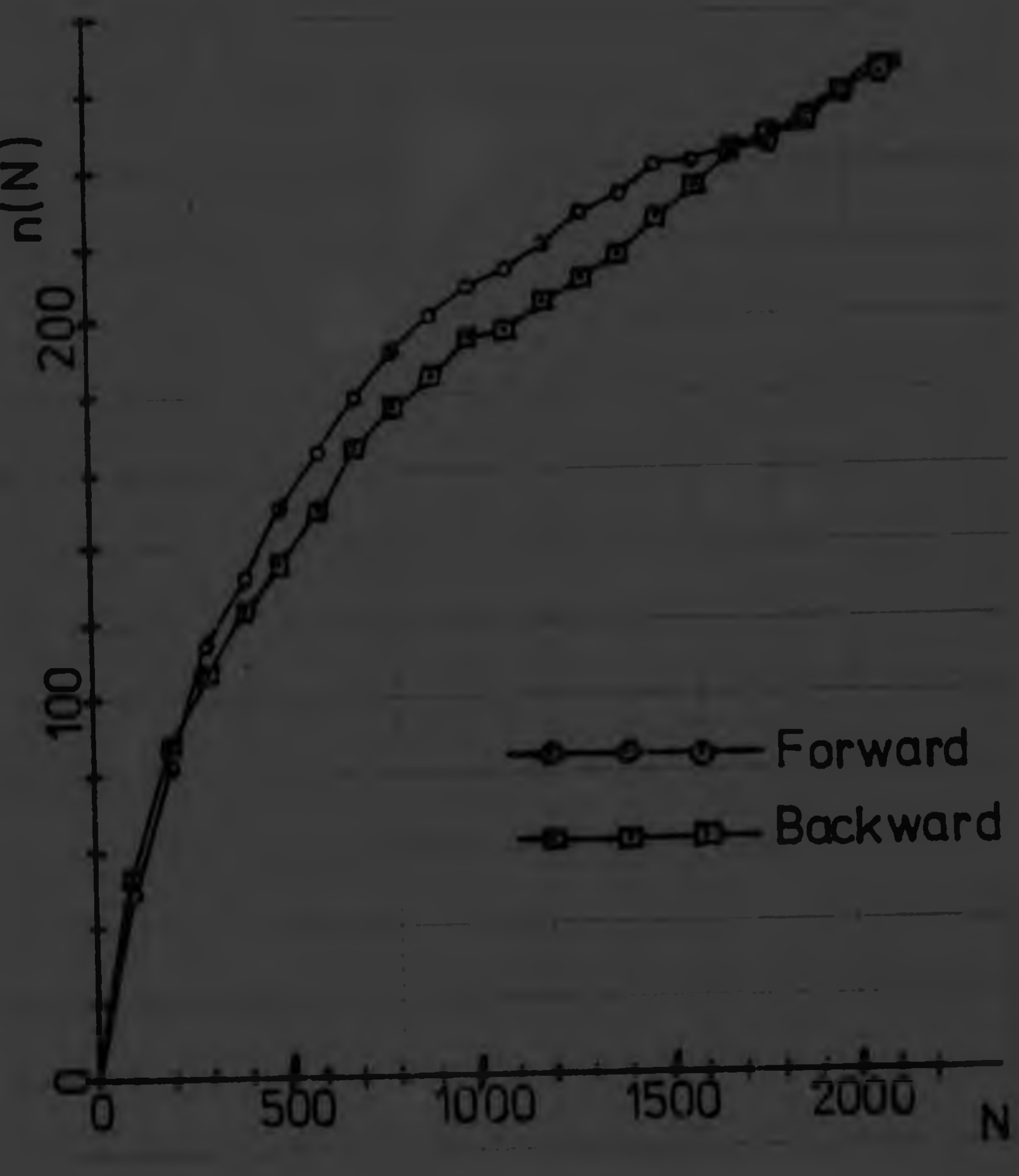


FIGURE 79 KERRY 1 $\bar{n}_N$ vs N RANDOM FORWARD/BACKWARD HYSTERESIS

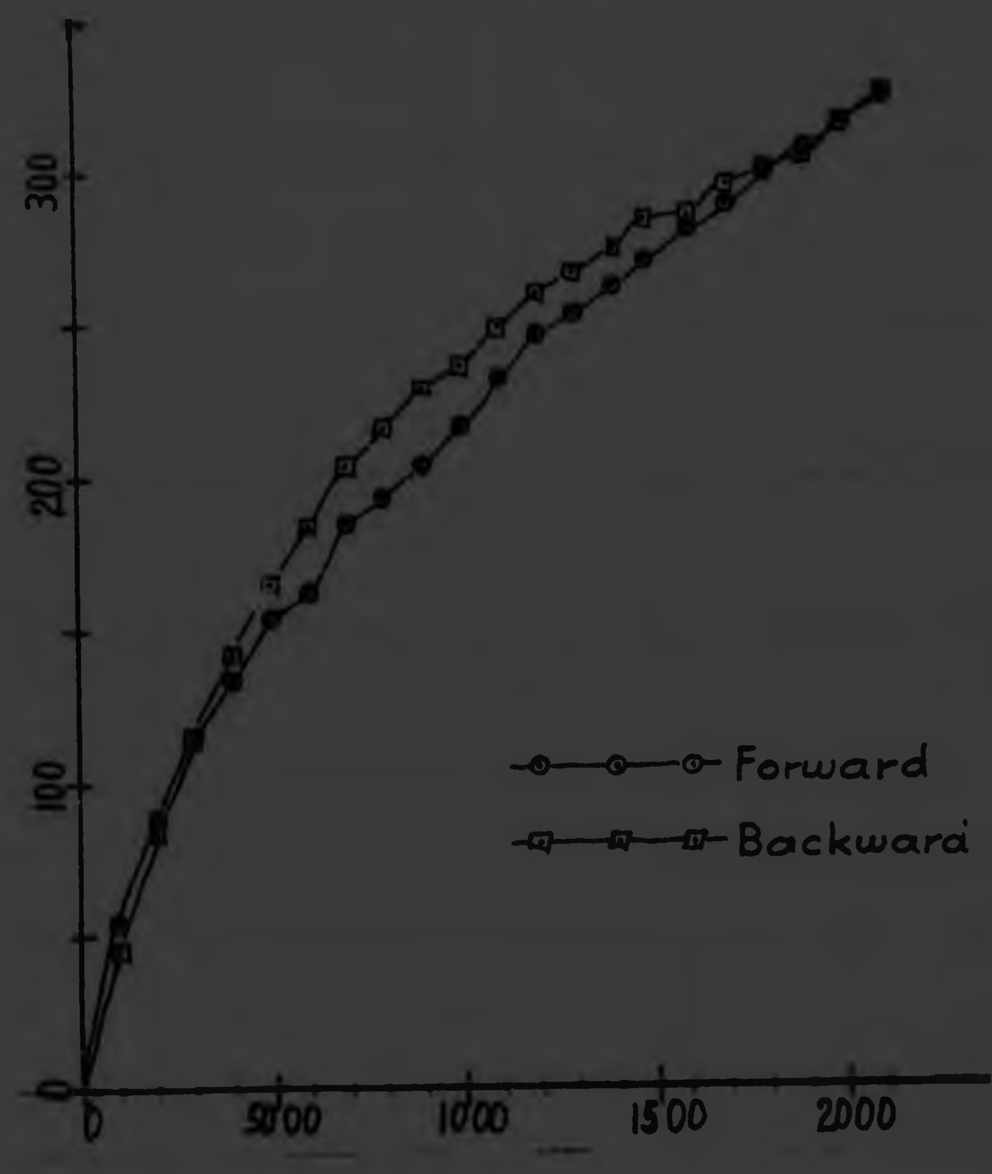


FIGURE 80 KERRY 2 θ_N vs N FORWARD/BACKWARD HYSTERESIS

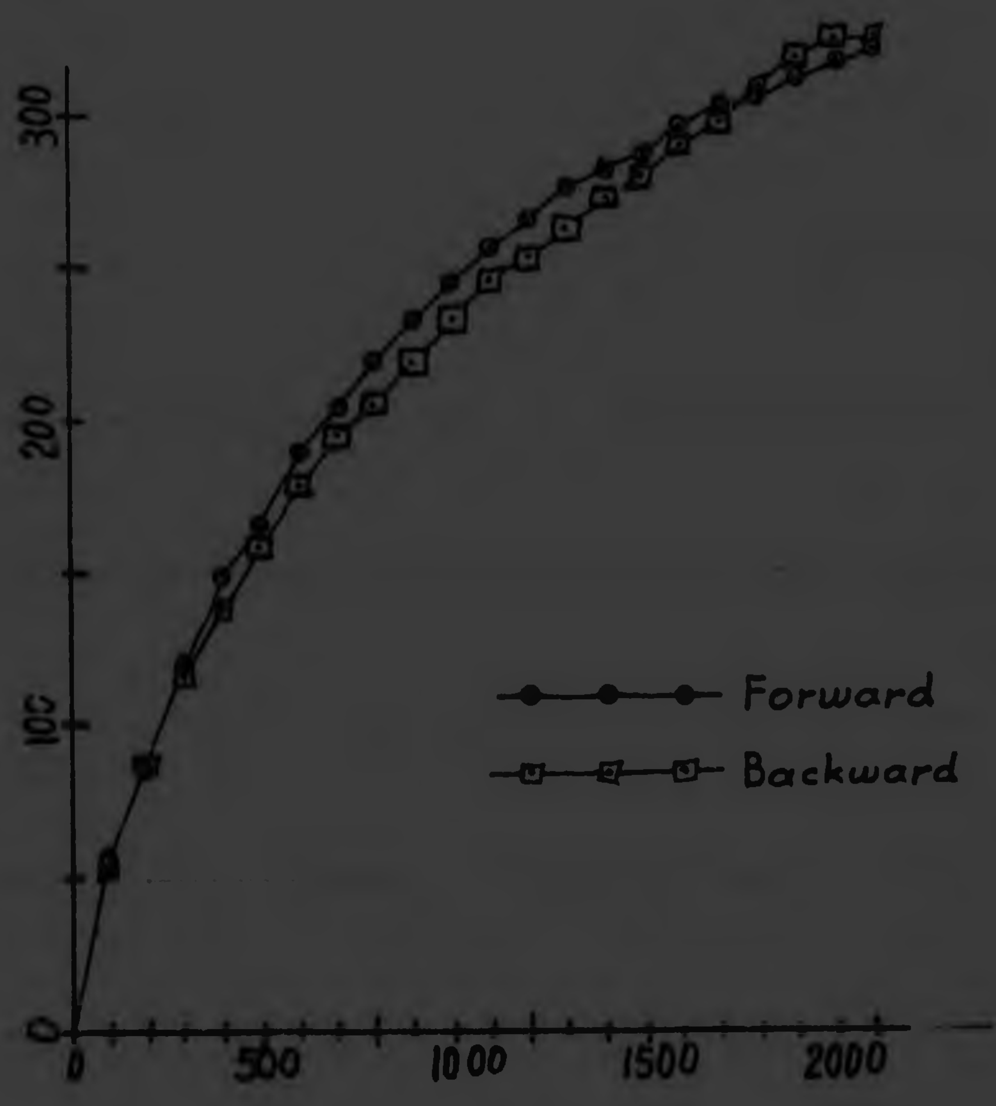


FIGURE 81 KEF¯Y 2 $\bar{n}_N$ vs N RANDOM FORWARD/BACKWARD HYSTERESIS

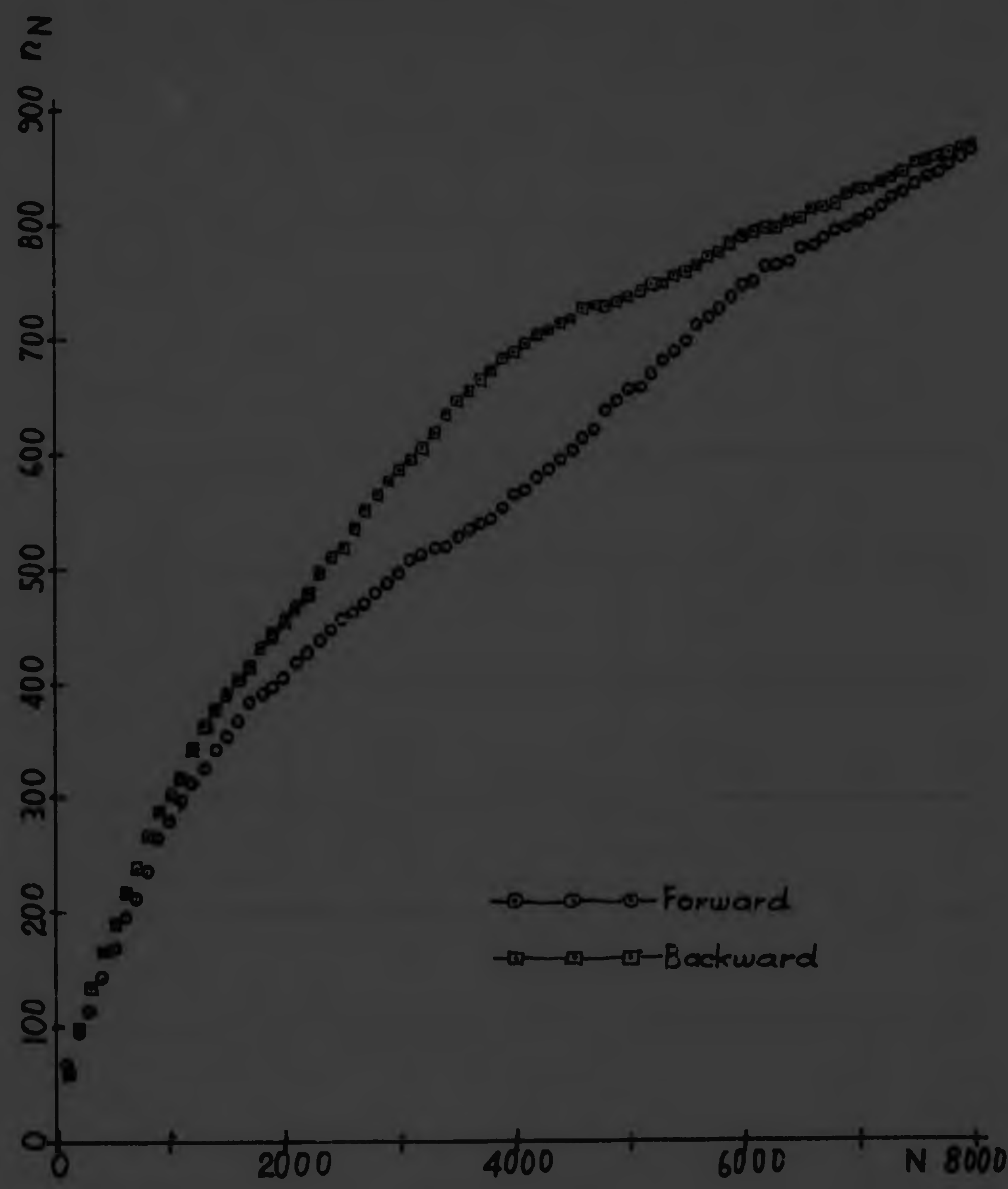


FIGURE 82 BRONWYN 1 FORWARD/BACKWARD HYSTERESIS

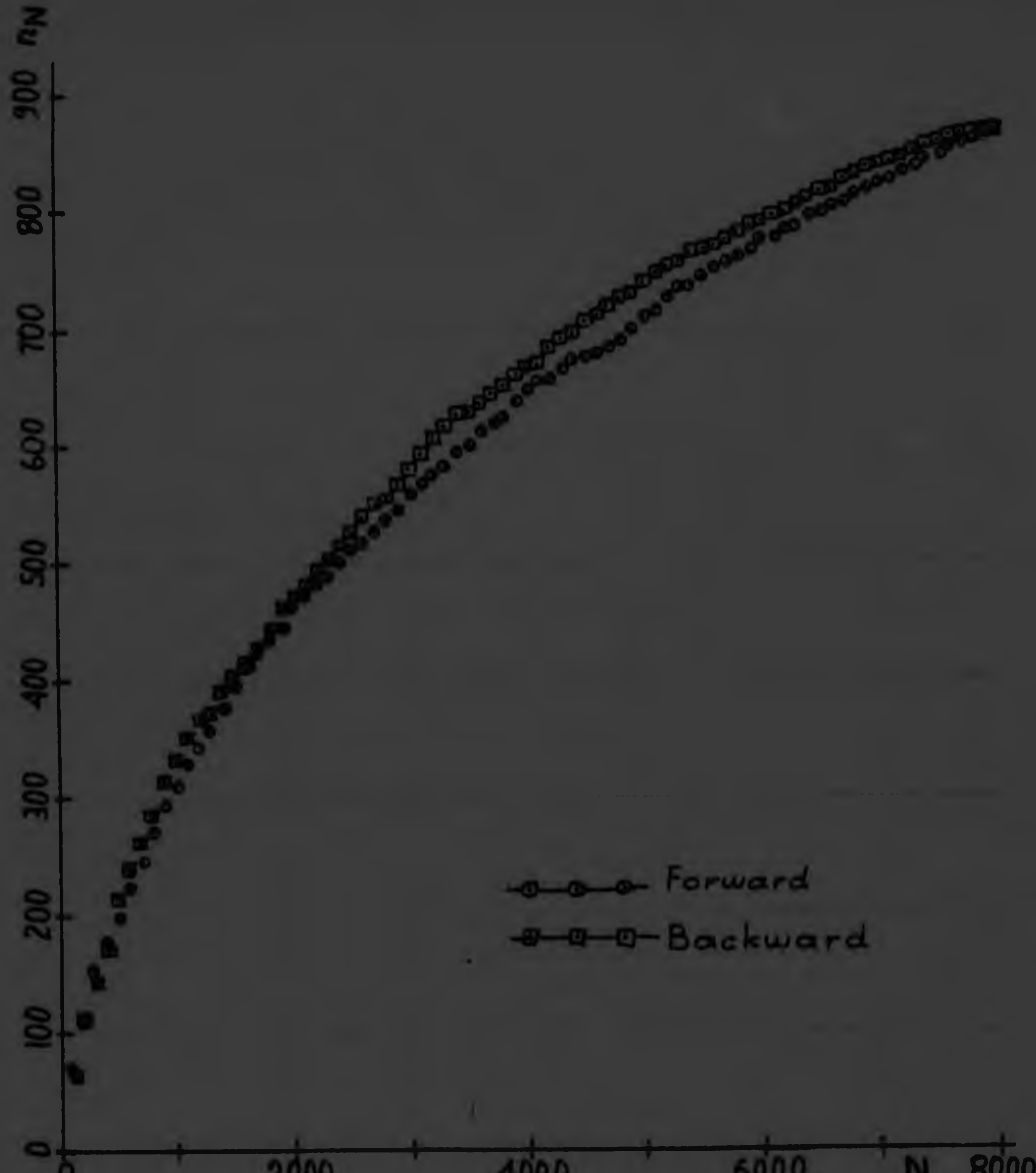


FIGURE 83 BRONWYN 1 RANDOM $\bar{n}_N$ vs N FORWARD/BACKWARD HYSTERESIS

TABLE 3[illegible] SANDRé FORWARD

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$T_{\hat{n}}$ $\frac{N}{\hat{n}_N}$
200	120	90	16	4	0,178	0,250	0,750	0,133	0,033	1,667
400	200	141	34	7	0,241	0,206	0,705	0,170	0,035	2,000
600	272	180	46	23	0,256	0,500	0,662	0,169	0,085	2,222
800	332	214	56	20	0,262	0,357	0,645	0,169	0,060	2,416
1000	395	248	69	24	0,278	0,348	0,628	0,175	0,061	2,532
1200	455	287	78	28	0,272	0,359	0,631	0,[illegible]	0,062	2,637
1400	520	328	90	30	0,274	0,333	0,631	0,173	0,058	2,692
1600	565	342	105	38	0,307	0,362	0,605	0,186	0,067	2,832
1800	604	357	110	48	0,308	0,345	0,591	0,182	0,079	2,980
2000	639	362	127	43	0,351	0,339	0,567	0,199	0,067	3,130
2200	680	382	129	53	0,338	0,411	0,562	0,190	0,078	3,235
2400	727	411	140	55	0,341	0,393	0,565	0,193	0,076	3,170
2600	769	425	150	62	0,353	0,413	0,553	0,195	0,081	3,381
2800	807	438	158	66	0,361	0,418	0,543	0,196	0,082	3,470
3000	833	445	160	65	0,360	0,406	0,534	0,192	0,078	3,601
3200	871	466	162	69	0,348	0,426	0,535	0,186	0,079	3,674
3400	904	483	161	75	0,335	0,466	0,534	0,178	0,083	3,761
3600	931	491	167	82	0,340	0,491	0,527	0,179	0,088	3,867
3800	967	510	171	82	0,335	0,480	0,527	0,177	0,085	3,930
4000	1007	529	182	80	0,344	0,440	0,525	0,181	0,079	3,971
4200	1044	544	190	88	0,349	0,463	0,521	0,182	0,084	4,023
4400	1072	549	202	88	0,368	0,436	0,512	0,180	0,082	4,104
4600	1100	561	196	100	0,349	0,510	0,510	0,178	0,091	4,182
4800	1124	571	200	99	0,350	0,495	0,508	0,178	0,083	4,270

TABLE 10 SANDRé FORWARD (CONTD.)

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	T_N $\frac{N}{\hat{n}_N}$
5000	1144	578	201	107	0,349	0,532	0,505	0,176	0,094	4,371
5200	1172	592	203	106	0,342	0,522	0,506	0,173	0,090	4,437
5400	1188	596	207	106	0,347	0,512	0,502	0,174	0,089	4,545
5600	1225	620	208	112	0,335	0,538	0,506	0,170	0,091	4,571
5800	1239	621	201	119	0,324	0,592	0,501	0,162	0,096	4,681
6000	1261	629	201	126	0,320	0,627	0,493	0,159	0,100	4,758
6200	1275	628	210	118	0,334	0,562	0,493	0,165	0,093	4,865
6400	1289	625	217	120	0,347	0,553	0,485	0,168	0,093	4,965
6600	1311	629	219	130	0,350	0,594	0,480	0,167	0,099	5,034
6800	1359	662	224	133	0,369	0,594	0,487	0,165	0,098	5,004
7000	1380	672	227	134	0,338	0,590	0,483	0,164	0,097	5,072
7200	1423	702	234	136	0,333	0,581	0,493	0,164	0,096	5,060
7400	1447	715	232	144	0,324	0,621	0,494	0,160	0,100	5,114
7600	1486	741	236	147	0,318	0,623	0,499	0,159	0,099	5,114
7800	1507	754	237	142	0,314	0,599	0,500	0,157	0,094	5,176
8000	1532	773	235	144	0,304	0,613	0,505	0,155	0,094	5,222
8200	1549	778	240	141	0,308	0,598	0,502	0,155	0,091	5,294
8400	1575	786	251	141	0,319	0,562	0,499	0,159	0,090	5,333
8600	1598	793	259	144	0,327	0,556	0,496	0,162	0,090	5,382
8800	1614	798	266	142	0,333	0,534	0,494	0,165	0,088	5,452
9000	1626	793	276	139	0,348	0,504	0,488	0,170	0,085	5,535
9200	1680	820	289	146	0,352	0,505	0,488	0,172	0,097	5,476
9400	1740	872	289	149	0,331	0,516	0,501	0,166	0,086	5,402

TABLE 30 SANDRé FORWARD (CONTD.)

N	$\hat{n}_N$	$\hat{f}_1(1\|N)$	$\hat{f}_1(2\|N)$	$\hat{f}_1(3\|N)$	$\frac{\hat{f}_1(2\|N)}{\hat{f}_1(1\|N)}$	$\frac{\hat{f}_1(3\|N)}{\hat{f}_1(2\|N)}$	$\hat{\phi}(1\|N)$ $\frac{\hat{f}_1(1\|N)}{\hat{n}_N}$	$\hat{\phi}(2\|N)$ $\frac{\hat{f}_1(2\|N)}{\hat{n}_N}$	$\hat{\phi}(3\|N)$ $\frac{\hat{f}_1(3\|N)}{\hat{n}_N}$	$\bar{r}_N$ $\frac{N}{\hat{n}_N}$
9600	1777	895	298	144	0,333	0,498	0,504	0,168	0,081	5,402
9800	1833	941	303	146	0,322	0,482	0,513	0,165	0,080	5,346
10000	1857	955	301	153	0,315	0,508	0,514	0,162	0,082	5,385
10200	1882	963	310	155	0,322	0,500	0,519	0,165	0,082	5,420
10400	1900	965	318	156	0,330	0,487	0,508	0,167	0,082	5,474
10600	1927	985	324	153	0,331	0,472	0,509	0,168	0,079	5,501
10800	1946	989	327	153	0,331	0,468	0,513	0,168	0,079	5,550
11000	1975	1003	333	157	0,332	0,471	0,508	0,169	0,079	5,570
11200	1995	1014	335	159	0,330	0,475	0,508	0,168	0,080	5,614
11400	2002	1011	338	160	0,334	0,473	0,505	0,169	0,080	5,694
11600	2022	1022	342	160	0,335	0,468	0,505	0,169	0,079	5,737
11800	2060	1047	347	166	0,331	0,478	0,508	0,168	0,081	5,728
12000	2069	1047	342	175	0,327	0,512	0,506	0,165	0,085	5,800
12200	2089	1051	346	178	0,329	0,514	0,503	0,166	0,085	5,840
12400	2118	1066	351	181	0,329	0,516	0,503	0,166	0,085	5,855
12600	2139	1074	358	182	0,333	0,508	0,502	0,167	0,085	5,891
12800	2151	1072	358	189	0,334	0,528	0,498	0,166	0,088	5,951
13000	2161	1075	355	193	0,330	0,544	0,497	0,164	0,089	6,016

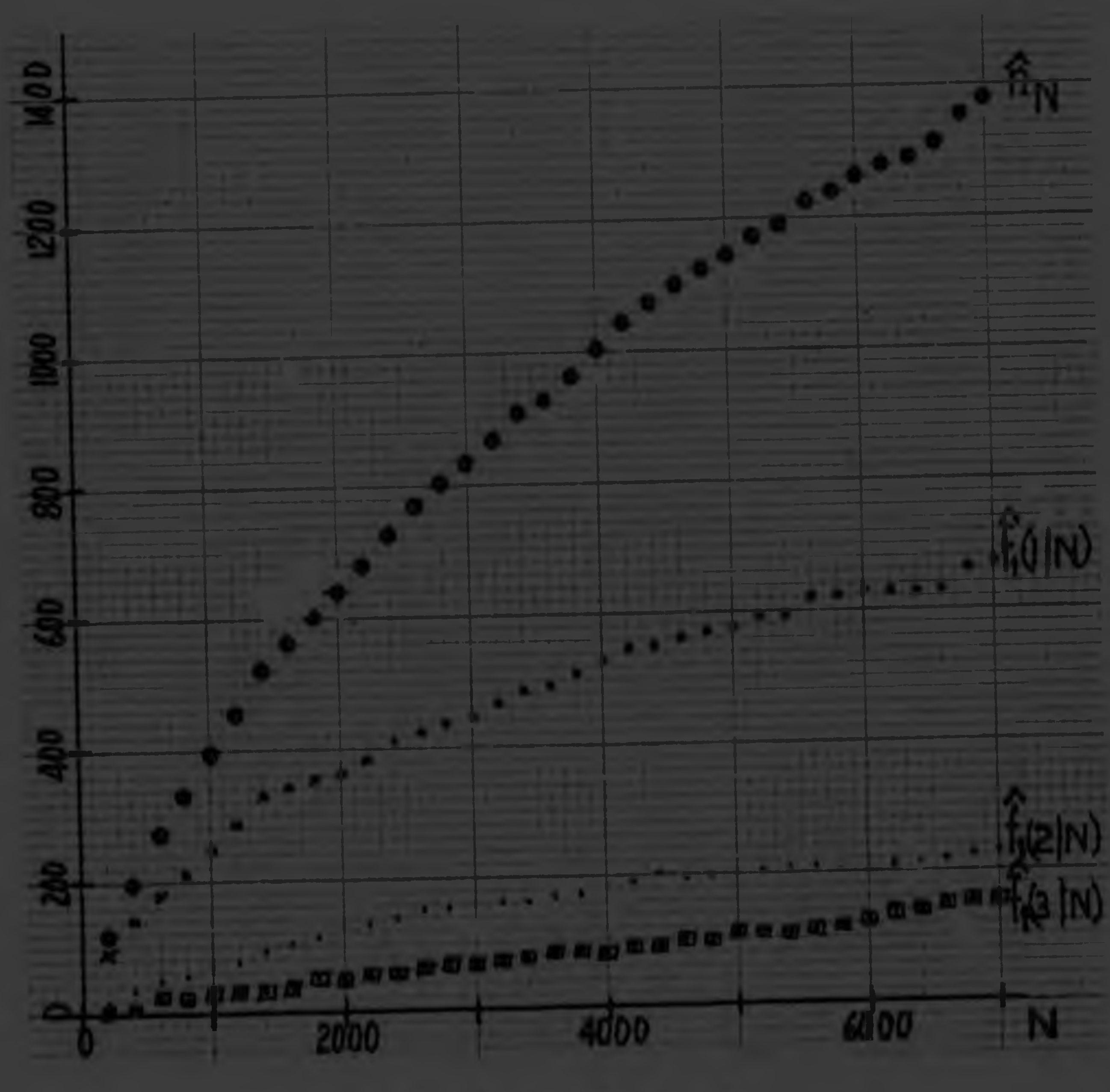


FIGURE 84 SANDRé $\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{f}_1(2|N)$, $\hat{f}_1(3|N)$ vs N

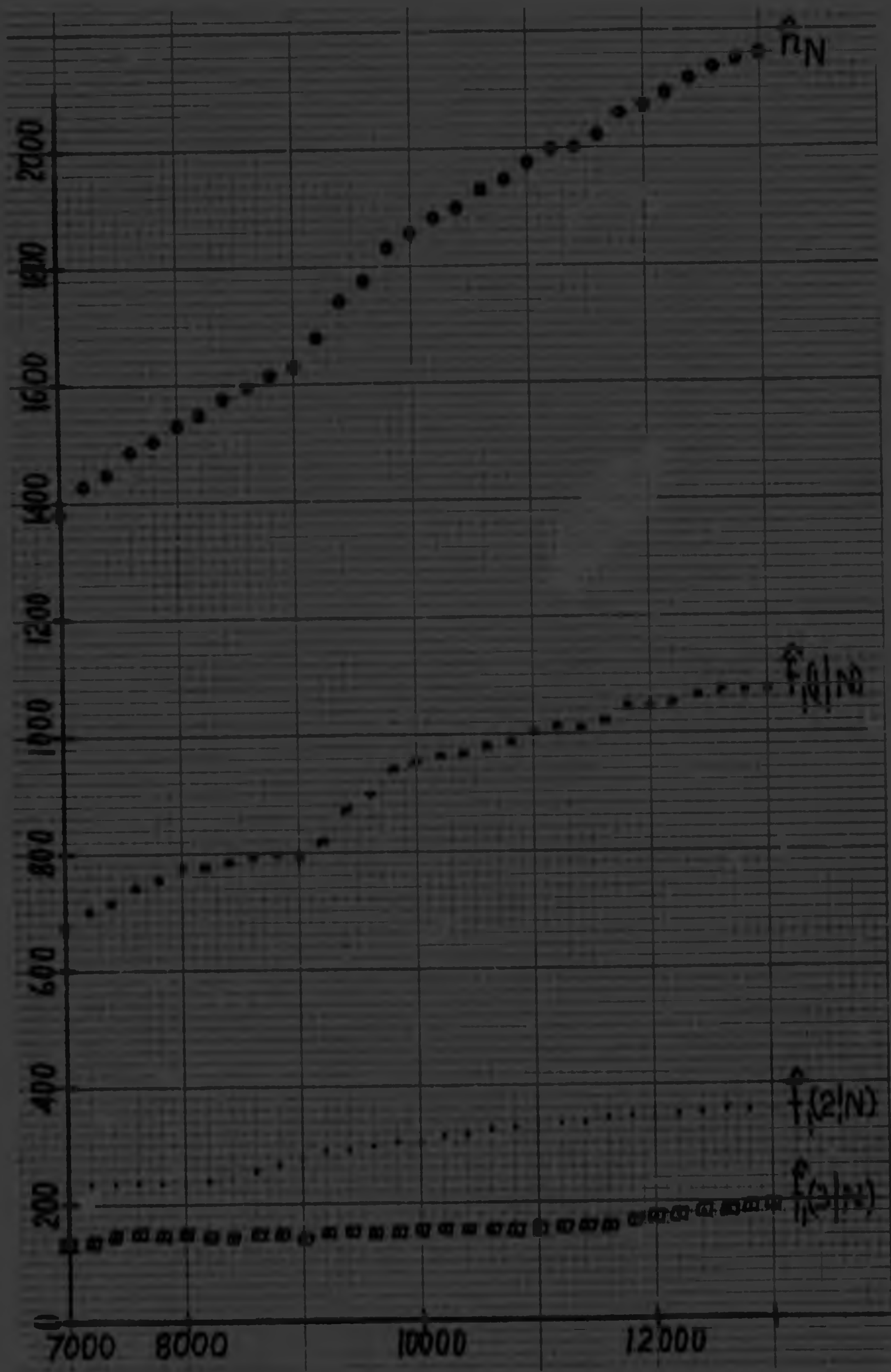


FIGURE 84 contd.

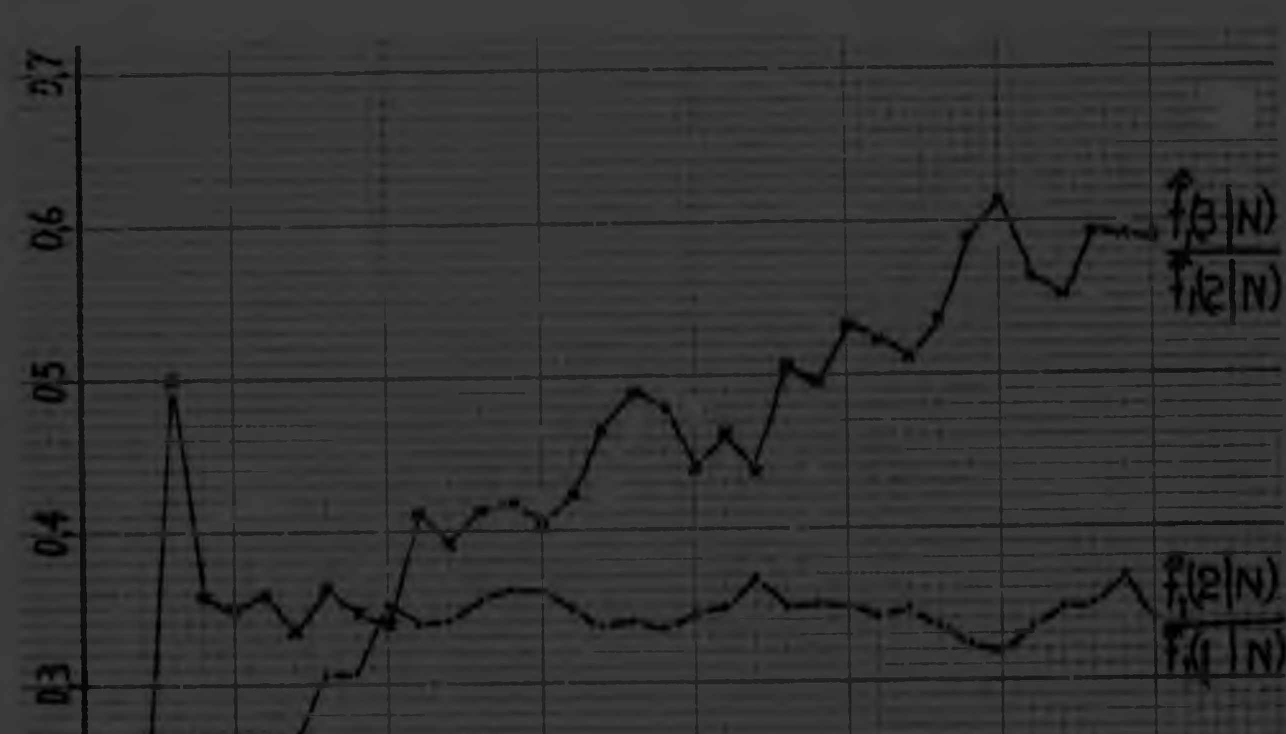


FIGURE 85 SANDRE $\frac{\hat{f}_1(2|N)}{\hat{f}_1(1|N)}$, $\frac{\hat{f}_1(3|N)}{\hat{f}_1(2|N)}$ vs N

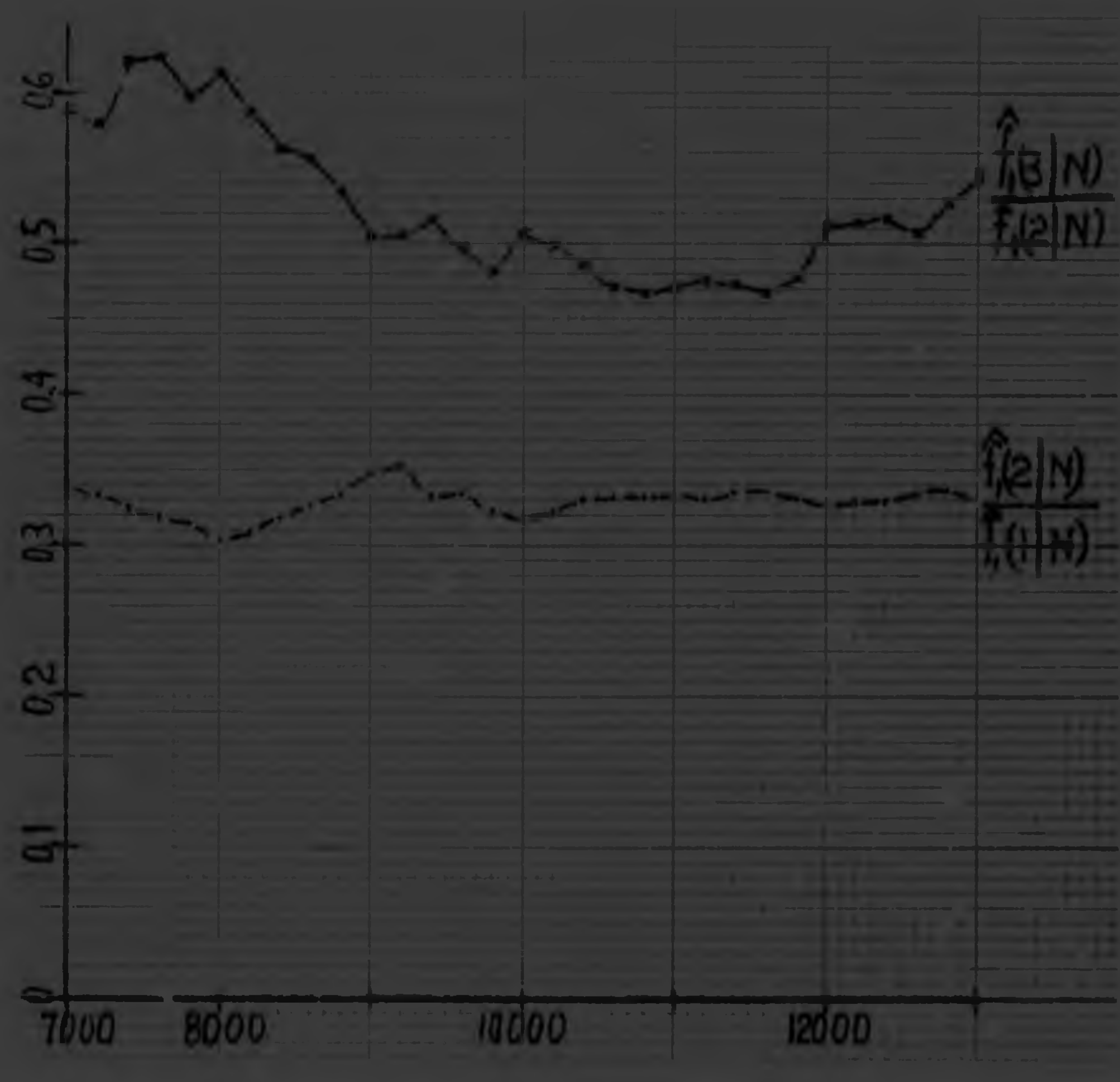


FIGURE 8a contd.

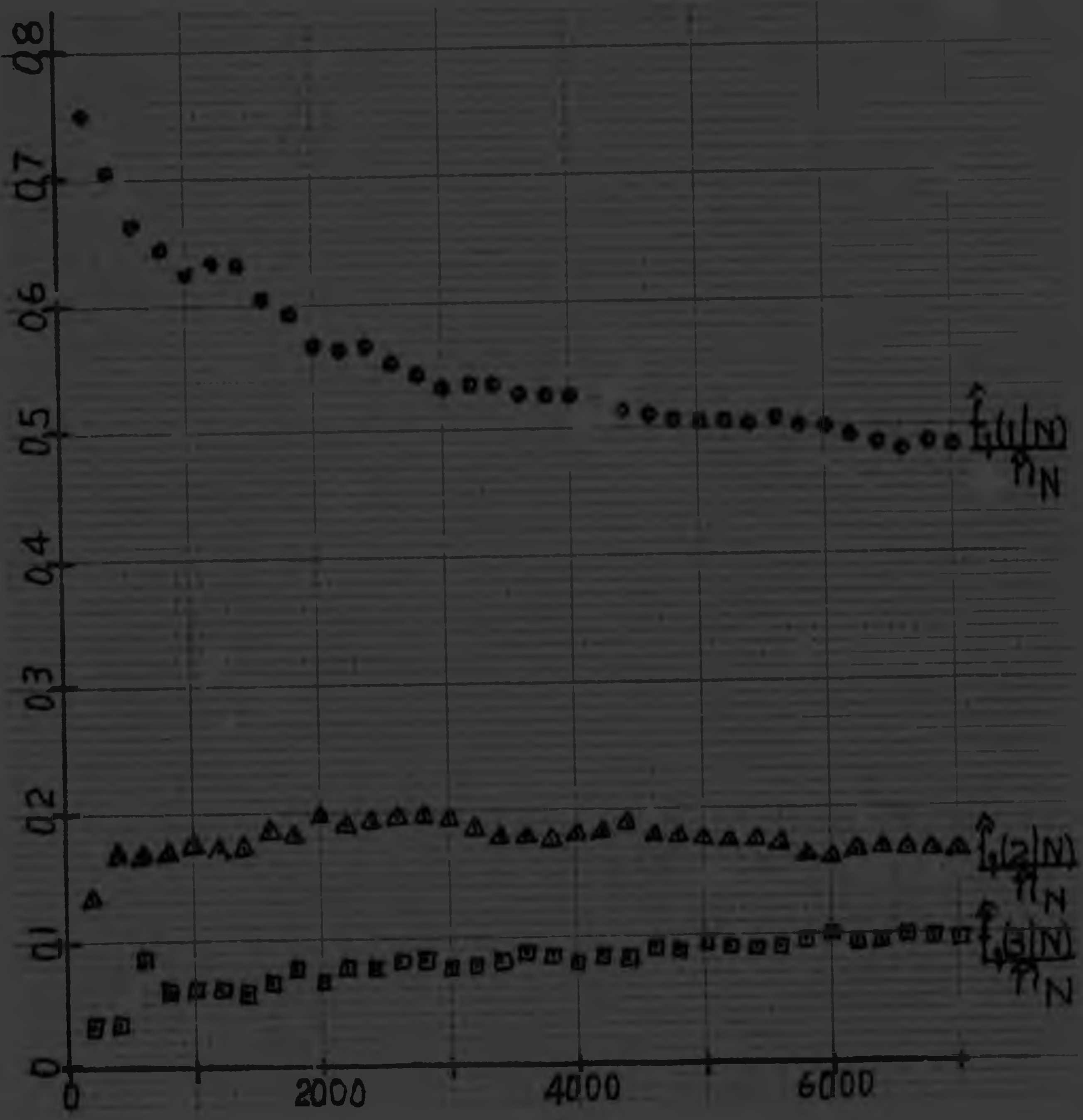


FIGURE 86 SANDRé $\frac{\hat{f}_1(1|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(2|N)}{\hat{n}_N}$, $\frac{\hat{f}_1(3|N)}{\hat{n}_N}$ vs N

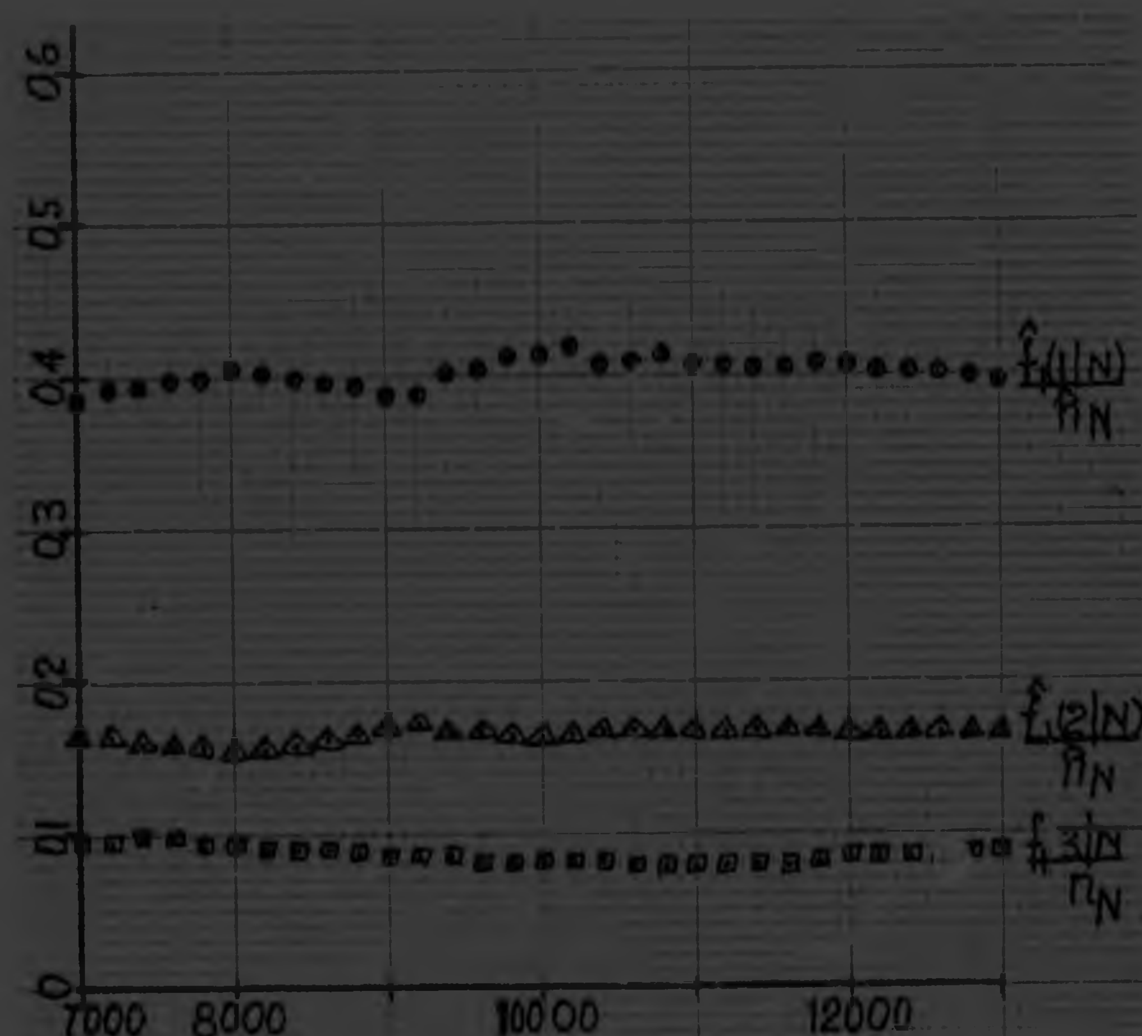


FIGURE 86 contd.

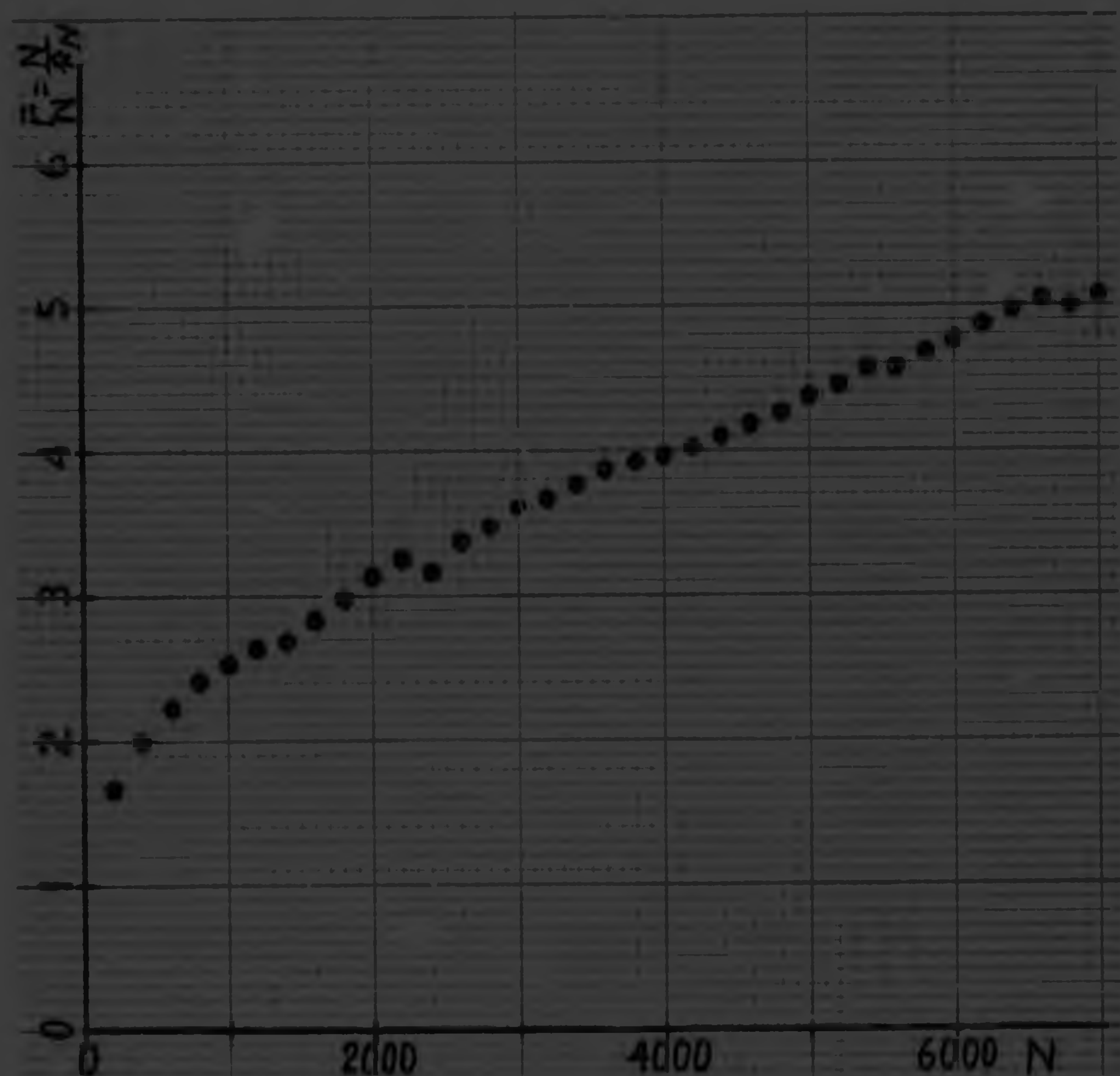


FIGURE 87 SANDRé $\vec{r}_N = \frac{N}{\bar{a}_N}$ vs N

FIGURE 87 contd.

TABLE 31 SANDRé (CALCULATIONS)

N	$\bar{\theta}_N$	$\bar{\tau}_N$	$\bar{d}_N$	$\bar{c}_N$	$\bar{v}_N$
200	0.79750	0.00002	0.019692	0.000007	∞
400	0.87574	0.00002	0.017619	0.000007	∞
600	0.90998	0.00005	0.016849	0.000014	∞
800	0.92842	0.00012	0.016214	0.000033	∞
1000	0.93849	0.00014	0.015258	0.000034	∞
1200	0.94279	0.00004	0.013733	0.000008	∞
1400	0.94510	0.00010	0.012296	0.000023	∞
1600	0.95355	0.01053	0.012830	0.002270	68684
1800	0.95729	0.05195	0.012453	0.010735	14961
2000	0.95829	0.15176	0.011488	0.030993	5617
2200	0.96129	0.15542	0.011289	0.030576	5794
2400	0.96415	0.12397	0.011205	0.023472	7604
2600	0.96438	0.17706	0.010414	0.033415	5747
2800	0.96527	0.21504	0.009926	0.040075	5028
3000	0.96741	0.23963	0.009896	0.043258	4672
3200	0.96932	0.22385	0.009874	0.039208	5166
3400	0.97117	0.21477	0.009908	0.036465	5535
3600	0.97238	0.23688	0.009781	0.039364	5194
3800	0.97358	0.22861	0.009699	0.037155	5550
4000	0.97409	0.23430	0.009398	0.037717	5642
4200	0.97445	0.25018	0.009082	0.039987	5507
4400	0.97472	0.28822	0.008764	0.045825	4980
4600	0.97572	0.29026	0.008736	0.045228	5062
4800	0.97683	0.29086	0.008783	0.044275	5143
5000	0.97774	0.30398	0.008784	0.045354	5020
5200	0.97887	0.283[illegible]	0.008908	0.041284	5439
5400	0.97976	0.29[illegible]03	0.00[illegible]963	0.042120	5298
5600	0.98047	0.[illegible]065	0.008967	0.037820	5898
5800	0.98121	0.28629	0.009004	0.039241	5660
6000	0.98178	0.29215	0.008982	0.039433	5647
6200	0.98225	0.31623	0.008927	0.042128	5318
6400	0.98256	0.3478[illegible]	0.008801	0.045936	4947
6600	0.98275	0.36924	0.008631	0.048498	4778
6800	0.98308	0.33321	0.008544	0.043344	5401
7000	0.98365	0.32852	0.008595	0.042005	5539
7200	0.98403	0.29692	0.008558	0.037521	6228
7400	0.98452	0.28892	0.008593	0.035950	6474
7600	0.98484	0.26602	0.008551	0.032747	7142
7800	0.98541	0.25333	0.008661	0.030596	7548
8000	0.98602	0.22888	0.008816	0.027062	8383
8200	0.98634	0.23631	0.008805	0.027621	8224
8400	0.98639	0.24978	0.008628	0.029140	7955
8600	0.98652	0.260[illegible]3	0.008509	0.030284	7761
8800	0.98684	0.26559	0.008522	0.030467	7703
9000	0.98692	0.29438	0.008385	0.033664	7085
9200	0.98659	0.29585	0.008000	0.034254	7299
9400	0.98695	0.23523	0.008044	0.026873	9251
9600	0.98710	0.22289	0.007970	0.025316	9912
9800	0.98734	0.17829	0.007962	0.020056	12525
10000	0.98762	0.17180	0.007978	0.019114	13115
10200	0.98767	0.18239	0.007854	0.020250	12574
10400	0.98776	0.19785	0.007759	0.021891	11776
10600	0.98795	0.19296	0.007732	0.021185	12209

TABLE 31 SANDRé (CONTD.)(CALCULATIONS)

N	$\bar{\theta}_N$	$\bar{\tau}_N$	$\hat{d}_N$	$\bar{c}_N$	$\hat{V}_N$
10800	0.98819	0.19230	0.007745	0.020902	12354
11000	0.98827	0.19319	0.007658	0.020925	12481
11200	0.98852	0.18850	0.007689	0.020195	12879
11400	0.98875	0.20036	0.007707	0.021254	12209
11600	0.98897	0.19633	0.007730	0.020619	12548
11800	0.98907	0.18330	0.007569	0.019163	13609
12000	0.98929	0.19040	0.007699	0.019703	13185
12200	0.98935	0.20241	0.007601	0.020907	12586
12400	0.98940	0.20076	0.007531	0.020664	12851
12600	0.98950	0.20486	0.007483	0.020987	12735
12800	0.98959	0.21996	0.007424	0.022446	12001
13000	0.98981	0.22136	0.007477	0.022338	11973

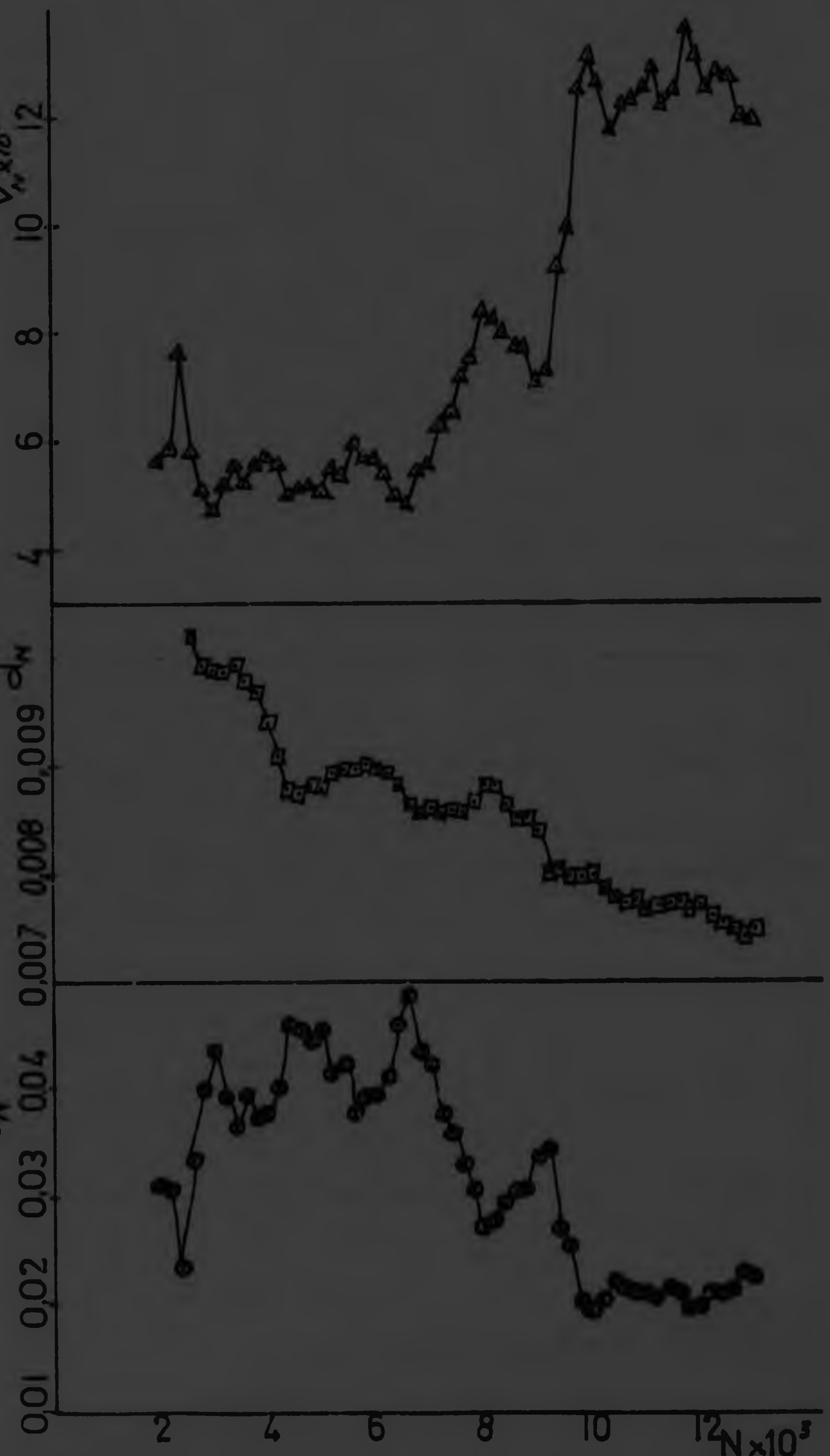


FIGURE 88 SANDRE $\hat{c}_N$, $\hat{d}_N$ AND $\hat{V}_N$ vs N

TABLE 32 SANDRE. LISTING

WORD INTERVAL	TOPIC	NO. OF WORDS IN LIST	INCREASE IN $f_1(1\|N)$, $\Delta f_1(1\|N)$	INCREASE IN $\bar{r}_N$, $\Delta\left(\frac{N}{\bar{n}_N}\right)$	INCREASE IN $\hat{V}_N$, $\Delta \hat{V}_N$
6401 to 6600	Flowers	4	24	0,069	-169
6601 to 6800	Animals	18			
	Insects	12	23	-0,030	623
6801 to 7000	No Lists		10	0,068	138
7001 to 7200	Dogs	19	30	-0,012	689
7201 to 7400	No Lists		13	0,054	246
7401 to 7600	Jam Flavours	8	26	0,000	668
7601 to 7800	No Lists		13	0,062	406
7801 to 8000	Languages	7	19	0,046	835
8001 to 8200	No Lists		5	0,072	-159
8201 to 8400	No Lists		8	0,039	-269
8401 to 8600	No Lists		7	0,049	-194
8601 to 8800	No Lists		5	0,070	- 58
8801 to 9000	No Lists		-5	0,083	-618
9001 to 9200	Kitchen	43	27	-0,059	214
9201 to 9400	Labour Savers	7			
	Medicines	12			
	Travel:Countries	42	52	-0,074	1952
9401 to 9600	Travel:Sights	19	23	0,005	661
9601 to 9800	Travel: Aero-planes	46	46	-0,056	2613
9801 to 10000	No Lists		14	0,039	590
10001 to 10200	No Lists		8	0.035	-541
10201 to 10400	No Lists		2	0,054	-798
10401 to 10600	School Subjects	18	15	0,027	433
10601 to 10800	No Lists		9	0,049	145
10801 to 11000	School Sports	19	14	0,020	127
11001 to 11200	No Lists		11	0,044	398
11201 to 11400	No Lists		-3	0,080	-670
11401 to 11600	No Lists		11	0,043	339
11601 to 11800	Cars	23	25	-0,009	1061
11801 to 12000	No Lists		0	0,072	-424
12001 to 12200	No Lists		4	0,040	-599
12201 to 12400	Fishes	17	15	0,015	265
12401 to 12600	No Lists		8	0,046	-116
12601 to 12800	No Lists		-2	0,060	-734
12801 to 13000	No Lists		3	0,065	- 28

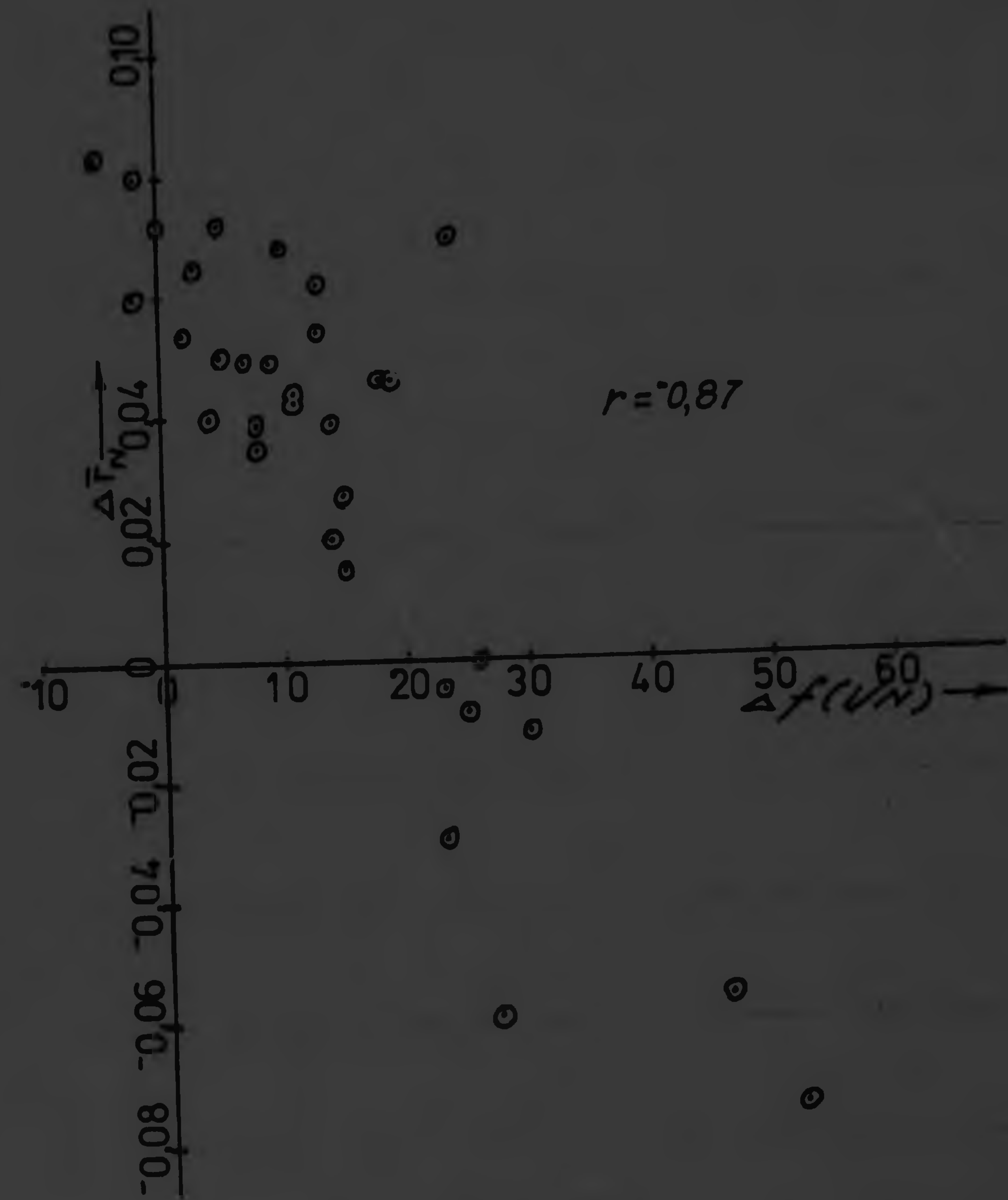


FIGURE 89 SANDRÉ $\Delta\bar{r}_N$ vs $\Delta\bar{f}_1(1|N)$

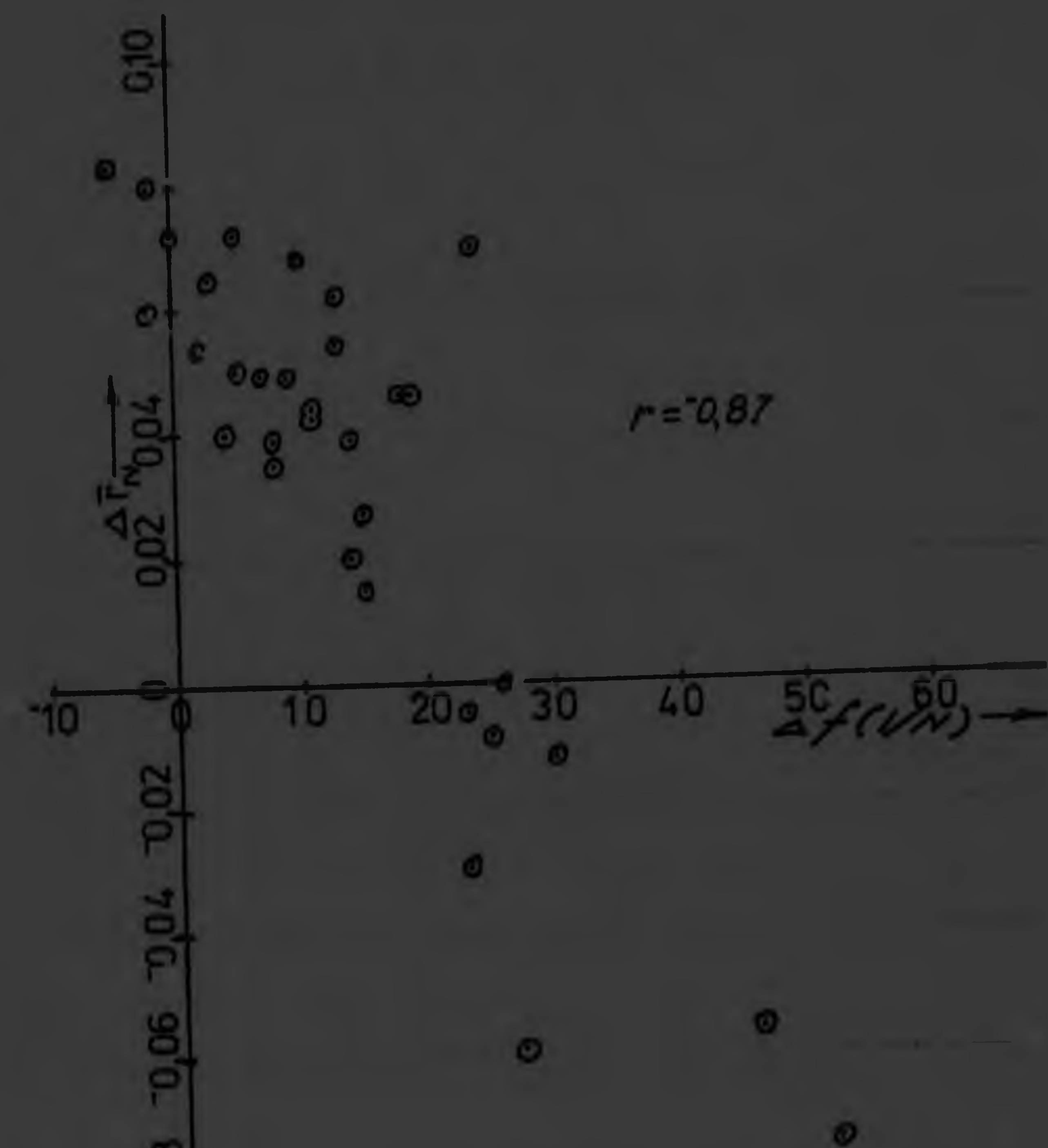


FIGURE 89 SANDRé $\Delta\bar{r}_N$ vs $\Delta\bar{f}_1$(1|8

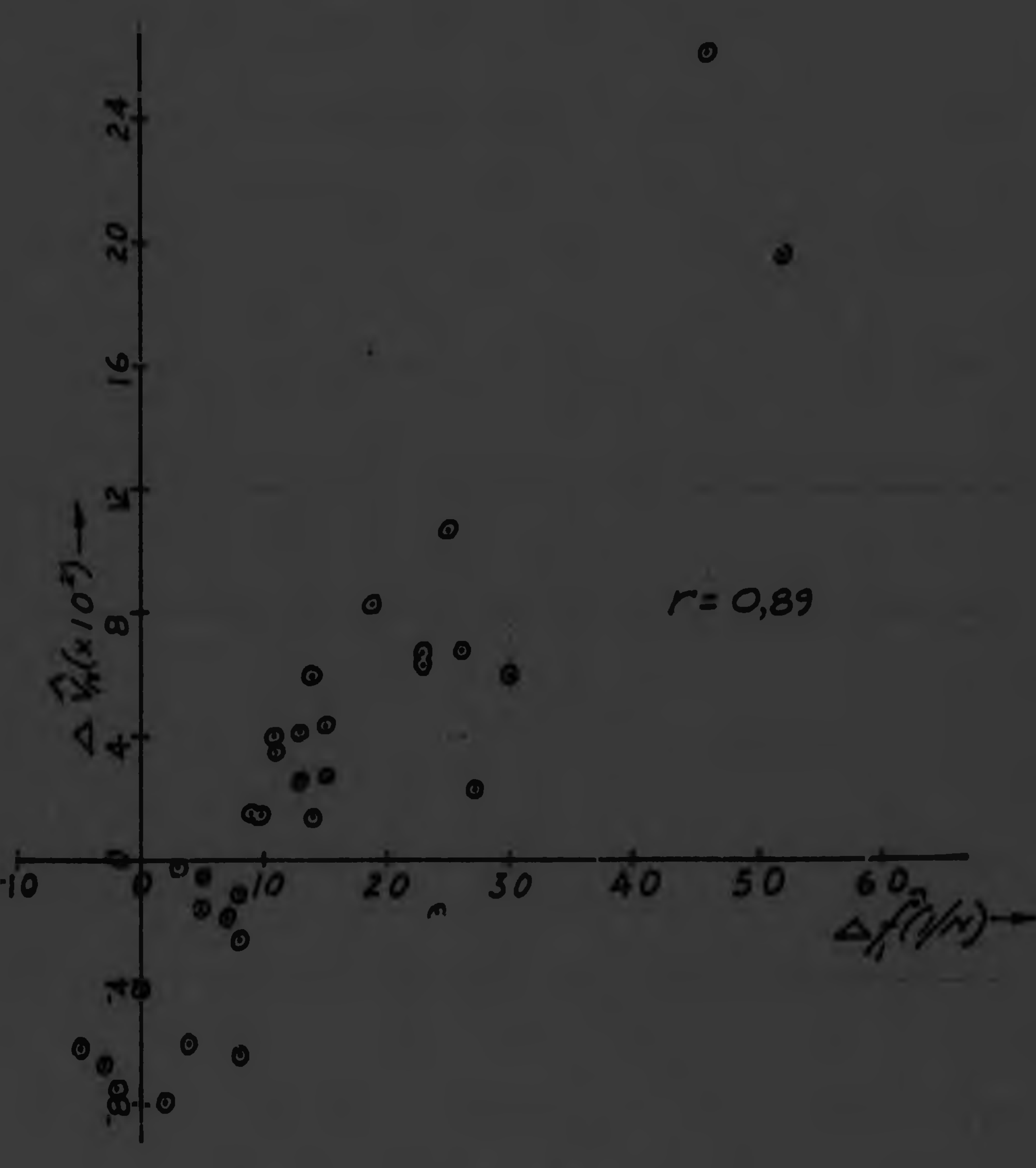


FIGURE 90 SANDRé $\hat{V}_N$ vs $\Delta \hat{f}_1(1|N)$

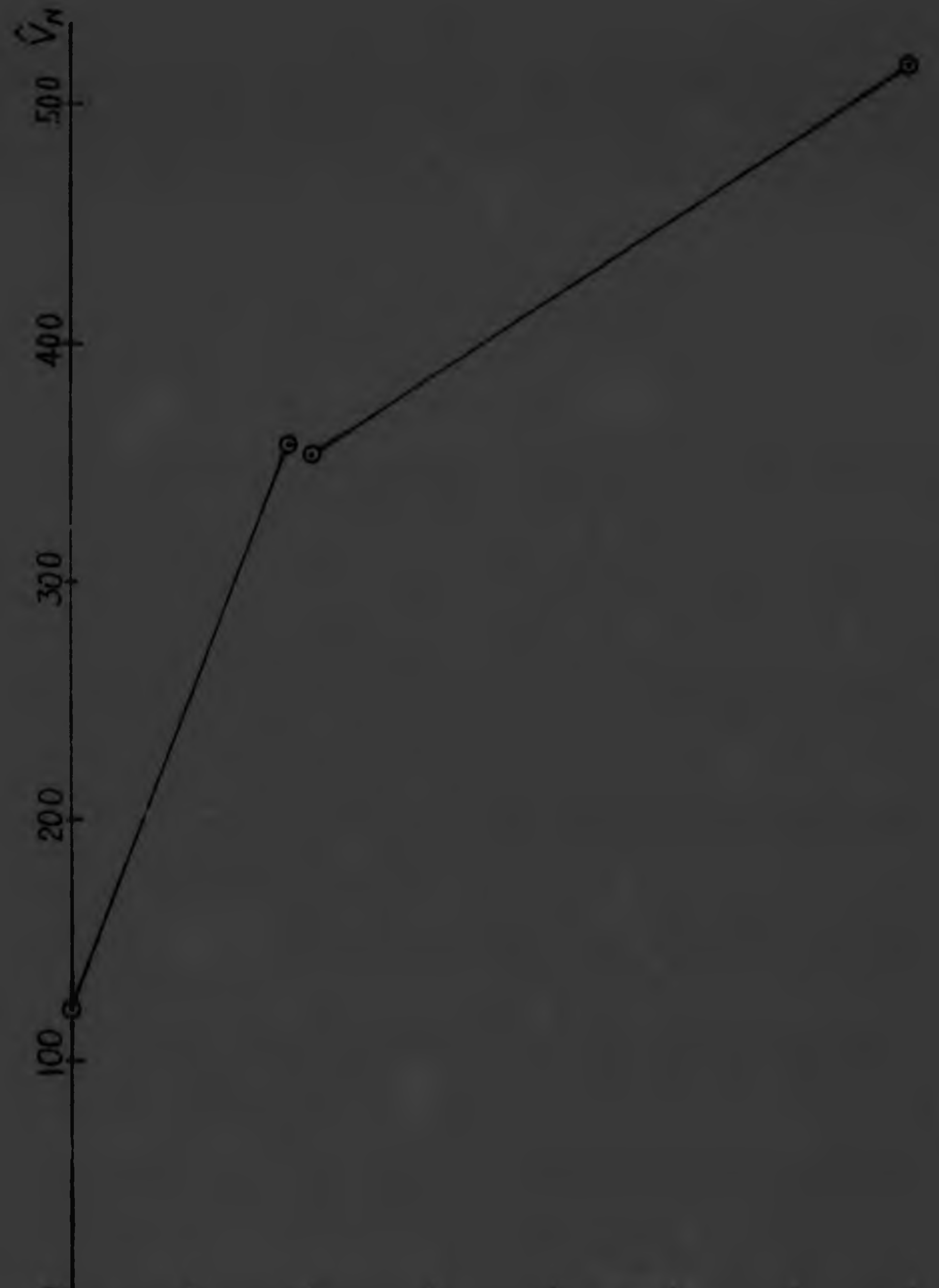


FIGURE 91 KEPRY 1 AND 2 $\hat{V}_N$ vs TIME

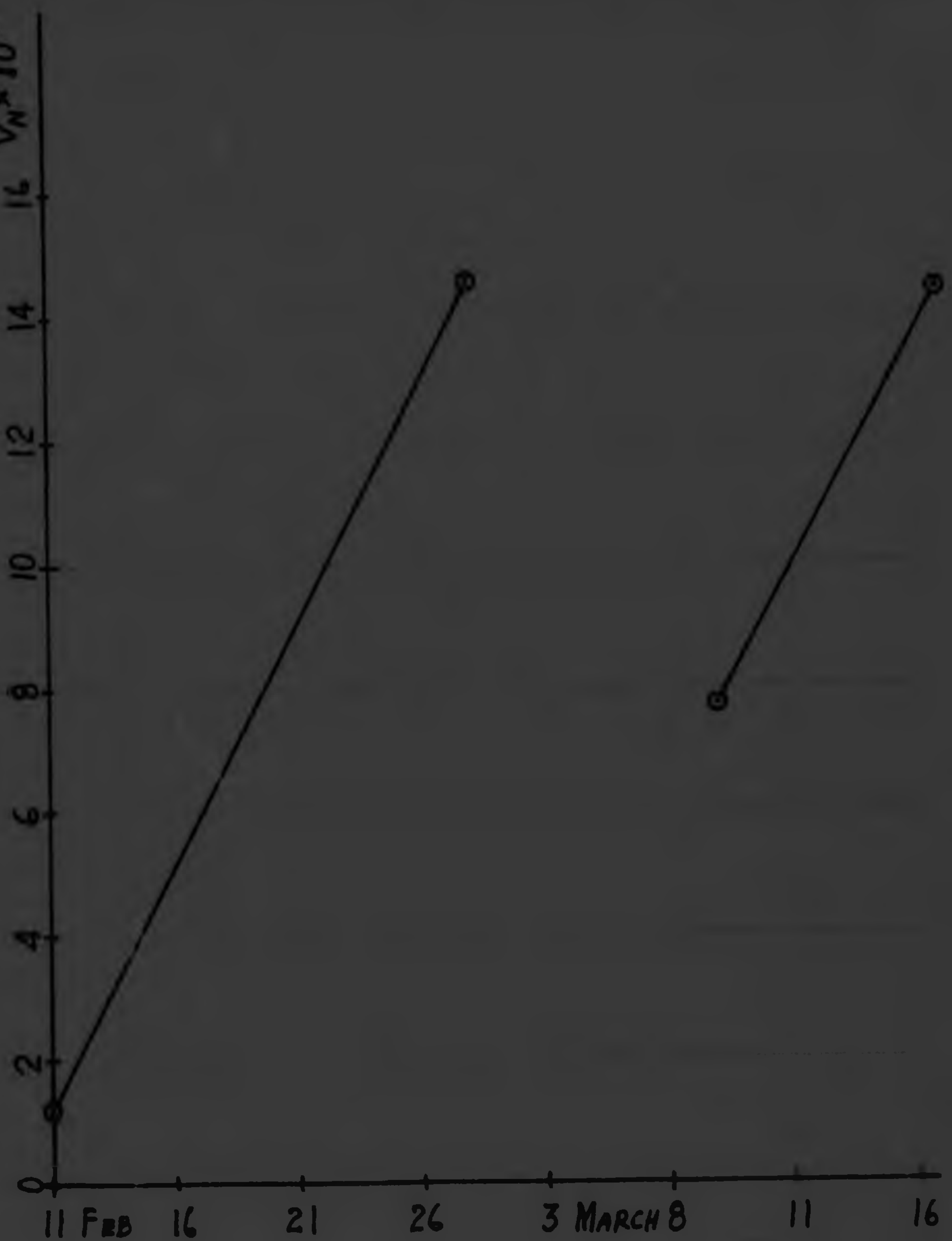


FIGURE 92 BRONWYN 1 AND 2 $\bar{V}_N$ vs TIME

CHAPTER 7 Testing the model

7.1 The idea

The basic idea behind the model is that a vocabulary is merely a set of probabilities and that an individual in writing or speaking was merely sampling at random from his/her set of probabilities.

7.2 Distributions

To test the hypothesis the set of data known as Kerry I was divided into four independent sets of five hundred words each and a frequency table for each was prepared.

The results can be seen side by side in Table 33 on page 225.

As can be seen, the distributions are remarkably similar.

At the foot of each column is shown the number of types V_N appearing in the sample and the average number of times the words appear, $\bar{r}$. The similarities are remarkable.

7.3 Increasing text Kerry 1 forward

As the size of the text increases so the development of the distribution of word frequencies can be traced.

Table 35 on page 227 shows the numbers of words appearing once, twice, thrice, etc., for a steadily increasing text length taken from Kerry I forward. The data are analysed for N=100, 500, 1000, 2000.

Column 2 shows the distribution of frequencies in the first 100 words. The distribution is reverse J-shaped and drops very sharply from Hapax Legomena of 24 to Hapax Dislegomena of 6, thereafter there is a long tail stopping at r = 11.

Column 3 shows the distribution of frequencies for N = 500. The drop from Hapax Legomena of 56 to Hapax Dislegomena of 22 is not as sharp as it is for N = 100. The tail is even more pronounced with a maximum frequency of r = 55.

Column 4 shows the distribution of frequencies for N = 1 000. The distribution is still reverse J-shaped but less sharply descending. Hapax Legomena is 65 and Hapax Dislegomena is 23. The tail stretches out to r_{max} = 110.

Column 5 shows the distribution of frequencies for N = 2 000. The distribution is still reverse J-shaped but is not incurved all the way along its range, there is a slight convexity at r = 2 heralding the eventual appearance of a mode but a much,

TABLE 33 KERRY 1 FORWARD (NATURAL DATA)

FOUR SETS OF N = 500 WORDS EACH

r	TOKENS 1-500 $\bar{f}_1(r\|N)$	TOKENS 501-1000 $\bar{f}_1(r\|N)$	TOKENS 1001-1500 $\bar{f}_1(r\|N)$	TOKENS 1501-2000 $\bar{f}_1(r\|N)$
1	54	51	45	52
2	23	33	30	27
3	13	12	20	16
4	14	9	12	9
5	4	3	8	4
6	1	3	5	3
7	3	1	4	5
8	3	4	3	1
9	-	3	3	1
10	2	1	1	4
11	2	1	1	2
12	1	2	-	-
13	1	-	-	1
15	1	-	-	1
16	1	1	-	-
17	-	-	1	1
18	-	-	-	1
19	-	2	-	-
21	1	-	1	1
22	-	1	-	-
29	1	-	-	-
30	1	-	-	-
36	-	1	-	-
55	-	1	-	-
56	1	-	-	-
74	-	-	-	1
79	-	-	1	-
$\bar{n}_N$	127	129	135	130
$\bar{r}_N$	3,93701	3,87597	3,70370	3,84615

much larger sample would be required. The tail extends to $r_{max} = 263$.

7.4 "me"

The most popular word was "me" and the relative frequency of its appearance is remarkably constant, as table 34 shows.

TABLE 34 KERRY 1

Relative frequency of the word "me"		
Sample size	No of appearances	Relative frequency
100 500 1000 2000	11 56 110 263	0,11 0,11 0,11 0,1315

7.5 Kerry 2 Comparison, data vs estimate

Once having estimated the individual's parameters δ and $\hat{\theta}$ from a set of data it is a mere matter of routine to calculate the expected values for any word frequency.

7.5.1 Table of $\bar{f}_r$ and f_E vs r

Table 36 shows a comparison between the observed word frequency distribution $\bar{f}_r$ and the calculated word frequency f_E for the data set known as Kerry 2.

The estimates were based on the calculated values:

$$\hat{\theta} = 0,980198$$

$$\hat{\beta} = 0,556955$$

$$\hat{\alpha} = 1,136413$$

which were computed from the data values

$$\hat{\phi}(1|2000) = \frac{f_1(1|2000)}{f_{2000}} = \frac{106}{315} = 0,336508$$

$$\text{and} \quad \bar{r}_{2000} = \frac{2000}{315} = 6.34921$$

Since the frequencies for $r \geq 8$ are in general small, the cells have been combined from $r = 9$ onwards as can be seen in table 36 on page 228.

Column 1 shows the number of times each word appears, r. From $r = 8$ onwards the cells are combined as shown. The frequencies 9, 10 and 11 are combined as are 12, 13, 14, 15, etc. All frequencies from 41 onward are combined into one cell.

TABLE 35 KERRY 1 FORWARD (NATURAL DATA)

N = 100 500 1000 2000

r	N=100	N=500	N=1000	N=2000
1	24	56	65	66
2	6	22	29	49
3	3	12	20	28
4	7	15	18	19
5	2	4	12	17
6	1	1	5	15
7		3	2	8
8		3	4	8
9		1	3	3
10		1	2	4
11	1	2	1	2
12		1	1	6
13				
14		1	1	
15		1		5
16-20		1	3	5
21-30		3	6	9
31-50			2	7
≥ 51		1	2	5
$\bar{n}_N$	44	128	180	256
$\bar{r}_N$	2,27	3,91	5,56	7,81

TABLE 36

KERRY 2 WORD FREQUENCY FOR N = 2000 OBSERVED $\bar{f}_r$ vs FITTED f_E

r	$\hat{f}_r$	f_E	$\bar{f}_r - f_E$	χ^2
1	106	106,0	0,0	0,000
2	57	55,5	+1,5	0,040
3	33	32,7	+0,3	0,003
4	25	21,5	+3,5	0,570
5	18	15,2	+2,8	0,516
6	10	11,4	-1,4	0,172
7	11	8,9	+2,1	0,4[illegible]
8	6	7,1	-1,1	0,170
9 - 11	17	15,0	+2.0	0,267
12 - 15	8	11,8	-3,8	1.224
16 - 18	4	5,8	-1,8	0,559
19 - 22	4	5,5	-1,5	0,409
23 - 29	6	6,1	-0,1	0,002
30 - 40	4	5,1	-1,1	0,237
≥ 41	6	7,4	-1,4	0,265
Σ	315	315,0	0.0	4,930

$P(\chi^2 \geq 4,93|12) = 0,96$

$\gamma = -0,5$ (a priori)

$\hat{\alpha} = 1,136413$

$\bar{\theta} = 0,980198$

$\bar{\beta} = 0,556955$

Column 2 shows the number of words appearing with frequency shown in Column 1, $\hat{f}_r$. For example 106 words appeared once, 57 words appeared twice, etc. Further down Column 2, there appear 17 words which appear 9, 10 or 11 times, and finally there are 6 words which appear 41 times or more.

Column 3 shows the calculated or expected number of words appearing with frequency shown in Column 1, f_E. There is one exception and that is the number of words appearing once. Since $\hat{f}(1|N)$ is used in estimating the parameters there can be no calculated value of $\hat{f}(1|2000)$ and the first numbers shown in Columns 2 and 3 are identical.

Column 4 shows the difference between the values shown in Columns 2 and 3, $(\hat{f}_r - f_E)$ and is thus self explanatory.

Column 5 shows value of $\chi^2 = \frac{(\hat{f}_r - f_E)^2}{f_E}$ calculated from Columns 3 and 4.

Totals. At the bottom of the table appear the totals of Columns 2, 3 and 5.

The totals for Columns 2 and 3 are identical because the parameter estimates were calculated from

$$\bar{r}_N = \frac{N}{n_N}$$

The total for Column 5 gives the total for χ^2, namely

$$\Sigma\chi^2 = 4{,}95$$

which, for 12 degrees of freedom is acceptable at the 0,96 level of significance.

7.6 Graph of $\hat{f}_r$ and f_E vs r. Kerry 2. Figure 92 on page 230 shows the graphs of $\hat{f}_r$ and f_E vs r on the same system of axes. The combined frequencies are plotted as the mean for one frequency and appear at the mid abscissa. Thus for r=9, 10 or 11 the value plotted is $\frac{17}{3}$ and appears at r = 10. For the frequencies ≥ 41, the point is not shown on the graph.

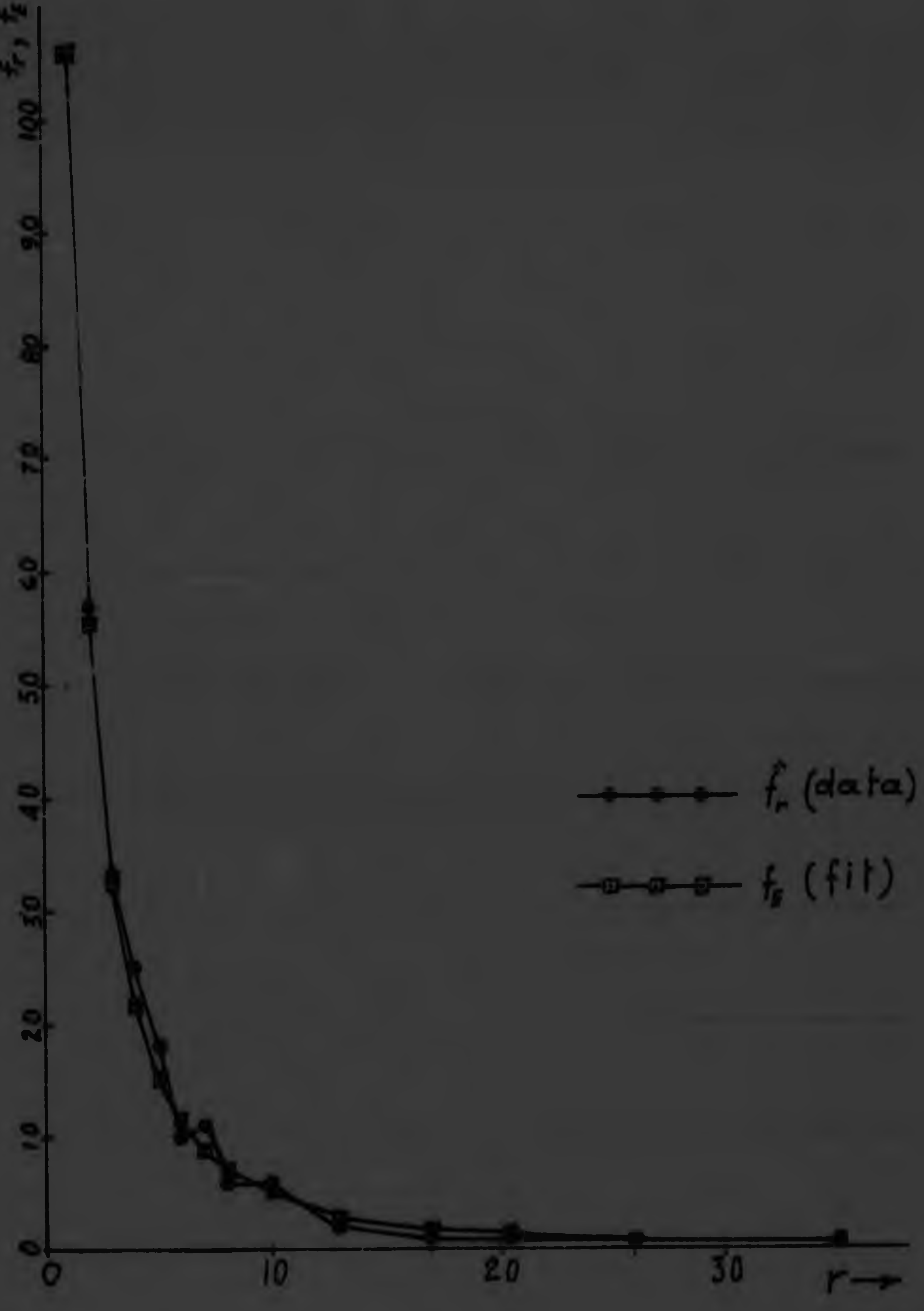


FIGURE 93 KERRY 2 $\hat{f}_r$ AND $\hat{f}_g$ vs r

7.7 The ultimate V

As has been pointed out in Chapter 5.4 the estimates of $\bar{d}_N$ show a general downward trend with increasing N, but the slope $\frac{d\bar{d}}{dN}N$ shows a generally decreasing trend.

It is thus assumed that $\bar{d}_N$ will tend to a lower limit as N tends to infinity.

An estimate of the limit of $\bar{d}_N$, $\bar{d}_\infty$ would then give an upper limit to $\bar{V}_N$.

7.7.1 Least squares estimate of $\bar{d}_\infty$

In an attempt to estimate $\bar{d}_\infty$ the equation

$$\bar{d}_N = \bar{d}_\infty + ae^{-bN} \quad \text{.........} \quad (90)$$

was fitted to the values of $\bar{d}_N$ obtained from Kerry 2, as can be seen in table 37 on page 233.

The method of least squares for three parameters was used, and the following results were obtained:

$\bar{d}_\infty$ = 0,021

a = 0,020811

b = 0,0008591

7.7.2 Kerry 2: $\bar{d}_N$ (data vs fit)

Having obtained values for the parameters $\bar{d}_\infty$, a and b, the corresponding fitted values of $\bar{d}_N$ were calculated. The results are shown in table 37 on page 233.

Column 1 shows the text size N at 100 unit intervals.

Column 2 shows the values of $\bar{d}_N$ obtained by the series of calculations as detailed in equations 73 to 84.

Column 3 shows the fitted values of $\bar{d}_N$ namely, d_N as calculated by the equation

$$d_N = 0,021 + 0,020811e^{-0,0008591N}$$

7.7.3 Graph of $\bar{d}_N$ and d_N vs N (Kerry 2)

The graphs of $\bar{d}_N$ and d_N vs N are plotted on the same set of axes and are shown in Figure 94 on page 234.

The downward trend of both data and fit can be clearly seen. The asymptote $\hat{d}_\infty = 0{,}021$ is also shown.

7.7.4 Estimating V_∞

As has been shown in Chapter 5. The equation for $\hat{V}_N$ is

$$\hat{V}_N = \frac{2}{\hat{c}_N \hat{d}_N}$$

and thus the expression for $\hat{V}_\infty$ is

$$\hat{V}_\infty = \frac{2}{\hat{c}_\infty \hat{d}_\infty}$$

Since $\hat{c}_N$ vacillates about a mean line and does not show any trend, an adequate estimate of $\hat{c}_\infty$ is the arithmetic mean, $\bar{c}$, which in the case of Kerry 2 is 0,155.

Thus finally, for Kerry 2

$$V_\infty = \frac{2}{0{,}155 \times 0{,}021} = 614$$

7.8 Kerry 1 and Kerry 2. Data and fit $\hat{n}_N$, $\hat{f}_1(1|N)$, $\hat{\phi}(1|N)$

Having made all the calculations it was a matter of routine to estimate the expected values of any of the variables.

A comparison of the data and fitted values for $\hat{n}_N$, $\hat{f}_1(1|N)$ and $\hat{\phi}(1|N)$ from Kerry 1 was made. The results appear in table 38 and 39 on pages 235 and 239.

Column 1 shows text length N

Column 2 shows the calculated or expected values of $\hat{n}_N$. The values are given to one decimal place.

Column 3 shows the data values of $\hat{n}_N$.

Column 4 shows the expected or calculated values of $\hat{f}_1(1|N)$ to three figures.

Column 5 shows the data values of $\hat{f}_1(1|N)$

Column 6 shows the expected or calculated values of $\phi(1|N)$ the values are given to four figures.

Column 7 shows the data values of $\hat{\phi}(1|N)$ to three figures.

7.9 Graphs of Kerry 1 and Kerry 2. Data and fit $\hat{n}_N$, $\hat{f}_1(1|N)$ $\hat{\phi}(1|N)$ vs N

The graphs appear in figures 95 to 100 on pages 236 to 238 and 240 to 242. As can be seen the agreement between data and fit appears to be reasonably good.

TABLE 37 KERRY 2 $\bar{d}_N$ DATA vs FIT

N	DATA d_N	FIT d_N
100	0,04498	0,04010
200	0,03833	0,03852
300	0,03694	0,03708
400	0,03856	0,03576
500	0,03721	0,03454
600	0,03427	0,03343
700	0,03198	0,03241
800	0,03406	0,03147
900	0,02850	0,03060
1000	0,02750	0,02981
1100	0,02659	0,02909
1200	0,02802	0,02842
1300	0,02935	0,02781
1400	0,03121	0,02725
1500	0,02801	0,02674
1600	0,02744	0,02626
1700	0,02655	0,02583
1800	0,02366	0,02543
1900	0,02514	0,02507
2000	0,02475	0,02473
2100	0,02442	0,02442

$$d_N = \hat{d}_\infty + ae^{-bN}$$

where $\hat{d}_\infty$ = 0,021

a = 0,020811

b = 0,0008591

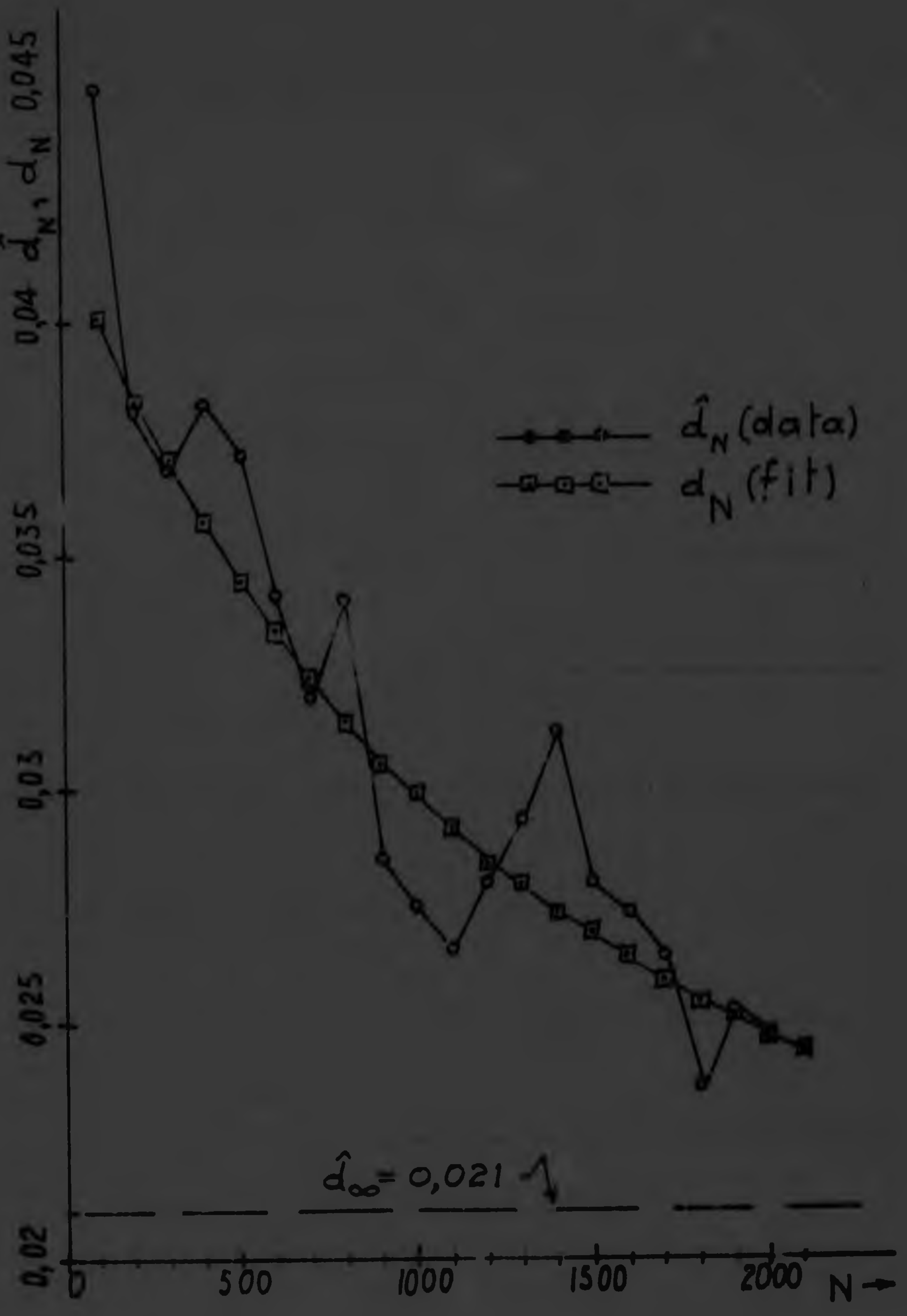


FIGURE 94 KERRY 2 $\bar{d}_N$ AND d_N vs N

TABLE 38 KERRY 1 DATA AND FIT $\hat{n}_N$, $\bar{f}_1(1|N)$, $\hat{\phi}(1|N)$

N	$\hat{n}_N$		$\bar{f}_1(1/N)$		$\hat{\phi}(1\|N)$	
	Fit	Data	Fit	Data	Fit	Data
100	49.6	44	29.8	24	.6005	.545
200	74.6	74	38.1	35	.5110	.473
300	93.2	92	43.0	40	.4616	.435
400	108.7	108	46.6	44	.4282	.407
500	122.5	128	49.4	56	.4034	.438
600	135.1	144	51.9	62	.3839	.431
700	146.8	155	54.0	64	.3681	.413
800	158.0	161	56.0	58	.3548	.360
900	168.6	170	57.9	62	.3435	.365
1000	178.8	180	59.6	65	.3336	.361
1100	188.6	189	61.3	62	.3248	.328
1200	198.1	197	62.8	60	.3169	.305
1300	207.2	212	64.2	62	.3098	.292
1400	216.1	215	65.5	66	.3032	.279
1500	224.6	229	66.7	67	.2971	.293
1600	232.8	235	67.9	69	.2915	.294
1700	240.8	239	68.9	72	.2861	.301
1800	248.4	245	69.8	66	.2810	.270
1900	255.8	249	70.6	66	.2762	.265
2000	262.9	256	71.4	66	.2716	.258
2100	269.7	263	72.0	71	.2671	.270

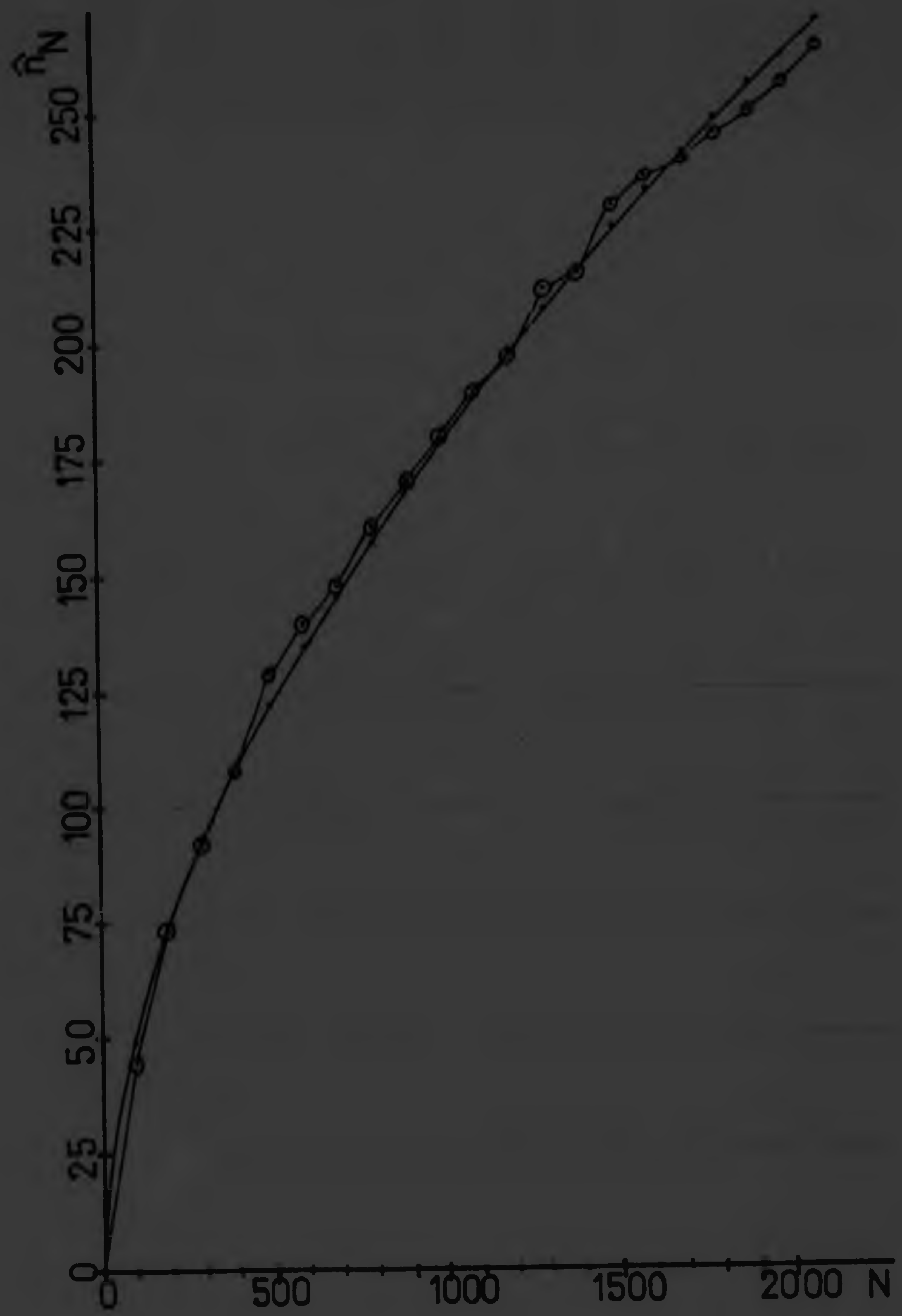


FIGURE 95 KERRY 1 $\hat{n}_N$ vs N DATA AND FIT

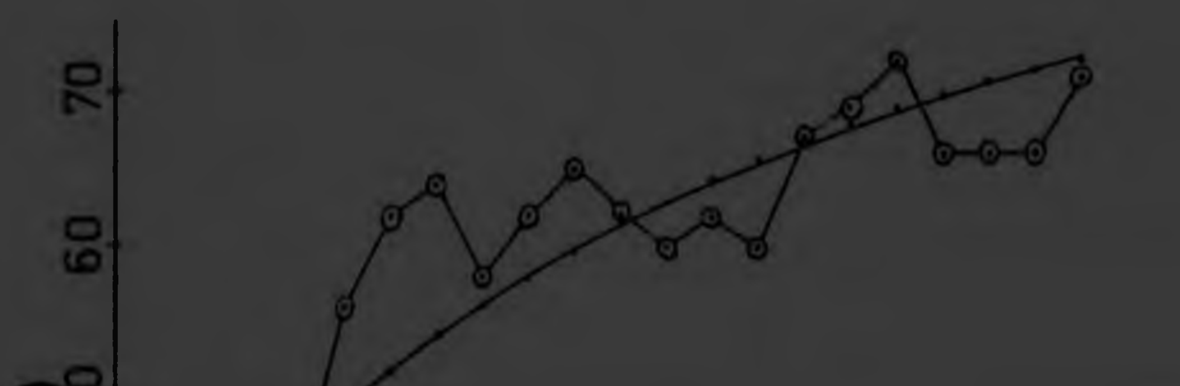


FIGURE 96 KERRY 1 $\hat{f}_1(1|N)$ vs N DATA AND FIT

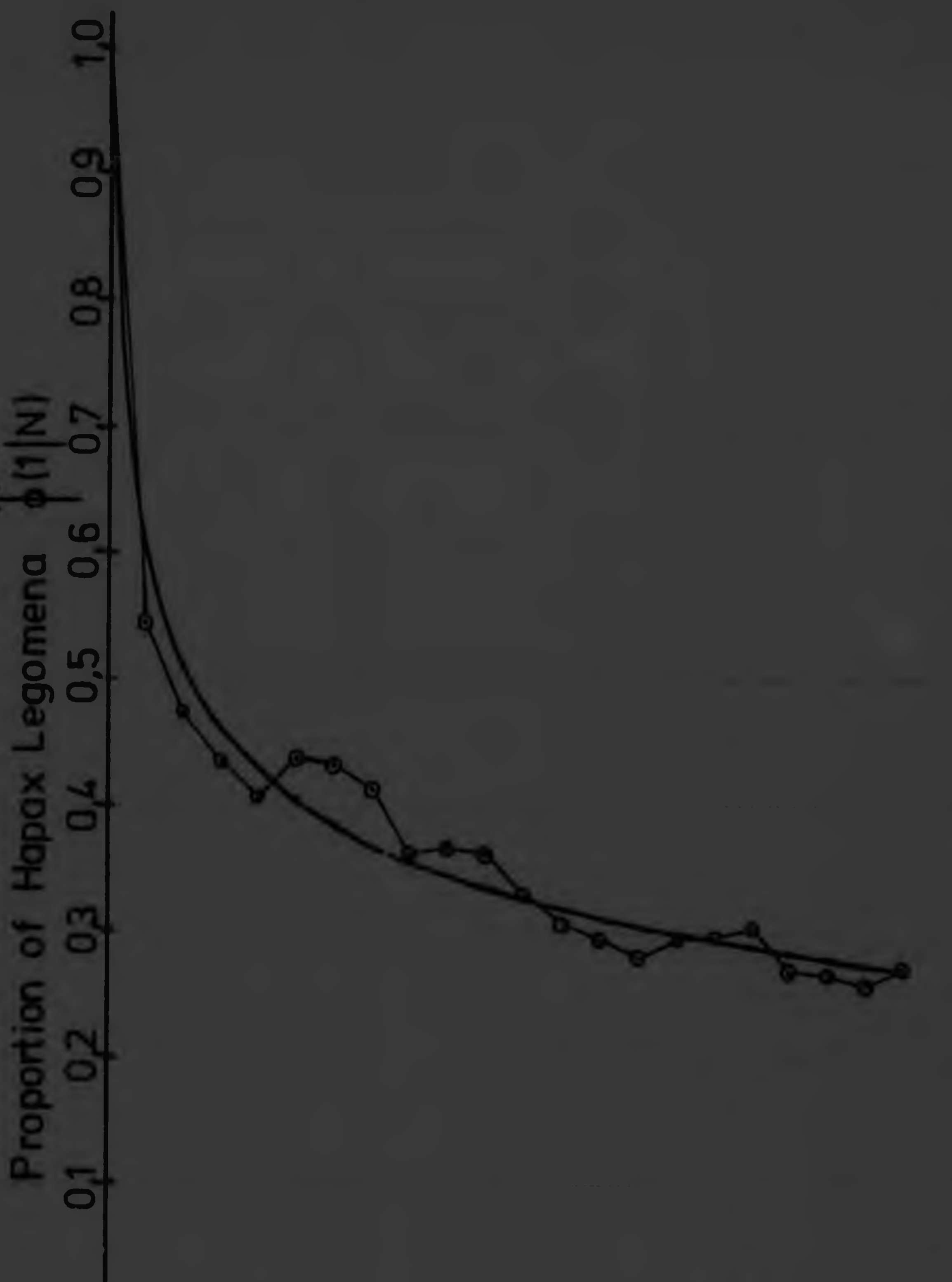


FIGURE 97 KERRY 1 $\hat{\Phi}(1|N)$ vs N DATA AND FIT

TABLE 39 KERRY 2 DATA AND FIT $\bar{n}_N$ $\hat{f}_1(1|N)$ $\hat{\phi}(1|N)$

N	$\bar{n}_N$		$\hat{f}_1(1|N)$		$\hat{\phi}(1|N)$	
	Fit	Data	Fit	Data	Fit	Data
100	56,2	55	36,9	[illegible]	0,6566	0,655
200	87,4	89	50,1	52	0,5732	0,584
300	111,1	115	58,[illegible]	63	0,5274	0,548
400	131,1	134	65,1	70	0,4966	0,522
500	148,7	154	70,4	78	0,4734	0,506
600	164,7	162	75,0	73	0,4554	0,451
700	179,5	185	79,0	84	0,4401	0,454
800	193,4	193	[illegible]	85	0,4271	0,440
900	206,6	204	8[illegible],9	81	0,4158	0,397
1000	219,0	217	89,0	[illegible]	0,4064	0,387
1100	231,0	232	91,3	89	0,3974	0,384
1200	242,4	246	94,4	97	0,3894	0,394
1300	253,4	255	96,8	99	0,3820	0,391
1400	263,9	262	99,1	104	0,3755	0,397
1500	274,0	270	101,1	100	0,3690	0,370
1600	283,8	281	103,1	103	0,3633	0,367
1700	293,1	288	104,8	102	0,3576	0,354
1800	302,2	299	106,5	106	0,3524	0,355
1900	310,8	307	107,9	105	0,3472	0,342
2000	319,2	315	109,3	106	0,3424	0,337
2100	327,2	324	110,6	108	0,3380	0,333

FIGURE 98 KERRY 2 $\hat{R}_N$ vs N DATA AND FIT

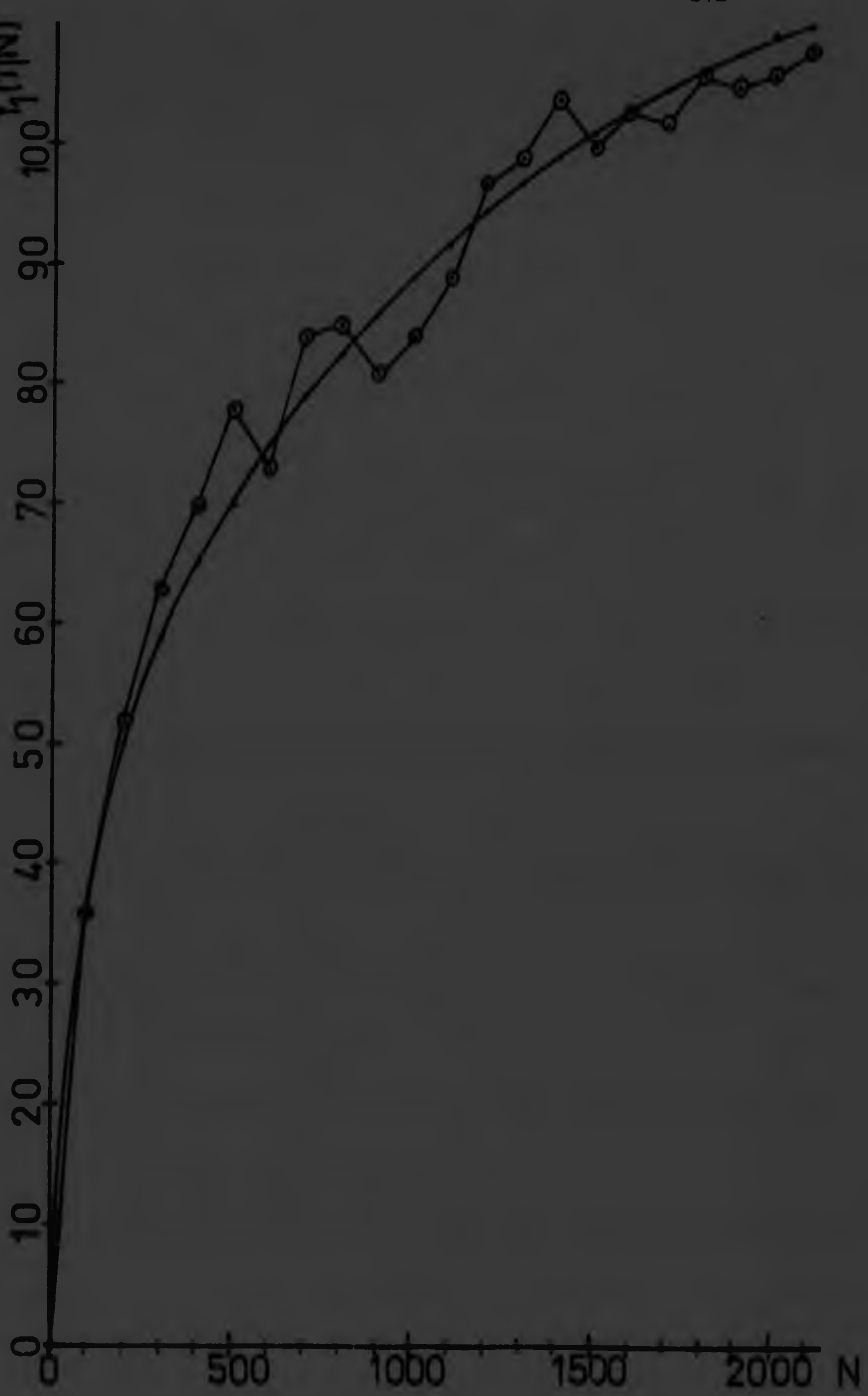


FIGURE 99 KERRY 2 $\hat{f}_1(1|N)$ vs N DATA AND FIT

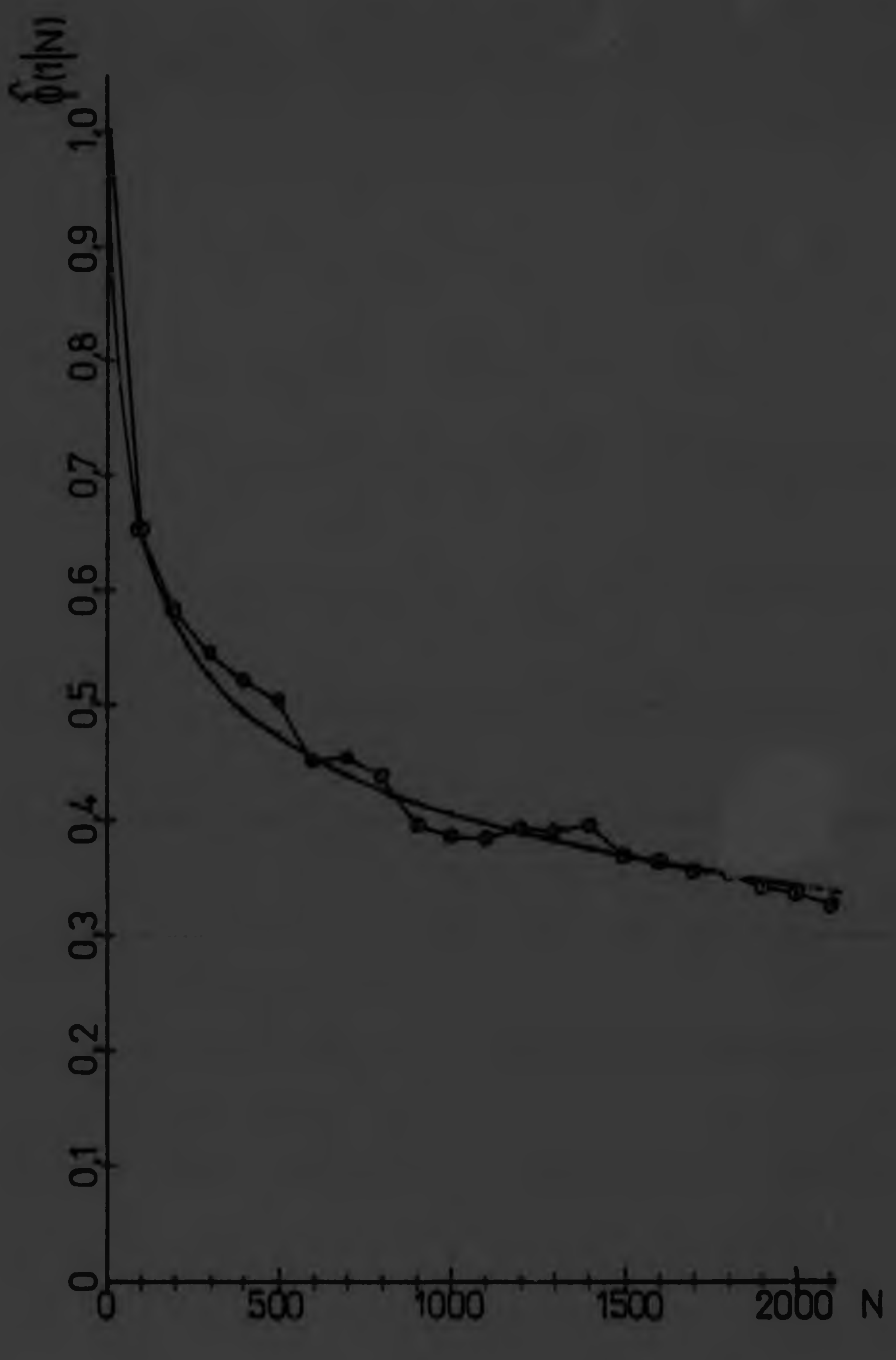


FIGURE 100 KERRY 2 $\hat{\phi}(1|N)$ vs N DATA AND FIT

CHAPTER 8 SUMMARY AND CONCLUSIONS

The inadequacy of word list tests, sampling the dictionary and recording to exhaustion as a means of estimating the number of words known by an individual, was shown.

An extrapolation method was proposed taking the length of text as the independent variable and the vocabulary contained in it as the dependent variable.

A preliminary investigation showed that the rate at which new words (vocabulary) appeared per text word was monotonic decreasing with increasing text length. A mathematical model was proposed which gave the total vocabulary of the individual as the upper bound of the vocabulary as length of text tended to infinity.

The model was based on the assumption that an individual has at his/her disposal a finite set of words, each appearing with a fixed, long term probability.

Efforts to estimate the upper bound were unsuccessful. A feature of the results was that the upper bound increased with increasing text length.

An improved model was supplied by Sichel[9]. This model expressed the number of words appearing r times in a text of length N as a function of three parameters. Estimates of the parameters could be made from the number of words appearing once only in the text and the average number of times each word appeared.

The data assembled for the first investigation were inadequate and new data were sought. It was decided to use young children and to collect as large a text as possible in a short time because Kerrich[18] has suggested that in young children the vocabulary was growing during the test period.

Two children both girls, aged two years and a few months, one an older sister, the other a younger sister were found and the mothers very kindly recorded their children's conversations in their normal environment.

Two recordings of each child were made. A complete analysis was made of word frequencies at hundred word intervals through the text. The point here being that goodness of fit on one sample was no test at all of the acceptability of the model; that for an increasingly large text a succession of estimates of the total vocabulary size should show no drift, simply a closer and closer vacillation about a fixed value.

Two samples from each child were collected and analysed by computer. For each sample, at one hundred word intervals, the words were first lemmatised and a printout was given showing how many times each word had appeared and a summary showing the number of words appearing once, twice, etc. and the total of different words.

To put the matter of vocabulary growth during test period beyond doubt, the text was reversed and the analysis made backwards. Then to test the random model (of speech being a sampling from a set of probabilities) the data cards containing approximately thirteen words each were randomised and then analysed forwards and backwards. No significant difference could be detected between any one of the four analyses and any other.

A set of data collected during the preliminary investigation was still extant. It consisted of essays written by a fourteen year old girl on topics of her own choice. The total number of words was in excess of thirteen thousand. Fortunately or unfortunately depending on the point of view, on being reminded to use as many words as possible, she made lists of words which rapidly increased Hapax Legomena and the average number of times the words were used. This listing produced a dramatic increase in the estimate of total vocabulary.

Fortunately between lists there was a long enough absence of lists which showed a steady decline in the estimate of the total vocabulary.

The final effect was to show the over-dependence of the estimate of the total vocabulary on Hapax Legomena.

The calculations were based on the number of words appearing once (Hapax Legomena) and on the average number of times the words appeared.

A series of calculations was made culminating in the estimate of the vocabulary being equal to twice the reciprocol of the product of two intermediate parameters.

If the theory was acceptable the two parameters with increasing text size should have shown a steadily decreasing vacillation about a mean result. No drift should have been evident.

One of the parameters behaved as expected but the other showed a steadily decreasing trend and a consequent steadily increasing trend in the total estimate Fortunately the parameter trend showed a decreasing slope. It appeared to be tending towards an ultimate asymptote with increasing text size.

An exponential decay curve was fitted and the ultimate asymptote of the parameter was estimated using the method of least squares for three parameters.

With this "ultimate value" of the drifting parameter and the arithmetic mean of the other (after the wild initial vacillations had subsided) an estimate of the "ultimate vocabulary" of the individual, at that stage of development could be made.

It was concluded that the method held promise as a psychometric tool for development and further research.

three parameters.

From this "ultimate value" of the shifting parameter and the arithmetic mean of the other (after the wild initial oscillations had subsided) an estimate of the "ultimate vocabulary" of the individual, at that stage of development could be made.

It was concluded that the method held promise as a psychometric tool for development and further research.

CHAPTER IX

SUGGESTIONS FOR FURTHER RESEARCH

9.1 A useful tool

In view of the obvious potential of the method it is obvious that it could be used as a very useful psychometric device.

Since the method is a measurement "de natura" it offers, for the first time, a method of reliably estimating an individual's vocabulary.

9.2 Standardisation of word list tests

It is suggested that word list multiple choice tests be standardised against vocabulary estimates obtained from this method for a test group. This would eliminate such arbitrary criteria as "This is a good test because for the group, it has a mean of one hundred and a standard deviation of fifteen".

9.3 Skewness

Distribution of vocabulary, because it is a natural measurement, must of necessity be skewed towards the high side, just as the natural measurements of mass and height are. This should be investigated.

9.4 Comparison and evaluation of various teaching methods

An obvious application is the comparison of vocabulary developed by pupils taught by the two extreme methods of:

(a) "You are allowed, nay even actively encouraged, to form your own opinion and express yourself",
and

(b) "Shut up and memorise your notes for tomorrow's test".

In other words self-discovery and development vs 'repetitio mater studiorum est."

9.5 Comparison of various methods of testing

Comparing the results of vocabulary estimates with teacher's essay marks may not only be interesting, but bring biases, associated with schools of thought, to light.

9.6 Evaluation and control of remedial education

In the current climate of remedial education the method could provide an objective, unconscious control of progress as opposed to practice effect during the remedial process and after the remedying has ceased. Neither the pupil nor the remedial teacher need be aware that anything whatsoever is being done.

books and test papers need only be photocopied and returned without any mark or remark being made which could influence the pupil's performance.

9.7 Sociology

For those who are so inclined the potential for research in the field of the correlation between sociological factors and vocabulary, from the point of view of vocabulary number alone is very large, without even mentioning type of vocabulary.

9.8 Comparison between socio-economic groups

Comparisons between vocabulary estimates and socio-economic backgrounds and environments of testees could involve a researcher in a maze of correlation coefficients.

E.g. Town vs country
Rich vs poor
Professional vs artisan etc.

9.9 Immigrants

With so many new arrivals in the country, language progress investigations and the dangerous ground of comparisons between the various language or cultural groups could be indulged in freely, again, without any evidence of any measurement having been made.

9.10 Indigenous Groups

For those who are brave enough the question of which of the indigenous groups has the larger vocabulary could be tackled.

9.11 Authorship

The method should provide a welcome and obvious addition to the tools already in use in authorship tests.

9.12 Characters by the same author

It would be most interesting to investigate whether Shakespeare for example assigned a larger vocabulary to Hamlet than to MacBeth or Julius Caesar or a smaller vocabulary to the youthful Romeo.

9.13 Lemma ratio and vocabulary

From the observation in Chapter 2.5 that in the case of very young children lemmatisation was unnecessary because the children knew and used only one form of the word. It would appear that as vocabulary grows so does the use of lemmas. It would be very interesting to see how the two develop and how well they are correlated because it may emerge that one is a function of the other and that the lemma ratio alone could be measured and the vocabulary calculated from it.

9.14 Relief from dependence on Hapax Legomena

As has been shown in Chapter VI the estimate of vocabulary is very largely dependent on Hapax Legomena i.e. the number of words appearing once. This is a disturbing feature of the method because it offers an opportunity for corruption. "Make lists of words and you'll get a good score".

It is, therefore, suggested that the possibility of estimating the vocabulary from, say Hapax Dislegomena and Hapax Trilegomena rather than Hapax Legomena and $\bar{r}_N$ be investigated.

9.15 The Drift

The drift towards higher estimates with increasing text could be taken to indicate a defect in the test but as has been shown in 7.7.1 and 7.7.4 on pages 231 and 232 adequate compensation can be made. However, investigation of the drift could provide a fruitful field for further research.

GLOSSARY OF TERMS AND SYMBOLS

CHAPTERS II AND III

Since the project began as an exercise in raw, creative thinking, without any 'a priori' knowledge of Statistics, symbols and functions were defined 'ad hoc.' These symbols have been retained because they convey the thinking employed in Chapters II and III.
The list of the symbols and their meanings follows below:

t : t is used with the following meanings:

1. The length of text analysed;
2. The number of words (repetitions included) in the length of text analysed;
3. The number of tokens;
4. The size of the sample analysed.

All these four meanings are identical. It must be clearly understood that t is considered as a variable. t is a natural number.

$v(t)$: $v(t)$ is used with the following meanings:

1. The vocabulary contained in t;
2. The number of types;
3. The number of different words (contained in the individual's vocabulary store) in a sample of size t.

x : The probability of occurrence of any individual word. x is a random variable such that $0 < x < 1$

$f(x)\delta x$: The number of words in the individual's vocabulary (or store of words), which have, associated with them, a probability of appearance of x to $x + \delta x$.
$f(x)$ is also known as the frequency distribution.

V : The total number of words known by the individual and available for use.

$$V = \int_0^1 f(x)dx$$

V is a fixed number, a constant, and

$$\int_0^1 x\, f(x)dx = 1.$$

$v(x,t)\delta x$: The number of words with probability of occurrence from x to $x + \delta x$ appearing in a sample of size t (or a text of length t).

Thus

$$\int_0^1 v(x,t)\ x = v(t) \text{ as defined above.}$$

<u>CHAPTERS IV et seq.</u>

The symbols in these chapters follow the conventions of Statistics. However, since their meanings are specific and somewhat specialised they are defined below. The symbols are not used consistently and alternates appear. It is essential that the alternate symbols be clearly understood.

V : The total number of words known by the indivudual.

N : The length of text analysed, <u>or</u> the number of tokens. N is a natural number.

r : The number of times which a word appears in a given sample.
r is a natural number.

$f_1(r|N)$: The number of words appearing r times each in a text N words long.
The subscript 1 is used to indicate that the distribution begins at $r = 1$ because in this case the number of words not appearing at all in the sample cannot be observed, although it does exist.

n_N : The vocabulary appearing in a text N word long <u>or</u> the number of types.

$$n_N = \sum_{r=1}^{N} f_1(r|N)$$

$\phi(r|N)$: The ratio $\dfrac{f_1(r|N)}{n_N}$ i.e. the proportion of types appearing r times each in the text. Both symbols $\phi(r|N)$ and $\dfrac{f_1(r|N)}{n_N}$ are used.

Π_i : The probability with which the i th word in the vocabulary appears.

$\Psi(\Pi)$: The probability distribution of Π

$\Psi(\Pi)\delta\Pi$ is that fraction of the vocabulary appearing with probability Π to $\Pi + \delta\Pi$.

$$\int_0^1 \Psi(\Pi)d\Pi = 1$$

$g(\Pi)$: The frequency distribution of probability Π.

$g(\Pi)\delta\Pi$ is the number of words appearing with probability Π to $\Pi + \delta\Pi$.

$$\int_0^1 g(\Pi)d\Pi = V$$

and $g(\Pi) = V\Psi(\Pi)$

$p(r|N)$: The probability with which any individual word will appear r times in a text N words long.

$\bar{r}_N$: The average number of times the vocabulary words contained in the text appear

$$\bar{r}_N = \frac{N}{n_N}$$

Both symbols are used.

γ, θ, α : The parameters of the generalised inverse Gaussian distribution

$$\Psi(\Pi) = C\,\Pi^{\gamma-1}\,e^{-(\frac{1}{\theta}-1)\Pi - \frac{\alpha^2\theta}{4\Pi}}$$

a : $a = 2\bar{r}_N - \ln|\bar{r}_N / \hat{\phi}(1|N)|$

b : $b = 2\bar{\phi}(1|N) + \ln|\bar{r}_N / \bar{\phi}(1|N)|$

$\hat{g}_N$: $\hat{g}_N = \sqrt{1-\hat{\theta}_N}$

$\bar{g}_N$: Root $\hat{g}_N$ is the root of the equation

$(1 + \hat{g}_N)\ln \hat{g}_N - a\hat{g}_N + b = 0$

c	: $c = \alpha\sqrt{1-\theta}$
d	: $d = \frac{\theta}{1-\theta}$

Any symbols or terms not listed in this Glossary have the usual Mathematical meaning and are also defined where they appear.

REFERENCES

1. TERMAN, Lewis M — "The Measurement of Intelligence" (Houghton Niffler Co Boston 1916)
2. HARTMAN, G W — "A critique of the common method of estimating vocabulary size together with some data on the absolute word knowledge of educated adults" (J. Educ. Psychol. 1941 32 351-364)
3. BURROUGHS, G E R — "A study of the vocabularies of young children" (Oliver & Boyd) 1957
4. PALG — "The Vocabulary of a Bengalee Girl" (Indian J. Psychol. 1926 1 203-213)
5. HAGGERTY, L C G — "What a two and a half year old child said in one day" (J. Genet Psychol. 1930 37 75-101)
6. KERRICH, J E — Emeritus Professor of Statistics, Department of Mathematics University of Wits - Private Communication
7. RIDLEY, J — Department of Mathematics University of Wits - Private Communication
8. SICHEL, H S — Professor of Statistics, Graduate School of Business Administration University of Wits - Private Communication
9. SICHEL, H S — On a Distribution Law for Word Frequencies J.A.S.A. Sept. 1975 Vol. 70 Number 351 p 543
10. SICHEL, H S — "The Natal Paper", "A Model for Vocabulary Studies", Paper read at the annual Conference of Statistics Association of S.A. ca 1975
11. SICHEL, H S — J.A.S.A. op cit, p543
12. SICHEL, H S — op cit, p543
13. SICHEL, H S — op cit, p543
14. SICHEL, H S — op cit, p544
15. SICHEL, H S — op cit, p544
16. SICHEL, H S — op cit, p546
17. SICHEL, H S — Private Communication
18. KERRICH, J E — Private Communication