# SOME STATISTICAL MODELS FOR BUYING BEHAVIOUR

by

Christopher Chatfield

Submitted as a thesis for the Ph.D. degree in the Faculty of Science at the Imperial College of Science and Technology in the University of London.

June, 1967.

## ABSTRACT

Several models are constructed to describe the pattern of purchases of non-durable consumer goods such as cocoa, margarine and soap powder.

The first model is based on the Negative Binomial Distribution (NBD) but certain systematic discrepancies from this model lead to an examination of other models, and in particular to a model based on the logarithmic series distribution (LSD).

The section on the Logarithmic distribution shows that within this model the distribution of the number of purchases made in any time-period can be completely described by just one statistic, namely, the rate of buying per buyer, where buyer refers to a member of the population who buys at least one unit in the time-period in question.

Prediction formulae are derived for both the Negative Binomial and Logarithmic models so that given data for some time period it is possible to make forecasts about the way the sample will behave in subsequent time-periods.

Finally a model based on the Beta-Binomial distribution is proposed to describe the distribution of the number of weeks in a time-period in which members of the population buy at least one packet. This model gives some insight into the reason for the discrepancies from the previous models.

ACKNOWLEDGEMENT

It is a pleasure to acknowledge my indebtedness to Dr. D. J. G. Farlie, my supervisor, and to Mr. A. S. C. Ehrenberg for their guidance, encouragement and constructive criticisms.

# CONTENTS


| | | Page |
|---|---|---|
| CHAPTER 1. | The Negative Binomial model | 7 |
| 1.1 | Introduction | |
| 1.2 | The Negative Binomial model | |
| 1.3 | The Negative Binomial distribution | |
| 1.4 | Other developments | |
| 1.5 | General theory of the NBD | |
| 1.6 | Models for the NBD | |
| 1.7 | Methods of estimating the NBD parameters | |
| 1.8 | Special cases of the NBD | |
| 1.9 | Tabulation of the NBD | |
| CHAPTER 2. | Repeat Buying | 24 |
| 2.1 | Amounts bought by repeat and lost buyers | |
| 2.2 | Deductions from the repeat buying formulae | |
| 2.3 | Practical results | |
| 2.4 | Repeat buying formulae by considering interpurchase times | |
| 2.5 | The bivariate NBD | |
| CHAPTER 3. | Problems arising | 47 |
| 3.1 | Introduction | |
| 3.2 | The Variance discrepancy | |
| 3.3 | Bunching | |
| 3.4 | Shelving | |
| 3.5 | The variance as a measure of fit | |

| | | Page |
|---|---|---|
| CHAPTER 4. | Alternative ways to fit the data | 55 |
| 4.1 | Introduction | |
| 4.2 | Adjustments | |
| 4.3 | Other frequency distributions | |
| 4.4 | Mixtures | |
| 4.5 | A general model | |
| 4.6 | The model based on the truncated Gamma distribution | |
| CHAPTER 5. | Fitting the truncated NBD | 89 |
| 5.1 | Introduction | |
| 5.2 | Methods from the literature | |
| 5.3 | A graph of $f_r/f_{r-1}$ against $1/r$ | |
| 5.4 | Geometric prediction for a | |
| 5.5 | Estimation from the mean and $f_1$ | |
| 5.6 | Moment method | |
| 5.7 | Brass's method | |
| CHAPTER 6 | The fit of the truncated NBD | 113 |
| 6.1 | Results of fitting the truncated NBD | |
| 6.2 | Appraisal of the truncated NBD | |
| 6.3 | Summary of position | |
| 6.4 | The shelving phenomenon | |
| CHAPTER 7. | The LSD model | 124 |
| 7.1 | Zeros and the NBD | |
| 7.2 | The LSD model | |
| 7.3 | The Logarithmic distribution | |
| 7.4 | The fit of the LSD | |

Page

CHAPTER 8. Prediction formulae for the LSD model 140

8.1 Introduction

8.2 Repeat buying formulae

8.3 Predictions over a longer period

8.4 Investigation of lost buyers

8.5 Cumulative tables

8.6 The standard error of w

8.7 Alternative ways of deriving the repeat buying formulae

CHAPTER 9. An application to a non-stationary situation 164

9.1 Introduction

9.2 The data

9.3 Repeat buying predictions

9.4 Results

CHAPTER 10. The distribution of Occasions and 'Weeks' 168

10.1 The distribution of Occasions

10.2 The distribution of 'Weeks'

10.3 The Beta Binomial model

10.4 Predictions from the BB model

10.5 The connection between the BB distribution and the NBD

10.6 The shelving effect

REFERENCES

APPENDIX

# CHAPTER 1

## THE NEGATIVE BINOMIAL MODEL

1.1. Introduction. This thesis is concerned with the purchases of non-durable consumer goods such as breakfast cereals, cocoa, detergents, margarines and soups. The most common technique for gathering data about such purchases is to have a large sample of housewives who keep a record every week of what they buy. This sample is called a continuous consumer panel.

Most analyses are made over periods of 1, 4, 12, 13, 24 or 26 weeks. For any such period of time we know how many consumers in the sample bought 0, 1,2 or in general r units or packets of the given product. We also know the number of occasions and the number of weeks in which each member of the sample bought at least one such packet.

Given this data we would like to be able to predict the way in which this sample will behave in subsequent periods. Thus we shall attempt to construct a model which adequately describes the buying behaviour of the population.

The model will be constructed to apply when the data is stationary; that is without any overall trend from one period to the next. However the model can also be used as a yardstick when the data is non-stationary.

The model will usually deal with one brand at a time so that no account will be taken of whether or not a consumer has also bought one or more other brands in the same product-field.

A more technical restriction for the distribution of packets is that the brand must be bought as multiples of a single unit pack-size. This restriction ensures that the resulting frequency distributions are integral valued. Thus if a brand is marketed in two or more pack-sizes, the distribution of packets for each pack-size is analysed separately.

## 1.2. The Negative Binomial model

The major portion of this thesis will be concerned with the distribution of packets or purchases. The distributions of occasions and 'weeks' will not be considered until Chapter 10. Ehrenberg (1959) noted the application of the Negative Binomial distribution to the analysis of stationary consumer purchasing over successive equal time-periods. He postulated the following compound Poisson model:-

(i) Purchases of a given consumer in successive time-periods are independent and follow a Poisson distribution with a constant mean

(ii) The average long-run rates of buying of different consumers should differ, the distribution being a Gamma distribution (strictly speaking a Pearson type III distribution).

This model is best illustrated diagrammatically.

Table 1a

| Consumer | I | II | III | IV | Long run Averages | Distributions Horizontally |
|---|---|---|---|---|---|---|
| | Periods of time | | | | | |
| A | x | x | x | x | $\mu_A$ | Poisson |
| B | x | x | x | x | $\mu_B$ | '' |
| C | x | x | x | x | $\mu_C$ | '' |
| D | x | x | x | x | $\mu_D$ | '' |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |
| Mean | m | m | m | m | m | |
| Distributions Vertically | NBD | NBD | NBD | NBD | Type III | |

NBD is used here (and throughout the thesis) as an abbreviation for Negative Binomial Distribution.

If this model is applicable, the distribution of purchases in one time-period will follow a NBD. Note that the converse of this statement is not true.

## 1.3 The Negative Binomial Distribution

The Negative Binomial distribution is a 2-parameter discrete distribution which has many useful applications. The probability of observing any non-negative integer r is given by

$$P_r = (1 + \frac{m}{k})^{-k} \frac{\Gamma(k+r)}{r!\ \Gamma(k)} (\frac{m}{m+k})^r$$

The two parameters are the <u>mean</u> m and the <u>exponent</u> k. It is also common to refer to m as the <u>scale</u> parameter and k as the <u>shape</u> parameter.

It is often convenient to use instead of m the parameter $a = m/k$.

Then $P_r = (1+a)^{-k} \frac{\Gamma(k+r)}{r!\ \Gamma(k)} (\frac{a}{1+a})^r$

The distribution has one mode, which is at zero for the fairly small values of m and k which occur with consumer purchasing data, and then the distribution is reverse J-shaped.

The variance of the distribution is

$$m(1 + m/k) = m(1+a).$$

The NBD can be fitted to an observed consumer purchasing distribution by equating theoretical and observed means and proportion of zeros.

$\hat{m}$ = sample mean = $\bar{x}$

$\hat{k}$ is the root of

$(1 + m/k)^{-k} = 1 - b$ where b is the observed proportion of buyers in the particular time-period.

$= f_o/N$ where N is sample size, and $f_o$ is the number of non-buyers.

The NBD was fitted to a large number of consumer purchasing distributions and a good fit was obtained in most cases. A typical fit is shown in Table 1b.

Table 1b

A typical example of the fit of a NBD

| Number of Units Bought | Frequencies Observed | Frequencies Theoretical | Number of Units Bought | Frequencies Observed | Frequencies Theoretical |
|---|---|---|---|---|---|
| 0 | 1612 | 1612 | 10 | 6 | 4.8 |
| 1 | 164 | 156.9 | 11 | 3 | 3.8 |
| 2 | 71 | 74.0 | 12 | 3 | 2.9 |
| 3 | 47 | 44.2 | 13 | 5 | 2.3 |
| 4 | 28 | 29.2 | 14 | 0 | 1.8 |
| 5 | 17 | 20.3 | 15-18 | 2 | 4.4 |
| 6 | 12 | 14.7 | 19-22 | 3 | 1.8 |
| 7 | 12 | 10.7 | 23-26 | 3 | 0.8 |
| 8 | 5 | 8.2 | 27+ | 0 | 0.9 |
| 9 | 7 | 6.2 | | | |

$m = 0.636 \qquad k = 0.115 \qquad a = 5.53$

$s = 2.12 \qquad \sqrt{m(1+a)} = 2.04$

The data were taken from a 2000 household sample over 26 weeks.

The sample standard deviation was compared with the theoretical value $\sqrt{m(1+a)}$ to test goodness of fit. For most distributions there is good agreement.

## 1.4 Other developments

In two later papers (1963, 1964) Ehrenberg deals with the prediction of the proportion of buyers over a longer period and also with repeat buying.

Suppose that we have detailed sales data for some time-period. We need only 2 statistics from this information to fit a NBD, namely,

$b$ = proportion of buyers in that period

$m$ = mean rate of purchasing in that period per informant.

Estimate $m$ = observed mean

Estimate $k$ by solving

$$(1 + m/k)^{-k} = 1 - b.$$

Buying over a longer period. We will define the market penetration to be the proportion of buyers in the relevant time-period. Given information about a unit time-period the stochastic model enables us to predict what will happen in a time-period $c$ times as long. In particular it is easy to predict what the market penetration will be in the longer period.

Under stationary conditions the average quantity bought in the longer period is given by

$$m_c = cm$$

The shape parameter $k$ remains constant so that

$$k_c = k.$$

This parameter k is an intrinsic property of the population sampled. When the time period changes the mean of the underlying Gamma distribution changes proportionately but the second parameter k does not change.

The penetration $b_c$ in the longer period is given by

$$1 - b_c = (1 + {}^{m_c}/k_c)^{-k_c}$$

$$= (1 + \frac{cm}{k})^{-k}$$

The predicted values of $b_c$ were compared with the observed values for a variety of brands over different time-periods and good agreement was found under stationary conditions.

Repeat-Buying. If we consider buying activity over two successive, equal periods of time, the population can be divided into the 4 sub-groups set out in Table 1c.

TABLE 1c

Buying activity in two successive time-periods

| Definition | 1st Period | 2nd Period |
|---|---|---|
| Lost Buyers | + | 0 |
| Repeat Buyers | + | + |
| New Buyers | 0 | + |
| Non-Buyers | 0 | 0 |

A + indicates the purchase of at least one unit.

We will use small letters to denote quantities in the first period and capital letters to denote quantities in the combined period which is twice as long.

In this longer period we also have an NBD with $K = k$ $M = 2m$ $A = 2a$.

$$\text{Also } B = 1 - (1+A)^{-M/A}$$

$$= 1 - (1+2a)^{-m/a}$$

Denote the proportions of repeat, lost and new buyers by $b_R$, $b_L$, and $b_N$ respectively.

$$B = b_R + b_L + b_N$$

$$b = b_R + b_L.$$

On the stationarity assumption $b_L = b_N$.

$$b_L = b_N = B - b$$

$$= (1+a)^{-k} - (1+2a)^{-k}$$

$$b_R = 2b - B$$

$$= 1 - 2(1+a)^{-k} + (1+2a)^{-k}$$

Good agreement was found between the observed and predicted values of $b_R$, $b_L$.

## 1.5 General Theory of the NBD

As the NBD will appear throughout the thesis we present an account of the relevant aspects of its theory.

The NBD has a long history. According to I. Todhunter's 'history' the earliest general statement of the NBD was given by Montmort in 1714. This is at almost exactly the same time as the first known derivation of the binomial distribution which is ascribed to Jakob Bernoulli - 'Ars Conjectandi' (1713). Other pioneers in the investigation of the NBD were Yule (1910), McKendrick (1914) and Polya(1923).

Reviews of the distribution are given by Bartko (1961) and Gurland (1959).

The NBD can be obtained by expanding the function

$$\left(\frac{k}{m+k}\right)^{k}\left[1-\frac{m}{m+k}\right]^{-k} = \left(1+\frac{m}{k}\right)^{-k}\left[1-\frac{m}{m+k}\right]^{-k}$$

Thus the p.g.f. is $\left(1+\frac{m}{k}\right)^{-k}\left[1-\frac{mt}{m+k}\right]^{-k}$

The m.g.f. is $\left(1+\frac{m}{k}\right)^{-k}\left[1-\frac{me^{t}}{m+k}\right]^{-k}$

Thus the c.g.f. is $-k\log\left(1+\frac{m}{k}\right)-k\log\left[1-\frac{me^{t}}{m+k}\right]$

$$= +k\log k - k\log\left[m+k-me^{t}\right]$$

$$= -k\log\left[1-\frac{m}{k}\left(e^{t}-1\right)\right].$$

Then the first 4 cumulants are

$$k_1 = m$$

$$k_2 = m(1 + \frac{m}{k})$$

$$k_3 = m(1 + \frac{m}{k})\ (1 + \frac{2m}{k})$$

$$k_4 = m(1 + \frac{m}{k})\ (1 + \frac{6m}{k} + \frac{6m^2}{k^2})$$

---

Notice that $k_2 > k_1$ whereas for the ordinary Binomial $k_2 < k_1$ and for the Poisson $k_2 = k_1$.

## 1.6 Models for the NBD

The NBD can be derived from several different models, only two of which will concern us.

(1) Compound Poisson If the mean $\lambda$ of a Poisson distribution varies randomly according to some probability distribution then a compound Poisson distribution results. The NBD results if $\lambda$ has a Pearson type III (Gamma) distribution. (Greenwood and Yule (1920)).

The need for such a compound Poisson distribution first arose in connection with accident figures. Thus Kendall and Stuart (p.129) give some accident figures for which the Poisson distribution gives a bad fit but the NBD gives a good fit. A plausible reason for this is that liability to accident varies from person to person, and this leads naturally to the concept of accident proneness.

We will derive the NBD in the context of consumer purchasing.

We suppose that each consumermakes Poisson purchases in successive (equal) time periods with an average rate of buying $\lambda$. We will also suppose that $\lambda$ varies from person to person and is distributed as

$$dF = \frac{1}{a^k \Gamma(k)} e^{-\lambda/a} \lambda^{k-1} d\lambda \qquad 0 \leq \lambda \leq \infty$$

$$a > 0$$

[Note: The Pearson type III distribution is obtained from the Gamma distribution $dF = \frac{1}{\Gamma(k)} e^{-x} x^{k-1} dx$ by putting $x = {}^{\lambda}/a$].

Thus in a particlar time period the proportion of people buying r units is given by

$$\int_0^{\infty} \frac{1}{a^k \Gamma(k)} e^{-\lambda/a} \lambda^{k-1} e^{-\lambda} \frac{\lambda^j}{j!} d\lambda$$

$$= \frac{(1+a)^{-k} \Gamma(k+j)}{\Gamma(k)\ j!} \left(\frac{a}{1+a}\right)^j \qquad \text{This is the NBD}$$

where $a = m/k$.

(2) <u>Generalised Poisson</u> Quenouille (1949), following Lüders (1934), has given another derivation which will be discussed more fully later on in the thesis.

If the number of bacterial colonies per field follows a Poisson distribution and the number of bacteria per colony follows a logarithmic distribution, then the distribution of bacteria per field follows a NBD.

Let the number of colonies per field follow a Poisson distribution mean $\mu$. Then its probability generating function is given by

$$G_1(t) = \exp\left\{\mu(t-1)\right\}$$

Let the number of bacteria per colony follow a Logarithmic distribution with parameter q. (This distribution is obtained by expanding $-\ln(1-q)$. Thus $P_r = -\frac{1}{\ln(1-q)}\frac{q^r}{r}$)

Its p.g.f. is given by

$$G_2(t) = \frac{\ln(1-qt)}{\ln(1-q)}.$$

Then it can be shown (e.g. Feller) that the number of bacteria per field has the compound generating function

$$G_1(G_2(t)) = G_3(t)$$

Putting $k = -\mu/\ln(1-q)$

$$a = q/(1-q).$$

we get $G_3(t) = (1+a)^{-k}\left[1 - \frac{a}{1+a}t\right]^{-k}$ which is the p.g.f. of the NBD.

There are two other well-established models which have not been considered in this thesis. Inverse Binomial sampling, which is the most widely known model, was discussed by Yule (1910). Most textbooks include an account of this model (e.g. Kendall and Stuart p.130) but it does not appear to be relevant.

The NBD can also be derived from a model of population growth in which there are constant rates of birth and death per individual and a constant rate of immigration. Williamson and Bretherton (1964) give a good précis of Kendall's original paper (1949). A similar model was considered by Yule (1924), Furry (1939) and McKendrick (1914).

## 1.7. Methods of estimating the NBD parameters

The maximum likelihood estimates of m,k were derived by Fisher (1941).

$\hat{m}$ = sample mean = $\bar{x}$

$\hat{k}$ is the root of

$$N \log (1 + \hat{m}/k) = \sum_{r=1}^{\infty} f_r \sum_{i=o}^{r-1} \frac{1}{k+i}$$

Anscombe (1950) gives the variances of these estimates for large samples as

$$\text{Var } \hat{m} = (m + m^2/k)/N = m(1+a)/N$$

$$\text{Var } \hat{k} \simeq \left\{\frac{2k(k+1)}{N(\frac{m}{m+k})^2}\right\} \Big/ \left\{1+2 \frac{\sum_{j=2}^{\infty}\left(\frac{j}{j+1}\right)\left(\frac{m}{m+k}\right)^{j-1}}{\binom{k+j}{j-1}}\right\}$$

$$\text{Cov } (\hat{m},\hat{k}) \simeq 0.$$

Because the maximum likelihood equation for k is hard to solve several other methods of estimation have been proposed.

(1) Method of moments. Estimate m,k by equating

observed and theoretical means and variances.

$$\hat{m} = \bar{x} \qquad \bar{x}(1 + \frac{\bar{x}}{\hat{k}}) = s^2$$

(2) Zero frequency. Estimate k from the observed proportion of zeros.

$$P_o = f_o/N = (1 + \frac{\bar{x}}{\hat{k}})^{-\hat{k}}$$

This equation can easily be solved by iteration especially if written as a - c log(1+a) = 0

where $c = -m/\log P_o$.

A suitable starting value for the iteration is given by $k = f_1/f_o$.

Evans (1953) has prepared a graph in which $\sqrt[3]{a}$ is plotted against log c and hence values of a can be obtained directly.

Anscombe (1950) gives a table showing the efficiency of these two methods for various values of m,k. Generally speaking, if the distribution is reverse J-shaped with more zeros than ones (as for consumer purchase distributions) then the zeros method will be very efficient. But if the distribution is more symmetric with less zeros than ones, then the method of moments has a higher efficiency.

There is an additional compelling reason for choosing the zeros method. For a particular sample, size n, denote the probability that the ith person will buy at least once in a specific time-period by $p_i$. Then the expected proportion of the sample who buy in this time-period is given by

$$\frac{\sum_{i=1}^{n} p_i}{n} = P$$

But for each person the statistic

$$T_i = \begin{cases} 1 & \text{if person buys} \\ 0 & \text{if person does not buy} \end{cases}$$

is an unbiased estimate of $p_i$.

Thus for the sum of n independent Binomial variates we have that

the proportion of buyers in the sample = b = $\dfrac{\sum_{i=1}^{n} T_i}{n}$

is an unbiased estimate of P. Thus b is an unbiased estimate of the proportion in the <u>sample</u> who will buy in a succeeding equal time-period.

Thus when we want to make unbiased predictions about a particular sample we will fit the NBD by zeros. But if we wanted to make predictions about the whole population then predictions based on maximum likelihood estimates will have a smaller variance and the extra effort in fitting the distribution might be worthwhile. We will only be concerned with predictions for a particular sample.

## 1.8 Special Cases of the NBD

(1) Geometric The recurrence relationship for the NBD is

$$P_r = \left(\frac{a}{1+a}\right)\left(1 - \frac{a-m}{ar}\right) P_{r-1}.$$

When $a = m$ $(k = 1)$

$$P_r = \frac{a}{1+a} \quad P_{r-1},$$

so that $P_r = \frac{1}{1+a}\left(\frac{a}{1+a}\right)^r \qquad r \geq 0.$

---

This is a Geometric distribution with mean a, and variance $a(1+a)$.

(2) Poisson $P_r = \left(\frac{a}{1+a}\right)\left(1 - \frac{a-m}{ar}\right) P_{r-1}$

keep m fixed and let $k \to \infty$ so that $a \to 0$.

Then $P_r \to \frac{m}{r} P_{r-1}$ which is the recurrence relationship for a Poisson distribution.

$$P_r = e^{-m} m^r/r!$$

Then variance $= m(1+a) \to m$.

---

(3) Logarithmic $P_r = \left(\frac{a}{1+a}\right)\left(1 - \frac{a-m}{ar}\right)P_{r-1}$

Let both m,k tend to zero in such a way that their ratio a stays finite.

Then $P_r \to \frac{a}{1+a} \quad \frac{r-1}{r} P_{r-1}$ for $r \geq 2$.

But $P_1 \to 0 \quad P_o$.

In other words all the probability is concentrated in the zero class. Thus we are led to consider the conditional distribution of positive integers with the zero class missing. Then we find that the relative NBD probabilities of observing a positive (non-zero) integer tend to the Logarithmic probabilities

$$P_r = \frac{1}{\log_e(1+a)} \left(\frac{a}{1+a}\right)^r / r$$

---

## 1.9 Tabulation of the NBD

The tabulation of the NBD presents problems because the cumulative distribution function cannot be expressed as a simple function. However it can be evaluated from Pearson's tables of the incomplete Beta function.

Thus

$$\sum_{x=r}^{\infty} P_x = \frac{1}{B(k,r+1)} \int_0^{\frac{k}{m+k}} u^{k-1}(1-u)^r \, du$$

where $P_x = (1 + \frac{m}{k})^{-k} \frac{\Gamma(k+x)}{x!\ \Gamma(k)} \left(\frac{m}{m+k}\right)^x$

A full account is given by Patil (1960) following Pearson and Fieller (1933).

---

In addition, Williamson and Bretherton (1964) have published tables of the NBD. However Steck (1965) in reviewing these tables points out that Pearson's tables are more useful.

# CHAPTER 2. Repeat-Buying

## 2.1. Amounts bought by repeat and lost buyers

We have already derived formulae for the proportions of repeat, lost and new buyers in 2 successive equal time-periods. We shall now obtain formulae for the amounts bought by repeat and lost buyers.

It seems reasonable to expect the repeat buyers to buy at a different (higher) rate than the lost buyers.

Now the NBD model assumes that the mean rate $(\lambda)$ at which people purchase follows a Pearson type III distribution

$$dF = \left(\frac{1}{a}\right)^k e^{-\lambda/a} \lambda^{k-1}/\Gamma(k) \quad d\lambda \qquad 0 \leq \lambda \leq \infty .$$

and that the purchases of an individual in successive equal time-periods follow a Poisson distribution.

Consider those people who bought $j$ units in the 1st period. The distribution of the mean rates of purchasing of this subgroup is altered once we know they have bought $j$ units.

Posterior distribution $\alpha$ Prior distribution x Likelihood.

Likelihood of a person buying $j$ units when he has a mean rate of purchasing $\lambda$ $\alpha$ $e^{-\lambda} \lambda^j / j!$

Therefore posterior distribution is given by

$$dF = \frac{(\frac{1}{a})^k \, e^{-\lambda/a} \, \lambda^{k-1} \times e^{-\lambda} \, \frac{\lambda^j}{j!} \, d\lambda}{\int_0^\infty (\frac{1}{a})^k \, e^{-\lambda(1+\frac{1}{a})} \, \lambda^{j+k-1}/j! \qquad d\lambda} \qquad 0 \leq \lambda \leq \infty$$

$$= \frac{e^{-\lambda(1+\frac{1}{a})} \, \lambda^{j+k-1} \, (1+\frac{1}{a})^{j+k} \, d\lambda}{\Gamma(j+k)}$$

It is interesting to notice that this is also a Pearson type III distribution. In other words the conditional frequency distribution of purchases in any period for the subgroup who bought j units in the first period is also a NBD. This valuable property of the NBD is considered in more detail by Chatfield, Ehrenberg and Goodhardt (1966).

Now the prob (0 units in 2nd period | j units in 1st period)

$$= \int_0^\infty \frac{e^{-\lambda(1+\frac{1}{a})} \, \lambda^{j+k-1} \, (1+\frac{1}{a})^{j+k} \times e^{-\lambda} \, d\lambda}{\Gamma(j+k)}$$

$$= \frac{(1 + \frac{1}{a})^{j+k} \Gamma(j + k)}{(2 + \frac{1}{a})^{j+k} \Gamma(j + k)}$$

$$= (\frac{1+a}{1+2a})^{j+k} = P\ (0|j)\ .$$

and $P(j)$ = Prob (j units in 1st period)

$$= (1+a)^{-k} \frac{\Gamma(k+j)}{j!\ \Gamma(k)} \quad (\frac{a}{1+a})^{j}$$

Thus the mean quantity bought by lost buyers (in the 1st period)

$$= m_L = \sum_{j \geq 1} P(0|j)\ P(j)\ j$$

$$= (\frac{1}{1+2a})^{k} \sum_{j=1}^{\infty} (\frac{a}{1+2a})^{j} \frac{\Gamma(j+k)}{j!\ \Gamma(k)} \cdot j$$

$$= \left(\frac{k}{1+2a}\right)^{k} \quad \frac{a}{1+2a} \quad \left[1 - \frac{a}{1+2a}\right]^{-(k+1)}$$

$$= \frac{ak}{(1+a)^{k+1}} \quad = \quad \frac{m}{(1+a)^{k+1}}$$

This quantity can also be obtained in the following easier manner.

Consider those people who bought 0 units in the first period. This subgroup consists of non-buyers in both periods and new buyers who buy in the second period but not in the first.

But the purchases in successive time periods are independent Poisson variates so that a person whose average rate of buying is $\lambda$ will buy $\lambda$ units on average in the second period regardless of the fact that he bought 0 units in the first period.

Thus the mean of the quantity bought by this subgroup which is also the mean of the quantity bought by the new buyers is given by

$$m_N = m_L = \int_0^{\infty} \frac{\left(\frac{1}{a}\right)^{k} e^{-\lambda/a} \lambda^{k-1} \times e^{-\lambda} \lambda \, d\lambda}{\Gamma(k)}$$

$= \text{proportion of population who buy 1 unit}$

$$= \frac{m}{(1+a)^{k+1}} \quad .$$

So rate of buying per lost buyer = $w_L$ = $m_L/b_L$

$$= \frac{ak(1+a)^{-(k+1)}}{(1+a)^{-k} - (1+2a)^{-k}}$$

$$= \frac{ak}{(1+a)\left[1 - \left(\frac{1+a}{1+2a}\right)^k\right]}$$

Mean of the quantity bought by repeat buyers in 1st period = $m_R$ = $m - m_L$

$$= m\left[1 - (1+a)^{-(k+1)}\right] .$$

Rate of buying per repeat buyer = $w_R$ = $m_R/b_R$

$$= \frac{m\ \left(1 - (1+a)^{-(k+1)}\right)}{1 - 2(1+a)^{-k} + (1+2a)^{-k}}$$

<u>Example</u> Date from table 1b.

$m = 0.636$ $b = 0.194$ $k = 0.115$ $a = 5.53$

$w$ = rate of buying per buyer = $\frac{0.636}{0.194}$ = 3.28

$$w_L = \frac{0.636}{6.53 \left[ 1 - \left( \frac{6.53}{12.06} \right)^{0.115} \right]} = 1.43$$

$$w_R = \frac{0.636 \quad 1 - 6.53^{-1.115})}{1 - 2 \times 6.53^{-0.115} + 12.06^{-0.115}} = 4.0$$

$$b_L = 6.53^{-0.115} - 12.06^{-0.115} = 0.055$$

$$b_R = 1 - 2 \times 6.53^{-0.115} + 12.06^{-0.115} = 0.139$$

<u>CHECK</u> $m_L + m_R = w_L b_L + w_R b_R$

$= 0.635 = m$ as required.

## 2.2 Deductions from the repeat buying formulae

An inspection of the repeat buying formulae yields several simple deductions. In particular we shall be interested to see what happens to the formulae as the time period changes.

Thus given data for 1 time-period we can calculate m, a, k as before. This enables us to predict repeat-buying in this and the following equal period.

Also for a time period $c$ times as long the distribution will theoretically have parameters $m_c$, $a_c$, $k_c$ where

$$m_c = c\,m$$

$$a_c = c\,a$$

$$k_c = k$$

This enables us to predict repeat-buying in this longer period and the following equal longer period.

Rate of buying/lost buyer

We have $$w_L = \frac{m}{(1+a)\left[1 - \left(\frac{1+a}{1+2a}\right)^k\right]}$$

$w_L$ was computed for a range of values of $a$, $k$

Table 2a. Rate of buying/lost buyer

| | | k: 0.01 | 0.1 | 0.5 |
|---|---|---|---|---|
| | ∞ | 1.45 | 1.50 | 1.72 |
| | 10 | 1.42 | 1.45 | 1.65 |
| a | 5 | 1.39 | 1.41 | 1.59 |
| | 3 | 1.34 | 1.37 | 1.53 |
| | 1 | 1.22 | 1.22 | 1.41 |

This table covers the range of values found in practice.

The following points can be noted:-

1. $w_L$ increases with the length of time-period (that is with $a$) for all values of $k$.

2. For $k < 0.1$ $w_L$ is always less than 1.5.

3. For short time periods $w_L$ decreases towards 1.

4. However for $a > 3$, $w_L$ is always within 10 per cent of the value

$$\lim_{a \to \infty} w_L = \frac{k}{1 - (\frac{1}{2})^k} .$$

Thus this quantity, although it has no practical meaning, is useful as a good approximation for $w_L$ for values of $a$ greater than 3.

Table 2b.

| $k$ | $\frac{k}{1 - (\frac{1}{2})^k}$ |
|---|---|
| 0.01 | 1.45 |
| 0.1 | 1.5 |
| 0.5 | 1.72 |
| 1 | 2.0 |
| 2 | 2.66 |
| 5 | 5.1 |

In practice $k$ is rarely greater than 0.15.

For very small $k$ we find

$$\lim_{k \to o} \frac{k}{1 - \left(\frac{1}{2}\right)^{k}} = \frac{1}{\ell n\, 2} = 1.44.$$

Thus $w_L$ is usually within 10 per cent of a quantity which ranges from 1.44 to 1.5.

Thus $w_L = 1.4$ is a useful approximation for all distributions.

<u>Rate of buying/repeat buyer</u>

We will now consider repeat buyers.

$$w_R = \frac{m \left[1 - (1+a)^{-(k+1)}\right]}{1 - 2(1+a)^{-k} + (1+2a)^{-k}} .$$

It is easy to show that $w_R \longrightarrow \infty$ as $a \longrightarrow \infty$. In other words $w_R$ increases monotonically with the length of time period. (This is intuitively obvious.)

Let us look at the rate of increase of $w_R$ compared with the rate of increase of $w$ as the time period increases.

$$\text{Now} \quad {}^{w_R}/_{w} = \frac{\left[1 - (1+a)^{-k+1}\right] \left[1 - (1+a)^{-k}\right]}{1 - 2\,(1+a)^{-k} + (1+2a)^{-k}} .$$

This ratio was computed for various values of $a$, $k$.

Table 2c. $^{w_R}/_w$

| | | k | | |
|---|---|---|---|---|
| | | 0.01 | 0.1 | 0.5 |
| | 10 | 1.22 | 1.21 | 1.08 |
| | 5 | 1.26 | 1.24 | 1.14 |
| a | 3 | 1.29 | 1.24 | 1.16 |
| | 1 | 1.25 | 1.24 | 1.13 |

For given $k$, $^{w_R}/_w$ is virtually constant. In other words $w_R$ and $w$ increase at the same rate compared with the rate of increase of time-period.

For $k < 0.1$, $^{w_R}/_w$ is virtually constant at 1.25.

Now $$\frac{w(2)}{w(1)} = \frac{w \text{ in double period}}{w \text{ in single period}}$$

$$= \frac{2\left[1 - (1+a)^{-k}\right]}{1 - (1+2a)^{-k}} \quad .$$

This function which always lies between 1 and 2 has been tabulated for various values of $a$, $k$.

Table 2d. $^{w(2)}/_{w(1)}$

| | | k | | |
|---|---|---|---|---|
| | | 0.01 | 0.1 | 0.5 |
| | 10 | 1.58 | 1.64 | 1.8 |
| a | 5 | 1.5 | 1.52 | 1.7 |
| | 1 | 1.24 | 1.28 | 1.4 |

The rate of increase of $w$ (and hence $w_R$) varies with the value of $a$. However, in the range of $k$ valid for consumer purchase data, that is $k < 0.15$, it is always true that $w_R$ will increase at a slower rate than the increase in time period.

Proportion of lost buyers

We will now look at the proportion of lost buyers.

Now $$^{b}L/_{b} = \frac{(1+a)^{-k} - (1+2a)^{-k}}{1 - (1+a)^{-k}} .$$

For a fixed $b$ (20 per cent) and various values of $m$ (and hence $w = {}^{m}/_{b}$) we can compute $a$, $k$ in the usual way from $m$ and $b$. Then we can calculate $^{b}L/_{b}$ for various values of $w$.

Table 2e. $^{b}L/_{b}$

| $w$ | $^{b}L/_{b}$ |
|---|---|
| 1 | 71 per cent |
| 1.5 | 52 |
| 2 | 42 |
| 5 | 24.5 |
| 10 | 18 |
| 20 | 13.5 |

The proportion of lost buyers decreases monotonically as $w$ increases. This result is intuitively obvious if we remember that a higher value of $w$ means that a smaller proportion of buyers buy a low number of units and that it is these people who are more likely to be lost buyers.

## 2.3. Practical results

The repeat-buying formulae were found to hold for a wide variety of brands over different time-periods even where a variance discrepancy occurs. Thus over 100 distributions in product field X were considered (see Aske Research (1964)). Some small discrepancies were observed in the predictions. For example, the observed value of $w_L$ averaged out at 1.5, as compared to the theoretical value of 1.4. This is probably due to some inevitable non-stationarity, and in any case the difference is fairly small. The other formulae give good predictions. For example the mean deviation between observed and theoretical values of $b_R$ was 10 per cent of $b_R$ for $b_R$ ranging from $1\frac{1}{2}$ per cent to 25 per cent. These predictions were unbiased. Similarly the mean deviation between observed and theoretical values of $m_R$ was 5 per cent of $m_R$ for $m_R$ ranging from .01 to over 1.

Overall the results were very encouraging, and show that the NBD model is a powerful aid to description and prediction.

## 2.4 Repeat buying formulae by considering interpurchase times.

We will now show how the repeat-buying formulae can be derived by considering interpurchase times rather than rates of buying.

The NBD model assumes that the average rate of buying $\lambda$ varies from person to person and follows a Pearson type III distribution.

$$dF = \frac{1}{a^k \Gamma(k)} e^{-\lambda/a} \lambda^{k-1} d\lambda \qquad 0 \leq \lambda \leq \infty .$$

This means that the mean time between purchases follows a distribution given by

$$dF = \frac{1}{a^k \Gamma(k)} e^{-\frac{1}{ya}} y^{-k-1} dy \quad \text{where} \quad y = \frac{1}{\lambda} .$$

Suppose the population make Poisson purchases in any 1 time period. Then the waiting time till the 1st purchase for any consumer is exponential.

Thus for a person whose mean time between purchases is y, let time until next purchase be $t'$.

Therefore $P(t' > t) = P((\text{No. of purchases by time } t) = 0)$

$$= e^{-t/y}.$$

Thus time till next purchase has a cumulative distribution function given by $1 - e^{-t/y}$. Therefore probability density function $= \frac{1}{y} e^{-t/y}$.

Thus distribution of the times to 1st purchase is given by

$$dF = \left[ \int_0^\infty \frac{1}{a^k \Gamma(k)} e^{-\frac{1}{ya}} y^{-k-1} \times \frac{1}{y} e^{-t/y} \, dy \right] dt$$

$$= \frac{1}{a^k \Gamma(k)} \left[ \int_0^\infty e^{-(\frac{1}{a}+t)\lambda} \lambda^k \, d\lambda \right] dt \quad (\text{putting } \lambda = \tfrac{1}{y})$$

$$= \frac{(\frac{1}{a} + t)^{-(k+1)} \Gamma(k+1)}{a^k \Gamma(k)} \, dt$$

$$= \frac{k(\frac{1}{a} + t)^{-(k+1)}}{a^k} \, dt$$

Consider 2 time-periods $(0,t)$, $(t, 2t)$.

Then $b$ = Proportion of buyers who buy in the period we are considering, which is $(0,t)$

= Proportion of people who buy for the first time in $(0,t)$

$b_L = b_N$ = proportion of buyers who buy for the first time in $(t,2t)$.

$$\text{Thus} \quad \frac{\mathcal{E}(b_N)}{\mathcal{E}(b)} = \frac{\int_t^{2t} K(\frac{1}{a} + t)^{-(k+1)}/a^k \quad dt}{\int_0^t K(\frac{1}{a} + t)^{-(k+1)}/a^k \quad dt}$$

$$= \frac{\left[(\frac{1}{a} + t)^{-k}\right]_t^{2t}}{\left[(\frac{1}{a} + t)^{-k}\right]_0^t}$$

Standardise times so that $t = 1$

$$\text{Then} \quad \frac{\mathcal{E}(b_N)}{\mathcal{E}(b)} = \frac{(1 + 2a)^{-k} - (1 + a)^{-k}}{(1 + a)^{-k} - 1}$$

But $\mathcal{E}(b) = 1 - (1+a)^{-k}$

$$\therefore \mathcal{E}(b_N) = \mathcal{E}(b_L) = (1+a)^{-k} - (1+2a)^{-k}$$

as obtained previously.

Then $b_R$ = proportion of buyers who buy in both $(0,t)$ and $(t,2t)$

$$= b - b_L$$

$$= 1 - 2(1+a)^{-k} + (1+2a)^{-k}$$

Consider a person whose mean time is y.

Then Prob (j purchases in (0, t))

$$= e^{-t/y} \frac{\left(\frac{t}{y}\right)^j}{j!} \qquad \text{as he makes Poisson purchases.}$$

Thus P(j purchases in (0, t) and 0 purchases in (t, 2t))

$$= e^{-t/y} \frac{\left(\frac{t}{y}\right)^j}{j!} \quad e^{-t/y} .$$

$m_L = m_N$ = mean amount bought by lost buyers

$$= \int_0^\infty \left[ \sum_{j\geq 1} P(j \text{ and } 0/y)\ j \right] P(y)\ dy$$

$$= \int_0^\infty \left[ \sum_{j\geq 1} \frac{e^{-2t/y} \left(\frac{t}{y}\right)^j j}{j!} \right] P(y)\ dy$$

$$= \int_0^\infty e^{-2t/y}\, e^{t/y}\, \frac{t}{y}\, \frac{1}{a^k \Gamma(k)}\, e^{-\frac{1}{ya}}\, y^{-k-1} dy.$$

$$= \frac{1}{a^k \Gamma(k)} \int_0^\infty e^{\frac{-\left(\frac{1}{a}+t\right)}{y}}\ t\ y^{-k-2}\ dy$$

$$= \frac{t}{a^k \Gamma(k)} \int_0^\infty e^{-(\frac{1}{a}+t)\lambda} \lambda^k d\lambda \quad \left(\text{putting } \lambda = \frac{1}{y}\right)$$

$$= \frac{t \; (\frac{1}{a} + t)^{-(k+1)} \quad \Gamma(k+1)}{a^k \Gamma(k)}$$

But m = mean amount bought by all buyers

$$= \int_0^\infty \left[ \sum_{j \geq 1} P(j|y) \; j \right] P(y) dy$$

$$= \int_0^\infty \frac{t}{y} \frac{1}{a^k \Gamma(k)} e^{-\frac{1}{ya}} y^{-k-1} dy$$

$$= \int_0^\infty \frac{t}{a^k \Gamma(k)} e^{-\lambda/a} \lambda^k d\lambda \quad \left(\text{putting } \lambda = \frac{1}{y}\right)$$

$$= \frac{a.t \; \Gamma(k+1)}{\Gamma(k)}$$

$$m_L/m = \frac{(\frac{1}{a} + t)^{-k+1)}}{a^{k+1}}$$

$$= (1+at)^{-(k+1)}$$

Standardise times so that $t = 1$

Then $m_L = m(1+a)^{-(k+1)}$ as obtained previously.

New and Lost Buyers. The consideration of the Poisson ~~interpurchase~~ property ~~times~~ also enables us to prove a relationship which has so far been assumed without proof, namely that

$$\mathcal{E}\left[P(\text{new buyers})\right] = \mathcal{E}\left[P(\text{lost buyers})\right]$$

not only for the whole population, but also for any sample.

Consider 2 time periods $(0,t)(t,2t)$ and a person whose mean interpurchase time is $y$. Then because he makes Poisson purchases

$$\begin{aligned} P(0,0) &= \text{Prob (he buys 0 units in 1st and 2nd periods)} \\ &= e^{-2t/y}. \end{aligned}$$

$$\begin{aligned} P(0,1) &= \text{Prob (0 units in 1st period, \underline{at least} 1 in 2nd period)} \\ &= e^{-t/y}\,(1 - e^{-t/y}) \\ &= p(1,0). \end{aligned}$$

Let the actual sample be such that the mean time between purchases has a distribution given by

$$dF = f(y)dy$$

Then
$$\begin{aligned} \mathcal{E}\left[P(\text{lost buyers})\right] &= \int e^{-t/y}(1-e^{-t/y})f(y)dy \\ &= \mathcal{E}\left[P(\text{new buyers})\right] \end{aligned}$$

Thus the proportion of new and lost buyers is expected to be the same whatever the sample and whatever the underlying distribution provided that the consumers make Poisson purchases.

## 2.5. The Bivariate NBD

G. Goodhardt(1965) has pointed out that the repeat buying formulae can be derived in yet another way by considering the bivariate NBD.

The multivariate NBD has been considered by Arbous and Kerrich (1951) and Bates and Neyman (1952). It possesses some remarkable properties similar to those of the multivariate normal distribution. In particular, it yields another multivariate NBD if 'cut' in a number of ways. For example we have already seen that the conditional distribution in some time period for those people who bought r units in some previous equal time period is itself a NBD with mean $(k+r)\frac{a}{(1+a)}$ and exponent $(k+r)$. Here a,k are the NBD parameters in the previous time-period. It is of interest to notice that this conditional mean can be rewritten as $m + (r-m)\frac{a}{(1+a)}$ in which case it can be seen that the regression curve of the conditional mean on the amount bought is linear.

Before considering the bivariate NBD we will obtain some general results for the generating functions of multivariate distributions.

Consider a time period $T$. Let $X$ be the random variable denoting the number of purchases in time $T$.

Suppose $T$ is subdivided into $\ell$ shorted time periods $t_1 \ldots t_\ell$ $(\sum t_i = T)$. Let $[X_1 \ldots X_\ell]$ be the purchases in each of the sub-periods $[\sum X_i = X]$.

We consider the case where the conditional distribution of $[X_1 \ldots X_\ell]$ given $X = r$ is a multinomial distribution with $P_i = {}^{t_i}/_T$.

This is certainly true for the stationary compound Poisson model of consumer purchase.

Lemma 1. If $X$ has p.g.f. $g(u; T)$ then $[X_1 \ldots X_\ell]$ is a multivariate distribution with p.g.f.

$$h(u_1 u_2 \ldots u_\ell \; ; t_1, t_2 \ldots t_\ell)$$

$$= g\left( \frac{\sum u_i t_i}{T} \; ; \; T \right). \qquad \text{(For proof see Feller (1957)).}$$

We will only consider the case $\ell = 2$ and in particular

Case 1. $t_1 = 1, \; t_2 = T - 1.$

$$h(u_1 u_2; \; 1, T-1) = g\left( \frac{u_1 + (T-1)u_2}{T} \; ; \; T \right) \qquad (1)$$

<u>Case 2</u>. $t_1 = 1, \; t_2 = 1$

$$h(u_1 u_2; \; 1, 1) = g(\frac{u_1+u_2}{2}; \; 2) \qquad - \quad (2)$$

<u>Lemma 2</u>. If the random variable $[X_1 \; X_2]$ has p.g.f. $h(u_1 \; u_2)$ then the conditional distribution of $X_1$ given $X_2 = 0$ has p.g.f.

$$C(u_1 | X_2 = 0) = \frac{h(u_1, 0)}{h(1, 0)} \qquad - \quad (3) .$$

<u>Lemma 3</u>. The marginal distribution of $X_1$ has p.g.f. $h(u_1, 1)$.

Now for case 1 the marginal distribution of $X_1$ is the distribution of purchases in unit time and its p.g.f. can be denoted by $g(u_1, 1)$

and $g(u_1, 1) = h(u_1, 1; \; 1, T-1)$

$$= g(\frac{u_1+T-1}{T}; \; T) \qquad - \quad (4)$$

<u>Bivariate NBD</u>. We now suppose that buying in a certain time period can be represented by a NBD.

The NBD is given by

$$g(u; \; T) = (1 + a_T - a_T u)^{-k} .$$

From (4) $$g(u, 1) = \left[1 + a_T - \frac{a_T}{T}(u_1 + T-1)\right]^{-k}$$

$$= \left[1 + \frac{a_T}{T} - \frac{a_T}{T} u_1\right]^{-k}$$

$$= \left[1 + a_1 - a_1 u_1\right]^{-k} \quad \text{since} \quad a_T = T_a.$$

All the repeat buying formulae follow from

$$C(u_1 \mid X_2 = 0) = \frac{h(u_1, 0)}{h(1, 0)} = \frac{g(\frac{u_1}{2}; 2)}{g(\frac{1}{2}; 2)}$$

$$= \frac{(1 + 2a_1 - a_1 u_1)^{-k}}{(1 + a_1)^{-k}} .$$

For example, the cumulant generating function of $X_1$ given $X_2 = 0$ is given by

$$\frac{(1 + 2a_1 - a_1 e^{u_1})^{-k}}{(1 + a_1)^{-k}}$$

$$= \frac{(1+2a_1 - a_1(1+u_1+\frac{u_1^2}{2!} + \ldots))^{-k}}{(1 + a_1)^{-k}}$$

$$= \frac{(1+a_1)^{-k}\left[1 - \frac{a_1}{1+a_1}\left(u_1 + \frac{u_1^2}{2!} + \ldots\right]^{-k}}{(1 + a_1)^{-k}} .$$

Then the mean of $X_1$ is the coefficient of $u_1$

$$= \frac{k\, a_1}{1 + a_1} .$$

Now the mean of $X_1 = \dfrac{m_L}{(1+a_1)^{-k}}$

where $m_L$ = mean of quantity bought by lost buyers

$(1 + a_1)^{-k}$ = proportion of people for whom $X_2 = 0$.

Also $k\, a_1 = m$ = mean of full distribution,

therefore $\dfrac{m_L}{m} = \dfrac{1}{(1 + a_1)^{k+1}}$ as previously obtained.

# CHAPTER 3

## Problems Arising

### 3.1 Introduction

We have seen that the NBD model generally gives a good fit and enables us to make useful predictions. However it soon became apparent that though the NBD fits well in most respects there are certain systematic discrepancies. The first such discrepancy is the one which was the starting point for much of the work in this thesis. It is the 'variance discrepancy'.

### 3.2 The Variance discrepancy

As already indicated, the sample standard deviation can be compared with the theoretical value $\sqrt{m(1+a)}$ as one measure of the goodness of fit. The fit is generally good for standard deviations up to about 2, but for larger values of the standard deviation the theoretical value is generally higher than the observed one. An investigation of 150 varied cases is summarised in the graph taken from Ehrenberg (1959).

Note that because the scales are logarithmic the discrepancy is worse than it appears from the graph. It seems immaterial as to whether we call this discrepancy a 'standard deviation discrepancy' or a 'Variance discrepancy'. We choose the latter as it is shorter.

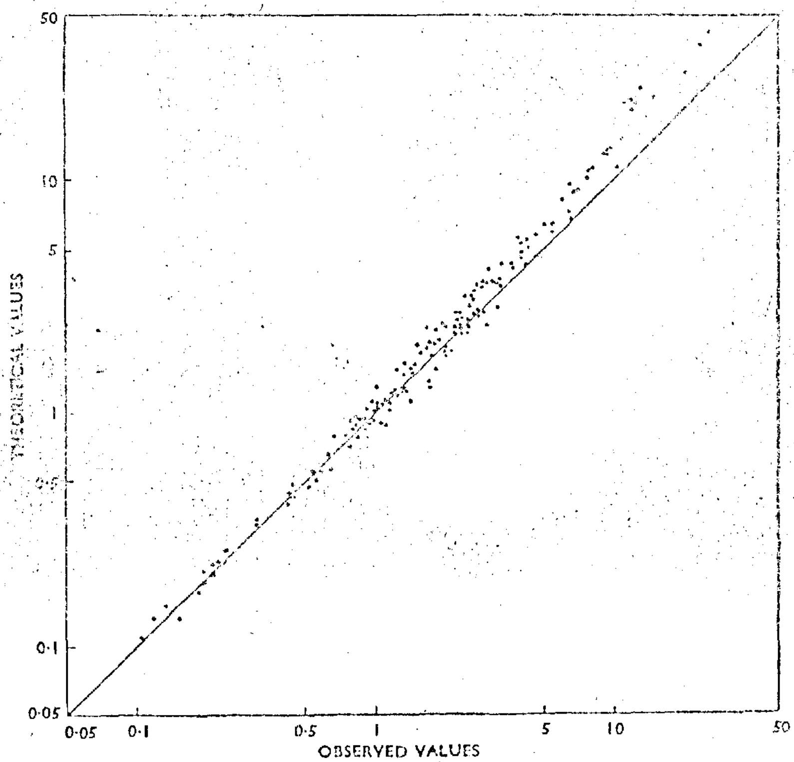



FIG. 1. Comparison of 'theoretical' and 'observed' values for the standard deviation of the frequency distributions of consumer purchases.

A high standard deviation is associated with a high rate of buying. In other words, the variance discrepancy seems to be connected with heavy buying. A detailed investigation of a large number of distributions revealed that another factor was involved. The variance discrepancy occurs for particular classes of goods, like margarines and soap powders, over any time period and hence for any rate of buying.

A table was constructed to show for various time-periods the corresponding values of w above which a variance discrepancy greater than 20°/o always occurred.

Table 3b

Table showing values of w above which a variance discrepancy >20°/o occurs

| Time Period (weeks) | w |
|---|---|
| 4 | 1.8 |
| 8 | 2.5 |
| 13 | 4 |
| 24 | 7 |

Thus the existence of a variance discrepancy depends not only on the rate of buying per buyer but also on the time-period. The table enables us to quantify the notion of 'heavy buying'. For example, we can decide to call a product 'heavily bought' if the rate of buying per buyer (w) is greater than the value shown in table 3b.

## 3.3. Bunching

One feature of consumer purchasing data which gives rise to discrepancies between the observed and theoretical distributions is a tendency for purchase frequencies to cluster or 'bunch' at or near the number of units equal to the number of weeks in the analysis period.

In the 26-weekly data of table 1b for example, there is a small amount of bunching at 25 and 26 units. There is also some bunching at 13 units, that is, half the number of weeks in the period.

Bunching occurs because a number of consumers report very regular purchasing habits, usually buying 1 unit every week. In Chapter ten we shall see how these regular purchases can be described by considering the distribution of 'weeks'.

## 3.4. Shelving

Another and more important feature of the frequency distributions, which is very noticeable in 'heavy buying' data is something which can be called shelving. It is related to bunching but seems to be present throughout the distribution. Shelf-like discontinuities occur at multiples of the time-period in weeks. Instead of the frequencies of purchases decreasing more or less steadily with the increasing size of purchase, they tend to remain more or less steady over a range of several units and then drop suddenly to a lower level just above multiples of the time-period in weeks.

A typical example of this shelving effect is provided by brand $P_3$ over 12 weeks. The effect is illustrated graphically (diagram 3c).

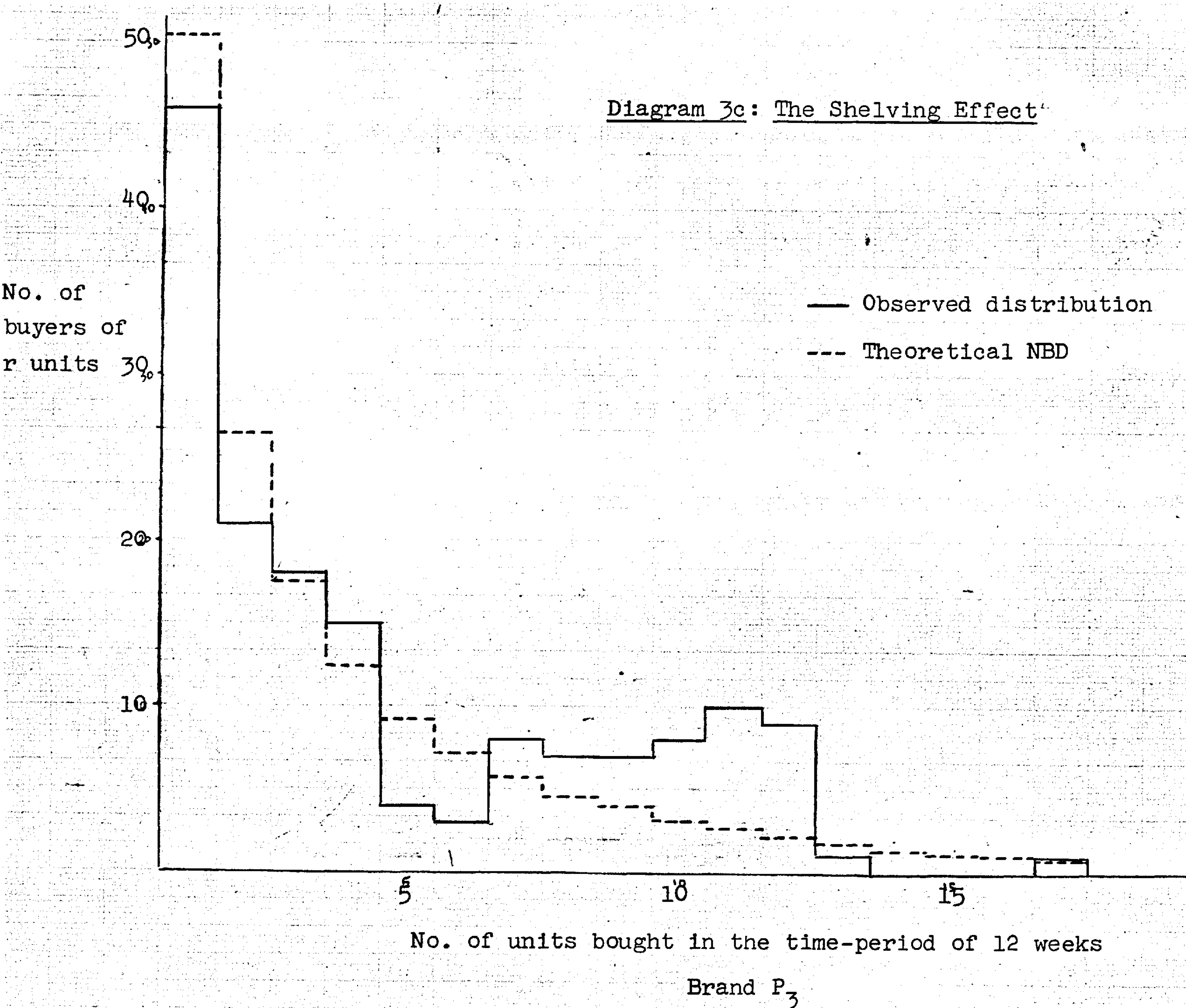



Diagram 3c: The Shelving Effect

In this particular example the 'shelf' extends from 7 to 12 units.

It is not immediately obvious how to measure the size of the shelving phenomenon. However it can be noted that the variance discrepancy and the shelving effect occur together and that a high variance discrepancy is associated with a 'large' shelving effect. Thus the variance discrepancy is in some sense a measure of the shelving effect.

The cause of the shelving effect is not apparent at this stage, but like the bunching effect we shall see in Chapter ten how it can be explained by considering the distribution of 'weeks'.

### 3.5. The Variance as a measure of fit

The variance discrepancy is, as yet, the only way the shelving effect can be quantified. Thus it is important to investigate the variance discrepancy more closely.

The first feature to notice is that it is, in a sense, rather artificial. Thus the distribution could have been fitted by the second method given in section 1.7, namely to estimate m,k from the first two moments.

In this case the theoretical and observed variances are, of course, equal. Instead there would be a systematic difference between the theoretical and observed number of non-buyers. In other words we would have a 'zeros discrepancy'.

But there are several good reasons for estimating m,k from the mean and the proportion of non-buyers. Firstly this method is statistically efficient for reverse J-shaped distributions (see Anscombe (1950)). Secondly, in order to make repeat buying predictions for the sample it is essential to have the theoretical and observed number of non-buyers equal. Otherwise we have seen that the predictions will be biased. Thirdly the bunching and shelving effects mean that most of the discrepancy occurs in the tail of the distribution and the variance statistic is very sensitive to changes in the tail. Thus consider the distribution of brand $P_3$ over 4 weeks (see appendix). The shelving effect is clearly visible with more people buying 4 units than 3 units and hardly anyone buying more than 4 units. In fact no one buys more than 6 units. This shelving effect produces a 'tail' to the fitted NBD of 3.9 buyers for r = 7+.

The variance effect of this tail is

$$> \frac{3.9 \times 7^2}{474} = 0.40$$

Thus $4^{\circ}/o$ of the buyers account for more than $26^{\circ}/o$ of the theoretical variance.

On the other hand when the full NBD gives a good fit, as in table 1b, the variance effect of the 'tail' is negligible.

As we know this effect exists, it seems sensible to use an estimation procedure, which will not be affected by the 'tail'.

# CHAPTER 4

## Alternative ways to fit the data

### 4.1 Introduction

In the previous chapter we discussed the bunching and shelving effects and the variance discrepancy. These all indicate that a general systematic deviation from the NBD model exists.

Various alternative methods of fitting the data were tried to see if an alternative model could be developed which was 'better' than the NBD model.

Any such model must be judged on several counts.

(i) Simple. The NBD model describes the data in terms of two parameters. Fitting can usually be improved by increasing the number of descriptive parameters (though not necessarily). But this will probably lead to a more complicated model. In a field such as market research where results must be used in many cases by amateur statisticians, simplicity is of considerable importance.

(ii) General. The NBD model holds (with certain systematic discrepancies) for a wide variety of brands, over different periods and for different populations. Preferably any alternative model should also hold under these general conditions, though it may be necessary to have one model for one type of data, and a second model for a second type of data. But this latter situation can only be justified by a big improvement in fit and in the predictions which result.

(iii) Useful. The NBD gives useful prediction (for example repeat-buying predictions). Any alternative model should also give such predictions.

(iv) Descriptive. Any model must describe the data reasonably. For example it is useful to have good agreement between the observed and theoretical frequency distributions, although this feature can be (and often is) overemphasised.

There are three main types of alternative models.

(1) Those formed by adjusting the NBD model.

(2) Those obtained from other frequency distributions.

(3) Those obtained from mixtures of distributions.

## 4.2 Adjustments

Four types of adjustment will be considered. These are spreading peaks, removing peaks, curtailment of the NBD at the upper end and truncation at the lower end.

a) Spreading Peaks. One of the basic assumptions in the NBD stochastic model is that each member of the population makes Poisson purchases in successive periods of time.

But it is known that many members of the population actually buy at a much steadier rate, which is often

1 unit/week. This will reduce the observed variance compared with the theoretical NBD variance. We will spread out these steady buyers and see what effect this has on the theoretical variance.

Consider the distribution for brand $P_2$ over 12 weeks. (see Appendix). There is a bump at $r = 12$ of 7 people who buy regularly at 1 unit/week. If these 7 buyers had instead made Poisson purchases with mean 12 then this spreading out would increase the observed variance by $\frac{7 \times 12}{474} = 0.18$. [Sample size = 474]. This is insignificant compared with the observed variance of 12.2, so that there is still a large difference from the theoretical variance of 20.7.

This method was tried on several distributions with a peak of steady buyers and similar results were obtained. Thus spreading the peaks seems to have little effect on the observed distributions and is not the answer to the problem.

b) <u>Removing Peaks</u>. A second adjustment is to remove any peaks from the data and see if the rest of the distribution is more closely NBD. Thus the distribution for brand $P_2$ over 12 weeks was altered by reducing $f_{12}$ from 7 to 3 to see how sensitive the distribution was to such changes. The variance ratio (theoretical/observed) changed from 1.69 to 1.58. This change is in the right direction but is nowhere near big enough. Further examples were tried and even more of the distribution was removed but it was still not possible to gain agreement between theoretical and observed variances.

c) Curtailment of the NBD

A third approach is to curtail (or truncate) the theoretical NBD at some upper point, U, say. I shall only use the term 'truncate' to refer to removing data from the lower end of the distribution. This curtailment can be done in several ways.

Firstly the tail can be spread over the whole distribution to give

$$P_r = \frac{(1+a)^{-k}\,\Gamma(k+r)\left(\frac{a}{1+a}\right)^r / \Gamma(k) \times r!}{\sum_{r=0}^{U} (1+a)^{-k}\,\Gamma(k+r)\left(\frac{a}{1+a}\right)^r / \Gamma(k) \times r!} \qquad r \leq U$$

$$P_r = 0 \qquad r > U$$

Secondly the tail can be spread over the positive part of the distribution to give

$$P_o = (1+a)^{-k}$$

$$P_r = \frac{(1+a)^{-k} \frac{\Gamma(k+r)}{\Gamma(k)r!} (\frac{a}{1+a})^r \left[1 - (1+a)^{-k}\right]}{\sum_{r=1}^{U} (1+a)^{-k} \frac{\Gamma(k+r)}{\Gamma(k)r!} (\frac{a}{1+a})^r} \qquad 1 \leq r \leq U$$

$$P_r = 0 \qquad r > U$$

Thirdly we can curtail the distribution at $r = U$ and put all the tail into $P_u$. Thus we have

$$P_r = (1+a)^{-k} \frac{\Gamma(k+r)}{\Gamma(k)r!} (\frac{a}{1+a})^r \qquad 0 \leq r \leq U-1$$

$$P_U = \sum_{r=U}^{\infty} (1+a)^{-k} \frac{\Gamma(k+r)}{\Gamma(k)r!} (\frac{a}{1+a})^r$$

$$P_r = 0 \qquad r > U.$$

Maximum Likelihood estimation of U

Given sample size N, consisting of $f_o, f_1, \ldots, f_s$ where s is the largest number of units bought, we want to estimate U.

$$\text{Likelihood } (U \mid \text{sample}) = P(0)^{f_o} \ldots . P(s)^{f_s}$$

We know $U \geq s$. Then likelihood is maximised when $P(0)$, $P(1), \ldots, P(s)$ are as large as possible, which is when $U = s$.

This method of estimation emphasises the artificiality of all the methods of cutailment. Given several distributions over similar time periods for a particular product the value of the highest observation, s, will vary considerably, so that there is no real upper limit to the distribution. In addition trouble arises when we try to extend these curtailed distributions over time-periods of different lengths, for we do not know how the upper limit, U, will change.

To consider a curtailed distribution which is not artificial, we must look at a different distribution, namely the distribution of the number of weeks in which at least one purchase was made. In a time period of n weeks, no one can buy on more than n weeks, so that the distribution vanishes for values larger than n. An examination of the distribution of 'weeks' will be made in Chapter ten.

d) <u>Truncation of the NBD</u>

A fourth method is to truncate the distribution at the lower end; that is remove the non-buyers. Then a truncated NBD can be fitted to the positive part of the distribution.

A possible model to justify this action is to assume that the population is split into 2 groups. Firstly people who never buy the product and secondly potential buyers whose purchases follow a NBD. The problem then is to estimate the NBD parameters from the truncated distribution. This will give an estimate of the potential buyers who are part of the NBD but who bought 0 units in the period in question.

Since our model is a mixture of 'never-buyers' and a NBD, truncation will be considered later on in the chapter as a mixture in section 4.5.

## 4.3 Other frequency distributions

A variety of other frequency distributions was considered. Anscombe (1950) gives a general review of 2-parameter distributions, and Gurland (1959) and Feller (1943) give a good account of contagious distributions.

There are two types of contagious distributions and it is important to understand the distinction between them.

a) Compound distributions. Whenever the population parameter of some distribution itself varies according to some known distribution then a compound distribution arises. For example, the NBD can be derived from a Poisson distribution in which the Poisson parameter $\lambda$ has a Gamma distribution with parameters a,k. It is then called a compound Poisson distribution and can be written as

Poisson • Gamma (a,k)

b) Generalised distributions. In contrast we have seen that the NBD can also be derived as follows:- If the number of bacterial colonies per field follows a Poisson distribution and the number of bacteria per colony follows a logarithmic distribution, then the distribution of bacteria per field follows a NBD. The NBD is then called a generalised Poisson distribution and can be written as Poisson ($\lambda$) x Logarithmic ($\theta$). Note that the parameters of the Poisson and Logarithmic distributions stay fixed.

We will consider generalised distributions first.

i) NBD. It has been assumed that the NBD was derived from a compound Poisson model. However one can also postulate a generalised Poisson model. Thus if the number of purchasing occasions in a particular time-period follows a Poisson distribution mean $\mu$ and the number of units bought per purchasing occasion follows a Logarithmic distribution with parameter q, then the number of purchases in this particular time-period follows a NBD with parameters given by

$$k = -\mu/\log_e(1-q)$$

$$a = q/(1-q).$$

In a time-period T times as long the Logarithmic parameter q qill stay constant but the Poisson parameter will change to $T\mu$. Thus according to this generalised model the NBD parameter a should remain constant with changes in time-period but the parameter k will change proportionately with the time-period.

However this is exactly the opposite of what occurs in practice. The parameter k is found to be invariant under changes of time-period whereas the parameter a changes proportionately with the time-period.

Thus this generalised model is inapplicable.

ii) Any generalised Poisson. A similar type of argument can be used to disprove any generalised Poisson model.

We suppose as before that the number of purchasing occasions in a particular time-period follows a Poisson distribution with mean $\mu$ and that the number of units bought per purchasing occasion follows some unspecified distribution with p.g.f. $G_2(t)$. [In the previous section this unspecified distribution was the Logarithmic distribution.]

The p.g.f. of the Poisson distribution is given by $\exp \mu(t-1)$ .

Then we have already seen in Section 1.6 that the p.g.f. of the generalised Poisson distribution is given by

$$\exp\left\{\mu(G_2(t) - 1)\right\}.$$

Thus the m.g.f. is given by

$$\exp\left\{\mu(G_2(e^t) - 1)\right\}$$

so that the c.g.f. is given by

$$\mu(G_2(e^t) - 1).$$

But $G_2(e^t)$ is the m.g.f. of the distribution of the number of units bought per purchasing occasion.

Thus $G_2(e^t) = 1 + \sum_{i\geq 1} \frac{\mu_i \; t^i}{i!}$

where $\mu_i$ = ith moment of this distribution.

Thus the c.g.f. of the generalised Poisson distribution is given by

$$\mu \left(\sum_{i\geq 1} \frac{\mu_i \; t^i}{i!}\right).$$

Thus the variance is given by $\mu . \mu_2$.

Now in a time-period c times as long the mean of the Poisson distribution will become $c\mu$ but the other distribution will still be specified by $G_2(t)$.

Thus the variance of the generalised Poisson distribution will become $c\mu \times \mu_2$.

---

But in practice we observe that the variance of consumer purchase distributions is proportional to the square of the length of time-period. Thus no generalised Poisson model can be applicable.

Other distributions. Anscombe (1950) compares the NBD with several other distributions. He tabulates the third and fourth cumulants to form a sequence of distributions with increasing skewness and tail length.

The variance discrepancy indicates that we should seek a distribution with a shorter tail and this led to consideration of the Polya-Aeppli distribution, which is immediately above the NBD in Anscombe's table. Unfortunately, although this sometimes gives a good fit for heavy-buying data, it is a generalised Poisson distribution which we have already shown to be inapplicable.

The remaining distributions above the NBD are too complicated to be of any practical use.

## 4.4. Mixtures.

Two mixtures will be considered. The first is a mixture of a Geometric and a Poisson distribution and the second a very much more general model.

a) Mixture of Geometric and Poisson distribution. The comparatively good fit of the Polya-Aeppli distribution which is Poisson ($\lambda$) $\times$ Geometric ($\tau$) led us to consider the following model:- Let us suppose that the observed distribution consists of a mixture of a Geometric and a Poisson distribution (both are special cases of the NBD).

Let the observed sample (size N) consist of a proportion p of people from the Geometric distribution (parameter $\tau$) and a proportion $(1-p) = q$ from the Poisson distribution (parameter $\lambda$).

Then the probability distribution is given by

$$P_o = q\ e^{-\lambda}$$

$$P_r = q\ e^{-\lambda}\ \lambda^r/r! + p(1-\tau)\ \tau^{r-1} \qquad \text{for } r \geq 1.$$

The system is a 3-parameter situation.

We have to estimate i) $\lambda$

ii) $\tau$

iii) p.

As $P_o$ is large, $\lambda$ will be small, so that for $r > 3$ (roughly speaking) the distribution will depend almost entirely on the geometric part. For example, if, say, $P_o = 0.8$ then $\lambda \simeq 0.2$ and the Poisson part for $r > 3$ is less than 0.0009.

Thus $P_r \rightarrow p(1-\tau)\,\tau^{r-1}$ for $r > 3$.

Then we can estimate $\tau$ from the tail of the distribution which is virtually all from the Geometric part. Such an estimating procedure will be inefficient as it is only using part of the distribution. However the proceudre is quick and so can be carried out on a large number of distributions. This in part will overcome the inefficiency.

Now $$\sum_{r=M}^{\infty} P_r \simeq p\,\tau^{M-1}$$

$$\mathcal{E}\left(\sum_{r=M}^{\infty} f_r\right) = Np\,\tau^{M-1}$$

$$\mathcal{E}(f_M) = Np(1-\tau)\,\tau^{M-1}$$

So an estimate of $\tau$ can be obtained from

$$\frac{f_M}{\sum_{r=M}^{\infty} f_r} = 1 - \hat{\tau}$$

---

This ratio, namely, $\dfrac{f_m}{\sum_{r\geq M} f_r}$, is sometimes known as MILL's ratio.

---

It is possible to reduce the variance of $\hat{\tau}$ by using a smoothed value for $f_M$.

$$\text{Thus } \hat{f}_M \simeq \frac{f_{M-2} + f_{M-1} + f_M + f_{M+1} + f_{M+2}}{5}$$

The choice of 5 terms is arbitrary and was obtained by balancing the arithmetic involved against the reduction in variance.

The use of equal weights to obtain the smoothed value of $f_M$ actually introduces a small biasing factor which is $(1-\tau)^2$. Thus when $\tau = 0.9$ for example, the bias is $1^o/o$. But in view of the approximations involved in the whole method there is no virtue in seeking an optimal choice of weighting factors.

---

A suitable value for M is 6, since $\hat{f}_m$ will then include terms down to $f_4$ which is the lower limit for the Geometric part.

---

However, instead of finding just one value for Mill's ratio it is better to plot it against 1/M. The ratio should converge to a constant which can be used to estimate $\tau$.

---

Having estimated $\tau$ we can estimate p from $\sum_{r>M} f_r = Np\tau^{M-1}$.

---

Now people who buy no units all come from the Poisson distribution.

So we can estimate $\lambda$ by equating

$$f_o = Nqe^{-\lambda}.$$

The theoretical distribution is easy to calculate as successive terms of the Poisson and Geometric series are calculated by multiplying respectively by $\lambda/r$ and $\tau$.

Example. Data Brand $P_3$ over 24 weeks. $N = 474$.

$$\sum_{r=6}^{\infty} f_r = 84.$$

Estimate $f_6$ from $\sum_{r=4}^{8} f_r/5 = 46/5$.

Hence $\tau$ by equating

$$\frac{46/5}{84} = 1 - \tau$$

giving $\hat{\tau} = 0.89$

---

$$\sum_{r=6}^{\infty} f_r = 84 = N_p \ \tau^{6-1}$$
$$= N_p \ . \ 0.89^5$$

giving $\hat{p} = 0.317$

$\hat{q} = 0.683$

---

Thus $f_o = 285 = Nq \ e^{-\lambda}$

giving $\hat{\lambda} = 0.13.$

---

We can now generate the theoretical distribution. Several heavy-buying distributions were fitted by this mixture and in each case the fit (as measured by the variance) improved.

TABLE 4a

| | Theoretical NBD | Observed | Mixture | Poisson Component | Geometric Component |
|---|---|---|---|---|---|
| 0 | 285 | 285 | 285 | 285 | - |
| 1 | 47.6 | 53 | 53.5 | 37 | 16.5 |
| 2 | 26.4 | 19 | 17.5 | 2.5 | 15 |
| 3 | 18.1 | 12 | 13 | - | 13 |
| 4 | 13.6 | 12 | 11.5 | | 11.5 |
| 5 | 10.7 | 9 | 10 | | 10 |
| 6 | 8.7 | 12 | 9.5 | | 9.5 |
| 7 | 7.2 | 7 | 8.3 | | 8.3 |
| 8 | 6.1 | 6 | 7.4 | | 7.4 |
| 9 | 5.3 | 5 | 6.6 | | 6.6 |
| 10 | 4.6 | 1 | 5.9 | | 5.9 |
| 11-12 | 7.5 | 8 | 9.8 | | 9.8 |
| 13-14 | 5.8 | 7 | 7.7 | | 7.7 |
| 15-16 | 4.6 | 5 | 6.2 | | 6.2 |
| 17-20 | 6.8 | 17 | 8.9 | | 8.9 |
| 21-24 | 5.0 | 14 | 5.6 | | 5.6 |
| 25+ | 11.6 | 2 | 9.0 | | 9.0 |

Mean (mixture) $= q\lambda + p/(1-\tau) = m_T$

Variance (mixture) $= q(\lambda+\lambda^2) + p^{(1+\tau)}/(1-\tau)^2 - m_T{}^2$

TABLE 4b

| | Observed | mixture | NBD |
|---|---|---|---|
| Mean | 2.96 | 2.96 | 2.96 |
| Variance | 35.5 | 40.2 | 52.3 |

The model which could be set up to account for the mixture would be as follows:-

The population consists of regular buyers who always buy something, who make purchases according to a Geometric distribution and irregular buyers who all have the same (very small) rate of buying. Thus the irregular buyers form a Poisson distribution. It is very easy to derive repeat-buying formulae because only the irregular buyers are possible lost buyers in the second equal time-period.

$$\text{Thus } b_L = \sum_{r>o} q\, e^{-\lambda}\, \lambda^r/r! \quad e^{-\lambda} .$$

$$= q\, e^{-2\lambda}\, [e^{\lambda}-1].$$

$$\text{Similarly } m_L = \sum_{r>o} q\, e^{-\lambda}\, \lambda^r/r! \times r \times e^{-\lambda}.$$

$$= q\lambda$$

$$w_L = {}^{m_L}/b_L = \frac{\lambda e^{2\lambda}}{e^{\lambda}-1} .$$

In the above example with $\lambda = 0.13$ we get $w_L = 1.2$.

This is a serious underestimate as we have already found that the NBD formula ($w_L \simeq 1.4$) gives good predictions for all data. The repeat-buying predictions were generally found to be inaccurate.

A second drawback is that although the variance discrepancy has been reduced, the mixture does not describe the shelving effect any better than the NBD. (The shelving effect is clearly visible between $r = 17$ to 24.)

A third drawback is that the method of estimating which has been suggested requires a distribution with a long tail. Thus it would be no good for distributions in short time-periods.

Thus this mixture compared badly with the NBD model, particularly with regard to predictions, and was rejected as a possible alternative model.

---

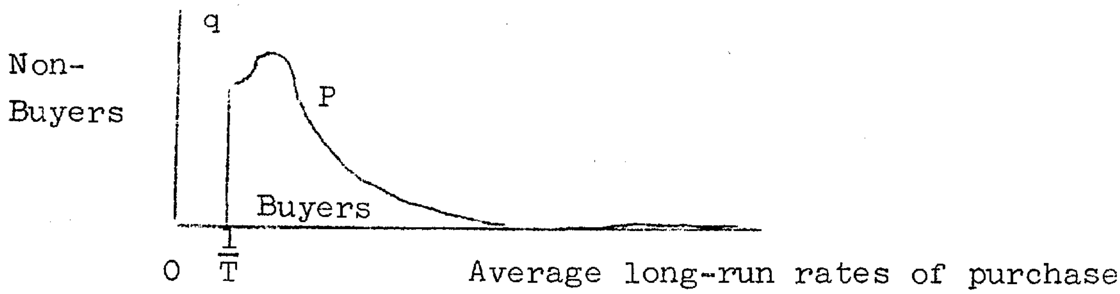
We will now consider a much more general model, derived from a simple mixture, which includes the simple NBD model as a special case.

## 4.5 General Model

Let us postulate a more general model for consumer purchases.

Firstly we suppose that there is a proportion q of 'never-buyers', i.e., people in the population who never buy.

Secondly we suppose that there is a proportion p of potential buyers whose average long run rates of buying follow a distribution whose p.d.f. is $f(\lambda)$. This distribution is truncated at the lower end at some point $1/T$ where T is a very long time interval




Thus the cumulative distribution function for the average long run rates of purchase is given by

$$F(x) = q \qquad 0 \leq x \leq 1/T$$

$$F(x) = q + p \int_{\frac{1}{T}}^{x} f(\lambda)d\lambda. \qquad x \geq \frac{1}{T}$$

Thirdly, as in the NBD model, we assume that the purchases of any one consumer in successive time-periods follow a Poisson distribution and are independent.

<u>Intuitive justification</u>. Intuitively this model is at least as good as the NBD model for it seems more plausible to hypothesise a group in the population who will never buy the commodity in question than to hypothesise infinitesimally low buying-rates. For in the NBD model everyone is a potential buyer. The truncation point is also plausible.

Thus we could take T to be the life-time of the product.

Repeat buying formulae

As before we are concerned with buying behaviour in two successive equal time-periods.

Note that throughout the following derivation we shall use the term average rate of buying to indicate the quantity bought by some subgroup averaged over that subgroup, but we shall use the term mean to indicate the quantity bought by a subgroup averaged over the whole population and not just averaged over the subgroup.

Now the probability that a person buys j units in some time-period is given by

$$P(j) = p \int_{\frac{1}{T}}^{\infty} f(\lambda)\, e^{-\lambda} \frac{\lambda^j}{j!}\, d\lambda \quad .$$

But the purchases in successive time-periods are independent so that a person whose average rate of buying is $\lambda$ will buy $\lambda$ units on average in the second period regardless of the fact that he bought, say, j units in the first period.

Thus if we consider the subgroup of people who bought j units in the first period, their average rate of buying in the second period is given by

$$\frac{p \int_{\frac{1}{T}}^{\infty} f(\lambda) e^{-\lambda} \frac{\lambda^j}{j!} \lambda d\lambda}{p \int_{\frac{1}{T}}^{\infty} f(\lambda) e^{-\lambda} \frac{\lambda^j}{j!} d\lambda}$$

$$= (j+1) \frac{P(j+1)}{P(j)}$$

Thus the mean of the quantity of this subgroup expressed as an average over the whole population is given by

$$(j+1)\ P(j+1)$$

In particular if we consider those people who bought 0 units in the first period then this subgroup consists of the non-buyers in both periods and new buyers who buy in the second period but not in the first.

The mean of the quantity bought by this subgroup in the second period is the same as the mean of the quantity bought by the new buyers in the second period and from the previous result is given by

$$m_N = 1.\ P(1)$$
$$= P(1).$$

Thus the mean of the quantity bought by repeat buyers is given by

$$m_R = m - P(1)$$

where m = mean of the distribution

$$= p \int_{\frac{1}{T}}^{\infty} f(\lambda)\ \lambda d\lambda.$$

These results depend mainly on the Poisson assumption but $f(\lambda)$ must be chosen to give good agreement between the observed and theoretical values of $P(j)$.

The proportion of buyers, b, is given by

$$b = p \int_{\frac{1}{T}}^{\infty} f(\lambda)(1-e^{-\lambda})d\lambda$$

The proportion of new buyers is given by

$$b_N = p \int_{\frac{1}{T}}^{\infty} f(\lambda)\ e^{-\lambda}\ (1-e^{-\lambda})d\lambda$$

This expression includes the quantity $\int_{\frac{1}{T}}^{\infty} f(\lambda)e^{-2\lambda}d\lambda$ which is not immediately meaningful.

Similarly the proportion of repeat buyers is given by

$$b_R = p \int_{\frac{1}{T}}^{\infty} f(\lambda)(1-e^{-\lambda})^2 d\lambda.$$

$$= b - b_N.$$

We can also consider the subgroup who bought in the first period.

In the second period the mean of the quantity bought by this subgroup is given by

$$p \int_{\frac{1}{T}}^{\infty} f(\lambda)\ (1-e^{-\lambda})\lambda\ d\lambda$$

$$= m - P(1) \qquad = m_R.$$

However the rate of buying of this subgroup in the second period is $\frac{m-P(1)}{b} = w - \frac{P(1)}{b}$. In other words this subgroup buys at a lower rate in the second period.

The other subgroup, the non-buyers in the first period, will buy at a correspondingly higher rate in the second period because of the effect of the new buyers.

Thus far we have made no assumptions about the distribution of the average long-run rates of buying.

First we consider the Beta distribution given by

$$dF = \frac{1}{B(p,q)} (1-x)^{p-1} x^{q-1} dx \qquad 0 \leq x \leq 1.$$

or more generally

$$dF = \frac{1}{B(p,q)} (1- \frac{\lambda}{d})^{p-1} (\frac{\lambda}{d})^{q-1} \frac{d\lambda}{d} \quad 0 \leq \lambda \leq d$$

Note that we have taken $T = \infty$, so that the distribution is defined for $\lambda \geq 0$. The distribution has three parameters p,q and d.

Then

$$P(j \text{ units}) = p \int_0^d \frac{1}{B(p,q)} (1- \frac{\lambda}{d})^{p-1} (\frac{\lambda}{d})^{q-1} \times e^{-\lambda} \frac{\lambda^j}{j!d} \quad d\lambda$$

$$= \frac{p}{j!B(p,q)d^q} \int_0^d (1- \frac{\lambda}{d})^{p-1} \lambda^{q-1} e^{-\lambda} d\lambda.$$

which can only be evaluated as the sum of a series of Gamma or Beta functions, which is very difficult to handle.

The F distribution is an associated distribution which is defined on $(0,\infty)$. It is obtained from the Beta distribution by the transformation $X/(1-X)$, but the integrals which result prove to be equally difficult.

The only integrals which give workable results are those resulting from a distribution based on a power of $\lambda$ or a power of $\lambda_x e^{-\alpha\lambda}$.

The first of these possibilities gives a distribution given by

$$dF = \frac{\alpha+1}{d^{\alpha+1}} \lambda^{\alpha} \, d\lambda \qquad 0 \leq \lambda \leq d$$

$$\text{giving } P(j \text{ units}) = p \int_0^d \frac{\alpha+1}{d^{\alpha+1}} \lambda^{\alpha} e^{-\lambda} \frac{\lambda^j}{j!} \, d\lambda$$

This is an incomplete Gamma integral and can be evaluated from the tables once d and $\alpha$ have been estimated. However the problems of estimating d have already been mentioned in section 4.2 when the curtailment of the NBD was considered. Similar problems arise here. Thus we are led to consider a distribution of the form

$$dF = Ce^{-\alpha\lambda} \times \lambda^{\beta} \qquad \frac{1}{T} \leq \lambda \leq \infty$$

where C is a constant depending on $\alpha$, $\beta$ and T.

This distribution is very similar to the distribution used to derive the NBD from a compound Poisson model. It is a truncated Gamma distribution.

4.6. The model based on the truncated Gamma distribution

We assume that the proportion p of potential buyers has average long run rates of buying which follow a truncated Gamma distribution. The truncation is at the lower end at some point 1/T where T is a very long time interval.

Thus the distribution is given by

$$dF = \frac{(\frac{1}{a})^k \, e^{-\lambda/a} \, \lambda^{k-1} \, d\lambda}{I(T,a,k)} \qquad \frac{1}{T} \leq x \leq \infty \ .$$

where $$I(T,a,k) = \int_{\frac{1}{T}}^{\infty} (\tfrac{1}{a})^k \, e^{-\lambda/a} \, \lambda^{k-1} \, d\lambda$$

Thus we now have a model which has four parameters, namely, p,T,a,k. In fact we will show later that the model is insensitive to T, provided that it is very large, so that there are really only three parameters.

Special cases

(i) When $T = \infty$ and $k > 0$ the distribution of purchases in a particular time-period is simply a mixture of zeros and a NBD.

For then $I(\infty,a,k) = \Gamma(k) \quad (k > 0)$

so that the c.d.f. of the average long run rate of buying is given by

$$F(x) = q \qquad\qquad x = 0$$

$$F(x) = q + p \, \frac{\int_0^x (\frac{1}{a})^k \, e^{-\lambda/a} \lambda^{k-1} \, d\lambda}{\Gamma(k)} \qquad\qquad x > 0$$

(ii) For $T < \infty$ and $k = 0$ we shall see in Chapter 7 that we get a mixture of zeros and a Logarithmic distribution.

(iii) For $T < \infty$ and $k > 0$ we shall see that, provided T is very large, the same prediction formulae result as in case (i) when $T = \infty$. The parameter T is only introduced so that case (ii) can be dealt with.

Lemma $I(T, f(a), j+k) \simeq \Gamma(j+k)$.

for $j \geq 1$, $k \geq 0$, $f(a) > 0$ and T very large.

Proof $e^{-\lambda f(a)} \leq 1$ for $\lambda \geq 0$

$\lambda^{j+k-1} \leq 1$ for $j \geq 1$, $k \geq 0$, and $0 < \lambda < 1$.

$$\int_0^{\frac{1}{T}} \left[f(a)\right]^{j+k} e^{-\lambda f(a)} \lambda^{j+k-1} \, d\lambda$$

$$\leq \left[f(a)\right]^{j+k} \times \frac{1}{T} \quad \left[\text{for } \frac{1}{T} < 1\right]$$

$$\rightarrow 0 \text{ as } T \rightarrow \infty$$

But $$\Gamma(j+k) = \int_0^{\infty} \left[f(a)\right]^{j+k} e^{-\lambda f(a)} \lambda^{j+k-1} \, d\lambda$$

$$= I(T, f(a), j+k) + \int_0^{\frac{1}{T}} \left[f(a)\right]^{j+k} e^{-\lambda f(a)} \lambda^{j+k-1} \, d\lambda$$

for $j \geq 1$, $k \geq 0$.

$I(T, f(a), j+k) \rightarrow \Gamma(j+k)$ as $T \rightarrow \infty$.

We will use this lemma repeatedly in the following derivations particularly in the form

$$I(T, f_1(a), j+k) = I(T, f_2(a), j+k)$$

for any $f_1(a)$, $f_2(a)$.

The lemma is of course 'obvious' when $f(a) = a$.

Repeat buying formulae. Throughout the following derivation it will be assumed that if $T$ is finite then $k \geq 0$, but if $T = \infty$ then $k$ is strictly greater than 0, for we have $I(\infty, a, k) = \Gamma(k)$ only exists for $k > 0$.

Note that the Gamma distribution is specified in terms of $a$ and $k$. The parameter $m$ is used as before to denote the mean of the whole population, but it is no longer equal to $a \cdot k$ as it is in the simple NBD.

The probability that a person buys $j$ units in some time-period is given by

$$P(j) = \frac{p \displaystyle\int_{\frac{1}{T}}^{\infty} \left(\frac{1}{a}\right)^k e^{-\lambda/a} \lambda^{k-1} e^{-\lambda} \frac{\lambda^j}{j!} \, d\lambda}{I(T,a,k)} \quad \text{for } j > 0$$

$$\simeq \frac{p\left(\frac{1}{a}\right)^k \left(\frac{a}{1+a}\right)^{j+k} \Gamma(j+k)}{I(T,a,k)\ j!} \quad \text{by lemma}$$

Then the mean of the quantity bought by new buyers (or by lost buyers) is given by

$$m_N = P(1)$$

$$= \frac{p\left(\frac{1}{a}\right)^k \left(\frac{a}{1+a}\right)^{k+1} \Gamma(k+1)}{I(T,a,k)}$$

But the mean of the quantity bought by all buyers is given by

$$m = p \int_{\frac{1}{T}}^{\infty} \frac{(\frac{1}{a})^k \, e^{-\lambda/a} \, \lambda^k \, d\lambda}{I(T,a,k)}$$

$$= \frac{p \, a \, \Gamma(k+1)}{I(T,a,k)}$$

$$m_N/m = \frac{(\frac{1}{a})^k \, (\frac{a}{1+a})^{k+1}}{a}$$

$= 1/(1+a)^{k+1}$ as obtained for the simple NBD model.

In particular

$$m_N/m = \frac{1}{1+a} \quad \text{when } k = 0.$$

We also have $m_R/m = 1 - m_N/m$

where $m_R$ = mean of quantity bought by repeat buyers.

The proportion of buyers b is given by

$$b = \frac{p \int_{\frac{1}{T}}^{\infty} (\frac{1}{a})^k \, e^{-\lambda/a} \, \lambda^{k-1} \, (1-e^{-\lambda}) \, d\lambda}{I(T,a,k)}$$

$$= \frac{p[I(T,a,k) - (\frac{1}{a})^k \, (\frac{a}{1+a})^k \, I(T,\frac{a}{1+a},k)]}{I(T,a,k)}$$

$$\simeq p[1 - (1+a)^{-k}] \qquad \text{for } k > 0.$$

The proportion of new buyers is given by

$$b_N = p \int_{\frac{1}{T}}^{\infty} \frac{(\frac{1}{a})^k e^{-\lambda/a} \lambda^{k-1} e^{-\lambda} (1-e^{-\lambda})d\lambda}{I(T,a,k)}$$

$$= \frac{p[(1+a)^{-k} I(T,\frac{a}{1+a},k) - (\frac{1}{a})^k (\frac{a}{1+2a})^k I(T,\frac{a}{1+2a},k)]}{I(T,a,k)}$$

$$\simeq p[(1+a)^{-k} - (1+2a)^{-k}] \qquad \text{for } k > 0$$

Thus $\frac{b_N}{b} = \frac{b_L}{b} = \frac{(1+a)^{-k} - (1+2a)^{-k}}{1 - (1+a)^{-k}}$ when $k > 0$

as obtained for the simple NBD model

But when $k = 0$ we have

$$b = p \int_{\frac{1}{T}}^{\infty} \frac{e^{-\lambda/a} \frac{1}{\lambda} (1-e^{-\lambda})d\lambda}{I(T,a,0)}$$

$$= p \int_{\frac{1}{T}}^{\infty} \frac{e^{-\lambda/a}(1 - \lambda/2! + \lambda^2/3! - \lambda^3/4! + \ldots)d\lambda}{I(T,a,0)}$$

$$\simeq p \frac{\sum_{j\geq 1} a^j \frac{\Gamma(j)}{j!} (-1)^{j+1}}{I(T,a,0)} \qquad (\text{for } T \text{ large})$$

$$= p \frac{\log_e(1+a)}{I(T,a,0)}$$

When $k = 0$ we also have

$$b_N = p \int_{\frac{1}{T}}^{\infty} \frac{e^{-\lambda/a} \frac{1}{\lambda} e^{-\lambda} (1-e^{-\lambda})d\lambda}{I(T,a,0)}$$

$$\doteq p \int_{\frac{1}{T}}^{\infty} \frac{e^{-\lambda(1+\frac{1}{a})} \left(1 - \frac{\lambda}{2!} + \frac{\lambda^2}{3!} \ldots\ldots\right)}{I(T,a,0)} d\lambda$$

$$= \frac{p \sum \left(\frac{a}{1+a}\right)^j \frac{\Gamma(j)}{j!} (-1)^{j+1}}{I(T,a,0)}$$

$$= \frac{p \log\left(1 + \frac{a}{1+a}\right)}{I(T,a,0)}$$

Thus $^{b}N/b = \dfrac{\log\left(\frac{1+2a}{1+a}\right)}{\log(1+a)} \quad = {}^{b}L/b$

We also have $^{b}R/b = 1 - {}^{b}N/b = 1 - {}^{b}L/b$

where $b_R$ = proportion of repeat buyers.

<u>Prediction over a longer period</u>

Over one time-period we have a frequency distribution specified by p, T, a, k. Over a time-period a times as long the mean of the distribution and hence the mean of the underlying truncated Gamma distribution will also multiply by c, (under stationary conditions).

We will show that the frequency distribution over the longer period is specified by p,T,ca,k.

The mean of the truncated Gamma distribution is given by

$$p \int_{\frac{1}{T}}^{\infty} \frac{\left(\frac{1}{a}\right)^k e^{-\lambda/a} \lambda^{k-1} \lambda d\lambda}{I(T,a,k)}$$

$$\simeq \frac{pa\,\Gamma(k+1)}{I(T,a,k)}$$

In a time-period c times as long k will remain constant. We will also assume that T remains constant. Let a increase to become A.

The mean of the new truncated Gamma distribution is given by

$$\text{New mean} \quad \frac{pA\,\Gamma(k+1)}{I(T,A,k)}$$

$$= c \times \text{mean of single period}$$

$$= \frac{cpa\,\Gamma(k+1)}{I(T,a,k)}$$

$$\text{Thus } A = \frac{ca\; I(T,A,k)}{I(T,a,k)}$$

$$\text{Now } \frac{I(T,A,k)}{I(T,a,k)} = \frac{\int_{\frac{1}{T}}^{\infty} (\frac{1}{A})^k\, e^{-\lambda/A}\, \lambda^{k-1}\, d\lambda}{\int_{\frac{1}{T}}^{\infty} (\frac{1}{a})^k\, e^{-\lambda/a}\, \lambda^{k-1}\, d\lambda}$$

$$= \frac{\int_{\frac{1}{AT}}^{\infty} e^{-x}\, x^{k-1}\, dx}{\int_{\frac{1}{aT}}^{\infty} e^{-x}\, x^{k-1}\, dx}$$

Case (1): If $T = \infty$ and $k > 0$

then $I(\infty, A, k) = I(\infty, a, k) = \Gamma(k)$.

$$A = c\,a.$$

Case (2): If $T < \infty$ and $k > 0$

$$\int_{\frac{1}{aT}}^{\infty} e^{-x}\, x^{k-1}\, dx \rightarrow \Gamma(k) \text{ as } T \rightarrow \infty$$

$$\int_{\frac{1}{AT}}^{\frac{1}{aT}} e^{-x}\, x^{k-1}\, dx \rightarrow 0 \quad \text{as } T \rightarrow \infty$$

$$I(T,A,k) \rightarrow I(T,a,k) \text{ as } T \rightarrow \infty$$

$$A \simeq ca$$

Case (3): If T is finite and $k = 0$.

Consider $\displaystyle\int_{\frac{1}{AT}}^{\frac{1}{aT}} e^{-x}/x\, dx$

$$e^{-x} < 1 \quad \text{for } x < 1$$

$$\frac{1}{x} \leq AT \text{ for } \frac{1}{AT} \leq x \leq \frac{1}{aT}$$

$$\int_{\frac{1}{AT}}^{\frac{1}{aT}} e^{-x}/x\, dx \leq AT\, \frac{1}{aT} = A/a.$$

But $\displaystyle\int_{c}^{\infty} e^{-x}/dx \rightarrow \infty$ as $c \rightarrow 0$

$$\frac{\displaystyle\int_{\frac{1}{AT}}^{\frac{1}{aT}} e^{-x}/x\, dx}{\displaystyle\int_{\frac{1}{aT}}^{\infty} e^{-x}/x\, dx} \rightarrow 0 \text{ as } T \rightarrow \infty$$

$I(T,A,k) \simeq I(T,a,k)$ for T large

$A \simeq ca.$

Now when $k > 0$ the rate of buying per buyer over 1 time-period is given by

$$w = m/b = \frac{pa\,\Gamma(k+1)}{I(T,a,k)} \qquad \frac{1}{p[1-(1+a)^{-k}]}$$

$$\simeq \frac{ak}{[1-(ka)^{-k}]}$$

as obtained for the simple NBD model.

Thus in a time-period c times as long

$$w_c = \frac{cak}{1-(1+ca)^{-k}}$$

or $$\frac{w_c}{w} = \frac{c[1-(1+a)^{-k}]}{[1-(1+ca)^{-k}]}$$

We also have $\dfrac{b_c}{b} = \dfrac{1-(1+ca)^{-k}}{1-(1+a)^{-k}}$

But when $k = 0$ we have

$$\begin{aligned} w &= m/b \\ &= \frac{pa\,\Gamma(1)}{I(T,a,0)} \quad \frac{I(T,a,0)}{p\,\log_e(1+a)} \\ &= \frac{a}{\log(1+a)} \end{aligned}$$

Then in a time-period c times as long

$$w_c = \frac{ca}{\log(1+ca)}$$

$$\text{Thus } \frac{w_c}{w} = \frac{c \log(1+a)}{\log(1+ca)}$$

$$\text{Similarly } \frac{b_c}{b} = \frac{\log(1+ca)}{\log(1+a)}$$

Insensivity to T. When T is finite all the formulae which have been derived are approximations. But provided T is 'very large' the approximations are all very close even for the case $k = 0$. Thus provided T is indeed very large the formulae are insensitive to small changes in T. In fact we never need to know what T is, nor do we need to estimate it to calculate any of the above predictions.

T is introduced to describe our intuitive knowledge that there is some very small rate of buying below which people are effectively 'never-buyers'. As a by-product it enables us to obtain predictions for the special case $k = 0$.

Thus, although the model apparently has four parameters, it really has only three which matter, namely p,k,a.

# CHAPTER 5. Fitting the truncated NBD.

## 5.1 Introduction

We will consider a special case of the general model proposed in the previous chapter, by putting $T = \infty$ and $k > 0$. Thus we now have a 3-parameter model which is simply a mixture of zeros and a NBD.

### Estimation of Parameters.

The problem is tackled by noting that the truncated part of the distribution (that is the distribution without the zeros) should be a truncated NBD under our model. So the problem reduces to estimating the NBD parameters a, k from the truncated distribution. This will give an estimate of the potential buyers who are included in the zeros, but are part of the NBD. This will give an estimate of p.

### Intuitive Motivation.

For some classes of goods it is not clear what the population total (and hence the number of non-buyers) is. Thus the potential market for a brand of tipped cigarettes may be all adults, all smokers, or merely all smokers of tipped cigarettes.

If it is found that the truncated distribution really is NBD then we will have a revised population total which may have the above-mentioned meaning; namely the potential market for the brand under consideration.

## Methods

Various ways of estimating the parameters of a truncated NBD are considered. Methods from the literature are reviewed and then some new methods are given. The methods are compared in terms of efficiency and simplicity.

The maximum likelihood method, though fully efficient, is very laborious to perform, but is probably worth adopting by the statistician if a computer is available. But in market research this is sometimes not so and a simpler method is to be preferred.

Eventually, Brass's first method is selected as being the 'best' (by balancing efficiency and simplicity) and is tried on a large number of distributions.

## Notation

$f_r$ = Number of people who buy r units in a certain time-period

$f_o^{x}$ = Adjusted number of non-buyers, estimated from the truncated distribution.

The NBD is given by

$$P_r = (1+a)^{-k} \frac{\Gamma(k+r)}{r!\,\Gamma(k)} \cdot \left(\frac{a}{1+a}\right)^r$$

$=$ Probability of observing $r$.

Let $m = a.k$

$=$ mean of the NBD part of the total distribution.

Note that $m$ is not the mean of the observed distribution. Thus $m$ is the mean of the distribution $(f_o^{*}, f_1, f_2, f_3 \ldots$

The truncated NBD is given by

$$P_r = (1+a)^{-k} \frac{\Gamma(k+r)}{r!\,\Gamma(k)} \left(\frac{a}{1+a}\right)^r \frac{1}{1-(1+a)^{-k}} \qquad r \geq 1$$

$$F_o = \sum_{r \geq 1} f_r \qquad F_1 = \sum_{r \geq 1} r f_r$$

$$F_2 = \sum_{r \geq 1} r^2 f_r \qquad F_3 = \sum_{r \geq 1} r^3 f_r.$$

In all the methods our sample consists of $f_1, f_2, f_3 \ldots$

We want to estimate the parameters $a, k$ of the distribution and also $f_o^{*}$.

## 5.2 Methods from the Literature.

The truncated NBD was discussed by David and Johnston (1952). They give the maximum Likelihood method for estimating the parameters which is unfortunately very cumbersome. They also give a method which involves the ratios of the first 3 product moments. This method, while providing explicit solutions, is very inefficient because of the use of the 3rd moment which is very sensitive to outlying values.

The equations are

$$1 + \hat{a} = \frac{F_3F_1 - F_2^2}{F_1(F_2 - F_1)}$$

$$\hat{k} = \frac{2F_2^2 - F_2F_1 - F_3F_1}{F_3F_1 - F_2F_1 + F_1^2 - F_2^2}$$

Sampford (1955) gives iterative methods for finding the moment and maximum Likelihood estimates. These are rather complicated.

Brass (1958) gives 2 much simpler methods. The first gives explicit solutions for the parameters in terms of the first 2 product moments and $f_1$.

The equations are

$$\hat{w} = \frac{1}{1+a}$$

$$= \frac{F_1}{F_2 - F_1^2/F_0} \left(1 - f_1/F_0\right)$$

and $$\hat{k} = \frac{\hat{w}\; F_1/F_0 - f_1/F_0}{1 - \hat{w}}$$

This method is more than 60 per cent efficient over a wide range of parameter values.

Brass's second method is a modification of the maximum Likelihood method. This, while more efficient, is still rather laborious.

Hartley (1958) also considers the maximum Likelihood method. He suggests guessing the missing zero value, fitting by maximum Likelihood, re-estimating the zero value and iterating. Unfortunately, the maximum Likelihood method is still very cumbersome even for the full distribution.

The best of these methods seems to be Brass's first method. But various new methods were considered to see if a better one could be found.

## 5.3 A graph of ${}^{f_r}/f_{r-1}$ against $\frac{1}{r}$

The recurrence relationship for the NBD is

$$P_r = \left(\frac{a}{1+a}\right)\left(1 - \frac{1-k}{r}\right) P_{r-1}$$

So, theoretically, plotting ${}^{f_r}/f_{r-1}$ against $\frac{1}{r}$ should give a straight line which intercepts the ${}^{P_r}/P_{r-1}$ axis at $\frac{a}{(1+a)}$ and has a gradient of $\frac{a(1-k)}{1+a}$.

In practice the ratios are very variable with successive readings highly negatively correlated. This correlation is less than $-\frac{1}{2}$ for all r. A typical series of ratios is shown for brand $C_1$ over 26 weeks, which had a sample size of 2000 (see ~~appendix~~ p 137). Even for this sample size the ratios were much too variable to fit a line satisfactorily. Thus, this method is too inefficient.

Table 5a.

| r | ${}^{f_r}/f_{r-1}$ |
|---|---|
| 2 | .38 |
| 3 | .64 |
| 4 | .75 |
| 5 | .73 |
| 6 | .88 |
| 7 | .71 |
| 8 | .47 |
| 9 | .71 |
| 10 | .20 |
| 11 | 3.0 |

## 5.4 Geometric Prediction for a.

The recurrence relationship for the NBD is

$$P_r = \left(\frac{a}{1+a}\right)\left(1 - \frac{1-k}{r}\right) P_{r-1} .$$

As $r$ gets large $\frac{1-k}{r} \longrightarrow 0$

so that $P_r \longrightarrow \frac{a}{1+a} P_{r-1}$ .

Thus for r large the tail of the distribution tends to a GEOMETRIC distribution.

Thus we can estimate a by a similar method to that proposed in section 4.4 when the model comprising a mixture of a Geometric and a Poisson distribution was considered.

That is, we consider Mill's ratio $= \dfrac{f_r}{\sum_{j\geq r} f_j}$

which for $r$ large tends to $\frac{1}{(1+a)}$ .

As before we can find this for 1 particular value for r by calculating a smoothed value

$$\tilde{f}_r = \frac{f_{r-2} + f_{r-1} + f_r + f_{r+1} + f_{r+2}}{5}$$

Then $\frac{1}{1+a} = \tilde{f}_r / \sum_{j\geq r} f_j$ .

Fifteen is a suitable value for $r$ as Mill's ratio is roughly constant for the observed distributions for values of $r$ greater than 12.

Alternatively, we can plot Mill's ratio against $\frac{1}{r}$. It should converge to a constant which can be found by taking the intercept of the graph when $\frac{1}{r}$ is zero.

Having determined a, we can now use an iterative method to find the mean $m$ (= a.k) of the NBD part of the distribution and the adjusted number of zeros.

Guess an initial value for the adjusted number of zeros - $f_o^{(1)}$. The choice $f_o^{(1)} = 2f_1 - f_2$ is a convenient starting point.

Then $N^{(1)}$ = guessed sample size

$$= F_o + f_o^{(1)}$$

$$m^{(1)} = a\,k^{(1)} = F_1/N^{(1)} \quad .$$

Hence $f_o^{(2)} = N^{(1)}(1+a)^{-k^{(1)}}$.

In general this will be different from the guessed number of zeros.

Calculate a 2nd iteration using $f_o^{(2)}$ as the starting point.

In general the $i^{th}$ iteration is

$$N^{(i)} = F_o + f_o^{(i)}$$

$$m^{(i)} = a\,k^{(i)} = F_1/N^{(i)}$$

$$f_o^{(i+1)} = N^{(i)}\,(1 + a)^{-k^{(i)}} \quad .$$

Example

Brand $P_2$ over 24 weeks. (See table 6a)

Now $\sum_{r \geq 15} f_j = 28$

$$\tilde{f}_{15} = \frac{f_{13} + f_{14} + f_{15} + f_{16} + f_{17}}{5} = \frac{9}{5}$$

Therefore $1 + \hat{a} = 28 \times \frac{5}{9} = 15.5$

$\hat{a} = 14.5.$

$$F_o = \sum_{r \geq 1} f_r = 103$$

$$F_1 = \sum_{r \geq 1} r f_r = 1040.$$

Choose $f_o^{(1)} = 2f_1 - f_2 = 37.$

Then the iteration proceeds as follows.

**Table 5b**

| $f_o$ | N | m | m/a = k |
|---|---|---|---|
| 37 | 140 | 7.42 | .512 |
| 34.5 | 137.5 | 7.57 | .522 |
| 32.8 | | | |

The process converges to a value

$f_o^{*} = 32$ with $m = 7.65$ and $a = 14.5.$

Criticism.

The trouble with this method is that it can only be used for frequency distributions with a long tail. In particular we need a substantial number of people to buy more than 15 units in order to be able to estimate $a$. But we are looking for a method which is suitable for all consumer purchase distributions, which means that the above method is not general enough. There are, of course, cases where different methods of estimation are used for different types of series, but if we can find an estimation method which is at least as good for longer time-periods and which is also suitable for shorter time-periods then it will be preferred to the above method.

## 5.5 Estimation from the mean and $f_1$

Another method of estimating the parameters of the truncated NBD is to use $w$ and $f_1$.

Equates the observed and theoretical values of these quantities. Thus we have

$$w = \frac{ak}{1 - (1+a)^{-k}} = \frac{F_1}{F_o} = \text{rate of buying/buyer}$$

$$f_1 = \frac{F_o(1+a)^{-k}}{1 - (1+a)^{-k}} \times \frac{ak}{1+a} = \text{number of people who buy 1 unit.}$$

These 2 equations can be solved iteratively for $a, k$. The easiest method of iteration is accomplished, as before, by completing the distribution with a guessed number of zeros - $f_o^{(1)}$. For consumer research data the observed number of non-buyers is a convenient starting point. The reason for choosing a different starting point to that used in the previous method is simply that for consumer research data the estimates of $f_o^{*}$ obtained by this method are consistently higher.

We have $N^{(1)} = F_o + f_o^{(1)}$.

Then $m^{(1)} = a\,k^{(1)}$

$$= F_1/N^{(1)}.$$

Then estimate a from the guessed proportion of zeros, that is, from the equation

$$f_o^{(1)} = N^{(1)} \; (1+a)^{-m/a}.$$

This equation can be solved iteratively in the usual way.

Put $z_1 = f_1 - N^{(1)} \; (1+a)^{-k} \; \dfrac{m}{1+a}$

= difference between observed and theoretical buyers of 1 unit.

The value of the 2nd guess of the number of zeros ($f_o^{(2)}$) depends on the sign of $z_1$.

If $z_1 > 0$ choose $f_o^{(2)} = 2f_o^{(1)}$

If $z_1 < 0$ choose $f_o^{(2)} = \frac{1}{2} f_o^{(1)}$.

Repeat the process and calculate $z_2$ where

$$z_2 = f_1 - (1+a)^{-k} \frac{m}{1+a} N^{(2)}$$

and $m$, $a$ are new estimates calculated from $f_o^{(2)}$.

We can now interpolate linearly for $z_3 = 0$.

This will give a value for $f_o^{(3)}$ and hence an estimate of $f_o$, $a$ and $k$.

Example

Brand $P_2$ over 24 weeks. (see table 6a. p114).

$F_o = 103$ $\quad F_1 = 1040$ $\quad f_1 = 25$.

Choose $f_o^{(1)} = 371$ = observed number of non-buyers.

Then $m^{(1)} = 2.19$. $\quad a^{(1)} = 30.5$

$z_1 = -1$.

As $z_1 < 0$, choose $f_o^{(2)} = \frac{1}{2}f_o^{(1)} = 185$.

Then $m^{(2)} = 3.61$ $\quad a^{(2)} = 27.3$.

$z_2 = 3.8$.

Interpolating linearly for $z = 0$ we get

$$f_o^{(3)} = 371 - {}^{186}/_{4.8}$$

$$= 332.$$

This gives estimates as follows:-

$$f_o^* = 332$$

$$m = 2.40$$

$$a = 30.5.$$

<u>Comments</u>

There are 2 snags connected with this method.

Firstly, this method will often increase the adjusted number of non-buyers above the observed number. Then it will be impossible to give any physical meaning (e.g. the potential market) to the revised population total.

Full NBD's were fitted in the normal way to 19 observed 'heavy-buying' distributions, which included the observed number of non-buyers and the ratio

$$y = \frac{\text{Theoretical value of } f_1}{\text{Observed value of } f_1} \text{ was calculated.}$$

The average value of $y$ was 1.02 and 14 of the ratios were in the range (0.8, 1.2). Thus despite the variance discrepancy, $f_1$ seems to fit quite well. Thus for these distributions the above method will give a theoretical distribution similar to the original NBD. Other values of $y$ were as far apart as 0.73 and 1.69 so that for these distributions the theoretical distribution would change quite a bit.

Low values of $y$ ($< 1$) lead us to the 2nd snag, for if the observed value of $f_1$ is bigger than some value which depends on $w$ and $F_o$, then this method will not have a solution. As we shall see in a later chapter a Logarithmic distribution can be fitted to the truncated distribution. Then if $f_1$ > theoretical value of $f_1$ in the Logarithmic distribution then this method will not have a solution.

(that is if $f_1 > F_o \dfrac{-q}{\log_e(1-q)}$ where $w = \dfrac{-q}{(1-q)\log_e(1-q)}$ ).

We need an estimation procedure which always gives a solution for heavy-buying distributions, so that this method is inapplicable.

## 5.6. Moment Method.

Another method of estimation is to use the first 2 moments of the truncated distribution. This method has already been considered by Sampford (1955), but his method is rather complicated. A much simpler method has been derived.

The technique is similar to that of Hartley (1958). After guessing the missing zero value, the full distribution is fitted by the method of moments. Then we re-estimate the zero value and iterate. With interpolation 3 steps usually gives good enough accuracy. We must have agreement between theoretical and adjusted zeros so that there is agreement between the theoretical and observed moments of the truncated distribution.

First we guess a suitable initial zero value - $f_o^{(1)}$.

A convenient starting point is given by

$$f_o^{(1)} = 2f_1 - f_2.$$

We then calculate the 2 parameters $m( = a.k)$ and $a$ by the method of moments.

$$N_1 = F_o + f_o^{(1)} = \text{sample size.}$$

$$m_1 = \text{mean of adjusted distribution}$$

$$= F_1/N_1$$

$$m(1+a) = \text{Variance of adjusted distribution}$$

$$= F_2/N_1 - m^2.$$

Therefore $a_1 = F_2/F_1 - m_1 - 1$

$$k_1 = {}^{m}1/a_1 \quad .$$

Then we can find a new estimate of the number of non-buyers, namely $N_1(1+a_1)^{-k_1}$ .

We could use this new estimate as the starting point for a 2nd iteration. But simple iteration in this way may require up to 20 steps. The process can usually be shortened to 3 steps by interpolation.

Scores $z_i$ are computed from trial values of $f_o^{(i)}$ in the equation

$$z_i = f_o^{(i)} - N_i(1+a_i)^{-k_i}.$$

The first score, $z_1$, is computed with the 1st trial value, $f_o^{(1)}$. The 2nd trial value depends on the sign of $z_1$. If $z_1$ is positive then choose $f_o^{(2)} < f_o^{(1)}$ ; if negative, choose $f_o^{(2)} > f_o^{(1)}$. Whichever it is choose $f_o^{(2)}$ far enough from $f_o^{(1)}$ preferably to give opposite signs to $z_1$ and $z_2$. This can usually be achieved by putting $f_o^{(2)} = f_o^{(1)}/2$ or $f_o^{(2)} = 2f_o^{(1)}$ according as to whether $z_1 > 0$ or $z_1 < 0$. The 3rd trial value $f_o^{(3)}$ is obtained by linear interpolation for $z = 0$. This 3rd trial value is adjudged to be sufficiently accurate if we find $z_3/f_o^{(3)} < 0.02$, a condition which is nearly always satisfied in practice.

Existence and convergence problems.

The iteration procedure is equivalent to

$$f_o^{(i+1)} = N_i(1+a_i)^{-k_i}$$

where

$$N_i = F_o + f_o^{(i)}$$

$$m_i = F_1/N_i$$

$$a_i = F_2/F_1 - m_i - 1$$

$$k_i = m_i/a_i \; .$$

Thus $f_o^{(i+1)} = g(f_o^{(i)})$ where $g$ is the above function.

Thus our estimate of $f_o^*$ is the root of the following equation in $x$.

$$x = g(x).$$

This equation will not always have a solution.

It is easy to show that $g(x) > x$ when $x = 0$ for all distributions. But distributions also exist where $g(x) > x$ for all $x$. A trivial example occurs when all the buyers buy the same number of units. A more important example occurs for a class of reverse J-shaped distributions.

If $F_2/F_1$ is greater than a certain value which depends on $w$ then no solution will exist. As we shall see later, a Logarithmic distribution can be fitted to the truncated distribution. In this case the theoretical value of ${}^{F_2}/F_1$ is $^1/_{(1-q)}$ where the parameter $q$ is estimated from $w = \frac{-q}{(1-q)\log_e(1-q)}$ . If the observed value of ${}^{F_2}/F_1$ is bigger than $^1/_{(1-q)}$ then this moment method will not have a solution. In consumer purchase distributions this only occurs when a full NBD is fitted to an observed distribution by the mean and zeros and the observed variance is higher than the theoretical variance, The reverse is usually true, particularly for heavy-buying data, when the variance discrepancy occurs and the theoretical variance is consistently higher than the observed value. In such cases as these a solution will exist.

We will show that $g(x)$ is continuous and monotonically increasing, so that, if a solution exists, the iterative procedure must converge to one of the roots.

Now if $f_o^{(i)}$ is monotonic increasing

$m_i$ is monotonic decreasing

Therefore $a_i$ is monotonic increasing

But $\log(1+a_i)/a_i$ is monotonic decreasing

But $(1+a_i)^{-m_i/a_i} = e^{-m_i/a_i \log(1+a_i)}$ .

Hence $(1+a_i)^{-m_i/a_i}$ is monotonic increasing.

Also $N = F_o + f_o^{(i)}$ is monotonic increasing

Therefore $f_o^{(i+1)} = N(1 + a_i)^{-m_i/a_i}$ is monotonic increasing.

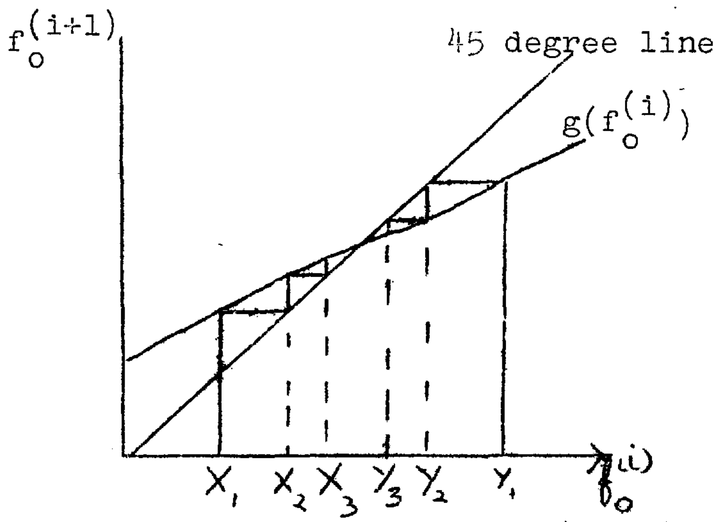
Consider $z = x - g(x)$.

$z$ is a continuous function for $x > 0$.

The monotonic property means that the iterative procedure must converge to one of the roots of $z = 0$, provided, of course, that a root exists. (We have already seen that this is not necessarily so.) Note that if $g(x)$ was not monotonic then the process would not necessarily converge.

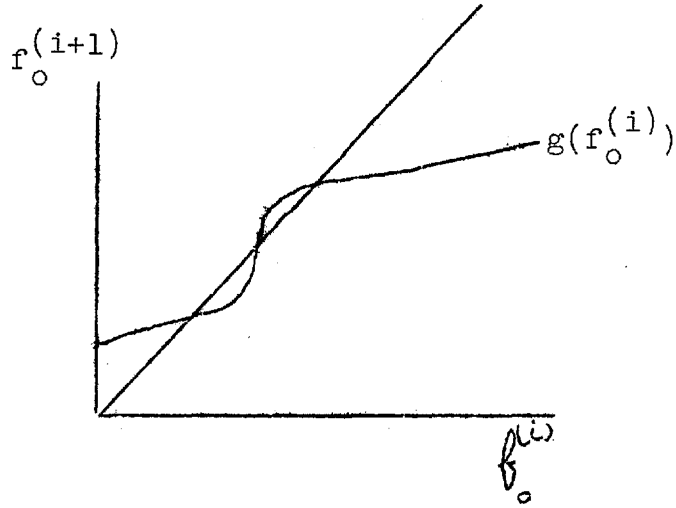
This is best illustrated graphically.

Diagram 1.




This diagram shows the situation if there is exactly 1 root. 2 typical iterative routes are shown, starting at $x_1$, $y_1$ respectively. The root of the equation occurs when $g(f_o^{(i)})$ cuts the 45 degree line; that is, when $f_o^{(i+1)} = g(f_o^{(i)})$.

Diagram 2.




This diagram show the uncommon case when the root is not unique. The iterative procedure will still converge to one of the roots.

Some distributions do give 3 roots. Thus for the distribution taken from Bliss (1953) (p.186) consisting of 128 ones, 37 twos, 18 threes, 3 fours and 1 five, we find roots at $x = 235$, $x = 250$ and $x = 255$. However, in the 3 examples where this was found to occur, the roots were relatively close together; that is, within 15 per cent of each other.

Worked Example

Our truncated distribution is taken from brand $P_1$ over 24 weeks (see appendix) When a full NBD was fitted, the variance discrepancy occurred, indicating that the method of moments would have a solution.

The calculation proceeds as follows.

(i) $F_o = 36 \quad F_1 = 347 \quad F_2 = 9139$

(ii) Choose $f_o^{(1)} = (2 \times 9) - 4 = 14.$

(iii) This gives

$$N_1 = 36 + 14 = 50$$

$$m = 347/50 = 6.94$$

$$a = 9139/347 - 6.94 - 1 = 18.46$$

$$k = {}^{m}/_{a} = 0.37$$

$$N(1+a)^{-k} = 17.1$$

Therefore $z_1 = 14 - 17.1 = -3.1$

iv) As $z_1 < 0$ take $f_o^{(2)} = 2 \times 14 = 28.$

This gives $z_2 = -0.2.$

As ${}^{z_2}/f_o^{(2)}$ this value of $f_o$ is sufficiently accurate. However, to complete the example we find linear interpolation gives

$$f_o^{(3)} = 14 + 3.9 \times \frac{28-14}{3.7}$$

$$= 29.$$

Hence $z_3 = +0.1.$

The calculation of $z_3$ also gives

$m = 5.32 \quad a = 20.1 \quad k = 0.265.$

Therefore $f_o^{*} = 29.$

Estimated NBD population total $= 29 + 36$

$= 65.$

Efficiency

Sampford (1955), who also considers the method of moments, but gives a more laborious method, derives formulae for the variances of the estimates which will also be applicable to my iterative method. They indicate that the method is more than 60 per cent efficient over a wide range of parameter values.

Criticism

This method seems to satisfy most of the criteria for a suitable estimation procedure. It is reasonably quick to carry out, is fairly efficient and will give solutions in all the distributions likely to be considered. Unfortunately, If compares badly with Brass's first method. Brass's method is quicker since it gives explicit estimates. It is also more efficient. The efficiency of a method of estimating two parameters is found by calculating the determinant of the variance-coveriance matrix of the estimates and comparing it with the corresponding determinant of the maximum Likelihood estimates. Brass gives a table which shows that the efficiency is greater than 60 per cent over a wide range of parameter values. In particular Brass shows that the efficiency is always better than that of the method of moments in the range of parameter values that we require.

Table 5b. Ratio of efficiency of estimation by Brass's method to the method of moments

| | | m | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.5 | 1 | 2 | 5 | 10 | $\infty$ |
| k | 0.5 | 1.13 | 1.18 | 1.22 | 1.27 | 1.28 | 1.00 |
| | 1 | 1.07 | 1.10 | 1.14 | 1.17 | 1.18 | 1.00 |

Thus Brass's method of estimation seems to be the best available.

## 5.7. Brass's method

The estimates of a,k are given by

$$1 + \hat{a} = \frac{F_2 F_o - F_1^2}{F_1 (F_o - f_1)}$$

$$\hat{k} = \frac{F_1 - (1+\hat{a}) f_1}{\hat{a} F_o}$$

$$f_o^* = \frac{F_o}{(1+\hat{a})^{\hat{k}} - 1}$$

As with the moment and maximum likelihood methods, it is quite easy to find distributions which will not give acceptable solutions. (An acceptable solution is $a > 0$, $k > 0$). In particular a similar restriction applies to reverse J-shaped distributions as for the method of moments.

However, for distributions of the 'heavy-buying' type where the variance discrepancy exists, solutions will exist.

Brass's method was used to fit a large number of distributions from the X-product field. As we shall see in the next chapter the resulting fit was not satisfactory in a number of important ways.

# CHAPTER 6

## The Fit of the Truncated NBD

### 6.1. Results of fitting the truncated NBD

A large number of distributions from product field X was fitted by Brass's method and the resulting fit was compared with that obtained by fitting a full NBD. All aspects of the fit were considered, and not just the fit of the frequency distribution. Certain systematic features were observed.

Firstly the adjusted number of non-buyers, $f_o^*$, was considerably less than the observed number of non-buyers, $f_o$. In most cases we found

$$f_o^* < f_o/10.$$

Secondly truncation causes a considerable increase in the value of the parameter k. In long time-periods k increases to about 0.5 but over shorter time-periods k increases to as much as 4. The other parameter a decreases but in such a way that the mean of the adjusted distribution ($m = a \, . \, k$) increases.

There are also systematic changes in the theoretical frequency distribution. The theoretical values of $f_1$ and $f_2$ are reduced as are the frequencies in the tail. This latter effect removes the variance discrepancy. Other frequencies are correspondingly increased.

The table shows a typical distribution from product-field X which has been fitted by a full NBD and also by Brass's method.

TABLE 6a

| r | Full NBD | Observed Distribution | Truncated NBD |
|---|---|---|---|
| 0 | 371 | 371 | 48.0 |
| 1 | 25.5 | 25 | 18.7 |
| 2 | 13.2 | 13 | 12.3 |
| 3 | 8.8 | 7 | 9.3 |
| 4 | 6.6 | 2 | 7.5 |
| 5-8 | 16.0 | 12 | 19.9 |
| 9-12 | 8.8 | 12 | 11.9 |
| 13-16 | 5.8 | 8 | 7.6 |
| 17-20 | 4.0 | 9 | 5.4 |
| 21-24 | 3.0 | 9 | 3.8 |
| 25+ | 11.3 | 6 | 6.6 |

On an 'overall' basis, the fit of the truncated distribution is improved by fitting a truncated NBD by Brass's method. For example, the value of $X^2$, calculated for the cells shown in the above table, is reduced from 27.2 to 19.5. Another way of seeing if the fit has improved is to look at the likelihood function.

We recall that our model is a mixture of zeros and a NBD. Thus the distribution is given by

$$P_o = q + p(1+a)^{-k}$$

$$P_r = p(1+a)^{-k} \frac{\Gamma(k+r)}{r!\,\Gamma(k)} \left(\frac{a}{1+a}\right)^r \qquad \text{for } r \geq 1.$$

Split the observed number of non-buyers, $f_o$, into 2 parts A and B where A are 'never-buyers' and B are part of the NBD. Then we can work out the likelihood of the sample for various values of A and B. When the likelihood is maximised, the value of B is the maximum likelihood estimation of $f_o^*$. This estimating procedure was mentioned in the previous chapter but was not adopted because it is too laborious.

For a particular value of A, and hence of B, we estimate

$$q = A/N \qquad \text{where } N = \text{population total}$$

$$m = F_1/(B+F_o) = \text{mean of NBD part.}$$

Estimate k is the maximum likelihood estimator of the NBD part.

Thus k is the root of

$$(B+F_o)\log\left(1+\frac{m}{k}\right) = \sum_{r=1}^{\infty} f_r \sum_{i=o}^{r-1} \frac{1}{k+i}$$

$$\text{Likelihood (sample } | B) = \prod_r P_r^{f_r}.$$

The likelihood when A is zero is taken as standard. Then the likelihood was calculated for various values of B. A computer program was written to do the necessary calculations. The table shows the likelihood of the distribution for brand $P_3$ over 4 weeks for various values of B. (See appendix).

TABLE 6b

| B | Likelihood |
|---|---|
| 387 | 1.00 |
| 348 | 1.46 |
| 309 | 1.81 |
| 270 | 2.33 |
| 232 | 3.17 |
| 193 | 4.57 |
| 154 | 6.95 |
| 116 | 10.67 |
| 96 | 12.56 |
| 77 | 13.25 |
| 58 | 10.90 |
| 38 | 5.26 |
| 19 | 0.79 |

$f_o$ = observed number of non-buyers = 3.87

The likelihood of the sample is maximised when about 77 of the non-buyers are assumed to be in the NBD.

Thus, in an overall way, the fit of the frequency distribution is improved when we reduce the number of zeros and fit the truncated NBD.

## 6.2. Appraisal of the truncated NBD

In Chapter 4 we mentioned 4 criteria by which any model should be judged, namely, simplicity, generality, usefulness and descriptiveness.

The better fit of the frequency distribution is just one aspect of this last criterion, namely, descriptiveness. We now consider other descriptive aspects of the truncated NBD.

Firstly let us consider the shelving effect which is visible in the example given in Table 6a. We have already noted that the full NBD does not describe this property. We now note that the truncated NBD does not describe it any better.

Secondly let us consider the adjusted number of non-buyers. If the revised population total is to have some meaning such as the potential market of the brand in question then this adjusted total should be roughly constant for a particular brand in different time-periods. In fact it increases steadily as the time-period increases. The adjusted population total is often less than the total number of buyers over a longer period. Thus it cannot be an estimate of the potential market for the brand in question.

Thirdly truncation caused a systematic reduction in the theoretical value of $f_1$. But we have already seen that the full NBD gives an unbiased estimate of $f_1$. So the truncated NBD will systematically underestimate $f_1$. This discrepancy is particularly serious as the people who buy only one unit form a large proportion of lost buyers and so this bias will seriously affect the repeat-buying formulae.

This brings us to another criterion, by which the model can be judged, namely usefulness. The most important aspect of this is the ability of the model to give useful predictions.

As mentioned in Chapter 2, a whole range of repeat-buying formulae have been developed. When a full NBD is fitted we have seen that these formulae give good predictions even when the variance discrepancy occurs. In particular the predictions are unbiased. However when the truncated NBD is fitted, $f_1$ is underestimated and the estimate of k increases so that the repeat-buying predictions change. For example the predicted number of lost buyers is considerably reduced. Thus the predictions will now be biased. Similarly predictions over different time-periods (for example market penetration) will also be biased. Thus, judged by its usefulness, the truncated NBD model is worse than the full NBD.

The third criterion which we will consider is that of simplicity. The full NBD model has two parameters but the truncated NBD model has three parameters, so that in this respect the latter model is not as simple as the full NBD model. Now the addition of a third parameter, with a consequent reduction in simplicity, can only be justified if it means that the model is better in other respects. But we have already seen that in a purely descriptive capacity the model, while improving the overall fit of the frequency distribution, leads to certain systematic discrepancies which decrease the usefulness of the model. Thus the addition of a third parameter cannot be justified.

## 6.3. Summary of position

Certain systematic discrepancies from the NBD model, notably shelving, bunching and the variance discrepancy, led us to consider a variety of alternative models, none of which was judged to be better than the NBD model.

This is perhaps not too surprising. For we have seen that the NBD model gives good predictions for a wide variety of brands over different time-periods even when systematic discrepancies occur in the frequency distribution. And from the market research point of view it is these predictions which are of the prime importance, provided that they are general and reasonably simple. The descriptive aspect, while important, is secondary.

## 6.4. The shelving phenomenon

It seems appropriate at this point to have a closer look at the shelving effect which is at the heart of the discrepancy from the NBD model. What is the cause of this effect?

First let us look at the assumption in the NBD model that the average long run rates of buying of different consumers follows a Gamma distribution. If this is incorrect then there may be some other compound Poisson distribution which will describe the data adequately. The observed frequency distributions contain a definite discontinuity which is usually very sharp indeed. Thus we require that the distribution of the average long run rates of buying should also have a definite discontinuity. As an example we will consider a 12 week period and assume that the average long run rates of buying follow a distribution which has a discontinuity between 12 and 13 units. A suitable distribution is the uniform distribution on (0,12). Then we expect a fairly sharp drop in the resulting frequency distribution from 12 to 13 units.

$$\text{Now } P(j \text{ purchases}) = \frac{1}{12} \int_0^{12} e^{-\lambda} \frac{\lambda^j}{j!} \, d\lambda$$

$$\frac{P_{13}}{P_{12}} = \frac{1}{12} \frac{\int_0^{12} e^{-\lambda} \lambda^{13} \, d\lambda}{\int_0^{12} e^{-\lambda} \lambda^{12} \, d\lambda}$$

$$= \frac{1}{12} \frac{\Gamma(14,12)}{\Gamma(13,12)}$$

$$= \frac{0.32}{0.43} = 0.74 \quad \text{(from } \chi^2 \text{ tables)}$$

However the observed drop is usually much more drastic than this and is of the order

$$\frac{P_{13}}{P_{12}} = 0.1$$

Thus a compound Poisson distribution is too smooth to give such a discontinuity.

Another way to show that no compound Poisson distribution could give this shelving effect is to consider the result given in Section 4.5 that the average rate of buying over the subgroup who bought j units in the first of two successive equal time-periods is given by

$$(j+1)\,\frac{P(j+1)}{P(j)} \text{ in the second period.}$$

But if there is a sudden discontinuity at j units so that $\frac{P(j+1)}{P(j)}$ is small, this would mean that this subgroup should buy at a very low rate in the second period.

For example if $j = 12$ and $\frac{P(j+1)}{P(j)} = 0.1$, then the average rate of buying of the subgroup is 1.3 in the second period after buying at an average rate of 12 in the first period. This situation is contrary to our marketing knowledge.

This must mean that the other basic assumption in the NBD model, namely that the purchases of any one consumer in successive time-periods are independent Poisson variates must be incorrect. An investigation of the buying habits of a large number of consumers revealed that while most consumers follow the Poisson assumption reasonably well, a few consumers are too regular in their buying habits.

Thus we must accept the fact that the NBD model is nothing more than a useful approximation to the real situation, and that certain systematic discrepancies will occur. Further light will be shed on the shelving effect in Chapter ten, when we examine the distributions of 'weeks' and occasions.

Brand. $P_3$

Time period. 4 weeks

| Number of units bought | Frequencies: Observed | Frequencies: Fitted NBD |
|---|---|---|
| 0 | 376 | 375.9 |
| 1 | 40 | 50.3 |
| 2 | 24 | 21.1 |
| 3 | 14 | 10.9 |
| 4 | 17 | 6.1 |
| 5 | 1 | 3.6 |
| 6 | 2 | 2.2 |
| 7+ | 0 | 3.9 |
| Variance | 1.13 | 1.54 |

N = 474 $\quad$ m = 0.45

a = 2.39

k = 0.19

# CHAPTER 7

## The LSD Model

### 7.1. Zeros and the NBD

We will investigate the effect of altering the number of zeros of a distribution which is already well fitted by a simple NBD.

Example: Consider the distribution given in Table 1b

$N$ = sample size = 2000 $\qquad$ $f_o$ = number of non-buyers = 1612.

$m$ = mean = 0.636 $\qquad$ $k$ = 0.05

standard deviation = 2.12.

Keep the positive part of the sample the same and let $N$ range from 750 to 20,000. Thus we make $f_o$ range from 362 to 19,612. In each case a NBD is fitted and the two parameters $m,k$ are estimated from the mean of the new distribution and from the new number of buyers. This gives a theoretical distribution which can be compared with the observed distribution.

Apart from the lowest value, the fit of $f_1$ and the standard deviation (s.d.) is good over a wide range of $N$ (and $f_o$). In fact the positive part of the fitted NBD is hardly changing at all. We will show that as $f_o \nearrow \infty$, the positive terms tend to a well-known series called the LOGARITHMIC Series distribution.

TABLE 7a

| N | $f_o$ | $P_o$ | Obs. $f_1$ | NBD $f_1$ | Obs. s.d. | Theoretical s.d. | k | a |
|---|---|---|---|---|---|---|---|---|
| 750 | 362 | .485 | 164 | 136 | 3.2 | 2.76 | 0.48 | 3.50 |
| 1000 | 612 | .612 | 164 | 145 | 2.86 | 2.6 | 0.29 | 4.34 |
| 1500 | 1112 | .740 | 164 | 153 | 2.42 | 2.29 | 0.16 | 5.14 |
| 2000 | 1612 | .806 | 164 | 156 | 2.12 | 2.04 | 0.11 | 5.53 |
| 5000 | 4612 | .922 | 164 | 162 | 1.38 | 1.36 | 0.04 | 6.23 |
| 10000 | 9612 | .961 | 164 | 164 | .98 | .97 | 0.02 | 6.45 |
| 20000 | 19612 | .981 | 164 | 165 | .70 | .69 | 0.01 | 6.57 |

Theorem. Consider any distribution $f_1, f_2 \ldots f_j$. Then for any $f_o$, estimate the two parameters of the NBD from the mean and the proportion of zeros. Then as $f_o \nearrow \infty$, $k \rightarrow 0$, and the positive part of the resultant theoretical distribution tends to a LOGARITHMIC series. (Hereafter the Logarithmic series distribution will be abbreviated to LSD).

Proof. Calculate $F_o = \sum_{r \geq 1} f_r$

$$F_1 = \sum_{r \geq 1} r f_r$$

For any $f_o$, $N = F_o + f_o$ = total sample size.

The estimate $\hat{m}$ of m is given by

$$\hat{m} = F_{1/N}$$

The estimate $\hat{k}$ of k is given by

$$(1 + \hat{m}/k)^{-k} = f_o/N.$$

Assume that $k \not\to 0$ as $f_o \nearrow \infty$

Then as k is positive, $k \geq c$ for some positive constant c.

Then $\exists$ N, such that $\frac{1}{Nk}$ is small.

Then $(1 + \hat{m}/k)^{-k} = (1 + {}^{F}1/Nk)^{-k}$

$$\rightarrow 1 - {}^{F}1/N, \text{ as } N \nearrow \infty$$

But $(1 + \hat{m}/k)^{-k} = f_o/N$

$$= (N-F_o)/N$$

$$= 1 - F_o/N$$

Hence $F_1 \rightarrow F_o$ which is not true.

$\therefore$ $k \rightarrow 0$ as $f_o \nearrow \infty$.

Now the recurrence relationship for the NBD is given by

$$P_r = \left(\frac{a}{1+a}\right)\left(1 - \frac{1-k}{r}\right) P_{r-1}$$

Thus $P_r \rightarrow \frac{a}{1+a} \frac{r-1}{r} P_{r-1}$ as $f_o \nearrow \infty$.

In other words, $P_o \rightarrow 1$ and $P_r \rightarrow 0$ for $r \geq 1$.

We will only consider the positive distribution, when we note that the recurrence relationship is that of the Logarithmic series distribution. This distribution is given by

$$P_r = \frac{1}{\ln(1+a)} \left(\frac{a}{1+a}\right)^r \frac{1}{r} \quad \text{for } r = 1,2,3,\ldots.$$

= Probability of observing a positive integer.

The parameter a changes as $N \nearrow \infty$. We will show that it tends to a positive quantity given by $\frac{a}{\ln(1+a)} = w =$ rate of buying per buyer.

a is estimated from the equation

$$(1+a)^{-\frac{\hat{m}}{a}} = 1 - b \text{ where b is proportion of buyers}$$

$$\frac{\hat{m}}{a} \ln(1+a) = -\ln(1-b)$$

$$= b + b^2/2 + b^3/3 + \dots.$$

As $f_o \nearrow \infty$, $b \searrow 0$

$$\frac{w}{a} \ln(1+a) = 1 + b/2 + b^2/3 + \dots.$$

Thus $\lim_{f_o \to \infty}$ a is given by

$$\frac{a}{\ln(1+a)} = w$$

## 7.2. The LSD model

Fisher (1943) derives the LSD by letting the population size tend to infinity and m,k tend to zero in such a way that their ratio $a = m/k$ stays finite. But this limiting process is meaningless as applied to the distribution of consumer purchases.

Instead we will derive the LSD as a special case of the general model proposed in Section 4.5.

We suppose that the population contains a proportion (1-p) of 'never-buyers' and a proportion p of buyers whose mean rates of buying follows a truncated Gamma distribution.

Thus $F(x) = 1-p \qquad 0 \leq x \leq \frac{1}{T}$

$$F(x) = (1-p) + \frac{p \int_{\frac{1}{T}}^{x} e^{-\lambda/a} \frac{1}{\lambda} \, d\lambda}{I(T,a,0)} \qquad \frac{1}{T} \leq x \leq \infty$$

In other words we consider the special case $k = 0$, with T finite.

This model is intuitively acceptable as one feels that there certainly are some people in the population who will never buy. Also one feels that there is a certain maximum time-period in which one is interested. (e.g. the life of the product).

As before we assume that each member of the population makes Poisson purchases in successive time-periods.

Then

$$\text{Prob}(j \text{ purchases}) = \frac{p \int_{\frac{1}{T}}^{\infty} e^{-\lambda/a} \frac{1}{\lambda} e^{-\lambda} \frac{\lambda^j}{j!} \, d\lambda}{I(T,a,0)}$$

$$\simeq \frac{p\,\Gamma(j)/j! \; (1 + \frac{1}{a})^{-j}}{I(T,a,0)} \quad \text{for } j \geq 1.$$

Thus $P_j = (\frac{a}{1+a}) \frac{j}{j-1} \; P_{j-1} \qquad$ for $j \geq 2$.

This is the recurrence relationship for the LSD.

Thus the people who actually make a purchase form a distribution given by

$$\frac{1}{\ln(1+a)} \left\{ \frac{a}{1+a} \, , \; (\frac{a}{1+a})^2/2 \, , \; (\frac{a}{1+a})^3/3 \, , \; \ldots\ldots \right\}$$

Thus for any time-period the population consists of a block of non-buyers (some of whom are potential buyers) and those members of the population who make one or more purchases in the time-period considered. The latter form a LSD.

<u>The parameter a</u>. It is important to realise that the estimates of a obtained by fitting an NBD and an LSD model will be different.

It can be shown that the maximum likelihood estimation of a for the LSD is given by equating observed and theoretical means. Let $a_L$ be the estimate from the LSD model. Then $a_L$ is obtained from the equation

$$w = \frac{a_L}{\ln(1+a_L)}$$

Let $a_N$ be the estimate of a obtained from the NBD model.

$$(1 + a_N)^{-m/a_N} = 1 - b$$

$$\frac{m}{a_N} \ln(1+a_N) = -\ln(1 - b)$$

$$= b + b^2/2 + b^3/3 \; \ldots .$$

$$w = \frac{m}{b} = \frac{a_N}{\ln(1+a_N)} \left[1 + b/2 + b^2/3 + \ldots .\right]$$

For b small enough, the two equations, and hence the two estimates, will be virtually the same.

## 7.3 The Logarithmic distribution

We will now present the main properties of the Logarithmic distribution which are relevant to the investigation of consumer purchasing.

The LSD is usually written

$$P_r = -\frac{1}{\ln(1-q)} \; q^r/r \qquad r = 1,2,3....$$
$$0 < q < 1.$$

= Probability of observing a positive integer r.

Thus $q = \frac{a}{1+a}$ .

Some authors use $\Theta$ instead of q.

It is customary to put $\alpha = \frac{-1}{\ln(1-q)}$.

The Probability generating function is given by

$$\phi(t) = -\alpha \ln(1 - qt)$$

The moment generating function is given by

$$M(t) = \phi(e^t)$$
$$= -\alpha \ln(1 - qe^t)$$

Hence mean $= \alpha q/(1 - q)$.

Variance$= \alpha q(1 - \alpha q)/(1 - q)^2$.

Reviews of the LSD are given by Williamson and Bretherton (1964) and Patil et al (1964).

Models. Fisher (1943) derived the LSD in connection with some work by Corbet and Williams on the distribution of species of butterflies caught in a light trap in a given period. As already mentioned, he obtained it as the limit of a NBD by letting $k \to 0$ and removing the zero class.

C. B. Williams (1944, 1947) published a series of papers on biological applications of the distribution.

He uses the notation

$f_r$ = No. of species with r individuals

$= \alpha q^r/r$ [This is a <u>different</u> $\alpha$]

Then $S$ = total number of species

$= -\alpha \ln(1 - q)$.

$N$ = total number of individuals inspected

$$= \frac{\alpha q}{1-q} .$$

He estimates $\alpha$ and $q$ from $S$ and $N$ giving the impression that there are two parameters.

He calls $\alpha$ the index of diversity and found that this was constant over different time periods.

Comparing purchasing results with the biological derivation we note

$$w \longleftrightarrow N/S$$

$$-\frac{F_o}{\ln(1-q)} \longleftrightarrow \text{index of diversity where}$$

$F_o$ = Number of buyers.

Thus $F_o/\ln(1-q)$ should be constant over ranging time-periods.

Three distributions from product field X were examined to see if this quantity was constant for purchase distributions.

TABLE 7b

| Time-Period (weeks) | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|
| 24 | 28.5 | 58 | 28.5 |
| 12 | 25 | 58 | |
| 8 | 22.5 | 60 | 21.5 |
| 4 | 27 | 67 | 20.3 |
| 2 | 44 | 94 | 51 |

Apart from the 2-week period, the statistic is reasonably constant.

Estimation

Maximum likelihood estimator of q

This is found by equating theoretical and observed means.

$$w = -\frac{q}{(1-q)}\ \frac{1}{\ln(1-q)}$$

This cannot be solved explicitly for q but a solution can be found iteratively from

$$q = \frac{-w\ \ln\ (1-q)}{1-w.\ln(1-q)}$$

or directly from a table given by Williamson and Bretherton (1964). A graph plotting (1-q) against w is given in the appendix.

The variance of this estimate is given by

$$\frac{q^2}{Nw\left(\frac{1}{(1-q)} - w\right)} = \frac{q^2}{N\mu_2}.$$

Other methods of estimating q have been given by Anscombe (1950) and Patil (1962).

These include

(1) $\hat{q} = 1 - {}^{f}1/F_1$

This is a fairly efficient method. The efficiency $\geq 74^{o}/o$ for $q < 0.9$.

(2) $q = {}^{2f}2/f_1$.

This is a useful quick quess.

Tables of the Logarithmic distribution

Williamson and Bretherton (1964) have published tables of the LSD for values of the mean as follows

mean = w = 1.1 (0.1) 2.0 (0.5) 5.0 (1.0) 10.0.

However as q,or w, is a continuous variable it will usually be necessary to interpolate to fit a particular distribution, in which case it will probably be quicker to work out the distribution from scratch rather than use these tables.

7.4. The fit of the LSD

A large number of distributions was fitted by an LSD and also by an NBD. Those distributions which are already well fitted by an NBD give an equally good fit to the LSD. This can be seen either by visual inspection or by calculating a statistic such as $\chi^2$ or the likelihood. Sometimes this statistic will indicate that the NBD gives a better fit and sometimes that the LSD gives a better fit.

When systematic discrepancies from the NBD occur (such as for 'heavy-buying' data), systematic discrepancies from the LSD will also occur. Again there seems to be little difference in the goodness of fit.

Four examples are given of the fit of the LSD.

The variance of the LSD is

$$\frac{-q}{\ln(1 - q)} \frac{[1 + q/\ln(1-q)]}{(1-q)^2}$$

$$= w[1 + a_L + w].$$

From this we can calculate the theoretical LSD variance of the whole distribution by combining the block of zeros and the LSD.

This variance is given for comparison.

The values of the parameter a, calculated from the NBD and from the LSD are also given.

Discussion. The investigation of the zeros led to the discovery that for data which is already well fitted by a simple NBD the LSD gives an equally good fit. In other words the distribution consists of a block of non-buyers together with the positive part of the distribution which is well fitted by the 1-parameter LSD.

This is quite reasonable as the method of estimating the parameters of the NBD is bound to give a positive (i.e. non-zero) k regardless of whether the truncated distribution is LSD or not. In practice fitting a NBD always gives a very small k (< 0.15) and in many cases k is so small as to be indistinguishable from zero.

Thus the data can be described by the one-parameter LSD together with one other parameter, the proportion of buyers, b, or the proportion of non-buyers. But whereas the NBD model mixes its two parameters with relatively complex formulae, for example $b = 1 - (1+a)^{-k}$, the LSD model gives two parameters which are in the main independent and which are immediately meaningful. Thus the LSD parameter a is linked simply with w by the relation

$$
\begin{aligned}
w &= \text{rate of buying per buyer} \\
&= \text{mean of LSD} \\
&= \frac{a}{\ln(1+a)}
\end{aligned}
$$

TABLE 7c

Data from Table 1b

| Number of units bought | Observed $f_r$ | LSD | NBD |
|---|---|---|---|
| 0 | 1612 | 1612 | 1612 |
| 1 | 164 | 165.5 | 156.9 |
| 2 | 71 | 72 | 74 |
| 3 | 47 | 41.7 | 44.2 |
| 4 | 28 | 27.4 | 29.2 |
| 5 | 17 | 19.1 | 20.3 |
| 6 | 12 | 14 | 14.7 |
| 7 | 12 | 10.4 | 10.8 |
| 8 | 5 | 7.9 | 8.2 |
| 9 | 7 | 5.6 | 6.2 |
| 10 | 6 | 4.4 | 4.8 |
| 11-12 | 6 | 6.2 | 6.7 |
| 13-14 | 5 | 4.1 | 4.1 |
| 15-16 | 0 | 2.7 | 2.7 |
| 17-20 | 3 | 3.1 | 2.8 |
| 21+ | 5 | 3.9 | 2.4 |
| Mean | 0.636 | 0.636 | 0.636 |
| Variance | 4.50 | 4.53 | 4.18 |

Parameters $q = 0.869$ $k = 0.115$
$a_L = 6.62$ $a_N = 5.53$

TABLE 7d

Brand $C_1$ 26 Weeks

| Number of units bought | Observed $f_r$ | Logarithmic distribution | NBD |
|---|---|---|---|
| 0 | 1568 | 1568 | 1568 |
| 1 | 182 | 175 | 163.6 |
| 2 | 69 | 78 | 79.5 |
| 3 | 44 | 46 | 48.8 |
| 4 | 33 | 31 | 33.1 |
| 5 | 24 | 22 | 23.7 |
| 6 | 21 | 14.6 | 17.5 |
| 7 | 15 | 11.2 | 13.3 |
| 8 | 7 | 8.7 | 10.3 |
| 9 | 5 | 7.0 | 8.1 |
| 10 | 1 | 5.6 | 6.4 |
| 11-12 | 6 | 8.2 | 9.2 |
| 13-14 | 4 | 5.5 | 6.0 |
| 15-16 | 6 | 3.7 | 4.0 |
| 17-18 | 4 | 2.6 | 2.7 |
| 19-20 | 3 | 1.8 | 1.8 |
| 21+ | 8 | 8.1 | 4.3 |
| Mean | 0.79 | 0.79 | 0.79 |
| Variance | 6.63 | 6.57 | 6.01 |

Parameters $q = 0.89$ $a_N = 6.59$

$a_L = 8.1$ $k = 0.12$

TABLE 7e

Brand Q 13 weeks

| r | Observed $f_r$ | LSD | NBD |
|---|---|---|---|
| 0 | 1886 | 1886 | 1886 |
| 1 | 53 | 53 | 51.6 |
| 2 | 18 | 22 | 22.2 |
| 3 | 11 | 12.2 | 12.5 |
| 4 | 9 | 7.6 | 7.9 |
| 5 | 7 | 5.1 | 5.3 |
| 6 | 2 | 3.5 | 3.7 |
| 7 | 4 | 2.5 | 2.7 |
| 8 | 4 | 1.9 | 1.9 |
| 9 | 4 | 1.4 | 1.4 |
| 10 | 1 | 1.0 | 1.1 |
| 11+ | 1 | 3.8 | 3.7 |
| Mean | 0.16 | 0.16 | 0.16 |
| Variance | 0.85 | 0.98 | 0.98 |
| | Parameters | $q=0.835$ | $k=0.03$ |
| | | $a_L=5.05$ | $a_N=4.98$ |

The estimate of k for the NBD is so small that the two theoretical distributions are virtually identical.

TABLE 7f

Brand $P_2$ 8 weeks

This is an example of a 'heavy-buying' distribution. The variance discrepancy does get slightly worse for the LSD but otherwise there is again virtually no difference between the two distributions

| r | Observed $f_r$ | LSD | NBD |
|---|---|---|---|
| 0 | 396 | 396 | 396 |
| 1 | 22 | 28.7 | 27.3 |
| 2 | 9 | 13.2 | 13.3 |
| 3 | 8 | 8.1 | 8.4 |
| 4 | 7 | 5.6 | 5.9 |
| 5 | 5 | 4.2 | 4.3 |
| 6 | 7 | 3.2 | 3.3 |
| 7 | 7 | 2.5 | 2.6 |
| 8 | 7 | 2.1 | 2.1 |
| 9-10 | 0 | 2.9 | 3.1 |
| 11-12 | 1 | 2.0 | 2.2 |
| 13-14 | 0 | 1.5 | 1.5 |
| 15-16 | 4 | 1.1 | 1.1 |
| 17+ | 1 | 2.9 | 2.7 |
| Mean | 0.75 | 0.75 | 0.75 |
| Variance | 5.47 | 8.45 | 8.18 |

Parameters $q=0.918$ $k=0.08$

$a_L=11.2$ $a_N=9.89$

# CHAPTER 8

## Prediction Formulae for the LSD Model

### 8.1 Introduction

The LSD model is a special case of the general model proposed in Section 4.6, with $k = 0$ and $T$ finite. Thus we already have a whole range of prediction formulae which, as we have already seen, do not involve the parameters $p$ or $T$ of the general model.

Thus given only the average rate of buying per buyer, $w$, and the proportion of buyers, $b$, we can calculate all the repeat buying formulae. We shall see that the predictions which result are very close to those obtained by fitting a NBD.

The only snag of the LSD model is that the LSD parameter, $q$ or $a$, cannot be expressed explicitly in terms of $w$. But tables for $q$ are readily available. In any case the NBD also involves the estimation of a paratmer, $a$ or $k$, which cannot be obtained explicitly and for which tables are not available.

### 8.2. Repeat-buying formulae

These have all been derived in Section 4.6. The LSD model is the special case when $T$ is finite and $k = 0$.

As before we divide the population into 4 subgroups, namely, repeat buyers, lost buyers (lapsed buyers), new buyers and non-buyers.

Then we have

$b_R$ = Proportion of repeat buyers

$$= b\left[1 - \frac{\ln\left(\frac{1+2a}{1+a}\right)}{\ln(1+a)}\right]$$

$$= b\left[1 + \frac{\ln(1+q)}{\ln(1-q)}\right]$$

$b_L$ = Proportion of lost buyers

$$= b\,\frac{\ln\left(\frac{1+2a}{1+a}\right)}{\ln(1+a)}$$

$$= -b\,\frac{\ln(1+q)}{\ln(1-q)}$$

Both the quantities ${}^{b}R/b$ and ${}^{b}L/b$ are functions of q (or a) only and hence functions of w only.

TABLE 8a

| w | ${}^{b}R/b$ | ${}^{b}L/b$ |
|---|---|---|
| 2 | 0.57 | 0.43 |
| 4 | 0.73 | 0.27 |
| 6 | 0.77 | 0.23 |
| 8 | 0.80 | 0.20 |
| 10 | 0.81 | 0.19 |
| 15 | 0.84 | 0.16 |

We also have

$m_L$ = mean of quantity bought by lost buyers

$$= \frac{m}{1+a} \qquad \text{where } m = \text{mean of whole distribution.}$$

$$= m(1 - q)$$

$m_R$ = mean of quantity bought by repeat buyers

$$= \frac{ma}{1+a}$$

$$= mq$$

Thus $w_L$ = rate of buying/lost buyer

$$= {}^{m}L/b_L$$

$$= \frac{m(1 - q)}{-b\,\frac{\ln(1+q)}{\ln(1-q)}}$$

But $m = wb = -\frac{1}{\ln(1-q)}\ \frac{q}{1-q}$

$$w_L = \frac{q}{\ln(1+q)}$$

$w_R$ = rate of buying/repeat buyer

$$= {}^{m}R/b_R$$

$$= \frac{qm}{b[1+\frac{\ln(1+q)}{\ln(1-q)}]}$$

$$= \frac{qw}{1+\frac{\ln(1+q)}{\ln(1-q)}}$$

Thus ${}^{m}L/m$, ${}^{m}R/m$, $w_L$, $w_R$ are all functions of q only and hence functions of w only.

TABLE 8b

| $w$ | $w_R$ | $w_L$ |
|---|---|---|
| 2 | 2.53 | 1.33 |
| 4 | 4.95 | 1.40 |
| 6 | 7.39 | 1.42 |
| 8 | 10.3 | 1.425 |
| 10 | 12.0 | 1.43 |
| 15 | 17.5 | 1.435 |

Example of predictions  The repeat buying prediction for the data given in Table 1b were made by fitting both a NBD and a LSD.

TABLE 8c

| | $b$ | $w$ | $b_L$ | $b_R$ | $w_L$ | $w_R$ |
|---|---|---|---|---|---|---|
| NBD | 0.194 | 3.3 | .055 | .139 | 1.43 | 4.0 |
| LSD | 0.194 | 3.3 | .058 | .136 | 1.42 | 4.1 |

The predictions are within 5°/o of each other.

## 8.3. Predictions over a longer period

Assume that we know the values of $w$ and $b$ in a certain time-period. Then in a time-period $c$ times as long we have the following results from Section 4.6.

$$\frac{w_C}{w} = \frac{c\ln(1+a)}{\ln(1+ca)}$$

where $w_c$ = rate of buying per buyer in the longer period.

$$\text{and } \frac{b_c}{b} = \frac{\ln(1+ca)}{\ln(1+a)}$$

where $b_c$ = market penetration in the longer period.

Procedure. Given w for a particular period, look up the corresponding value of q in the tables. Hence

$$a = q/(1-q).$$

$$\text{Hence } w_c = \frac{wc\ln(1+a)}{\ln(1+ca)}$$

$$\text{Hence } b_c = \frac{cwb}{w_c}$$

Note that this is much simpler than the corresponding method for the NBD.

Example of predictions. These predictions are made for the data given in Table 1b by both the NBD and LSD methods. Predictions are made for 52 and 104 weeks.

TABLE 8d

| Time Period | LSD w | NBD w | LSD b | NBD b |
|---|---|---|---|---|
| 26 weeks | 3.3 | 3.3 | 0.194 | 0.194 |
| 52 weeks | 5.0 | 5.25 | 0.256 | 0.244 |
| 104 weeks | 8.1 | 8.5 | 0.316 | 0.302 |

Even over a time-period four times as long the predictions still differ by less than $5^{o}/o$.

Practical results. We have already seen in Section 2.3 that the repeat-buying predictions for the NBD model generally give good unbiased results. As the LSD predictions are so similar we expect them to give equally good results and this in fact proves to be the case.

## 8.4. Investigation of lost buyers

This investigation was carried out to see if the repeat-buying formulae work, even when the variance discrepancy occurs, because nearly all the lost buyers buy only one or two units in the first period. Thus if nearly all the other buyers are repeat buyers, so long as the frequency distribution gives a good fit for the 1's and 2's any other discrepancy will not affect the repeat-buying formulae.

We first consider the conditional distribution of purchases in period II for people who bought r units in period I. This is a NBD with mean $(k+r)/\frac{a}{(1+a)}$ and exponent $(k+r)$.

We consider the LSD case with $k = 0$ and $\frac{a}{(1+a)} = q$. Our population is the buyers in the first period.

$$\text{Thus Prob}(0 \text{ units in II} \mid r \text{ units in I}) = (1+q)^{-r} = f(r) \text{ say.}$$

TABLE 8e

| w | q | f(1) | f(2) | f(3) |
|---|---|---|---|---|
| 2 | 0.71 | 0.58 | 0.34 | 0.20 |
| 10 | 0.97 | 0.51 | 0.25 | 0.13 |

But Prob (r units in I) = $-\dfrac{1}{\ln(1-q)} \; q^r/r$.

P(0 units in II and r units in I)

$$= P(0 \mid r)\, P(r)$$

$$= (1+q)^{-r}\, q^r/r \;.\; -\frac{1}{\ln(1-q)}$$

$$= g(r) \text{ say.}$$

TABLE 8f

| w | q | g(1) | g(2) | g(3) |
|---|---|---|---|---|
| 2 | 0.71 | 0.33 | 0.07 | 0.019 |
| 5 | 0.93 | 0.18 | 0.044 | 0.014 |
| 10 | 0.97 | 0.142 | 0.035 | 0.012 |

Now $^{b}N/b = -\dfrac{\ln(1+q)}{\ln(1-q)}$

When w = 10 $^{b}N/b = 0.194$

But g(1) + g(2) = 0.177.

Thus 91°/o of the lost buyers bought 1 or 2 units (when w = 10).

But it is not true to suppose that the remaining 9°/o of the lost buyers have a negligible effect, for we find that the average rate of buying for g(1) and g(2) is given by

$$\frac{142 + 2 \times 35}{142 + 35} = 1.2$$

The average rate of buying for all lost buyers is much higher, namely, 1.43. Thus the remaining 9°/o account for 24°/o of the purchases by lost buyers.

The average rate of buying for g(1), g(2) and g(3) is $\frac{142 + 2 \times 35 + 3 \times 12}{142 + 35 + 12} = 1.31$

Thus although it is most important that the frequency distribution should fit the 1's and 2's it should also fit the next few values reasonably well. But for values of r above about 4 any discrepancies in the fit will have a negligible effect on the repeat-buying formulae.

Approximation to the distribution of lost buyers

A useful approximation to the distribution of lost buyers can be obtained from the equation

$$\text{Prob}(0 \text{ units in II} \quad r \text{ units in I}) = (1+q)^{-r}$$

For reasonably large values of w, q is close to 1 and then

$$(1+q)^{-r} \rightarrow (\tfrac{1}{2})^{r} \text{ as } w \nearrow \infty$$

TABLE 8e1

| w | $\frac{1}{1+q}$ |
|---|---|
| 2 | .582 |
| 4 | .525 |
| 6 | .514 |
| 8 | .509 |
| 10 | .507 |
| 15 | .503 |

$\frac{1}{1+q}$ is close to $\frac{1}{2}$ for values of w bigger than about 4. Then a good approximation to the distribution of lost (or new) buyers can be obtained by taking a proportion $(\frac{1}{2})^r$ of the people who bought r units in the first of two equal periods to be lost buyers. ($r = 1,2,\ldots.$).

Conversely a proportion $1 - (\frac{1}{2})^r$ of the people who bought r units in the first of two equal time-periods also buys at least one unit in the second period.

For small values of w, the proportion of repeat buyers is somewhat smaller.

## 8.5. Cumulative tables

The LSD model also allows us to calculate the proportion of units which are purchased by people who buy j or more units.

$$\text{This proportion} = \sum_{r=j}^{\infty} rf_r \Big/ \sum_{r=1}^{\infty} rf_r$$

$$= \frac{\sum_{r=j}^{\infty} r \frac{-1}{\ln(1-q)} q^r/r}{\sum_{r=1}^{\infty} r \frac{-1}{\ln(1-q)} q^r/r}$$

$$= \frac{q^j/(1-q)}{q/(1-q)}$$

$$= q^{j-1}$$

This is a very useful formula. For example, it is known that a few heavy purchasers often account for a high proportion of purchases. But it had previously not been possible to quantify this fact. It also enables us, for example, to calculate the 50°/o point $\hat{j}$, that is the value of j above which half the purchases are made. This will vary with q and hence with w.

We have $q^{j-1} = \frac{1}{2}$

$$j = \frac{\ln\frac{1}{2}}{\ln q} + 1$$

Thus we can calculate j for various values of w.

TABLE 8g

Table of 50°/o points

| j | q | w |
|---|---|---|
| 2 | 0.5 | 1.44 |
| 3 | 0.707 | 1.96 |
| 4 | 0.793 | 2.41 |
| 5 | 0.841 | 2.87 |
| 6 | 0.87 | 3.28 |
| 8 | 0.917 | 4.5 |
| 10 | 0.933 | 5.1 |
| 20 | 0.966 | 8.4 |
| 30 | 0.977 | 11.5 |

Example of a cumulative frequency distribution

The table compares the proportion of purchases which are made by people who buy (r+1) or more units and $q^r$.

The data is for brand k over 24 weeks (see appendix). Some typical values for r are shown.

TABLE 8h

| r | Observed proportion | $q^r$ |
|---|---|---|
| 1 | 0.937 | 0.932 |
| 2 | 0.859 | 0.869 |
| 3 | 0.804 | 0.810 |
| 4 | 0.757 | 0.755 |
| 6 | 0.658 | 0.656 |
| 8 | 0.565 | 0.570 |
| 12 | 0.440 | 0.430 |
| 16 | 0.320 | 0.325 |
| 24 | 0.199 | 0.185 |
| 32 | 0.115 | 0.105 |
| 48 | 0.051 | 0.034 |

The fit of this cumulative frequency distribution is very good indeed.

When systematic discrepancies occur as in 'heavy-buying' data then they will show up clearly in the cumulative table. One cannot of course expect the cumulative table to fit this kind of data.

Lastly it is worth pointing out that explicit cumulative formulae cannot be obtained for the NBD model as it is not possible to express the necessary summations in a closed form.

## 8.6. The standard error of w

We often wish to compare rates of buying in two different time-periods in order, for example, to see if the purchasing behaviour is stationary.

Let $w_1$, $w_2$ be the rates of buying per buyer in two successive equal time-periods. If the same panel of consumers is used then the two rates of buying will be correlated.

Let $m_1$, $m_2$ be the mean rates of buying in the two periods over the whole sample and let $m = \frac{m_1 + m_2}{2}$

Let $b_1$, $b_2$ be the proportion of buyers in the two periods and let $b = \frac{b_1 + b_2}{2}$

Then $w_1 = {}^{m}1/b_1 \qquad w_2 = {}^{m}2/b_2$.

Let $w = m/b$.

Now the same panel of consumers is used and the purchases of any consumer in the two time-periods are independent Poisson variates.

Thus standard error $(m_1 - m_2) = \sqrt{\frac{m_1 + m_2}{2}} = \sqrt{\frac{2m}{n}}$

where n = sample size.

We also have

Standard error $(b_1 - b_2)$ = standard error $(b_L - b_N)$

$$= \sqrt{\frac{b_L + b_N}{n}}$$

where $b_L$ = proportion of lost buyers

$b_N$ = proportion of new buyers

In addition we require the covariance of $(m_1 - m_2)$ and $(b_1 - b_2)$ for the same panel.

Now as the same panel of consumers is used we know that any change in the number of buyers is caused by new or lost buyers whose average rates of buying we denote by $w_N$ and $w_L$ respectively.

Given a change $(b_1 - b_2)$ in the proportion of buyers we have

$$m_1 - m_2 = b_L\ w_L + b_R\ w_{R1} - b_N\ w_N - b_R\ w_{R2}$$

where $b_R$ = proportion repeat buyers

$w_{R1}$ = rate of buying per repeat buyer in first period

$w_{R2}$ = rate of buying per repeat buyer in second period.

Then if $b_1 > b_2$

$$m_1 - m_2 = b_R(w_{R1}-w_{R2}) + (b_1-b_2)\ w_L + b_N\ (w_L-w_N)$$

[The proof is exactly the same if $b_2 > b_1$ except that $w_N$ replaces $w_L$ in the second term on the right]

Then $\mathcal{E}\ (m_1 - m_2)(b_1 - b_2)$

$$= \mathcal{E}\,[b_R(w_{R1}-w_{R2}) + (b_1-b_2)w_L + b_N(w_L-w_N)](b_1-b_2)$$

$$= \mathcal{E}\,(b_1 - b_2)^2 w_L$$

as $(w_{R1}-w_{R2})$ and $(w_L-w_N)$ are uncorrelated with $(b_1-b_2)$.

In addition $w_L$ is uncorrelated with $b_1-b_2$ and $\mathcal{E}\,(b_1-b_2) = 0$.

Thus

$$\mathcal{E}\,(m_1-m_2)(b_1-b_2) = w_L.\mathrm{Var}(b_1-b_2).$$

Let $\rho_{mb}$ be the correlation coefficient between $(m_1-m_L)$ and $(b_1-b_2)$ for the same panel.

Then $\rho_{mb} \times$ S.E. $(m_1-m_2) \times$ S.E.$(b_1-b_2)$

$$= \mathrm{Cov}\ (m_1-m_L)(b_1-b_2)$$

$$= \mathcal{E}(m_1-m_2)(b_1-b_2)$$

$$\rho_{mb} = \frac{w_L\ .\ \mathrm{Var}(b_1-b_2)}{\mathrm{S.E.}(m_1-m_2).\mathrm{S.E.}(b_1-b_2)}$$

$$= w_L\sqrt{\frac{b_L+b_N}{2n}}$$

We now replace $w_L, b_L$ and $b_N$ by the estimates obtained from the LSD model. Thus the resulting correlation is obtained by averaging over fixed panels.

$$\rho_{mb} = \frac{q}{\log(1+q)}\sqrt{\frac{-2b\ \log(1+q)}{\log(1-q)m}}$$

$$= \sqrt{\frac{q(1-q)}{\log(1+q)}} \qquad \text{since } w = m/b = \frac{-q}{(1-q)\log(1-q)}$$

The table shows the correlation coefficient for various values of w.

These figures are intuitively acceptable as one expects the correlation to decrease as the number of purchases by loyal buyers, and hence w, increases.

TABLE 8i

| w | $\rho_{mb}$ |
|---|---|
| 2 | 0.61 |
| 4 | 0.37 |
| 6 | 0.28 |
| 8 | 0.21 |
| 10 | 0.19 |
| 15 | 0.15 |

Consider the variance of m/b.

If $m^*, b^*$ are true values of m,b we have

$$m/b = m^*/b^* + (m-m^*)\frac{1}{b^*} + (b-b^*)\frac{-m^*}{b^{*2}}$$

+ higher order terms.

$$\mathrm{Var}(m/b) = \mathcal{E}\left(\frac{m}{b} - \mathcal{E}\left(\frac{m}{b}\right)\right)^2$$

$$\simeq \mathcal{E}\left(\frac{m}{b} - m^*/b^*\right)^2$$

$$\simeq \frac{\mathrm{Var}(m)}{b^{*2}} + \frac{\mathrm{Var}(b)m^{*2}}{b^{*4}} - \frac{2\mathrm{Cov}(m,b)m^*}{b^{*3}}$$

For the same sample of consumers

$$\mathrm{Var}(m) = \frac{1}{2}\mathrm{Var}(m_1-m_2) = m/n$$

$$\mathrm{Var}(b) = \frac{1}{2}\mathrm{Var}(b_1-b_2) = {}^{b}L/n.$$

$$\mathrm{Cov}(m,b) = \mathcal{E}(m-m^*)(b-b^*)$$

$$= \frac{1}{2}\mathcal{E}(m_1-m^*+m^*-m_2)(b_1-b^*+b^*-b_2)$$

[As $\mathcal{E}(m_1-m^*)(b^*-b_2) = 0 = \mathcal{E}(m^*-m_2)(b_2-b^*)$

$$= \frac{1}{2}\,\mathrm{Cov}(m_1 - m_2)(b_1 - b_2)$$

$$= \frac{b_L \times w_L}{n}$$

The sample estimates of $m^*, b^*$ are $m, b$.

$$\mathrm{Var}\left(\frac{m}{b}\right) = \frac{m}{b^2 n} + \frac{b_L\, m^2}{b^4 n} - 2\,\frac{b_L\, w_L m}{b^3 n}$$

Let $N = nb$ = Average number of buyers.

$$\text{Then } \mathrm{Var}(w) = \frac{w}{N} + \frac{w^2\, b_L/b}{N} - 2\,\frac{b_L/b\; w_L\; w}{N}$$

Hence standard error of $(w_1 - w_2)$ is

$$\sqrt{\frac{2w + 2w^2\; b_L/b - 4\; b_L/b\; w_L\; w}{N}}$$

This expression was calculated for various values of $w$. Note that $b_L/b$ and $w_L$ have already been calculated for various values of $w$.

TABLE 8j

| $w$ | S.E. of $(w_1 - w_2) \times \sqrt{N}$ |
|---|---|
| 2 | 1.7 |
| 4 | 3.2 |
| 6 | 4.5 |
| 8 | 5.9 |
| 10 | 6.9 |
| 15 | 9.4 |

## 8.7 Alternative ways of deriving the repeat-buying formulae

As for the NBD model, the repeat-buying formulae can be derived in two completely different ways from the method used in Section 4.5. The first method uses the bivariate LSD and the second method considers the problem in terms of interpurchase times.

### [illegible] The bivariate LSD

The repeat-buying formulae can be derived in a similar way to that given in Section 2.5 for the bivariate NBD. We again note that the conditional distribution in a certain time-period for those people who bought r units in a previous equal time-period is NBD with mean ra/(1+a) and exponent r (since k = 0). For the particular case when r = 0, the people who buy for the first time in the second period form an LSD in this second period (since the exponent r will be zero). In other words the new and lost buyers form an LSD in their respective time-periods

Consider the buyers in some period T. We use the lemmas given in Section 2.5.

For the LSD $g(u,T) = \dfrac{\ln(1-q_T u)}{\ln(1-q_T)}$

$$g(u_1,1) = \frac{\ln(1-\, {}^{q_T}/T\ (u_1+T-1))}{\ln(1-q_T)}$$

$$= \frac{\ln(1-{}^{q_T}/T - q_T - q_T\ {}^{u_1}/T)}{\ln(1-q_T)}$$

$$= \alpha \ln \left(1 - \frac{u_1 \, q_T}{q_T(1-T)+T}\right) + \beta$$

where $\alpha, \beta$ do not contain $u_1$.

This is the marginal distribution of $X_1$ in a unit time-period (excluding any non-buyers). It is LSD with parameter $q_1 = \dfrac{q_T}{q_T(1-T)+T}$.

Putting $a_1 = q_1/(1-q_1)$ and $a_T = q_T/(1-q_T)$

we have $a_T = Ta_1$.

Then the conditional distribution of buyers in period T given that $X_2 = 0$ has p.g.f.

$$\mathcal{G}(u_1 | X_2 = 0) = \frac{g\left(\frac{u_1}{2};2\right)}{g\left(\frac{1}{2};\ 2\right)}$$

$$= \frac{\ln\left(1 - q_2\frac{u_1}{2}\right)}{\ln(1-q_2)} \cdot \frac{\ln(1-q_2)}{\ln(1-q_2/2)}$$

$$= \frac{\ln\left(1 - \frac{q_2}{2}u_1\right)}{\ln(1-q_2/u_1)}$$

Thus the lost buyers follow an LSD with parameter

$$\frac{q_2}{2} = \frac{q_1}{1+q_1}$$

Let us consider the proportion, B, of people who buy in at least one of two successive equal period. The bivariate distribution of purchases in the two periods is given by

$$h(u_1\ u_2;\ 1,1) = \frac{\ln(1 - q_2 \frac{u_1+u_2}{2})}{\ln(1 - q_2)}$$

Then $b_L$ = Proportion of lost buyers in whole population

$$= B \sum_{r=1}^{\infty} \text{coeff. of } u_1^r$$

$$= - B \frac{\sum_{r=1}^{\infty} (\frac{q_2}{2})^r \ \frac{1}{r}}{\ln(1-q_2)}$$

$$= B \frac{\ln\ (1 - \frac{q_2}{2})}{\ln(1-q_2)}$$

But $B = b + b_L$

$$\frac{b + b_L}{b_L} = \frac{\ln(1 - q_2)}{\ln(1 - q_2/2)}$$

giving $b_L/b = - \dfrac{\ln(1 + q_1)}{\ln(1 - q_1)}$ since $q_2 = \dfrac{2q_1}{1+q_1}$

Now we have just seen that the lost buyers follow an LSD with parameter $q_1/(1+q_1)$.

Thus the mean of the distribution of lost buyers

$$= - \frac{q_1/(1+q_1)}{\frac{1}{1+q_1} \ln(\frac{1}{1+q_1})}$$

$$= \frac{q_1}{\ln(1+q_1)} = w_L$$

All the other formulae follow from this result.

## Interpurchase times

These formulae can be derived in a similar way to that given in Section 2.4.

The LSD model assumes that the population is split into 'never-buyers' and potential buyers whose long-run average rates of buying $\lambda$, follow the truncated Gamma distribution.

$$dF = k\ e^{-\lambda/a}\ \frac{1}{\lambda}\ d\lambda \qquad \frac{1}{T} \leq \lambda \leq \infty .$$

Thus the mean times, y, between purchases follow the distribution given by

$$dF = k\ e^{-\frac{1}{ay}}\ \frac{1}{y}\ dy \qquad 0 \leq y \leq T$$

$$[\text{put } y = \frac{1}{\lambda}]$$

We suppose that the population makes Poisson purchases in any one time-period. Thus the waiting time till the first purchase is exponential. Hence, as in Section 2.4, it can be shown that the distribution of times to first purchases is given by

$$k\ \frac{e^{-\left(\frac{\frac{1}{a} + t}{T}\right)}}{-(\frac{1}{a} + t)}\ dt \qquad 0 < t < T$$

Consider two time-periods (0,t) (t,2t).

b = Proportion of buyers

= Proportion of people who buy for the first time in (0,t)

$b_L$ = Proportion of lost buyers

= Proportion of new buyers

= Proportion of people who buy for the first time in $(t, 2t)$

$$\text{Thus } \frac{\mathcal{E}(b_L)}{\mathcal{E}(b)} = \frac{\displaystyle\int_t^{2t} k \frac{e^{-\left(\frac{\frac{1}{a}+t}{T}\right)}}{-\left(\frac{1}{a}+t\right)} dt}{\displaystyle\int_o^{t} k \frac{e^{-\left(\frac{\frac{1}{a}+t}{T}\right)}}{-\left(\frac{1}{a}+t\right)} dt}$$

$$\left[\begin{aligned} \text{Put } \frac{1}{a} + t &= z \\ dt &= dz \end{aligned}\right] = \frac{\displaystyle\int_{t+\frac{1}{a}}^{2t+\frac{1}{a}} e^{-z/T}/z \; dz}{\displaystyle\int_{\frac{1}{a}}^{t+\frac{1}{a}} e^{-z/T}/z \; dz}$$

$$\simeq \frac{\displaystyle\int_{t+\frac{1}{a}}^{2t+\frac{1}{a}} \frac{1}{z} dz}{\displaystyle\int_{\frac{1}{a}}^{t+\frac{1}{a}} \frac{1}{z} \; dz} \qquad \text{for T large}$$

$$= \frac{\ln\left(\frac{2t+\frac{1}{a}}{t+\frac{1}{a}}\right)}{\ln\left(\frac{t+\frac{1}{a}}{\frac{1}{a}}\right)}$$

Standardise times so that $t = 1$

Then $^{b}L/b = \dfrac{\ln\left(\frac{1+2a}{1+a}\right)}{\ln(1+a)}$ as obtained previously.

Consider a person whose mean time between purchases is y.

Then $P\big(j \text{ purchases in } (0,t)\big) = \dfrac{e^{-t/y}\left(\frac{t}{y}\right)^j}{j!}$

$P\big(j \text{ purchases in } (0,t) \text{ and } 0 \text{ purchases in } (t,2t)\big)$

$$= e^{-2t/y}\left(\frac{t}{y}\right)^j / j!$$

$$\sum P(j \text{ and } 0)j = \sum_{j\geq 1} \frac{e^{-2t/y}\left(\frac{t}{y}\right)^j}{(j-1)!}$$

$$= \frac{t}{y}\, e^{-t/y}$$

$m_L$ = mean amount bought by lost buyers

= mean amount bought by new buyers

$$= \int_0^T \left\{ \sum_{j\geq 1} P(j \text{ and } 0|y) \times j \right\} \times P(y)dy$$

$$= \int_0^T \frac{t}{y}\, e^{-t/y}\, k\, e^{-\frac{1}{ay}} \frac{1}{y}\, dy.$$

Similarly $m = \displaystyle\int_0^T \left\{ \sum_{j\geq 1} P(j|y) \times j \right\} P(y)dy$

$$= \int_0^T k\, e^{-\frac{1}{ay}} \frac{1}{y}\, t/y\, dy$$

giving $\frac{m_L}{m} = \frac{\frac{1}{a}}{\frac{1}{a} + t}$ when T is large

Standardise times so that t = 1.

Then ${}^{m}L/m = \frac{1}{1+a}$ as obtained previously.

<u>Prediction over longer periods</u>

Consider the two periods (0,t), (0,ct).

Let $b_c$ = Proportion of buyers in (0,ct)

= Proportion of people who buy for the first time in (0,ct).

$$\text{Then } {}^{b}c/b = \frac{\int_0^{ct} k \frac{e^{-\frac{(\frac{1}{a}+t)}{T}}}{-(\frac{1}{a}+t)} \, dt}{\int_0^{t} k \frac{e^{-\frac{(\frac{1}{a}+t)}{T}}}{-(\frac{1}{a}+t)} \, dt}$$

Put $\frac{1}{a} + t = z$

$dt = dz$

$$\simeq \frac{\int_{\frac{1}{a}}^{ct+\frac{1}{a}} \frac{1}{z} \, dz}{\int_{\frac{1}{a}}^{t+\frac{1}{a}} \frac{1}{z} \, dz} \quad \text{for T large}$$

$$= \frac{\ln\left(\frac{ct + \frac{1}{a}}{\frac{1}{a}}\right)}{\ln\left(\frac{t + \frac{1}{a}}{\frac{1}{a}}\right)}$$

Standardise times so that $t = 1$.

$$b_c/b = \frac{\ln(1 + ac)}{\ln(1 + a)} \quad \text{as previously obtained.}$$

Let $w_c$ be the average rate of buying over the longer period. In this period c times as many units are bought as in the unit period.

$$w_c \, b_c = cwb$$

$$w_c/w = c \; b/b_c = \frac{c \, \ln(1+a)}{\ln \, (1+ca)}$$

# CHAPTER 9

## An Application to a non-Stationary Situation

### 9.1 Introduction

As an example of the effectiveness of the LSD model, we shall examine a case study in which a household product experienced a large increase in sales from one period to the following equal period. This increase was not shared by other brands in the same field so that the sales increase was not a seasonal phenomenon. In fact the increase was associated with a sales promotion campaign which was offered to consumers of that particular brand in the second period. The data is taken from Goodhardt and Ehrenberg (1966) who analyse it by using the bivariate NBD. We shall see that the LSD repeat buying formulae give the results much more simply.

The following information was taken from a representative sample of the population:-

a) which consumers bought the brand in each period

b) how much each consumer bought in each period.

We shall see if the increase in sales has been caused by repeat-buyers buying more, by getting more new buyers than expected or by a combination of both effects.

The repeat buying formulae which have developed in the previous chapter give us a 'norm' by which to judge the sales data. For one of the problems of analysing data is that even when the average rate of buying is constant and the data is stationary, there are still substantial changes from one period to the next. Some buyers buy only in the first period (lost buyers), and some only in the second period (new buyers).

## 9.2. The data

The following data was available from a sample size 1000.

Period I. 78 buyers bought 320 packets giving

rate of buying per buyer = $w$ = 4.1

Also proportion of buyers = $b$ = .078.

Fit a LSD to the frequency distribution of purchases in the usual way by equating observed and theoretical means. This gives $q = 0.906$.

Period II. The sales show the effect of the promotion campaign.

146 buyers bought 570 packets giving

rate of buying per buyer = 3.9.

These buyers can be split into two groups: new and repeat buyers.

80 buyers bought for the first time in the second period so that $b_N = .080$. The number of packets bought by new buyers = 188 so that $w_N = 2.3$.

66 buyers bought in both periods so that $b_R = .066$. The number of packets bought by the repeat buyers in the

second period was 382 so that the rate of buying per repeat buyers in the second period = 5.8.

## 9.3. Repeat buying predictions

Under stationary conditions we can make the following predictions about what would have happened without the sales campaign. This will enable us to see how the sales campaign has affected the different classes of buyers.

All the repeat buying predictions are simple functions of w or q (see Section 8.2).

Thus

$w_R$ = rate of buying per repeat buyer = 5.2

$w_N$ = rate of buying per new buyer = 1.4

$b_R$ = proportion of repeat buyers = .056

$b_N$ = proportion of new buyers = .022.

Thus the predicted number of packets bought by repeat buyers = $w_R \; b_R \quad 1000$

$= 290$

The predicted number of packets bought by new buyers

$= w_N \; b_N \quad 1000$

$= 30$

## 9.4. Results

The four quantities $b_N$, $w_N$, $b_R$ and $w_R$ are all higher than predicted. The results are given in Table 9.1.

TABLE 9.1

| | | Buyers | | Number of packets bought in second period | |
|---|---|---|---|---|---|
| Observed | | 1st period | | 1st period | |
| | | 0 | + | 0 | + |
| 2nd | 0 | 842 | 12 | 0 | 0 |
| period | + | 80 | 66 | 188 | 382 |
| Predicted | | | | | |
| 2nd | 0 | 900 | 22 | 0 | 0 |
| period | + | 22 | 56 | 30 | 290 |

The difference between the observed and predicted number of packets bought by new buyers

$$= 188 - 30$$

$$= 158$$

The difference between the observed and predicted number of packets bought by repeat buyers

$$= 382 - 290$$

$$= 92$$

The total increase in sales = 250 packets.

Thus the trend among repeat buyers accounts for 36°/o of the total sales effect of this particular promotion campaign.

# CHAPTER 10

# THE DISTRIBUTION OF OCCASIONS AND 'WEEKS'

## 10.1. The distribution of Occasions

In this chapter we will investigate two new models for consumer purchasing behaviour by considering the distribution of occasions and 'weeks', rather than the distribution of packets or purchases.

The marketing man is interested not only in how much a consumer buys but also on how many occasions these purchases are made. For example he may want to know if his product is habitually bought at a rate of just one packet per purchasing occasion or if a substantial amount is bought at more than one packet per occasion.

In addition there are some products, of which petrol is the obvious example, which give data of a different type from that previously considered. Petrol is bought in any amount with 4 gallons a particularly popular choice. But it is also possible to buy £1 worth, which means that it is not possible to construct a frequency distribution in the usual way by recording the number of people who bought 0,1,2,.... gallons in a particular time-period.

Instead we can construct the frequency distribution of the number of people who bought petrol on exactly 0,1,2,.... occasions in a particular time-period. The

NBD was fitted to several petrol distributions of this type and a reasonable fit was obtained in all cases. A typical fit for brand S is shown in table 10a. In this particular case there is a variance discrepancy of 24°/o and over all the petrol distributions there was an average variance discrepancy of about 20°/o. The existence of a variance discrepancy is typical of heavily bought products. However the discrepancy was much smaller than might have been expected for such a heavily bought product.

Thus the possibility arose that a better fit might be obtained with other brands by fitting the NBD to occasions rather than packets. This did not prove to be the case.

For example for brand $P_3$ over 24 weeks (see table 10b) the distribution of packets gave a variance discrepancy of 48°/o while the distribution of occasions gave a variance discrepancy of 50°/o. The two distributions are very similar indeed and the shelving effect occurs in both. This similarity means that there is no advantage in fitting the NBD to occasions rather than packets. However we have obtained the useful information that most purchases are made at the rate of just one packet per purchasing occasion.

Table 10a

Brand: S. Distribution of Occasions

Time Period: 4 weeks

| Number of Occasions on which a purchase was made | Frequencies | |
|---|---|---|
| | Observed | NBD |
| 0 | 493 | 492.9 |
| 1 | 100 | 110.4 |
| 2 | 55 | 57.2 |
| 3 | 39 | 35.2 |
| 4 | 24 | 23.4 |
| 5 | 15 | 16.2 |
| 6 | 20 | 11.6 |
| 7 | 11 | 8.4 |
| 8 | 10 | 6.2 |
| 9-12 | 11 | 12.7 |
| 13-16 | 2 | 4.3 |
| 17+ | 1 | 2.5 |
| Variance | 5.06 | 6.27 |

N = 781 $m = 1.186$

$a = 4.29$

$k = 0.276$

Table 10b

Brand: $P_3$

Time-Period : 24 weeks

| | Frequencies | | | |
|---|---|---|---|---|
| | Packets | | Occasions | |
| No. | Observed | NBD | Observed | NBD |
| 0 | 285 | 284.9 | 285 | 285.0 |
| 1 | 53 | 47.6 | 52 | 48.1 |
| 2 | 19 | 26.4 | 21 | 26.7 |
| 3 | 12 | 18.1 | 13 | 18.2 |
| 4 | 12 | 13.6 | 11 | 13.7 |
| 5 | 9 | 10.7 | 10 | 10.7 |
| 6 | 12 | 8.7 | 10 | 8.7 |
| 7 | 7 | 7.2 | 8 | 7.3 |
| 8 | 6 | 6.1 | 7 | 6.1 |
| 9-12 | 14 | 17.4 | 12 | 17.3 |
| 13-16 | 12 | 10.4 | 12 | 10.4 |
| 17-20 | 17 | 6.8 | 17 | 6.6 |
| 21-24 | 14 | 4.6 | 16 | 4.6 |
| 25+ | 2 | 11.6 | 0 | 10.9 |
| Variance | 35.61 | 52.71 | 33.04 | 49.54 |

| Packets | Occasions |
|---|---|
| $m = 2.97$ | $m = 2.89$ |
| $a = 16.77$ | $a = 16.14$ |
| $k = 0.18$ | $k = 0.179$ |

## 10.2. The distribution of 'Weeks'

In the distribution of occasions for brand $P_3$ over 24 weeks (see table 10b), we notice that no one buys on more than 24 occasions and the possibility arose that all such distributions vanish for values above the length of the time-period in weeks. If this were so then we could try to fit a new type of distribution, which would vanish naturally above this point, unlike the artificially curtailed NBD which was considered in Chapter 4. Unfortunately the distribution of brand P over 24 weeks (see table 10c) shows that some people do buy some products on more than one occasion per week on average. Thus the distribution must be modified further to enable a distribution defined on a finite set of integers to be fitted. Because of this attention was drawn to the distribution of 'weeks'.

In a time-period of, say, n weeks we can construct the frequency distribution

$$f_o, f_1, \ldots, f_n$$

where $f_r$ = number of people in the sample who buy at least one packet in r out of n weeks.

An example of such a distribution for brand P is given in table 10c where it can be compared with the distributions of occasions and packets for the same product. The distribution of 'weeks' is of course only

TABLE 10c

Brand: P all sizes combined

Time Period: 24 weeks

| No. | Observed frequencies Packets | Occasions | Weeks |
|---|---|---|---|
| 0 | 202 | 202 | 202 |
| 1 | 45 | 48 | 48 |
| 2 | 22 | 23 | 24 |
| 3 | 15 | 14 | 15 |
| 4 | 16 | 14 | 14 |
| 5 | 11 | 12 | 10 |
| 6 | 13 | 11 | 12 |
| 7 | 10 | 13 | 13 |
| 8 | 8 | 7 | 7 |
| 9 | 10 | 11 | 11 |
| 10 | 3 | 5 | 4 |
| 11-12 | 11 | 10 | 10 |
| 13-14 | 11 | 9 | 11 |
| 15-16 | 11 | 12 | 12 |
| 17-18 | 13 | 14 | 15 |
| 19-20 | 11 | 13 | 14 |
| 21-22 | 19 | 15 | 16 |
| 23-24 | 24 | 31 | 36 |
| 25-28 | 6 | 3 | |
| 29-32 | 3 | 4 | |
| 33-36 | 1 | 2 | |
| 37-40 | 0 | 0 | |
| 41-44 | 5 | 1 | |
| 45+ | 4 | 0 | |

defined for integers less than or equal to the number of weeks in the time-period.

It is worth emphasizing that the distribution of 'weeks' is of no direct interest to the marketing man, so far as is known. Our interest in the subject is motivated by a desire to gain insight into the discrepancies from the NBD model which is of prime importance.

## 10.3. The Beta Binomial model

We now look for a model which will describe consumer purchasing behaviour in terms of the distribution of 'weeks'. We cannot of course consider the NBD model as the NBD is defined for all positive integers.

The NBD model depends on the Poisson and Gamma distributions. It is well known that the Poisson distribution can be obtained as the limit of the Binomial distribution, by letting $p \longrightarrow 0$ and $n \rightarrow \infty$. Similarly the Gamma distribution can be obtained as the limit of the Beta distribution. The Beta distribution can be expressed as

$$dF = \frac{p^{a-1} (1-p)^{b-1} \, dp}{B(a,b)} \qquad 0 \leq p \leq 1$$

$$= \frac{z^{a-1} \left(1 - \frac{z}{b-1}\right)^{b-1} dz}{(b-1)^a \, B(a,b)} \quad 0 \leq z \leq (b-1)$$

(putting $z = p(b-1)$)

$$\rightarrow \frac{z^{a-1}\, e^{-z}}{(b-1)^a B(a,b)} \qquad \text{as } b \rightarrow \infty$$

$$\left[ B(a,b) = \int_0^1 p^{a-1}\, (1-p)^{b-1}\, dp \right]$$

Thus the Binomial and Beta distributions are the discrete analogues of the Poisson and Gamma distributions, so we will consider the following compound Binomial model for the distribution of 'weeks':-

(i) The probability that a given consumer will buy in a particular week is a constant, p, which is independent of previous purchases. Thus in a time-period of n weeks, the number of weeks in which the consumer buys at least one packet will follow a Binomial distribution with parameters n,p.

(ii) The probability, p, varies from consumer to to consumer, the distribution being a Beta distribution given by

$$f(p) = \frac{1}{B(a,b)}\, p^{a-1}\, (1-p)^{b-1} \qquad 0 \leq p \leq 1$$
$$a > 0. \quad b > 0.$$

Then the overall distribution of 'weeks' will follow the Beta-Binomial distribution where

P(r) = proportion of population who buy on exactly r out of n weeks.

$$= \int_0^1 \binom{n}{r} p^r (1-p)^{n-r} f(p)dp$$

$$= \frac{B(a+r,n+b-r)}{(n+1)\ B(a,b)\ B(a+1,n+1-r)} \qquad r = 0,1,\ldots,n.$$

The Beta-Binomial distribution

The properties of the Beta-Binomial distribution (hereafter called the BB distribution) have been discussed by Raiffa and Schlaifer (1961) and Ishii and Hayakawa (1960). It is interesting to note that the BB distribution can also be derived from an inverse sampling model in a similar way to the NBD, by considering inverse sampling from a finite population. In this case the BB distribution is called the Negative Hypergeometric distribution.

Note that if we were to consider Beta Binomial models for shorter time-periods such as days, we would find that the Beta Binomial model tends to the NBD model as the time-period tends to zero. Thus the BB distribution is the discrete analogue of the NBD.

However while we suspect that purchases in successive weeks may be independent to a close approximation, it is reasonable to expect that purchases in successive days will be negatively correlated as once a purchase has been made the consumer is unlikely to require any more of that product for several days. This latter argument may be the reason for the discrepancies from the NBD model which have been noted and which have led us to consider the Beta-Binomial model.

For durable goods, such as furniture, it would probably be necessary to consider a much longer time-period as the unit to ensure that purchases in successive periods are approximately independent. However in this thesis we are only concerned with non-durable goods.

The mean of the BB distribution is given by $na/(a+b)$ and the variance by $nab\ (n+a+b)/(a+b)^2\ (1+a+b)$. A general discussion of methods of fitting the compound Binomial distribution is given by Wetherill (1957) following Vagholkar.

The standard method of estimating a and b is by the method of moments. This gives

$$\hat{a} = \frac{-m\ (s^2 - m(n-m))}{ns^2 - m(n-m)}$$

and $$\hat{b} = \frac{(n-m)\ \hat{a}}{m}$$

where $m$ = observed mean

$s^2$ = observed variance.

Shenton (1950) shows that the method of moments has high efficiency.

Fitting by zeros and mean

Another method of fitting the BB distribution is to use the mean and the proportion of zeros. This is analogous to the standard method of fitting the NBD.

Find initial estimates of a and b (say $a_1$, $b_1$) by the method of moments. Then compare

$$P_1(0) = \frac{b_1\ (b_1+1)\ldots\ (b_1+n-1)}{(a_1+b_1)\ldots\ (a_1+b_1+n-1)}$$

with the observed proportion of zeros, namely ${}^{f}o/N$.

If $P_1(0) > {}^{f}o/N$ then increase a and b in such a way that their ratio stays constant. A suitable increase is obtained by putting $a_2 = 1.1a_1$ and $b_2 = 1.1b_1$.

Then if $b_1 = \dfrac{a_1\ (n-m)}{m}$

we have $b_2 = \dfrac{a_2\ (n-m)}{m}$

so that theoretical and observed means are still equal.

Moreover we have $\dfrac{b_2}{a_2 + b_2} = \dfrac{b_1}{a_1 + b_1}$

but $\dfrac{b_2 + k}{a_2+b_2+ k} < \dfrac{b_1 + k}{a_1+b_1+ k}$ for $k = 1,2,\ldots.(n-1)$

so that $P_2(0) < P_1(0)$

Thus the change in a,b has altered P(0) in the desired direction.

Conversely if $P_1(0) < {}^{f}o/N$ then decrease a and b so that $a_2 = 0.9a_1$ and $b_2 = 0.9b_1$.

Put $z_1 = P_1(0) - {}^{f}o/N$

and $z_2 = P_2(0) - {}^{f}o/N$.

Interpolate linearly for $z_3 = 0$ to find $a_3$, $b_3$ i.e. choose $a_3 = a_1 + \frac{z_1\ (a_2 - a_1)}{(z_1 - z_2)}$

$$\text{Hence } b_3 = \frac{a_3\ b_1}{a_1}$$

These estimates of a and b will usually be sufficiently accurate.

<u>The fit of the BB distribution</u>

A large number of distributions was fitted by a BB distribution and a good fit was obtained in all cases, whether the distribution was fitted by moments or by zeros. When the distribution was fitted by zeros the variance is a convenient measure of the goodness of fit. No systematic variance discrepancy was found for heavily bought products. Not many distributions were available for products which were not heavily bought, but the BB distribution seemed to fit equally well.

The estimates of the parameter a were always less than 1 and often close to zero. The value of the parameter b, on the other hand, varied considerably from about 0.2 to over 10. However for the same brand over different time-periods the estimates were reasonably constant as required by the model, for under staionary conditions the Beta distribution for p is the same for all time-periods. Thus for brand P over 24 weeks we find a = 0.19 and b = 0.59. This time-period was divided into six 4-week periods and estimates of a and b were found for each of these periods. The estimates of a varied between 0.123 and 0.185 with an average of 0.152 while the estimates for b varied between 0.436 and 0.519 with an average of 0.45.

The distributions can be divided into two categories. Firstly those which give a fitted BB distribution, which decreases monotonically with the number of weeks (as in table 10d) and secondly those which give a fitted BB distribution which is U-shaped (as in table 10e). The significant point is that the BB distribution fits both types of distribution.

The two types of distribution can be distinguished by the value of the parameter b.

For we have

$$\frac{P(r+1)}{P(r)} = \frac{(a+r)\ (n-r)}{(n+b-r-1)(1+r)}$$

so that when r = n-1 we have

$$\frac{P(n)}{P(n-1)} = \frac{(a+n-1)}{b \times n} .$$

Now a is always less than 1 so that for reasonably large n we find

$$\frac{P(n)}{P(n-1)} \simeq \frac{1}{b}$$

Thus a U-shaped distribution with $P(n) > P(n-1)$ must have $b < 1$. Conversely those distributions which are monotonically decreasing will have $b > 1$.

If the distribution is U-shaped, with $0 < a < 1$, and $0 < b < 1$, then the smallest value in the distribution can be found by considering $\frac{P(r+1)}{P(r)}$ as a continuous function of r.

If we set $\frac{P(r_o+1)}{P(r_o)} = 1$

we find $r_o = n + \frac{b-1}{1-a}$

Thus the smallest value in the distribution occurs when r is the smallest integer greater than $r_o$.

Compound Binomial distributions in readership analysis

Hyett (1958) and Metheringham (1964) have noted the application of the BB distribution to readership studies. If an advertisement is inserted in several issues of the same publication, then the total number of people who will see the advertisement will be considerably greater than the number of people who see just one issue of the publication. One method of describing this situation is to suppose that the proportion of people who see exactly r out of n issues will follow a BB distribution.

A related compound Binomial model, which is also

Table 10d

Distribution of 'weeks'

Brand: all brands in product field Y combined

Time Period: 24 weeks.

| No. of Weeks | Observed (Frequencies) | BB fitted by zeros (Frequencies) | BB fitted by moments (Frequencies) |
|---|---|---|---|
| 0 | 278 | 278.1 | 308.1 |
| 1 | 112 | 124.8 | 120.3 |
| 2 | 94 | 87.6 | 81.7 |
| 3 | 65 | 68.9 | 63.2 |
| 4 | 70 | 56.9 | 51.9 |
| 5 | 59 | 48.4 | 44.1 |
| 6 | 51 | 41.8 | 38.2 |
| 7 | 40 | 36.5 | 33.5 |
| 8 | 33 | 32.1 | 29.7 |
| 9 | 32 | 28.4 | 26.5 |
| 10 | 18 | 25.2 | 23.7 |
| 11-12 | 26 | 42.0 | 40.4 |
| 13-14 | 18 | 32.7 | 32.6 |
| 15-16 | 19 | 25.0 | 26.0 |
| 17-18 | 16 | 18.4 | 20.2 |
| 19-20 | 11 | 12.6 | 14.9 |
| 21-22 | 9 | 7.6 | 10.0 |
| 23-24 | 19 | 3.0 | 4.8 |
| Variance | 30.3 | 27.6 | 30.3 |
| | | a = 0.468 | a = 0.403 |
| | | b = 2.034 | b = 1.749 |

Table 10e

Distribution of 'Weeks'

Brand: P. all sizes combined

Time Period: 4 weeks

| No. of weeks | Frequencies: Observed | Frequencies: BB fitted by zeros |
|---|---|---|
| 0 | 291 | 291.0 |
| 1 | 52 | 54.5 |
| 2 | 45 | 37.8 |
| 3 | 29 | 35.9 |
| 4 | 57 | 54.8 |
| Variance | 2.04 | 2.03 |

a = 0.165

b = 0.519

applicable to readership analysis, has been considered by Quenouille (1964). He assumes that the underlying distribution of p is given by

$$f(p) = (1-b)(a+1)p^a \qquad 0 \leq p \leq 1$$

$$a > -1.$$

$$\text{Prob } (p=1) = b$$

where p is the probability that a member of the population will see a particular issue of the publication, and b is the proportion of regular readers. This model gives somewhat similar results to those derived from a BB model where the BB parameter, b, is less than 1.

## 10.4. Predictions from the BB model

### Proportion of buyers in a longer period

The BB model, like the NBD model, enables us to make predictions about how the sample will behave in a longer time-period. Under stationary conditions, the Beta distribution of p is the same for all time periods, so that the parameters a,b stay constant.

Let b = proportion of buyers in the sampled period of n weeks

and B = proportion of buyers in a longer period of kn weeks

where k is some constant.

$$\text{Now } b = 1 - \frac{B\,(a,\, n+b)}{(n+1)\, B(a,b)\, B(1,n+1)}$$

$$B = 1 - \frac{B\,(a,\, kn+b)}{(kn+1)\, B(a,b)\, B(1,kn+1)}.$$

But $B(a,kn+b) = \frac{B(a,n+b)(kn+b-1)\ldots(n+b)}{(kn+a+b-1)\ldots(n+a+b)}$

and $B(1,kn+1) = \frac{B(1,n+1)\,kn(kn-1)\ldots(n+1)}{(kn+1)\ldots(n+2)}$

$$\frac{1-b}{1-B} = \frac{(kn+a+b-1)\ldots(n+a+b)}{(kn+b-1)\ldots(n+b)}$$

Hence a prediction of B can be made.

Repeat buying formulae

In addition we can also derive repeat-buying formulae for the BB model. In two successive equal time-periods of n weeks, the population can be divided into 4 sub-groups; namely lost, new, repeat and non-buyers.

Let B = proportion of buyers in the combined period

Then $b_N = B - b$

$$= (1-b)\left[1 - \frac{(2n+b-1)\ldots(n+b)}{(2n+a+b-1)\ldots(n+a+b)}\right]$$

Hence a prediction of $b_N$ or $b_L$ can be made.

We can also calculate quantities such as the mean number of weeks in which a purchase was made by a lost buyer averaged over the whole population. This will correspond to $m_L$ or $m_N$ in the NBD model and will be denoted by $m_L^*$. We expect $m_L^*$ to be slightly smaller than $m_L$.

A person who buys with probability p during any one week will buy on np weeks on average in the second of the two periods, regardless of the fact that he bought on say j weeks in the first period, because the number of weeks in which an individual buys in 2 successive time-periods are independent Binomial variates.

In particular if we consider those people who did not buy in the first period we obtain

$$m_L^{*} = \int_0^1 f(p)\ (1-p)^n\ np\ dp$$

$$= n \int_0^1 \frac{1}{B(a,b)}\ p^a\ (1-p)^{b+n-1}\ dp$$

$$= \frac{n\ B(a+1,\ n+b)}{B(a,b)}$$

$$= \frac{n\ B(a+1,\ n+b-1)}{B(a,b)} \frac{(n+b-1)}{(n+a+b)}$$

$$\text{But } P(1) = \frac{B(a+1,\ n+b-1)}{(n+1)B(a,b)B(2,n)}$$

$$= \frac{B(a+1,\ n+b-1) \times n}{B(a,b)}$$

$$\therefore\ m_L^{*} = \frac{(n+b-1)\ P(1)}{(n+a+b)}$$

For n large and $a < 1$ we notice that

$$m_L^{*} \simeq P(1) = m_L.$$

## 10.5. The connection between the BB distribution and the NBD

The position now is that whereas the distribution of 'weeks' is always well fitted by a BB distribution, we also know that there may be a discrepancy from the NBD for the distribution of packets. Given the distribution of 'weeks' in a time-period of c weeks, we would like to find the conditions on the BB parameters such that the frequency distribution of packets will be well fitted by a NBD.

Suppose that a member of the population has probability, p, of buying in one week, and that he buys $\lambda$ packets on average in the time-period. Then p and $\lambda$ are related by

$$p = 1 - e^{-\lambda/c}$$

$$dp = \frac{1}{c} e^{-\lambda/c} \, d\lambda.$$

Thus the underlying distribution for p which is given by

$$dF = \frac{1}{B(a,b)} \; p^{a-1} \, (1-p)^{b-1} \, dp$$

transforms to give the following underlying distribution for $\lambda$:-

$$dF_1 = \frac{1}{B(a,b)} \; (1-e^{-\lambda/c})^{a-1} \, e^{-\frac{\lambda b}{c}} \, \frac{1}{c} \, d\lambda.$$

We can compare this with the underlying Gamma distribution for $\lambda$ which is postulated to derive the NBD. This is

$$dF_2 = \left(\frac{1}{a_N}\right)^k \frac{e^{-\lambda/a_N} \lambda^{k-1}}{\Gamma(k)} \quad d\lambda$$

The parameter a of the NBD is given the suffix N to distinguish it from the BB parameter.

For $\lambda$ small we have

$$1 - e^{-\lambda/c} \simeq \lambda/c$$

so that $dF_1 \simeq \frac{1}{c.B(a,b)} \left(\frac{\lambda}{c}\right)^{a-1} e^{-\frac{\lambda b}{c}} d\lambda.$

Thus the part of the distribution where $\lambda$ is small is of Gamma type and as a large proportion of the population usually has a small value of $\lambda$ it is reasonable to compare the coefficients of $\lambda$ and $e^{-\lambda}$ in $dF_1$ and $dF_2$.

We find $(k-1)$ corresponds to $(a-1)$

and $\left(\frac{1}{a_N}\right)$ corresponds to $b/c$.

This correspondence is borne out in practice for we find that both k and a are usually close to zero, and we also know that $a_N$ and c are directly proportional to the length of the time-period.

For $\lambda$ large we have

$$1 - e^{-\lambda/c} \rightarrow 1 \text{ as } \lambda \rightarrow \infty$$

and $dF_1$ is dominated by $e^{-\lambda b/c}$.

Thus $dF_1$ will no longer be of the same Gamma type, and presuming that $dF_1$ and $dF_2$ are similar for small values of $\lambda$, then the relative difference between them will increase with $\lambda$. However if $e^{-\lambda b/c}$ tends to zero fairly rapidly with $\lambda$, then only a very small part of the distribution will be affected by this difference and the two distributions will be in close agreement. On the other hand if $e^{-\lambda b/c}$ tends to zero rather slowly with $\lambda$, then the discrepancy will become important.

For a fixed value of c, the larger b is, the quicker $e^{-\lambda b/c}$ will tend to zero. Thus for 'large' values of b we expect the NBD to fit the distribution of packets reasonably well, but for 'small' values of b we expect to find a discrepancy. The value of b below which a discrepancy occurs can be found by an inspection of distributions with values of b over a wide range. The value turns out to be somewhere between 1 and 2.

This result is intuitively acceptable because it appears harder to stretch out a U-shaped distribution into a NBD than a monotonically decreasing distribution.

For example products D and P are heavily bought but the BB distribution gives a good fit to the distribution of 'weeks' (see tables 10e, f, g), and we find $b < 1$ for both products indicating a U-shaped distribution.

For 6 distributions of 'weeks' each over a 4-week period we find

Brand D - average $b = 0.70$

Brand P - average $b = 0.45$.

Because b is small we are not surprised that the NBD does not give a good fit to the corresponding distribution of packets. A large variance discrepancy occurs for both brands and brand P, which has the smaller value of b, also has the larger variance discrepancy.

The NBD variance discrepancy is not confined to products with a U-shaped distribution of 'weeks', though these give the largest discrepancies, for a value of b which is slightly greater than 1 may still be associated with a small NBD variance discrepancy.

## 10.6. The shelving effect

One important result of the analysis of the distribution of 'weeks' was a deeper understanding of the shelving effect.

We have already seen several distributions of 'weeks' in which a similar effect is evident. For example in table 10 d the observed frequencies are relatively steady between 11 and 24 weeks, the latter period being the upper limit of the distribution.

It is easy to choose the BB parameters so that the BB distribution is fairly constant over a high proportion of the distribution. For example if b = 1 we have

$$\frac{P(x+1)}{P(x)} = \frac{(a+x)}{(n-x)}\frac{(n-x)}{(1+x)}$$

$$= \frac{a+x}{1+x}$$

$$\simeq 1 \text{ for } a < 1 \text{ and } x \text{ fairly large.}$$

## TABLE 10f

Distribution of 'weeks'

Brand: D, all sizes combined

Time Period: 4 weeks.

### Period 1

| No. of Weeks | Frequencies: Observed | BB fitted by moments | BB fitted by zeros |
|---|---|---|---|
| 0 | 390 | 390.2 | 390.0 |
| 1 | 38 | 37.8 | 38.1 |
| 2 | 21 | 20.7 | 20.7 |
| 3 | 14 | 14.3 | 14.3 |
| 4 | 11 | 10.9 | 10.9 |
| Variance | 0.77 | 0.77 | 0.77 |
| | | a = 0.097 | a = 0.098 |
| | | b = 1.014 | b = 1.023 |

### Period II

| No. of Weeks | Frequencies: Observed | BB fitted by moments | BB fitted by zeros |
|---|---|---|---|
| 0 | 399 | 403.6 | 399.0 |
| 1 | 36 | 26.3 | 31.0 |
| 2 | 14 | 15.8 | 17.7 |
| 3 | 6 | 13.1 | 13.5 |
| 4 | 19 | 15.2 | 12.7 |
| Variance | 0.84 | 0.84 | 0.79 |
| | | a = 0.060 | a = 0.074 |
| | | b = 0.655 | b = 0.815 |

Table 10g

Distribution of weeks

Brand: D. all sizes combined

Time Period: 24 weeks

| No. of weeks | Frequencies Observed | BB fitted by moments | BB fitted by zeros |
|---|---|---|---|
| 0 | 342 | 350.7 | 342.0 |
| 1 | 23 | 26.4 | 29.4 |
| 2 | 25 | 14.3 | 16.0 |
| 3 | 15 | 10.0 | 11.2 |
| 4 | 11 | 7.7 | 8.6 |
| 5-6 | 13 | 11.8 | 13.1 |
| 7-8 | 1 | 9.1 | 10.0 |
| 9-10 | 9 | 7.4 | 8.0 |
| 11-12 | 3 | 6.4 | 6.8 |
| 13-14 | 10 | 5.7 | 6.0 |
| 15-16 | 5 | 5.1 | 5.3 |
| 17-18 | 1 | 4.9 | 4.9 |
| 19-20 | 2 | 4.6 | 4.5 |
| 21-22 | 8 | 4.7 | 4.2 |
| 23-24 | 6 | 5.4 | 4.2 |
| Variance | 24.18 | 24.18 | 22.7 |
| | | a = 0.075 | a = 0.086 |
| | | b = 0.819 | b = 0.941 |

Thus if n = 12 and a = 0.5 we have

$$\frac{P(12)}{P(11)} = 0.96$$

$$\frac{P(6)}{P(5)} = 0.92$$

so that the distribution is fairly steady between x = 5 and x = 12. This effect is similar to the shelving effect described in Chapter 3.

Alternatively it is easy to choose the BB parameters so that the distribution is U-shaped and has a peak at the number of weeks in the time-period. This effect is similar to the bunching effect described in Chapter 3.

The importance of these two effects, which are very like the shelving and bunching effects, was realised when further analysis showed that the distribution of 'weeks' and packets was very similar. In other words few people buy more than one packet in any one week.

An example is given in Table 10c for brand P over 24 weeks, where the two distributions look very similar. This is confirmed by calculating

Grand total of packets bought = 3271

Grand total of weeks in which a purchase was made = 2784.

The difference is 18°/o.

For brand D, which is not so heavily bought, the similarity is even more striking, and we find a difference of only $13^{\circ}/\text{o}$ over a 24 week period. Brands which give $b > 1$ will have an even smaller difference.

Thus when the distribution of 'weeks' exhibits a shelving effect, the distribution of packets, being very similar, will also exhibit a shelving effect. As the distribution of 'weeks' is always well fitted by a BB distribution it follows that the BB model is more exact than the NBD model. Thus the average long-run rates of buying would better be described by the distribution derived in section 10.5, namely

$$dF_1 = \frac{1}{B(a,b)} \left(1 - e^{-\lambda/c}\right)^{a-1} \; e^{-\frac{\lambda b}{c}} \frac{1}{c} \, d\lambda.$$

However it was not practical to use this distribution to build a model to describe the distribution of packets, because the integrals which resulted could not be evaluated in a workable form.

Thus, as we are primarily interested in the distribution of packets, we will continue to use the NBD (or LSD) model as an aid to description and prediction. At the same time we must recognize that it is only a useful approximation and that the shelving effect will occur when the corresponding distribution of 'weeks' gives a BB distribution with a value of b less than or slightly greater than 1.

## REFERENCES


1. Anscombe, F.J. (1950) Sampling theory of the NBD and LSD. Biometrika 37, 358-382.
2. Arbous, A.G. and Kerrich, J.E. (1951) Accident Statistics and the concept of accident proneness. Biometrics 7, 340-432.
3. Aske Research (1964) X-buying in Lancashire. Unpublished report.
4. Bartko, J. (1961) The NBD. A review. Virginia Journal of Science 12 (1) 18-37.
5. Bates, G.E. and Neyman, J. (1952) Contributions to the theory of accident proneness. Univ. of California publications in statistics. 1, 9, 215-254.
6. Bliss, C.I. and Fisher, R.A. (1953) Fitting the NBD to biological data. Biometrics 9, 176-199.
7. Brass, W. (1958) Simplified methods of fitting the truncated NBD. Biometrika 45, 59-68.
8. Chatfield, C., Ehrenberg, A.S.C. and Goodhardt, G.J. (1966) Progress on a simplified model of stationary purchasing behaviour. J.R.S.S. A, 129. 317-368.
9. Cohen, A.C. (1960) Estimating the parameter in a conditional Poisson distribution. Biometrics 16, 203-211.
10. David, F.N. and Johnson, N.L. (1952) The truncated Poisson distribution. Biometrics 8, 275-285.

11. Eggenberger, F. and Polya, G. (1923) Über die Statistik verketter Vörgange. Z. Angewandte Math. Mech. Berlin, 1, 279-289.

12. Ehrenberg, A.S.C. (1959) The pattern of consumer purchases. Applied Statistics 8, 26-41.

13. Ehrenberg, A.S.C. (1963) Verified predictions of consumer purchasing patterns. Commentary 10.

14. Ehrenberg, A.S.C. (1964) Estimating the proportion of loyal buyers. J. Marketing Research. Feb. 56-60.

15. Evans, D.A. (1953) Experimental evidence concerning contagious distributions in ecology. Biometrika 40, 186-211.

16. Feller, W. (1943) On a general class of contagious distributions. Ann. Math. Stat. 14, 389-400.

17. Feller, W. (1957) Introduction to probability Theory. Vol. I. Wiley.

18. Fisher, R.A., Corbet, A.S. and Williams C.B.(1943) The relation between the no. of species and the no. of individuals in a random sample of an animal population. J. Animal Ecology. 12, 42-58.

19. Fisher, R.A. (1941) The NBD. Ann. Eugenics 11, 182-187.

20. Furry, W. H.(1939) On fluctuation phenomena in the passage of high energy electrons through lead. Phys. Rev. 52, 569-581.

21. Goodhardt, G.J. (1965) Private note on multivariate distributions.

22. Goodhardt, G.J. and Ehrenberg, A.S.C. (1966) Conditional trend analysis (read at ESOMAR Conference, Copenhagen).

23. Greenwood, M. and Yule, G.U. (1920) Frequency distributions of multiple happenings. J.R.S.S. A, 83, 255-279.

24. Gurland, J. (1958) A generalised class of contagious distributions. Biometrics 14, 229-249.

25. Gurland, J. (1959) Some applications of the NBD Amer. J. of Public Health. 49, 10, 1388-1399.

26. Hartley, H.O. (1958) Maximum likelihood estimation from incomplete data. Biometrics 14, 174-194.

27. Hyett, G.P. (1958) The measurement of readership. Statistics seminar at L.S.E.

28. Ishii and Hayakawa (1960). On the compound binomial distribution. Ann. Inst. of Stat. Math. 12, 69-80.

29. Kendall, D.G. (1948) On some modes of population growth to R.A. Fisher's LSD. Biometrika 14, 174-194.

30. Kendall, M.G. and Stuart, A.(1961) The Advanced Theory of Statistics Vol. I. Griffin.

31. Lüders, R. (1934) Die Statistik der seltenen reignisse. Biometrika 26, 108-128.

32. McKendrick (1914) Studies on the theory of continuous probabilities. Proc. London Math. Soc. 13, 401-416.

33. Metheringham, R. (1964) Measuring the net cumulative coverage of a print campaign. J. Advertising Res. Dec. 23-28.

34. Montmort, P.R. (1714) "Essai d'analyse sur les jeux de hazards."

35. Patil, G.P. (1962) Some methods of estimation for the LSD. Biometrics 18, 68-75.

36. Patil, G.P. (1960) Evaluation of the NBD in terms of the incomplete Beta function. Technometrics 2, 501-507.

37. Patil, G.P. (1965) Multivariate LSD as a probability model. (to be published).

38. Patil, G.P., Kamat, A.R. and Wani, J.K. (1964) The LSD and related tables. Aerospace Research Labs.

39. Pearson, K. (1934) Tables of the incomplete Beta function. C.U.P.

40. Pearson, K and Fieller, E.C. (1933) On the applications of the double Bessel function to statistical problems. Biometrika 25, 158-178.

41. Quenouille, M.H. (1949) A relation between the LSD, Poisson distribution and NBD. Biometrics 5, 162-164.

42. Quenouille, M.H. et al (1964) Cumulative readership. Report prepared by British Market Research Bureau Ltd.

43. Raiffa and Schlaifer (1961) Statistical Decision Theory. 237-241.

44. Rider, P.R. (1955) Truncated binomial and NBD. J.A.S.A. 50, 877-883.

45. Sampford, M.R. (1955). The truncated NBD. Biometrika 42, 58-69.

46. Shenton (1950) Maximum likelihood and the efficiency of fitting by moments. Biometrika 37, 111-116.

47. Steck, G.P. (1965) Review of reference 50. Technometrics 7, 83.

48. Wetherill, G.B. (1957) Ph.D. thesis.

49. Williams, C.B. (1947) The LSD and its application to biological problems. J. Ecology 34, 253-271.

50. Williamson and Bretherton (1964) Tables of the NBD. Wiley.

51. Williamson and Bretherton (1964) Tables of the LSD. Ann. Math. Stat. 35, 284-298.

52. Yule, G.U. (1910) J.R.S.S. A, 73, 26.

53. Yule, G.U. (1924) Mathematical theory of evolution. Phil. Trans. Roy. Soc. London, B, 213, 21-87.

# APPENDIX

## Tables of frequency distributions

Brand: $P_1$

Time Period: 24 weeks

| Number of Units bought | Frequencies: Observed | Frequencies: Fitted NBD |
|---|---|---|
| 0 | 438 | 438.0 |
| 1 | 9 | 9.6 |
| 2 | 4 | 4.7 |
| 3 | 4 | 3.1 |
| 4 | 2 | 2.3 |
| 5-8 | 3 | 5.4 |
| 9-12 | 3 | 3.0 |
| 13-16 | 3 | 1.8 |
| 17-20 | 2 | 1.3 |
| 20-24 | 4 | 0.9 |
| 25+ | 2 | 3.9 |
| Variance | 18.8 | 24.6 |
| | N = 474 | m = 0.73 |
| | | a = 32.53 |
| | | k = 0.02 |

Brand: $P_2$

Time Period: 12 weeks.

| Number of Units Bought | Frequencies Observed | Frequencies Fitted NBD |
|---|---|---|
| 0 | 395 | 395.0 |
| 1 | 21 | 23.3 |
| 2 | 7 | 11.7 |
| 3 | 6 | 7.6 |
| 4 | 4 | 5.5 |
| 5 | 1 | 4.2 |
| 6 | 7 | 3.4 |
| 7 | 2 | 2.8 |
| 8 | 7 | 2.3 |
| 9 | 2 | 2.0 |
| 10 | 4 | 1.7 |
| 11 | 3 | 1.5 |
| 12 | 7 | 1.3 |
| 13-16 | 2 | 3.8 |
| 17-20 | 0 | 2.4 |
| 21-24 | 5 | 1.6 |
| 25+ | 1 | 4.0 |
| Variance | 12.21 | 20.66 |
| | N = 474 | m = 1.10 |
| | | a = 17.72 |
| | | k = 0.06 |

Brand: $P_4$

Time Period: 8 weeks.

| Number of Units Bought | Frequencies: Observed | Frequencies: Fitted NBD |
|---|---|---|
| 0 | 435 | 435.0 |
| 1 | 19 | 17.7 |
| 2 | 5 | 7.7 |
| 3 | 3 | 4.3 |
| 4 | 2 | 2.7 |
| 5 | 2 | 1.8 |
| 6 | 3 | 1.3 |
| 7 | 4 | 0.9 |
| 8 | 1 | 0.7 |
| 9+ | 0 | 2.0 |
| Variance | 1.04 | 1.32 |
| | N = 474 | m = 0.23 |
| | | a = 4.70 |
| | | k = 0.05 |

Brand: $P_4$

Time Period: 24 weeks.

| Number of Units Bought | Frequencies Observed | Frequencies Fitted NBD |
|---|---|---|
| 0 | 403 | 403.0 |
| 1 | 34 | 24.9 |
| 2 | 11 | 12.1 |
| 3 | 2 | 7.6 |
| 4 | 3 | 5.3 |
| 5-8 | 9 | 11.1 |
| 9-12 | 4 | 4.8 |
| 13-16 | 2 | 2.3 |
| 17-20 | 4 | 1.3 |
| 21-24 | 2 | 0.7 |
| 25+ | 0 | 1.1 |
| Variance | 7.55 | 7.34 |
| | N = 474 | m = 0.67 |
| | | a = 9.90 |
| | | k = 0.07 |

Brand: $P_3$

Time Period: 4 weeks

| No. of Units bought | Frequencies Observed | NBD |
|---|---|---|
| 0 | 387 | 387.0 |
| 1 | 31 | 44.2 |
| 2 | 26 | 18.5 |
| 3 | 13 | 9.6 |
| 4 | 14 | 5.5 |
| 5 | 2 | 3.3 |
| 6 | 0 | 2.1 |
| 7 | 0 | 1.3 |
| 8 | 1 | 0.9 |
| 9+ | 0 | 1.7 |
| Variance | 1.08 | 1.50 |

m = 0.41

a = 2.62

k = 0.16

Brand: k

Time Period: 24 weeks

| Number of Units bought | Frequencies Observed | LSD | NBD |
|---|---|---|---|
| 0 | 1501 | 1500.9 | 1501.0 |
| 1 | 80 | 86.3 | 82.5 |
| 2 | 50 | 40.2 | 40.4 |
| 3 | 23 | 25.0 | 25.6 |
| 4 | 15 | 17.5 | 18.1 |
| 5-6 | 24 | 23.1 | 24.2 |
| 7-8 | 16 | 14.7 | 15.4 |
| 9-12 | 15 | 17.3 | 18.1 |
| 13-16 | 11 | 9.4 | 9.7 |
| 17-20 | 6 | 5.5 | 5.6 |
| 21-24 | 2 | 3.4 | 3.4 |
| 25+ | 7 | 6.5 | 6.0 |
| Variance | 10.62 | 10.18 | 9.63 |
| | $N = 1750$ | $q = 0.932$ | $m = 0.73$ |
| | | $a_L = 13.72$ | $k = 0.059$ |
| | | | $a_N = 12.24$ |

Graph for finding LSD parameter q from the mean w