CHAPTER3

THE REPEAT-BUYING STRUCTURE OF A MARKET

An Empirical Example

The purpose of Part II of this book is to try and describe the nature of repeat-buying theory for the general reader. The aim is not to propound an abstract theory but to help in describing and understanding empirical buyer behaviour itself – the forms which it takes and the factors on which it depends. That is where the theoretical work started and where its applications lie.

In the present chapter we describe the theory's major practical application, which is to help structure the observed facts of buyer behaviour in any given market. More specifically "problem-solving" orientated studies are illustrated in Part III. The theoretical results of the NBD/LSD model are outlined in Chapter 4, and the numerical calculations are described in Appendix A*.

The analysis in the present Chapter is closely based on a standardised type of report which has been used in the comprehensive analysis of many different product-fields in the last few years [e.g. Aske Research 1970]. The variety of product-fields covered so far in those intensive studies range from Instant Potatoes, Frozen Foods, Breakfast Cereals and Soup, through Confectionery, Cigarettes (which is rather different), "Take-home" Beer, Dentifrice and Detergents, to Petrol and Motor Oil. The results for one particular product-field are set out here in "coded" form. This product-field is typical not only in that the observed buying patterns mostly follow the theoretically predicted lines rather closely, but in that there are also a number of specific exceptions.

The analysis centres on repeat-buying and penetration growth for five leading brands, labelled A to E, during a 48-week period (i.e. effectively a year). The results are set out in ten tables or groups of tables, and commentary. The purchasing behaviour of each brand is studied here in isolation from that for all other brands, and corresponding results for "multi-brand" buying: relating purchasing of the different brands to each other, are given in Chapter 9.

* The theoretical calculations in this chapter are based on the NBD version of the theory.

The distinctive approach is that the analyses presented are already well understood so that the results usually follow simple and predictable patterns (as shown by the NBD/LSD theory which is set out in Chapter 4). This allows us to see' which of the findings are normal and which are exceptional. A wide range of tabulations can therefore be reduced to manageable proportions and interpretation can become more positive. The main structure of repeat-buying behaviour in this product-field is seen to be normal, and the exceptional features which may deserve special attention are readily isolated.

Table 3.1 Penetration and Penetration Growth

For branded consumer goods, the penetration of a brand is usually the feature which most clearly differentiates one brand from another in terms of sales level or market-share (as given in the first column of figures in Table 3.1). The "penetration" of an item is the proportion of the population who buy the item at least once in a given time-period. This is examined in the two facing tables for the total product-class ("Any Brand") and the five leading brands A to E (which account for 75% of the market).

In Table 3.1 we examine how many households buy in each of the four 12-week quarters of the 48-week "year" analysed here*. We see that the penetrations of the total product and of each of the five individual brands are relatively stable quarter by quarter. The variation is statistically significant but slight, and mainly reflects some seasonality in this market (more buyers in Quarters II and III)**. Such movements could form the basis of more detailed investigations (as was illustrated in the seasonal analysis in §2.6 of Chapter 2), but in the first instance the relative stability of the quarterly penetrations means that the "Average" column is a fair approximation to the penetration in any one quarter. Thus *about 62%* bought the product in each quarter, *about 42%* bought Brand A, *about* 17% bought Brand B, and so on***.

* In this and subsequent tables, results are usually set out to at most two significant figures. (An additional figure may occasionally be needed in detailed calculations.)

** On a 4-figure sample base, the standard error of an individual quarterly penetration is of the order of 1 percentage point, and with a continuous panel, the sampling error of a trend is reduced (see also §6.4 and Appendix A).

*** The fact that the values of the quarterly penetrations closely resemble the market-share values of the five brands is a numerical coincidence for *quarterly* periods; it does not occur in other length time-periods (see Table 3. la).

Table 3.1. Penetration

% of Households Buying in each Quarter

	% Market Share* (Annual)	QUARTER I	II	III	IV	Av.
		%	%	%	%	%
ANY BRAND	100%	58	63	65	61	62
Brand A	46	39	45	44	41	42
Brand B	12	15	17	18	16	17
Brand C	6	7	8	8	6	7
Brand D	5	5	6	8	6	6
Brand E	6	7	6	7	7	7
Average Brand	15	15	16	17	15	16

* In terms of purchase occasions, with Brands A to E accounting for 75% of the product-class ("Any Brand").

Table 3.1a. Penetration

% of Households Buying in Periods of Various Lengths

(Observed Values "0" and Theoretical Norms "T")

Period of length (in weeks)	1		4		12		24		48	
	0	T	0	T	0	T*	0	T	0	T
ANY BRAND	22	19	45	42	62	(62)	74	72	79	80
Brand A	12	10	28	26	42	(42)	55	52	62	61
Brand B	3	3	9	9	17	(17)	25	22	32	28
Brand C	2	1	4	4	7	(7)	12	10	17	13
Brand D	1	1	3	3	6	(6)	11	8	14	11
Brand E	2	1	4	4	7	(7)	9	9	12	11
Average Brand	4	3	10	9	16	(16)	22	20	27	24

* Used in fitting.

These quarterly averages are shown again in Table 3.1 a in the context of penetration growth, i.e. penetration figures for time-periods of lengths ranging from 1 week through 12 weeks to the full 48 weeks analysed here. Thus 22% of the population bought the product in a typical week, 62% in the average quarter as already noted, and 79% in the full 48-week year. Similarly, for Brand A the penetration growth is from 12% in a week to 62% in the year. And so on.

These observed figures ("0") are compared with theoretical norms ("T") which are predictions from the 12-week penetration and the corresponding 12-week buying frequency (given in Table 3.2a), the calculations being based on the NBD/LSD theory (see § 4.8 and Appendix A for numerical details).

These norms allow us to assess the penetration growth of each item against what *normally* happens. Table 3.1 b for example gives the results for the average brand in the last six product-fields analysed in this way at the time of writing, and the fit clearly is good. In the case of Table 3. la, given that 42% of the population buy the product-field in the typical quarter and that they do so on average 5 times during that quarter (see Table 3.2), the prediction is that 19% of them will buy the product in the average week and that 80% will buy during the "year". For Brand A the predictions are that about 10% should buy Brand A in a week and 61% in the year. And so on.

Table **3.1b.** Typical Penetration Growth: The Average Brand in 6 **Product-Fields**

% of Households Buying in Periods of Various Lengths

(Observed Value "0" and Theoretical Norms "T")

Six Product-Fields	Period of length (in weeks) 1		4		12		24		4%	
	0	T	0	T	0	**T***	0	**T**	**0**	**T**
The Average Brand	3	2	6	**6**	11	**(11)***	17	**16**	22	**21**

* Used in fitting.

The agreement between the predicted penetration figures and the observed ones is mostly close. (For 24 predictions ranging from 1% to 80%, the average discrepancy is about 1½ percentage points, and only 2 values are out by more than 3 percentage points.) This agreement shows the extent to which penetration growth from a week to a year largely follows a predictable or "normal" pattern. There is therefore little in the longer-term (e.g. annual) penetration figures which is not already implicit in (or predictable from) the shorter-term buying patterns. The amount and complexity of the information that has to be considered is therefore greatly reduced.

The exceptions are not only fairly small but also rather systematic. They occur mainly for Brands B, C and D. These have penetrations in 48 weeks which are several percentage points higher than is predicted from the shorter-term buying patterns. The reason is thought to be the seasonality of the market (noted in Table 3.1) which has brought in some extra buyers for Brands B, C and D. This does not occur for the product-field as a whole, so that the implication is one of greater brand-switching for these three brands during the peak-season (a specific kind of finding which can be followed up in more detail by further analysis). The brand-leader A however gained virtually no extra buyers in this way. Brand E also shows virtually no excess annual penetration but Table 3.1 also showed no seasonal trend for this brand; Brand E is in fact somewhat "different" in a number of other respects as well, as will be seen in subsequent tables.

Table 3.2 The Frequency of Buying per Buyer

In Table 3.2 we turn to the average rates of buying per quarter, i.e. the number of occasions on which the average buyer of an item in a given quarter bought the item during that quarter. There is little variation quarter by quarter, and the "Average" column provides a good approximation to the buying frequencies in any one quarter.

In examining the relationship between the average purchase frequency of different brands, a general finding is that the average buying frequency does not differ greatly from one brand to another (as is discussed more fully in § 10.2 of Chapter 10). Typically this also occurs here: on average about 3 purchases of a brand are made per quarterly buyer of the brand, irrespective of the fact that the penetration levels or market-shares of Brands A to D or E differ by a factor of almost 10. This uniformity in buying rates imposes a major constraint on marketing action – it is unlikely that the existing buyers of a brand will turn (or can be turned) into much heavier buyers of it: any *major* change in average purchase frequency would be out of line with the general pattern.

Specific exceptions to this regularity occasionally arise. These are sometimes due to the "growing pains" of new brands, or relate to some physical segmentation of the product-field. More generally however, this apparent "constancy" of the average purchase frequencies is in fact not quite the whole story. There tends to be a trend, usually quite a small one, with penetration level: the more people there are who buy a brand the more often (or *slightly* more often) they tend to buy it. The trend is not very clear in Table 3.2 but the brand-leader with an average purchase-frequency of 3.7 certainly stands out.

Table 3.2. The Frequency of Buying the Brand

The Average Number of Purchases of a Brand in each Quarter per Buyer of the Brand in that Quarter

	% Market Share* (Annual)	QUARTER I	II	III	IV	Av.
ANY BRAND	100	4.9	5.1	5.0	5.1	5.0
Brand A	4	3.7	3.8	3.6	3.7	3.7
Brand B	1	2.5	2.4	2.5	2.5	2.5
Brand C		3.1	3.0	2.6	2.7	2.9
Brand D		2.4	2.6	2.2	2.8	2.5
Brand E		2.7	3.2	3.2	3.1	3.0
Average Brand	1	2.9	3.0	2.8	3.0	2.9

* In terms of purchase occasions.

Table 3.2a. The Frequency of Buying the Brand

The Average Number of Purchases per Buyer, in Periods of Various Lengths

(Observed Values "0" and Theoretical Norms "T")

	1: 0	1: T	4: 0	4: T	12: 0	12: T*	24: 0	24: T	48: 0	48: T
ANY BRAND	1.16	1.39	2.3	2.4	5.0	(5.0)	8.8	8.6	15.7	15.7
Brand A	1.02	1.26	1.8	2.0	3.7	(3.7)	6.0	6.0	10.1	10.3
Brand B	1.01	1.14	1.5	1.5	2.5	(2.5)	3.3	3.6	5.0	5.8
Brand C	1.01	1.19	1.7	1.7	2.8	(2.8)	3.9	4.2	5.1	6.7
Brand D	1.01	1.17	1.5	1.6	2.5	(2.5)	3.1	3.7	4.3	5.7
Brand E	1.01	1.21	1.6	1.7	3.0	(3.0)	4.9	4.6	6.8	7.4
Average Brand	1.01	1.19	1.6	1.8	2.9	(2.9)	4.2	4.4	6.3	7.2

Periods of length (in weeks)

* Used in fitting.

This kind of trend normally takes the form that it is the average purchase frequency per buyer multiplied by the proportion of *non-buyers* which does not vary greatly from brand to brand (see § 10.2 of Chapter 10 and §§ 11.4 and 11.5 in Chapter 11). The effect is quite marked in longer time-periods, as is discussed overleaf, but when penetrations are low (as in the quarterly periods here other than for Brand A

– see Table 3.1), the "correction-factor", i.e. the proportion of non-buyers, is itself close to unity and is therefore almost negligible. Thus for the average quarterly results in Table 3.2 we have, using the complement of the percentage buying figures in Table 3.1, that

Average Quarterly Purchase Frequency

Brand	Observed	X Non-buyers
A	3.7	2.1
B	2.5	2.1
C	2.8	2.6
D	2.5	2.3
E	3.0	2.8
Average	2.9	2.4

The variability of the figures is generally reduced (from a range of 1.2 – i.e. 3.7 to 2.5 – to one only half as big at .7).

In particular, the high purchase frequency of Brand A is seen to be accounted for by its high penetration – it is nothing intrinsic to Brand A as such. Brands C and E however now stand out a little as having attracted a somewhat high frequency of purchase amongst their buyers, and these exceptions merit further attention. For Brand C the explanation lies with the quarter-by-quarter trend in the figures of Table 3.2, it being the results in the first two quarters which are too high. The picture for Brand E is that this brand in fact differs in product-formulation from the other brands (rather as in the U.K., Ribena or Lucozade differ from other soft drinks, or All Bran or Stergene differ from other cereals or detergents). We begin to see to what extent such a difference in potential product-appeal is reflected in actual consumer response: Brand E has no seasonal trend (see Table 3.1) and attracts somewhat more frequent buyers, but the difference is not vast.

In Table 3.2a we turn to the average buying frequencies per buyer in periods of varying length. The differences between brands just discussed become more marked in the longer periods such as the 48-week one, but in the main (i.e. other than the Brand E effect), they remain predictable simply in terms of the relation with the proportion of non-buyers, as can readily be checked by using the penetration figures in Table 3.1 a.

Another general feature of the average purchase frequencies shown in Table 3.2a is that they increase less than "pro rata" with the increasing length of the analysis-period. This is because extra buyers in a longer period are generally lighter (or less frequent) buyers – they buy in the *longer* period but not in the initial shorter period. This process tends to follow a general pattern. Theoretical NBD norms can therefore be calculated from the quarterly results and compared with the observed figures. Thus given that the quarterly buyer of Brand A bought it on average 3.7 times during the quarter, the prediction is that the larger number of buyers in the full year will buy the brand on average 10.3 times during the year.

As in the comparison with the penetration norms in Table 3.1 a, the agreement between the observed and theoretical buying rates in Table 3.2a tends to be close. Looking at the 48-week figures for example, the observed and theoretical figures are identical for the product-field as a whole and almost so for the leading brand. For the other brands, the observed figures are somewhat lower than the theoretical norms, which ties in with their above-normal annual penetrations in Table 3. la. The implication is mostly that there were extra "seasonal" buyers for each of these brands, and that these bought less often than normal (i.e. in the peak-season only).

In the shorter time-periods, especially the single week ones, the theoretical norms consistently overstate the observed figures. This is a general finding, namely that in short time-periods buying patterns are different from those in longer periods, and hence the NBD/LSD extrapolations do not apply, as discussed more fully in Chapter 4 and in Part IV.

Table 3.3 Packs per Purchase

The simplest results in examining buyer behaviour are generally obtained by working in terms of *purchase occasions,* as already mentioned in § 1.2 of Chapter 1. Table 3.3 provides the link with *sales*, showing the number of packs bought per purchase occasion*.

With most products that have been studied, only a single pack is purchased on most occasions. This is the case with all the brands in

* The decomposition of the sales of an item in a given time-period is into the number of buyers in the period *times* the average number of purchases per buyer *times* the average number of packs per purchase (*times* the average size or value per pack). This decomposition is valuable in as far as none of the components other than penetration vary very much from brand to brand.

Table 3.3, and does not depend on whether people are frequent or infrequent buyers of the product-group. (In some product-fields, several units tend to be bought per purchase, e.g. gallons of petrol, and analyses as in this table tend then to carry more information.)

In some markets (such as the present one), each brand is sold in several pack-sizes. Detailed analyses of penetration growth and frequency of purchase for the different sizes are then also needed. However, it is a basic finding that repeat-buying loyalty can be successfully examined for each brand as a whole (i.e. irrespective of pack-size), if the ***purchase occasion*** is treated as the analysis unit, as in this chapter. The theoretical reason for this is discussed in § 11.5 of Chapter 11 and largely stems from the empirical finding (§ 10.3, Chapter 10) that average rates of buying a pack-size per buyer of that size vary little from size to size.

Table 3.3. Packs per Purchase

The Average Number of Packs bought per Purchase Occasion

		Buyers of Product *		
48 WEEKS	ALL	Light	Medium	Heavy
ANY BRAND	1.05	1.03	1.03	1.08
Brand A	1.05			
Brand B	1.03			
Brand C	1.13	Not calculated		
Brand D	1.04			
Brand E	1.04			

* Light = 1 – 12 purchases of the product per year (55% of all buyers).
Medium = 13 – 25 purchases of the product per year (25% of all buyers).
Heavy = 26 + purchases of the product per year (20% of all buyers).

Table 3.4 Light and Heavy Buyers

The purchasing rates referred to in Table 3.2 were *averages.* In Table 3.4 we therefore consider the way in which individual purchase frequencies are distributed about these averages, over the 48-week period analysed here. (Similar results are obtained in shorter time-periods.)

Of all the households buying Brand A during the year, 18% bought only once, 9% twice, and so on, 52% buying six or more times during the year. And so on for other brands.

These observed frequencies are compared in the table with theoretical NBD norms which are calculated from the percentage of buyers of the brand in the period and their average frequency of purchase. Thus, given that 62% of the population bought Brand A at an average frequency of about 10 purcliases in the year (as shown in Tables 3.1 a and 3.2a), we predict that about 19% of these buyers will prove to be once-only buyers, 12% to have bought twice, and so on.

The agreement for the various brands is generally close. Something of a discrepancy however occurs for the product-field as a whole. This is often found for products which most of the population buy (here 79%

Table 3.4. Light and Heavy Buyers

The % of Buyers Making 1,2,3,.... Purchases in the Year

(Observed Values "0" and Theoretical Norms "T")

48 WEEKS	% of Population buying at all (T3.1a)		Number of Purchases in the Year 1	2	3	4	5	6+	Average*
ANY BRAND	79 = 100%	0:	8	8	6	4	4	70	16
		T:	12	9	7	6	5	62	
Brand A	62 = 100%	0:	18	9	8	7	6	52	10
		T:	19	12	9	7	6	49	
Brand B	32 = 100%	0:	32	18	11	6	5	29	5
		T:	31	17	11	8	6	28	
Brand C	17 = 100%	0:	38	16	13	7	3	23	5
		T:	33	16	10	7	6	28	
Brand D	14 = 100%	0:	45	20	9	3	0	23	4
		T:	36	17	11	7	5	24	
Brand E	12 = 100%	0:	36	12	6	5	6	36	7
		T:	30	15	10	7	5	34	
Average Brand	27 = 100%	0:	34	15	9	6	4	33	6
		T:	30	15	10	7	6	33	

* Average Number of purchases per buyer (Table 3.2a).

buying in the year):. buying is somewhat more regular for the product than for individual brands (or for the NBD model). There is here an excess of frequent buyers (e.g. 6+ times in the year) compared with the norm, but the difference is never very large.

Another small discrepancy occurs for Brands C, D and E. These show an abnormally high proportion of once-only buyers. This suggests that there was a higher than normal level of switching between brands on a very *occasional* basis (i.e. during the peak-season, as noted earlier).

Brand E was earlier noted to be a brand with an unusually high average frequency of purchase, for its penetration level. Table 3.4 now shows that this is not due to any particular excess of very heavy buyers as such. Other than the slight excess of *light* buyers just noted, the distribution of purchasing frequencies about the average of 7 is in fact closely predictable along the standard lines.

Table 3.5 The Sales Importance of Light and Heavy Buyers

Having examined the numbers of lighter and heavier buyers of each brand, we turn in Table 3.5 to the proportion of the total sales of the brand which they account for.

Table 3.5. The Sales Importance of Light and Heavy Buyers

The Percentage of the total Purchases of an Item going to People Who Bought the item Once, Twice, Three times etc. in the Year

(Observed Values "0" and Theoretical Norms "T")

48 WEEKS	% Market share (Purchases)		Number of Purchases in the Year: 1	2	3	4	5	6+
ANY BRAND	100 = 100%	0:	0.5	1.0	1.2	1.0	1.4	94.9
		T:	1.5	1.5	1.5	1.5	1.4	92.6
Brand A	46 = 100%	O:	2	2	2	3	3	88
		T:	2	2	3	3	3	88
Brand B	12 = 100%	0:	7	7	6	5	5	70
		T:	6	7	7	6	6	68
Brand C	6 = 100%	O:	7	6	8	6	3	70
		T:	6	6	6	6	5	71
Brand D	5 = 100%	O:	10	9	6	3	0	72
		T:	8	8	7	7	6	64
Brand E	6 = 100%	O:	5	4	2	3	4	82
		T:	4	5	4	4	4	79
Average Brand	15 = 100%	O:	6	6	5	4	3	76
		T:	5	6	5	5	5	74

For the product-field as a whole, the 30% of buyers who bought less than six times in the year (Table 3.4) account for only about 5% of sales.

For the average brand, something like 80% of its annual sales are accounted for by the heavier buyers, i.e. those who made six or more purchases of the brand in the year (about 25-50% of its buyers – Table 3.4). These figures agree closely with the corresponding NBD norms.

The ability to predict such figures from theory illustrates the extent to which the sales structure of a market is understood. Such use of theory can be quicker and cheaper than fresh tabulation of raw data. For example, similar results for periods shorter than a year can also readily be estimated. However, the more important outcome is the ability to *interpret* the observed data by judging which figures are normal and which require special attention.

Table 3.6 The Incidence of Repeat-Buyers

People who buy an item more than once in a given time-period are repeat-buyers, and the average frequency of purchase considered in Table 3.2 was therefore a measure of repeat-buying for that time-period. We now turn to another form of repeat-buying, namely the incidence of repeat-buying from one time-period to another. This is a particularly powerful form of analysis, as has already been illustrated in the practical application to a seasonal trend given in §2.5 of Chapter 2.

Table 3.6 sets out the incidence of repeat-buyers quarter by quarter, both for the product-field as a whole and for the five leading brands within it. Thus, of the 58% of the population who bought the product in the first quarter (see Table 3.1), 87% bought it again in the second quarter. Similarly, 87% of the second-quarter buyers of the product bought it also in the third quarter, and so on. In the absence of marked trends, the incidence of repeat-buyers should be the same for different pairs of quarters. The slight seasonality seen in Table 3.1 produces a fractionally lower repeat-buying level between Quarters III and IV, but despite this the average column in Table 3.6 reflects the general quarterly repeat-buying levels quite well.

Of themselves, such repeat-buying figures do not have a clear meaning. Thus for Brand A, something like 78% of those who bought it in one quarter bought it again in the subsequent quarter, but as already noted in Chapter 1, the question is whether such a repeat-buying level is high or low, good or bad.

Table 3.6. The Incidence of Repeat-Buyers

The Percentage of Buyers of an Item in One Quarter who bought it again in the Next Quarter

	QUARTERS			
	I/II	II/III	III/IV	Av.
ANY BRAND	87	87	81	85
Brand A	84	77	73	78
Brand B	61	58	58	59
Brand C	49	57	45	50
Brand D	55	58	46	53
Brand E	65	73	69	69
Average Brand	63	65	58	62

Table 3.6a. The Incidence of Repeat-Buyers

Percentage of Buyers in One Period Who bought again in the Next Period for Periods of Various Lengths

Observed Values "0" and Theoretical Norms "T")

	Period of Length (in weeks):							
	1		4		12		24	
	0	T	O	T	O	T	0	T
ANY BRAND	45	na	76	71	85	84	91	89
Brand A	40	na	69	58	78	77	83	04
Brand B	32	na	46	47	59	65	62	73
Brand C	37	na	52	52	50	67	52	73
Brand D	20	na	58	45	53	64	54	70
Brand E	34	na	65	48	69	69	63	75
Average Brand	33	na	58	50	62	68	63	75

na = Theory not applicable in very short time-periods.

To reach an answer, we compare the observed figures with what normally occurs, as reflected by the theoretical NBD norms. These are calculated from the average penetration and rate of buying in the *first* period of each pair. We have that 42% of the population bought Brand A in the typical quarter (Table 3.1) and made an average of 3.7 purchases each (Table 3.2). From these two figures, the theoretical model leads us to expect that about 77% of these buyers will buy again in the following quarter – almost exactly as was observed in this case. Table 3.6a gives such comparisons for various lengths of time-periods*.

Discrepancies between the observed and theoretical values help us to describe and to understand any peculiarities of the market. Considering first repeat-buying 12-weeks by 12-weeks, the agreement for the total market and Brands A and E is very close,. but the observed incidence of repeat-buyers of Brands B, C and D is substantially below the normal level**. This shortfall seems to link up with the seasonality of these brands (Table 3.1) and will be further examined in Tables 3.9 and 3.10. The 24-week by 24-week pattern is similar, except for an additional 12-point shortfall of repeat-buyers for Brand E which goes with an abnormally heavy rate of buying, as is shown in Table 3.7.

Table 3.7. The Buying-Frequency per Repeat-Buyer

The Average Number of Purchases per Repeat-Buyer in Period of Various Lengths

(Observed Values "O" and Theoretical Norms "T")

	Period of Length (in weeks)							
	1		4		12		24	
	0	T	0	T	0	T	0	T
ANY BRAND	1.2	na	*2.5*	*2.8*	*5.7*	*5.7*	*9.5*	*9.5*
Brand A	1.0	na	2.0	2.1	4.2	4.3	6.8	6.9
Brand B	1.0	na	1.9	1.8	2.9	3.0	4.4	4.2
Brand C	1.1	na	2.1	2.1	3.8	3.5	4.9	4.8
Brand D	1.0	na	1.8	1.8	3.6	3.1	4.9	4.1
Brand E	1.0	na	1.8	1.9	3.9	3.8	6.6	5.5
Average Brand	1.0	na	1.9	1.9	3.7	3.5	5.5	5.1

* The form taken by the theoretical norms here differs from that in Tables 3.1a and 3.2a, although both stem from the same theory. Given the number of buyers and their average purchase frequency in one period, we predict either how many will buy ***again*** in the next equal period (Table 3.6a) or how many buy ***at all*** in a period of differerit length (Table 3.1a).

**Tests of statistical significance for sampling errors are not easy to carry out rigorously when picking on the larger exceptions in large bodies of data (see also Appendix A). But the discrepancies for Brands B, C and D appear fairly large compared with the kinds of sampling fluctuations that can occur in these data.

In very short time-periods, repeat-buying patterns are different and the theoretical model does not apply in any product-field. Thus within a week here, almost no-one makes more than one purchase of the product (an average of 1 .O 1 purchases as shown in Table3.2a); it would then follow from the NBD/LSD model that there should be virtually no repeat-buyers in the next week. But in practice, Table 3.6a shows that about a third of the weekly buyers of any one brand buy it again in the next week. Weekly shopping habits and "dead-period" effects (the need to more or less use up one purchase before another is made) tend to dominate over the longer-term questions of brand-choice which the NBD theory models. (This is discussed more fully in Chapters 4 and 7.) Even for the 4-weekly periods in Table 3.6a, the repeat-buying estimates tend still for this reason to be somewhat below the observed values.

Table 3.7 The Buying-Frequency per Repeat-Buyer

Tables 3.6 and 3.6a gave the percentage of the buyers of a brand in one period who bought it again in the next, equal-length, period. Table 3.7 now shows how often these repeat-buyers bought the brand in the second period.

The agreement with the theoretical norms is close and shows that there is generally nothing unusual about the repeat-buyers' average frequency of buying, even in those cases where the *number* of repeat-buyers was unusual. This is a typical example of the power of the present approach – to pin-point a deviation in a neat and simple manner.

The largest discrepancy occurs for Brand E in the 24-week periods, where the observed buying frequency of the repeat-buyers is high; this mirrors the shortfall in the 24-week incidence of repeat-buyers in Table 3.6a and shows that total sales accounted for by repeat-buyers (i.e. their number times their purchase frequency) was just about on the mark.

Table 3.8 The Buying-Frequency per "New" Buyer

Table 3.8 gives the corresponding buying frequencies for people who bought in the second period but not in the first. Under no-trend conditions, such "new" buyers are usually infrequent buyers, whose rate of buying is almost invariably low*. In most cases, the theoretical norm is about 1.4 purchases per "new" buyer in the second period.

In the present instance the agreement with the theoretical rates is close enough in the 4-weekly periods, but in the longer periods the observed buying frequencies are higher than the norm. This is an unusual finding. It is however very systematic here for all brands, and is even more marked for the product-field as a whole.

These discrepancies might have been explained by the seasonality of the market, with some "new" buyers being peak-season-only buyers and relatively heavy ones at that, but in fact, the phenomenon occurs all the year round. Instead, we seem to have a situation where some people have an intense but short-lived enthusiasm for a particular brand, but then stop buying it (the "jag" phenomenon, as denoted some time ago by Dudley Ruch). This might seem a normal enough form of buying behaviour, but experience shows that it does *not* occur in most product-fields (see for example Tables 2.1 and 5.1 in Chapters 2 and 5). At this stage, we have therefore established that "new" buyers in this particular product-class tend to be unusually heavy buyers, thus pointing up a phenomenon where further examination can be undertaken.

Table 3.8. The Buying-Frequency per "New" Buyer

The Average Number of Purchases per "New" Buyer in Period of Various Lengths

(Observed Values "0" and Theoretical Norms "T")

	Period of Length (in weeks):							
	1		4		12		24	
	0	T	0	T	0	T	0	T
ANY BRAND	1.1	na	1.5	1.5	2.1	1.6	3.2	1.6
Brand A	1.0	na	1.3	1.4	2.0	1.5	2.3	1.6
Brand B	1.0	na	1.3	1.3	1.8	1.4	2.3	1.4
Brand C	1.0	na	1.3	1.3	1.6	1.4	1.7	1.4
Brand D	1.0	na	1.3	1.3	1.5	1.4	2.3	1.4
Brand E	1.0	na	1.1	1.3	1.5	1.4	2.0	1.4
Average Brand	1.0	na	1.3	1.3	1.7	1.4	2.1	1.4

* As already mentioned when introducing the concept of "new" buyers in § 2.2 of Chapter 2, a "new" buyer is defined here as someone who has not bought the item in question in the preceding equal period but may have bought it before that.

Table 3.9 Repeat-Buying in Non-Consecutive Periods

A symptom of one form of "jag-buying" would be a decay of repeat-buying loyalty over longer periods of time. More generally, a repeat-buying rate which falls away in non-consecutive periods might indicate increasing brand-switching and a possible erosion of the brand's franchise. If however buying behaviour is in an equilibrium state, the proportion of repeat-buyers in non-consecutive quarters such as from Quarter I to Quarter III should be the same as that found for consecutive periods: failure to repeat-buy would simply be attributable to infrequent buying and not to any fundamental change in buying habits.

Table 3.9 compares the incidence of repeat-buying in non-consecutive quarters with the figures for consecutive quarters and with the theoretical norms*. Some special turnover of buyers might be expected here because of the seasonality of the market, but in fact there is little difference between the consecutive and non-consecutive figures, irrespective of whether repeat-buying is at the normal level (as for the product-field as a whole and for Brands A and E) or somewhat below (as for Brands B and D). Only Brand C shows some erosion of repeat-buying and this is linked with the drop in the quarterly purchasing rate

Table 3.9. Repeat-Buyers in Non-Consecutive Periods

The *Percentage of Buyers in One Quarter who Also Buy Two Quarters Later*

(Average of Quarters I/III and II/IV)

	Non-Consec, Quarters	Consecutive (See T3.6)	Theoretical Norm
	%	%	%
ANY BRAND	84	85	84
Brand A	75	78	77
Brand B	54	59	65
Brand C	40	50	67
Brand D	52	53	64
Brand E	67	69	69
Average Brand	58	62	68

* According to the underlying NBD/LSD model as described in the next chapter, the same theoretical norms should apply in consecutive and non-consecutive periods.

of Brand C noted in Table 3.2. The evidence therefore points mainly to a steady level of repeat-buying even over extended time-periods, as is found in most product-fields.

Table 3.10 Repeat-Buying By Light and Heavier Buyers

The tables below present a more detailed analysis of repeat-buying than attempted so far. They consider separately non--buyers, lighter, and heavier buyers in one period and their buying in the subsequent period. This analysis provides a sensitive method of investigating trends in repeat-buying. It is frequently used in the detailed study of specific trend situations (as illustrated in § § 5.5 and 6.2 in Part III).

Table 3.10. Repeat-Buying by Light and Heavier Buyers

% Buying in QIII by purchasing level in QII

	Buying of stated item in Quarter II					
	Non-buyers		Once only		More than once	
	0 %	T %	0 %	T %	0 %	T %
NORMAL						
Any Brand	29	26	62	**61**	93	92
Brand A	18	**18**	**51**	**55**	87	89
Brand E	2	2	40	47	92	86
Average	**16**	**15**	51	**54**	91	89
BELOW NORMAL						
Brand B	10	7	39	47	78	82
Brand C	4	3	25	47	80	**85**
Brand D	5	2	35	46	87	83
Average	6	4	33	47	82	83

Table 3.10 deals with Quarters II and III, treating separately the total product (i.e. "Any Brand") and Brands A and E, for each of which the incidence of repeat-buying in Table 3.6a had been normal, and Brands B, C and D. For each item, the households in the sample are firstly grouped into those who did not buy the item (i.e. the product or

specific brand) in Quarter II, those who bought it only once, and those who bought it more than once. Thus the table shows what proportion of each group bought it at least once in the *next* quarter (i.e. Quarter III).

The observed buying levels in Quarter III are then compared with the appropriate theoretical NBD norms (see §7.6). We see close agreement at all points for the product-field as a whole ("Any Brand") and for Brands A and E. These are the three items for which the more aggregated repeat-buying levels 12-weeks by 12-weeks in Table 3.6a had already matched up to the theoretical norms.

For Brands B to D the total incidence of quarterly repeat-buyers in Table 3.6a was however below expectations. This could have arisen either from an excess of "new" and "lapsed" buyers moving in and out of the market periodically, or from an actual shortage of repeat-buyers as such – two alternatives with very different implications.

Table 3.10a. Repeat-Buying by Light and Heavy Buyers

Average number of purchases per buyer in Quarter III

	Buying of stated item in Quarter II					
	Non-buyers		Once only		More than once	
	0	T	0	T	0	T
NORMAL						
Any Brand	2.1	1.6	2.6	2.1	6.1	6.5
Brand A	1.9	1.5	1.9	2.0	4.6	5.1
Brand E	1.6	1.4	2.3	1.9	4.5	4.8
Average	1.9	1.5	2.3	2.0	5.1	5.5
BELOW NORMAL						
Brand B	1.8	1.4	2.2	1.8	3.4	3.5
Brand C	1.4	1.4	2.6	1.9	3.8	4.4
Brand D	1.6	1.4	2.1	1.8	3.8	3.9
Average	1.6	1.4	2.3	1.8	3.7	3.9

The analysis in Table 3.10 now shows that the repeat-buying franchise amongst ***heavier*** buyers of Brands B, C and D is almost precisely at the predicted (normal) level. The shortfall of repeat-buyers in fact occurs amongst ***light*** buyers of each brand – i.e. those who bought only once in the first period. (There is a compensating excess of "new" buyers in the second period – i.e. people who did not buy in the first period.)

This evidence therefore points once more to seasonal variation with a somewhat higher than normal level of brand-switching amongst *light* buyers. It also shows that there is no fundamental weakness in the repeat-buying loyalty of the heavier buyers of these brands.

Table 3.1 Oa gives the corresponding buying frequencies in Quarter III for the groups shown in Table 3.10. The results for heavier buyers are again close to the norms, but those for the "new" buyers and for the light buyers tend to be somewhat higher than expected. This reflects the unusually high buying rates for "new" buyers which were already noted in the 12-weekly results in Table 3.8.

Summary and Conclusions

The preceding analyses show that penetration growth and repeat-buying for the different brands of the product-field covered here are largely predictable.

One brand may have more buyers than another, but repeat-buying and the growth of penetration over different time-periods generally have the same form. In many respects this is as occurs in other product-fields (as is summarised by the theoretical NBD model), but there are also certain systematic differences from these more general norms. These discrepancies tend to be fairly small and can therefore only be spotted by analyses such as those illustrated here, rather than by direct examination of the observed data as such.

Thus for Brands B, C and D there is some excess of penetration in the longer periods (Table 3.1 a), and a substantial shortage of repeat-buyers of each brand from one quarterly (or longer) period to the next (Table 3.6a). For Brands C, D and E there is a small excess of once-only buyers, and more generally there is quite a marked tendency here for "new" or "lapsed" buyers to buy in abnormally heavy bursts or "jags" (Table 3.8).

The analysis of repeat-buying in non-consecutive periods (Table 3.9) shows that there is no marked tendency for the incidence of repeat-buyers (whether low or normal) to be eroded as time progresses. The indications are not so much that "jag-buyers" stop buying a given brand altogether, but that they stop temporarily and that they mostly return to the brand subsequently.

These deviations are partly linked to the seasonal trend in the market but there seems here also to be evidence that a few people temporarily "get tired" of a brand and this is something which could be subjected to further analysis. Thus the analysis in Table 3.10 already shows that from one quarter to the next it is the *light* buyers of each brand (not the *heavier* ones) who tended not to buy again to the normal extent.

For the product-field in total the number of repeat-buyers is normal. The relatively low incidence of reseat-buying for some of the individual brands is therefore a matter of extra switching between brands, rather than a question of moving out of or into the market altogether. A more direct attack on brand-switching behaviour in this product-field is described in Chapter 9.

The analyses have shown that the apparent short-fall of repeat-buyers for Brands B, C, and D is not due to any weakness in repeat-buying at all, but reflects an excess of occasional buyers – something with radically different marketing implications.

Despite these various exceptions, the outstanding characteristic of the data analysed here remains their basic regularity. Even the various departures from the norms are too systematic to hide this. The findings are therefore simple enough to lead to a better understanding of buyer behaviour and to provide a basis for evaluating specific marketing problems.

CHAPTER 4

BASIC THEORY

4.1. Three Forms of Repeat-Buying

In this chapter the basic concepts and working formulae of the NBD and LSD theory of repeat-buying will be set out *. Repeat-buying may be regarded as any situation where a consumer buys more than one unit of a particular item, such as a particular brand or pack-size of a consumer product. Three cases of repeat-buying can be distinguished.

Firstly, if a consumer buys the item at all in the given time-period, he may buy it on more than one purchase occasion. Different buyers differ in how often they repeat-buy in this sense. The resultant frequency distribution of purchases (i.e. the number of consumers making 0 or 1 purchases, or, if repeat-buyers, making 2, 3, 4 etc. purchases) can then generally be modelled by the Negative Binomial or Logarithmic Series Distributions. This is discussed in §§4.2 to 4.4 of this chapter.

Secondly, a consumer may buy the item in more than one time-period. In § 4.5 a certain underlying model of buyer behaviour is introduced which could represent this form of repeat-buying under stationary no-trend conditions, and which in practice yields formulae which successfully do describe the incidence of period-to-period repeat-buying (§§ 4.6 and 4.7) as well as related phenomena like penetration growth over different length periods (§ 4.8). This model also explains the basic NBD or LSD frequency distributions which occur in a *single* period.

The third form of repeat-buying is that more than one unit may be bought on the same purchase occasion. For many frequently-bought branded consumer goods this hardly ever happens (most buyers buy one unit on most occasions – see for example Table 3.3 in Chapter 3). But even for product-fields where multi-unit purchases do occur (e.g. dog foods or, in a sense, petrol), the tendency is for the average number

* The mathematical formulae can be passed over quickly by the reader, especially on a first or second reading: all one needs to note is that certain formulae exist for the stated purpose. The more mathematical details and proofs are in any case left over to Chapters 7 and 8, for the more technical reader. A worked numerical example for the calculations involved is given in Appendix A.

of units bought per purchase not to vary much from brand to brand. This form of repeat-buying therefore seems relatively easy to deal with by straightforward analysis. We therefore by-pass it here through using the ***purchase occasion*** as analysis unit, as has already been indicated in Chapter 1.

4.2. The Number of Purchases in a Time-Period

For most kinds of consumer products there tends to be something like a minimum time-interval between one purchase and another. This is a week for many types of household products (which are normally bought ***at most*** once a week), or a day or so for cigarettes. We now consider the frequency of purchase in time-periods which are relatively long compared with this minimum (e.g. months or quarters or years, rather than a week). Special problems arise for periods near the minimum and these will be discussed in § 4.9 and in Chapter 7.

Even in a relatively long time-period such as a month or a year, many buyers of any particular brand buy it only once, some buy it twice, fewer buy it three times, and so on, with smaller frequencies of still heavier buyers. Furthermore, the total number of consumers who buy the brand at all in the analysis-period is often a relatively small proportion of the total population of potential buyers, so that the largest single observed frequency is usually that of ***non-buyers*** of the item.

Typically, we therefore have a highly skew distribution (as illustrated by the column of "observed" figures in Table 4.1). The theoretical finding then is that such distributions can generally be fitted by one particular mathematical function, namely the Negative Binomial Distribution (or NBD for short)*. A typical example – the first one published – is given in Table 4.1 from a certain 26-week period, with 16 12 households out of a sample of 2,000 not buying the item at all,

* The need to fmd some such theoretical model for the observed frequency distributions of purchases was the start of the present work on repeat-buying. It arose specifically from a suggestion by Mr. D.A. Brown in dealing with a practical data-handling problem in the late 1950s which concerned a possible excess of "heavy buyers" in some observed data. The most obvious theoretical distribution to try and fit to the distribution of relatively rare events like consumer purchase is the Poisson, but this was found to be not skew enough for the observed data. The NBD is more skew than the Poisson. It was one of several theoretical distributions which was tried next and it gave a good fit, both in the initial studies and subsequently (see also § 11.4 in Chapter 11).

Table 4.1. The Earliest Published Example of the Fit of the Negative Binomial Distribution to PurchasingData

(26-week data for a 2,000 household sample)

Number of Purchases	No. of Households Observed	No. of Households Theoretical
0	1,612	(1,612)*
1	164	157
2	71	74
3	47	44
4	28	29
5	17	20
6	12	15
7	12	11
8	5	8
9	7	6
10	6	5
11-15	11	12
16+	8**	7

Proportion of non-buyers: $p_0 = .806$*
Proportion of buyers: $b = 1 - p_0 = ,194$

Average number of purchases *per household:* $m = 0.64$*
Average number of purchases *per buyer:* $w = m/b = 3.3$

Negative Binomial exponent: $k = 0.115$, and $a = m/k = 5.53$

Standard deviations: Observed (root-mean square) = 2.12
Theoretical $\sqrt{\{m(1+a)\}}$ = 2.04

* Used in fitting the theoretical distribution.
** Actual values: 17, 17, 20, 22, 22, 25, 26, 26.

164 households buying it once, 7 1 twice, and so on [Ehrenberg 1959]. The theoretical NBD purchase frequencies are 157 households buying once, 74 buying twice, and so on, and the agreement is clearly good.

Two coefficients or "parameters" have to be calculated in order to fit the theoretical distribution. This is usually done by using the observed proportion of zeros (the number of non-buyers) and the average number of purchases per household (the mean of the distribution) as inputs. These two quantities are usually denoted by p_0 and m, and the technical details are set out in the next few paragraphs.

The Mathematics of the NBD. The Negative Binomial Distribution is based on the non-negative integers 0, 1, 2, 3, 4 and in general *r*. The distribution is specified by two parameters, the mean, *m*, and the exponent *k*. The probability p_r 'of observing *r* purchases is then given by

$$p_r = \frac{(1+m/k)^{-k}\,\Gamma(k+r)}{\Gamma(r+1)\,\Gamma(k)}\left(\frac{m}{m+k}\right)^r ,$$

which follows from expanding the binomial expression $(1-m/m+k)^{-k}$, in which the exponent *k* has a negative sign. The negative binomial distribution is always positively skewed, and has one mode, which is at zero for the fairly small values of *m* and *k* which generally occur with consumer purchasing data. The variance of the distribution is $m(1 + m/k)$ or $m(1 + a)$, where the quantity $a = m/k$ is another useful way of expressing the parameters of the NBD.

In fitting a negative binomial distribution to empirical data, the two parameters *m* and *k* have therefore to be estimated. The best estimate of the mean *m* is simply the observed mean (for sample data it is the maximum-likelihood estimator and is also unbiased).

Estimating the second parameter *k* (or *a*) is less straightforward. (The maximum-likelihood equations for the second parameter are for example very cumbersome to solve.) Various ways of estimating *k* or *a* have however been developed [cf. Anscombe 1950]. An estimating procedure common in statistics is the method of moments, which here means estimating *a* by equating the observed sample variance to its expected value $m(1 + a)$. This is however not particularly efficient statistically for the NBD (sometimes less than 50%), although it would be adequate with large samples; but in practice it would often be laborious to have to compute the observed variance, since with market research data the basic frequency distribution is not usually tabulated.

An alternative method utilizes the number of non-buyers. We equate the observed proportion of zero readings p_0 to its expected value, i.e. we write

$$p_0 = (1+m/k)^{-k}, \qquad \text{or} \qquad (1+a)^{-m/a}.$$

This equation cannot be solved directly for *k* or *a*, from given values of p_0 and *m*, but various indirect procedures exist *. One is by iteration and this can readily be computerised if it is to be done routinely. Another is to work out the quantity $c = -m/\ln p_0$ from the observed values of *m* and p_0 and read off the corresponding value of *a* from Table B.3 in Appendix B, taken from Chatfield [1969]. (A transformation such as $c = -m/\ln p_0$ makes for ease of accurate interpolation; this is a simpler version, developed by G.J. Goodhardt, of an earlier suggestion by Evans [1953]†.)

In general, it is very convenient with consumer purchasing data to estimate the second parameter *k* or *a* of the NBD from the proportion of non-buyers p_0 in this way. Thus the mean m and the proportion of non-buyers p_0 – or the proportion of *buyers* $(1-p_0)$, usually denoted by *b* – are often all the observed figures that are routinely tabulated. Statistically, this method is in any case at least 90% efficient for most such data, and often a good deal more so [cf. Anscombe 195 0, for very low values of *k*].

Having estimated the two parameters *m* and *k* (or a), we need to calculate the theoretical proportion p_r of the sample who make *r* purchases. A relatively simple procedure is to use an

* Both the NBD and LSD theories appear mathematically more cumbersome than they really are because each involves, at its very earliest stage, a relationship between the theoretical parameters and the observed data which cannot be solved directly.

† Data for which $m < -\ln p_0$ cannot be fitted by an NBD.

iterative formula (starting with the observed value of p_0), namely

$$p_r = \left(\frac{a}{1+a}\right) (1 - y) \quad p_{r-1}\,.$$

The values tend to be tedious to compute by hand if r is at all large, but the procedure can be readily computerised for routine use.

The goodness of tit of the estimated distribution can be tested by calculating the chi-squared value for the observed and theoretical frequencies along standard statistical lines. A quicker test in practice is to compare the observed variance with the theoretical value $m(1 + a)$, the distribution having been fitted by the mean and the proportion of zeros. (A test of significance for this difference has been given by Evans [1953].)

4.3. The Fit of the Negative Binomial Distribution

In fitting the negative binomial distribution to consumer purchasing data, a good fit has been obtained in most stationary cases, i.e. when there is no marked trend in the aggregate sales level. A typical example, taken from the earliest data analysed, was shown in Table 4.1, the fitting as usual being by equating the number of zeros and the mean of the *theoretical* distribution to the corresponding *observed* values. The fit in this case was clearly close, as may be seen by eye. (A summary measure of the fit is obtained by comparing the standard deviations of the observed and theoretical distributions, at 2.12 and 2.04.) Many thousands of further cases with widely differing characteristics have now been successfully analysed, as is summarised in Table 4.2. Some recent examples were shown in Table 3.4 in Chapter 3, where certain specific discrepancies between the observed and theoretical frequencies are also discussed.

In general then, the fit of the theoretical NBD to observed purchasing data is good. Specific exceptions can however also occur. Some were already noted in the earliest work and include a clustering tendency at or near the number of weeks in the analysis unit [Ehrenberg 1959]. This is a manifestation of the minimum inter-purchase time-period effect which was referred to briefly at the beginning of § 4.2 and will be discussed more fully in § 4.9 and in Chapter 7. Thus for products which tend to be bought at most once a week, the number buying *more* than 26 times in a 26-week period, say, is less than the theoretical value. There is then something of a compensating excess of people who regulate their purchases to roughly the once-a-week cycle. Examples were given in Table 2.1 in Chapter 2, but this effect only shows up for products where the number of buyers. who *do* buy once a week is

Table 4.2. Conditions under which the Negative Binomial Distribution has Generally been Found to Hold *(updated as Table I in Preface)*

- For a variety of product-fields, *viz:*

  Breakfast Cereals, Butter, Canned Vegetables, Cat and Dog Foods, Cocoa, Coffee, Confectionery, Convenience Foods, Cooking Fats, Detergents, Disinfectants, Flour, Food Drinks, Household Soaps, Household Cleaners, Instant Potatoes, Jams and Marmalade, Margarine, Motor Oil, Petrol, Polishes, Processed Cheese, Refrigerated Dough, Sausages, Shampoos, Soft Drinks, Soup, Take-home Beer, Toilet Paper, Toilet Soap

- The leading brands in each product-field
- Large, medium and small pack-sizes and the brand as a whole
- Great Britain, Continental Europe, U.S.A.
- 1950-1970
- Various Demographic Subgroups
- Analysis Periods ranging from 1 week to 12 months

appreciable *. The shortfall of heavier buyers then leads to a lower value for the standard deviation (or variance) of the observed distribution as against that of the theoretical NBD. This is the so-called "variance discrepancy" which will be discussed more fully in Chapter 7 †.

There are also some products where the fit is not good in ways that are not yet fully understood. Essentially this seems to apply particularly to very regularly bought items such as perhaps bread, cigarettes and milk, or in certain cases also the total product-class (even when individual brands or pack-sizes give a good fit). Further discussion of discrepancies occurs at later stages in this book.

4.4. The LSD Approximation

A simplifying approximation to the Negative Binomial Distribution which holds under certain circumstances is the Logarithmic Series Distribution or LSD. (This is the model used in § 2.3 of Chapter 2.) It

* It hardly shows up in the example of Table 4.1, although the three heaviest buyers (out of the 8 buyers who made 16 or more purchases) bought 25 or 26 times in the 26-week period, and no one bought more often than 26 times.

† The title "variance discrepancy" arose historically but does not refer to what is now regarded as the most pertinent feature of this phenomenon.

applies when the proportion of the population buying the item in the analysis-period is relatively low – roughly that b, i.e. $(1 - p_0)$ is less than about 0.2 or 20%. It is then possible to fit the distribution of *buyers* (i.e. excluding the zeros or non-buyers) by a certain theoretical distribution, the LSD. The number of non-buyers is therefore treated separately from the buyers, and the LSD has only one parameter and hence is simpler than the NBD which has two (cf. Chatfield et al. 1966).

The LSD is not an *alternative* to the NBD, but merely a special case of it (see Chapter 8) which under certain conditions (i.e. roughly for $b < .2$) gives virtually the same results, but more simply. Since the LSD has only one parameter, many formulae for repeat-buying etc. are particularly straightforward in their LSD form, as was illustrated in Chapter 2 and is discussed more fully in §§ 4.7 and 4.8 of this chapter. An extreme example is the formula for the proportion of the total sales of an item which is accounted for by those relatively "heavy" buyers of the item who buy it more than r times in the analysis-period. In the LSD theory there is the very simple formula q^r which was used in § 2.3 of Chapter 2 and which is discussed below and in § 8.4 of Chapter 8. In the NBD theory there is *no* equivalent (simple) formula.

Conceptually the importance of the LSD is that it shows how buyer behaviour is independent, or at least largely independent, of the precise definition of the population of potential buyers, as long as this is large enough compared with the number of *actual* buyers (i.e. $b < .2$ or so).

The Mathematics of the LSD, We let p'_r stand for the proportion of all buyers in the given time-period who made r purchases of the item in that period. The values of p'_r for all $r \geq 1$ can then in practice be represented by the one-parameter Logarithmic Series Distribution, irrespective of the precise number of non-buyers (as long as this is large). It is important to note that p'_r is used here to stand for the proportion of *buyers in* the given period who made r purchase, whereas p_r without the dash stands for the proportion of the total sample (i.e. including the non-buyers) who made r purchases. We therefore have

$$p'_r = p_r/(1-p_0).$$

The mean of the distribution is denoted by w, where $w = m/(1-p_0)$ in terms of the mean m of the total distribution (including zeros).

The LSD probabilities for there being p'_r buyers making r purchases (for $r \geq 1$) are best expressed in terms of a certain parameter q:

$$p'_r = \frac{-1}{\ln(1-q)} \frac{q^r}{r}$$

the logarithm (ln) being to base e. (Tables of such "Natural" or "Naperian" logarithms to base e are reproduced in Appendix B.) The parameter q can be related to w, the mean of the observed distribution of purchases, by the implicit equation already noted in Chapter 2,

$$w = \frac{-q}{(1-q)\ln(1-q)}.$$

This equation cannot be solved directly for q in terms of w (see also the second footnote in §4.2). The numerical value of q for a given w can be obtained from Table B.2 given in Appendix B, of which Table 2.5 in Chapter 2 was a sample extract. Alternatively, q can be calculated by iteration (e.g. on a computer), or by the simple approximation

$$q \doteq (w-1.4)/(w-1.15), \qquad \text{which holds to within} \pm .01 \text{ for } w>2.$$

The value of q lies between 0 and 1. It quickly reaches high values close to 1 – e.g. it is as high as .9 for a relatively low value of w such as w = 4 – so that considerable accuracy is often needed in its arithmetical calculation. A better way of calculating q from w is through forming a new parameter $a = q(1+q)$ and using Table B.3 in Appendix B, as is discussed in § 8.3.

One aspect of the distribution which is of some special interest is its "tail", i.e. the purchases made by heavier buyers. This amounts to Σip_i for values of i greater than some number r. Expressed as a proportion of total purchases, this can easily be shown to reduce (see § 8.4) to the simple expression q^r, as already noted earlier.

4.5. An Underlying Model of Stationary Buyer Behaviour

So far we have examined the incidence of purchases in a single time-period of some given length. Under stationary conditions, the distribution of purchase frequencies then tends to follow the NBD or LSD and this applies for a period of any length (subject to the "minimum time-period" limitation touched on in § 4.2 and discussed further in § 4.9). We have however not yet considered how the results in periods of different lengths are related, nor have we yet dealt with repeat-buying from one period to another. For all this, a more elaborate model is needed.

The NBD approach is fruitful in this respect because there is a theoretical model of a stochastic (or "quasi-random") kind which not only yields an NBD in any single time-period, but also provides formulae relating the results in different length periods and for period-to-period repeat-buying. Furthermore, the model in question not only seems "reasonable" on *a priori* grounds, but has also been found to work well in practice.

The model is a twodimensional one, one dimension being time, and the other (an unordered one) being the individual consumers, as is

* In the modern literature, this type of model tends now to be referred to as a "mixed Poisson".

Table 4.3. A Stochastic Model over Time yielding the NBD in any given Period

	Successive Time-Periods						Long-run Averages	Distributions (horizontally)
	I	II	III	IV	.	.		
Consumer								
a	*x*	x	*x*	x	*x*	*x*	μ_a	Poisson
b	*x*	*x*	*x*	*x*	*x*	*x*	μ_b	Poisson
c	*x*	*x*	*x*	*x*	*x*	*x*	μ_c	Poisson
d	*x*	*x*	*x*	*x*	*x*	*x*	μ_d	Poisson
.	*x*	*x*	x	*x*	*x*	*x*	$\mu_.$	Poisson
	x	*x*	*x*	*x*	*x*	*x*	$\mu_.$	Poisson
Mean	m	*m*	*m*	*m*	*m*	*m*	*m*	
Distributions (vertically)	NBD	NBD	NBD	NBD	NBD	NBD	Gamma	

Note: The x-values in the body of the table represent varying observed numbers of purchases and are not intended to imply equality.

shown schematically in Table 4.3. The model is of a s-called "compound Poisson" type * [e.g. Anscombe 1950, Feller 1957, Haight 1965, Boswell and Patil 1970], the details being broadly as follows:

The Incidence of Purchases over Time. Any particular consumer makes some sequence of purchase of an item, e.g. 2, 0, 3, 1, 1, etc. purchases, in successive equal periods of time (see for example Table 1.2 in Chapter 1). The model then requires that these purchase frequencies should behave like independent random samplings from a system where the event (i.e. a purchase) has the same probability at any given point in time, and where these probabilities are independent of each other. This is then a so-called Poisson distribution.

This Poisson formulation for individual purchase sequences is a plausible *a priori* assumption under two conditions which should normally be more or less fulfilled, namely that

(a) Successive analysis-periods must not only be of equal length but also be "similar" to each other, e.g. weeks or longer periods measured in weeks for household-products (rather than days, because shopping patterns on Mondays tend to differ from those on Tuesdays or Saturdays and hence tend to be non-stationary, whereas in periods longer than a week, these short-term effects are in effect balanced) *. This "similarity" of the successive periods has to extend to actual purchasing behaviour itself. In other words, there must be no trend in the aggregate sales or penetration figures (or in the parameters of the underlying model). This is the "stationarity" assumption.

(b) The analysis-periods must not be too short, so that the purchases made in one period do not directly affect those made in the next. Periods of one week for instance may be too short for some products. For example, if a tin of cocoa is bought in one week, no such purchase is likely to be made until the initial purchase is more or less used up. This is part of the minimum time-interval problem which is discussed further in § 4.9.

In practice, these conditions are often likely to hold for periods which are long enough in terms of the minimum inter-purchase or usage habits for the product in question †. The basic test is of course how well a model based on such assumptions works in practice.

The Poisson distribution has one parameter, which may be denoted by μ, the consumer's average rate of purchasing "in the long run". We now consider the distribution of these mean values for different consumers.

Differences between Consumers. The second part of the model concerns the differences in the average purchasing rates of different consumers.

The model states that the frequency distribution of the long-run average purchasing rates μ_a, μ_b, μ_c, etc. of different consumers *a, b, c,* etc. should be proportional to a so-called "Gamma" distribution (with exponent k). This is a statistical probability distribution for non-negative values which is either reversed J-shaped or humpbacked, and always positively skewed. It is therefore of the right general shape (this being what the observed data generally look like), and is rather flexible, having two adjustable parameters. In fact, many different kinds of data

* At a more technical level, we note that purchasing behaviour is patterned in discrete time-periods, although the Poisson process is really continuous in time. When working in relatively long analysis-periods, this discrepancy does not greatly affect the fit of the model in most respects, although it underlies the problems already touched on in § 4.3.

† Something is known about the way in which the Poisson assumption does not quite hold in practice (cf. § 7.8), but this has little effect on most of the results here.

can be fitted by a Gamma-distribution, and this is therefore not a particularly stringent assumption *.

To summarise, the general model is one of Poisson distributions "compounded" by a Gamma-distribution, in that the *i*th consumer's purchases over time are to follow his own Poisson distribution with a mean μ_i, the means for different consumers following a Gamma-distribution, as was set out in Table 4.3. It is not necessary to assume that purchases actually follow this Poisson-Gamma model in successive periods, let alone in the long run. (The average "long-run" rates of purchasing μ_a, μ_b, μ_c, etc. which have been postulated for individual consumers are in fact not directly observable.) Instead, it is only necessary to suppose that for any one or more periods of time, consumers' purchases behave *as if* they were random samples from such a stochastic model. (In the next period, the parameters of the model – or the model itself – could have changed, due to some marketing disturbance or seasonal trend say, although usually this does not happen.)

A number of mathematical deductions can then be made from this formulation and tested empirically. The first deduction of this kind is that the distribution of purchases in any single time-period should follow the negative binomial distribution, (see § 7.2). This tends to be so in practice, as we have seen, and was of course the starting-point of the whole study.

Other deductions concern repeat-buying from one period to the next (to which we now turn), and the way in which the values of the penetration *b* and of the average purchase frequency w vary in periods of different lengths (as is discussed in § 4.8). It is the extent to which these deductions generally fit the observed facts which determines the practical validity and usefulness of this Gamma-Poisson model.

4.6. Period-to-Period Repeat-Buying

The repeat-buying pattern from one period to the next is of fundamental interest both for the theoretical understanding of buyer behaviour and for practical applications. There are four main questions to be

* By the same token, it is not possible (or necessary) to adduce any strong reasons why a Gamma-distribution should hold. No special consequences follow from the assumption other than the various repeat-buying formulae etc. whose validity is tested directly against the observed data anyway. (This has been superseded– See Goodhardt and Chatfield 1973 and Chapter 13, $13.2).

Table 4.4. The Definition of Repeat, Lapsed and "New" Buyers in Two Time-Periods

Definition	Period I	Period II
Repeat-buyers:	Buying	Buying
Lapsed buyers:	Buying	Not buying
"New" buyers:	Not buying	Buying
Non-buyers:	Not buying	Not buying
All buyers in period:	Repeat + Lapsed	Repeat + "New"

posed to any theoretical formulation, namely whether the theory can successfully predict:

(a) What proportion of the buyers of the item in one period also buy it in the next period and by implication,

(b) What proportion of buyers in each period do not buy in the other, i.e. are "lapsed" or "new" buyers, as is set out schematically in Table 4.4. Next,

(c) How often the repeat-buyers buy in each period, and finally,

(d) How often the "lapsed" buyers and the "new" buyers – i.e. the one-period-only buyers* – buy in the period in which they buy at all.

We now set out the answers to these questions which are given by the NBD/LSD theory under stationary (i.e. no-trend) conditions from one period to the other. The point is that if we are given data about purchasing behaviour in the *first* period, the theory predicts repeat-buying behaviour concerning the *next* period†.

4.7. Three Levels of Theory

The NBD/LSD repeat-buying theory gives results at three levels. These levels are of increasing mathematical simplicity but of decreasing generality. They are the NBD, the LSD, and certain numerical approximations to the latter.

* One-period-only out of the *pair* of periods being analysed.

† The NBD theory also copes with various elaborations of the above questions, such as the incidence of repeat-buying in the second period amongst light, medium, or heavy buyers in the first period, or the nature of repeat-buying in unequal time-periods, or in more than two periods, or in non-successive periods. The answers to these more elaborate questions are set out in Chapter 7 (with some practical applications being already given in Chapters 3, 5 and 6).

In using the NBD and the LSD theories to estimate how many repeat-buyers there should be in the second of two periods and how often they should buy, it is first of all necessary to calculate certain characteristic numbers or "parameters" for the *first* time-period. In the NBD there are *two* such parametes. One is m, which is simply the observed mean number of purchases per informant in the first period. The second is the special NBD parameter k (the negative binomial "exponent"). Its value can be calculated from the observed values of m and b (the observed proportion of buyers) in the first period. The NBD equation linking these observed values to k is set out in Table 4.5, and the calculations involved in solving this equation have already been outlined in § 4.2 earlier in this chapter. (A numerical example of this and the remaining calculations in this chapter is set out in Appendix A.)

In the LSD theory, there is only one special parameter value that needs to be worked out. This is the LSD parameter q which can be calculated, as already mentioned in § 4.4 and in Chapter 2, from the observed value of w, the average number of purchases per buyer in the first time-period. The relevant equation is also given in Table 4.5. Since the LSD theory requires only this single parameter it gives simpler results than the NBD. But the theory holds to a reasonably close degree of approximation – i.e. gives virtually the same results as the NBD – only for data where the proportion of buyers, b, is less than 20%. For items with a higher penetration, the LSD and NBD repeat-buying estimates diverge, and it is the NBD values which give a good fit.

The third type of repeat-buying formulae consists of numerical approximations to the LSD ones. Here there is no need to calculate any special parameter, as the formulae are expressed directly in terms of the

Table 4.5. The Basic Equation for the NBD and LSD parameters k and q, in terms of the Observed values of b, m and w

NBD:

$$b = 1-(1+m/k)^{-k}, \text{ or}$$
$$= 1-(1+a)^{-k}, \text{ where } a = m/k.$$

LSD:

$$w = -q/\{(1-q)\ln(1-q)\}, \text{ where } w = m/b.$$

b = the proportion of informants buying the item in the period
m = the average number of purchases per *informant*
w = the average number of purchases per *buyer* (w = m/b)

observed value of the average purchase frequency, w, in the first time-period. Arithmetically, these are therefore the simplest formulae, but they give close approximations to the NBD/LSD results only for a certain range of values of w, generally that w is greater than about 2 but less than about 20 (and of course that the penetration b is less than 20%, as for the LSD theory generally).

Given the appropriate values in the first time-period, each of the three levels of theory can now be used to give estimates of the various aspects of repeat-buying in the next equal period, under stationary conditions. (Whilst the LSD or approximation formulae are simpler – in those situations where they give the right answers at all – the more general NBD formulae are normally used when relatively large-scale or repetitive applications are involved, the calculations being simple enough when routinised on a computer.)

We start with the incidence of repeat-buyers, i.e. the proportion b_R of the population who buy in both periods. The three formulae for b_R are set out in Table 4.6 (the LSD and the "Approximation" formulae being expressed as b_R/b, i.e. as the proportion of the buyers in the first period who buy again in the next, as they are simpler in this form). The numerical values given by these formulae are illustrated in Table 4.15 at the end of this chapter.

The earliest published examples of the fit of the NBD formula to empirical data is reproduced in Table 4.7: there are some discrepancies (partly due to some minor non-stationarity and to sampling errors), but

Table 4.6. Three Formulae for b_R, the Proportion of the Population who are Repeat-Buyers in Two Equal (Stationary) Time-Periods, in terms of the Observed Parameters in Table 4.5.

NBD:

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

LSD (for $b < .2$):

$$\frac{b_R}{b} = 1 + \frac{\ln(1+q)}{\ln(1-q)}$$

Approximation (for $b < .2$ and $2 < w < 20$):

$$\frac{b_R}{b} \doteq \frac{2(w-1)}{2.3w-1}$$

Table 4.7. Example of the Fit of the NBD Formula for the Percentage of Repeat-Buyers

(The first published examples of NBD estimates of quarter-byquarter repeat-buying for 12 different items [Ehrenberg 1964])

	Brand or Pack-Size											
	L	M	N	O	P	Q	R	S	T	U	V	W
The Given Data in Quarter I:												
% buying = 100 *b*	1.3	3.8	4.5	4.9	6.5	8.9	lo	12	15	15	24	34
Mean buying rate = *m*	.03	.09	.14	.15	.20	.26	.28	.53	.67	.66	.80	1.5
% of Sample Buying in both Q.I and Q.II:												
Theoretical	.07	2.4	3.1	3.4	4.5	6.1	7	9	11	11	18	26
Observed	.06	1.4	2.2	2.5	4.1	4.5	7	9	10	8	18	26

there clearly was major agreement overall. This has been confirmed in many thousands of cases which have been examined since, some examples being given in Chapter 3 (Table 3.6) and in Chapters 5 and 6, where discrepancies are also further discussed (see also § 4.9 and Chapter 7).

Next, we consider the average frequency with which these repeat-buyers buy in the second period, expressed as m_R on a "per informant" basis in the NBD theory, or as w_R on a "per repeat-buyer" basis in the LSD, i.e. dividing through by the number of repeat-buyers. Table 4.8 gives the three formulae, the numerical approximation to the LSD giving the particularly simple result

$$w_R \doteqdot 1.23w\,,$$

for values of w such that $1.5 < w < 20$.

Table4.8. Three Formulae for w_R, the Average Purchase Frequency per Repeat-Buyer under Stationary Conditions

NBD:

$$m_R = m\,\{1 - (1+a)^{-k-1}\}\,,\ \text{where}\ m_R = w_R b_R$$

LSD (for $b < 0.2$):

$$w_R = -q^2/(1-q)\ln(1-q^2)$$

Approximation (for $1.5 < w < 20$):

$$w_R \doteqdot 1.23w$$

Turning to the "new" buyers in the second period, i.e. those buyers in the period who had not bought the item in the preceding period, it follows from Table 4.4 that b_N, the proportion of the population who are "new" buyers, is the difference between b, the total incidence of buyers in the period, and b_R, the incidence of repeat-buyers, i.e.

$$b_N = b - b_R\,.$$

The formulae for b_R or b_R/b which were given in Table 4.6 therefore indirectly provide estimates of b_N*.

* Under stationary conditions, b_L, the incidence of "lapsed" buyers – buying in the first but not the second period – is numerically equal to b_N. The value of b, the total number of buyers, in the first period is (by the definition of stationarity) equal to the value of b in the second period, and the values of b_R are necessarily identical. Under stationary conditions the average purchase frequencies per repeat-buyer in the first and second periods are also equal, and so are the average purchasing rates w_N and w_L per "new" and per lapsed buyer.

Table 4.9. Three Formulae for w_N, the Average Purchase Frequency per "New" Buyer under Stationary Conditions

NBD :

$$m_N = m(1+a)^{-k-1}, \text{ where } m_N = w_N b_N$$

LSD (for $b < 0.2$) :

$$w_N = q / \ln(1+q)$$

Approximation (for $w > 2$) :

$$w_N \doteqdot 1.4$$

Finally, w_N, the average purchase frequency per "new" buyer (or w_L, the average per "lapsed" buyer) is given by the formulae in Table 4.9. For items bought at an overall average frequency of at least 2 purchases per buyer (i.e. $w \geqslant 2$), the theoretical LSD value of w_N varies only between 1.35 and 1.44 and in the numerical approximation it can therefore be treated as a quasi-constant, i.e.

$$w_N \doteqdot 1.4,$$

the simplest result of all. (For $w < 2$, the theoretical value of w_N or w_L decreases from 1.3 down to 1.0; for $b > .2$, the NBD values of w_N and w_L are of course appropriate and these can increase beyond 1.4.)

The tendency for w_N (and w_L) to be virtually "constant" at 1.4 under a wide range of conditions is less surprising than it might seem at first sight. Thus in the underlying model of stationary buyer behaviour of §4.5, "new" buyers in a given period are essentially *infrequent* buyers, and mostly buy once or perhaps twice in the period. Hence it is not surprising that on average the "new" buyers buy *about* 1.4 times. In longer period, there are fewer "new" (or "lapsed") buyers than in shorter ones, but they are still mostly once-only or twice-only buyers, and hence it remains "intuitively" acceptable that the values of w_N and w_L need not vary with the length of analysis-periods *.

For the various other aspects of repeat-buying, this virtual independence from the length of the analysis-period does however not occur.

* Under no-trend conditions, people who buy *often* in the first period would in general not stop buying the item altogether in the second period, especially given that the choice of time-periods (i.e. the particular dividing-line between them) is essentially arbitrary anyway.

Thus the number of "new" (or "lapsed") buyers, the number of repeat-buyers, and their average purchase frequency, all vary with the length of the analysis-period, as does the *total* number of buyers (the penetration *b)* and their average purchase frequency w. For example, in a longer period there are not only more buyers than in a shorter period, but a higher proportion of them will buy again in the next equal period.

Since the various repeat-buying formulae depend at most on the two basic parameters '*b* and w for *all* buyers in the analysis-period, we only need to know how these two basic parameters vary with the length of analysis-period. This then provides the input for calculating the repeat-buying estimates according to the length of the analysis-period.

4.8. Time-Periods of Varying Length

The general form of the relationship between the penetration and the length of the time-period is implicit in the repeat-buying formula for two successive periods of a given length which was discussed in the previous section (Table 4.6). Thus under stationary no-trend conditions, the number of buyers in a time-period made up of two equal shorter sub-periods is the sum of the buyers in each sub-period *minus* the number of repeat-buyers (the latter being counted amongst the buyers in each sub-period – see Table 4.4). In symbols, if b_2 is the proportion of the population buying in a period of length 2, then

$$b_2 = 2b - b_R ,$$

where b is the proportion buying in the typical "unit" period and b_R is again the proportion of repeat-buyers from one of the two unit periods to the other. The formulae in Table 4.7 then allows us to calculate the penetration in the double period, b_2, from the parameters of the unit period.

More generally, for any time-period of length T, the proportion of the population, b_T, buying in period T can be expressed in terms of the observed data (or derived parameters such as k or q) for a "unit" length time-period. The alternative NBD and LSD formulae are set out in Table 4.10, and Table 4.11 gives the earliest empirical results for some 14 different brands or pack-sizes examined in this way, where the penetration in 6 months is successfully predicted from 4-weekly data (to within about ± 1 percentage point). More recent examples are illustrated in Table 3.1 a in the previous chapter.

Table 4.10. Three Formulae for b_T, *the* Proportion of the Population Buying in a Period of Length T

(**The** parameters ***b, m, a, k or q are*** for the "unit" period under stationary conditions – see Table 4.5)

NBD:

$$b_T = 1 - (1 + m_T/k_T)^{-k_T}, \text{ by definition of the NBD,}$$

$$= 1 - (1 + Tm/k)^{-k}, \text{ since } m_T = Tm \text{ and } k_T = k.$$

LSD (for $b_T < 0.2$) :

$$\frac{b_T}{b} = 1 - \frac{\ln \{1 + (T-1)q\}}{\ln(1-q)}$$

Approximation (for $w > 1.5$ and $w_T < 20$):

$$\frac{b_T}{b} \quad \frac{Tw}{\{1 + (w-1)T^{0.82}\}}$$

Table 4.12 sets out the corresponding relationships between w, the average number of purchases per buyer in "unit" period and w_T, the average number of purchases in period T made by the larger number (b_T) of buyers then. The approximation to the LSD formula here is rather simple. Thus

$$(w_T - 1) = T^{.82}(w-1),$$

as long as $w > 1.5$ and $w_T < 20$ (and $b < .2$). The quantity $T^{.82}$ (which also occurred for b_T *in* Table 4.10) can be readily calculated using log tables, but for simplicity, its values for some commonly occuring values of T are set out in Table 4.13. A case of specific interest is $T = 2$, when

$$(w_2 - 1) = 1.76(w-1).$$

We note that the expression (w-l) here in fact represents the average number of ***repeat-purchases*** made per buyer (i.e. after discounting the first purchase); it is a way of expressing the data which numerically leads to various other simplifications in the LSD theory (see Chapter 8).

These various "time-period" formulae have already been illustrated in Chapter 3 (Tables 3.1 and 3.2a) for monitoring the growth of penetration and of buying frequency, and are also used in § 6.3 of Chapter 6 for interpretative extrapolations to longer periods.

Table 4.11. Examples of the Fit of the NBD Formula for the Percentage of Buyers in a Longer Time-Period

(The first published examples of predicting the 24-week penetration from 4-weekly data for 14 different items [Ehrenberg 1962])

	Brand or Pack size													
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*
The Given Data in 4 Weeks:														
% buying $= 100b_4$	1.4	1.7	2.7	2.9	3.8	4.4	5.1	6	7	7	10	11	12	22
Mean buying rate $= m_4$	.03	.03	.06	.06	.12	.12	.11	.11	.15	.14	.22	.28	.36	.49
% of Sample Buying in 24 Weeks $= 100b_{24}$:														
Theoretical	4	4	6	7	7	8	11	15	16	17	20	22	20	41
Observed	4	4	5	8	8	8	13	15	16	18	19	22	22	40

Table 4.12. Three Formulae for w_T, the Average Purchase Frequency per Buyer in a Period of Length T under Stationary Conditions

(The parameters m, k, w and q refer to the time-period of "unit" length)

NBD:

$$w_T = Tm/b_T$$
$$= Tm/\{1-(1+Tm/k)^{-k}\}$$

LSD (for $b < 0.2$):

$$\frac{w_T}{w} = \frac{T \ln(1-q)}{\ln(1-q) - \ln\{1+(T-1)q\}}$$

Approximation (for $w > 1.5$ and $w_T < 20$):

$$(w_T - 1) = T^{0.82}(w-1)$$

Table 4.13. Values of $T^{0.82}$ for some Commonly Occurring Values of T, to Use in the LSD Approximations for b_T and w_T

T	2	3	4	5		12
$T^{0.82}$	1.76	2.46	3.12	3.74	4.14	7.67

4.9. Conditions of Fit

The technical details of the NBD/LSD repeat-buying theory outlined in this chapter are discussed more fully in Chapters 7 and 8 (including various elaborations such as the formulae for the "conditional" kind of repeat-buying analyses used in Table 3.10 in the previous chapter). The theory tends to fit the kinds of empirical regularities that were illustrated in Chapters 2 and 3, further examples being given in Chapters 5 and 6. Six common types of discrepancy can however also be distinguished.

Firstly, a major requirement for a good fit is of course the effective absence of a trend, i.e. **the "stationarity"** of the situation. In practice, it is unusual to observe *complete* stationarity, and most of the cases studied are ones of ***near-stationarity*** (e.g. changes in the sales-level m, in the penetration b, or in the rate of buying per buyer w, of not more than ± 10%, say). Such departures from strict stationarity will introduce some discrepancies between the observed and theoretical repeat-buying patterns, and this is one of the commonest causes for at least *minor* discrepancies between the theoretical and observed values.

Secondly, there is sampling error. With a panel of 1,000 households for example, and a "penetration" for the brand in question of perhaps 5% in the analysis-period, the sample base is only 50 buyers. However, because one is usually examining internal patterns within a given sample, the *effective* sampling errors tend to be reduced (for example, the sampling errors of the average purchase frequency w in one period and the incidence of repeat-buyers in the next period tend to be positively correlated and the sampling variations tend therefore to some extent to cancel out when estimating the one quantity from the other). Repeat-buying analyses are therefore more sensitive, and somewhat less subject to extraneous "noise", than might at first be thought. (Some sampling error formulae are discussed in $6.4).

Thirdly, discrepancies will arise if there are measurement errors in the data. Some case-history examples (especially for non-panel data) are given in Chapter 6.

Fourthly, there is the minimum time-period limitation to which reference has already been made earlier. Repeat-buying in time-periods of a length close to the shortest period between successive purchases in the product-field tends to be different. The minimum inter-purchase period is not necessarily something which can as yet be altogether rigidly defined, although for many grocery-products for example it tends to be a week. It is made up of more than one component, but basically, there is a tendency for purchasing acts to manifest themselves in discrete units of time like days or weeks, rather than in continuous time (as is assumed by the Poisson formulation of § 4.5). This can be partly a matter of shopping habits – many people shop for certain types of goods only at specific times, e.g. first thing each morning for milk from the milkman, or on a certain day of the week for meat and groceries (on Tuesdays for some people, or Fridays for others, and so on), and only at still longer intervals for other products (see for example § 5.3 in Chapter 5). This imposes a certain regularity on individual purchasing habits, as well as *non-stationary* aggregate behaviour in the short-term (e.g. by day of week).

A related short-term shopping pattern is that for any sub-sample of people who buy the given brand twice in a fortnight, almost all will buy it once in the first week and once in the next, and almost no one will buy it twice in one week (and not at all in the other). This is what generally occurs, as a matter of direct observation. In other words, the independent-random-events (or Poisson) hypothesis of § 4.5 does not apply in such short time periods (one would expect only half the buyers buying once each week, with the other half buying twice either in week 1 or twice in week 2). The traditional pattern of buying cigarettes day-by-day is an extreme example of such *spreading out* of purchasing, with small, rather regular acts, instead of some bulk-buying. This leads to an abnormally low average purchase frequency per buyer (i.e. a low w) in each week, and to an abnormally *high* incidence of repeat-buying from one week to the next. The combined effect is to produce a very striking discrepancy between a high observed repeat-buying and a low theoretical estimate (from the low *w*) in short time-periods. (No theoretical estimates for l-week periods were therefore given in Table 3.6.)

Linked to, but not necessarily identical with, such short-time shop ping habits is the fact that there is often a "dead-period" between one purchase and the next, where the initial purchase has first of all to be more or less used up before another purchase is made.

These various short-term shopping and usage habits have little if anything to do with people's *longer-term* repeat-buying and brand-switching behaviour, i.e. those aspects of buyer behaviour with which brand-loyalty concepts are primarily concerned. The short-term patterns tend to- "wash-out" in longer periods *. The basic finding is of course that the same NBD/LSD patterns tend to apply for periods of *any* length, as long as they are relatively long compared with the minimum inter-purchase interval, with only the numerical values in the pattern differing (predictably, as in § 4.8). In longer time-periods, only the so-called "variance discrepancy" phenomenon tends to remain as a general discrepancy problem (i.e. a short-fall of heavy buyers who buy more than once a week – see $4.3). This is in fact also an indirect manifestation of short-term buying patterns and is further discussed in Chapters 7 and 8. Here we note that it does not usually affect the fit of repeat-buying formulae such as those in $4.7.

The fifth general type of discrepancy to be noted is that as already mentioned in § 4.3 for the negative binomial frequency distribution as such, there is also some evidence – not yet altogether clear-cut – of excess regularity for certain very frequently bought items and for some

* For example, in a longer period, people who buy in just two successive weeks make up only a small proportion of all buyers, and the fact that their purchases are spread out over both weeks instead of some being bunched in one or the other week, has little effect on the overall purchasing patterns.

total product-classes (i.e. purchasing of ***any*** brand of detergent, or of margarine, or of petrol). Such purchasing may be somewhat more regular than the NBD/LSD model would imply, especially perhaps for apparently "saturated" markets with little or no growth-potential left. (The example treated in Chapter 3 is not such a case, the fit for the product-field – i.e. "Any Brand" – there being generally as good as for the individual brands, except possibly for the frequency distribution in Table 3.4.) In the earliest work, the study of repeat-buying was confined to each separate pack-size of an individual brand, and the results of aggregating different brands (or of aggregating the different pack-sizes of a brand) have only been examined more recently (on a "purchase occasion" basis – see § 1.3 of Chapter 1). The theoretical aspects of aggregating different NBD's are discussed in Chapter 7, but more empirical study is needed.

Finally, the wrong definition of the population of consumers is potentially another general factor which can lead to discrepancies. Thus for some products it is not very clear whether the "buyer" is a household or some specific individual. (This is quite apart from questions of measurement errors and biases, including vicarious purchasing, as when person X actually buys on behalf of person Y.) Furthermore some segments of the population may not be potential buyers of the product at all (or hardly at all), such as non-motorists for petrol and moter oil, non-owners of dogs or cats for pet foods, non-smokers for cigarettes, and so on. In other product-classes, the "never-buyer" may be less obvious to identify [see also Morrison 1969, Ehrenberg 1970a]. However, one outcome of the NBD/LSD theory has been to show that the precise definition of the population at risk does not usually affect the form of the observed repeat-buying patterns or the fit of the theoretical models, as long as the proportion of non-buyers is high enough. In the extreme case of the LSD version (i.e. where the proportion of buyers b has to be less than 20% or so for it to apply), the repeat-buying results in fact depend only on w, the average purchase frequency per ***buyer,*** and not on the number of non-buyers.

Apart from these more general types of discrepancies, more specific or isolated discrepancies tend also to occur. They are relatively rare, the analyses in Chapter 3 being typical in this respect (e.g. systematic but unusual discrepancies in Table 3.8, and otherwise a rather good fit apart from some seasonal non-stationarity). Such less general discrepancies are as yet of necessity largely unexplained.

Table 4.14 gives another example of a new, and hence as yet isolated and unexplained, discrepancy. Here the observed incidence of repeat-

Table 4.14. An Unusual Case of a REDUCING Incidence of Repeat-Buyers as the Length of Analysis Period Increases

(The percentage of buyers in one period who buy again in the next period, for periods of various lengths – Observed values "O" and Theoretical NBD norms "T" for the three leading brands in a certain product field X on pilot-scale data)

Product X	Period of length (in weeks) 1		2		5		10	
	O	T	O	T	O	T	O	T
Brand A	83	73	81	80	77	85	74	88
Brand B	85	70	83	78	80	a3	77	86
Brand C	80	73	74	79	73	84	60	86
Average	83	72	79	79	77	84	70	87

buying *decreased* from an average of 83% in one-week periods to 70% in lo-week periods, whereas the theoretical values increased (as usual) from about 72% to 87% *. This is the only occasion on which such a tendency – systematic for all the brands in question – has so far been observed. The data come from a pilot-study and there is some reason to suspect measurement biases. On the other hand, the product-field is unusual, i.e. it is on the fringe of the kind of empirical conditions covered so far (see Table 4.2) in terms of frequency of buying and related characteristics. Further data would need to be collected (on a larger scale) to establish first of all whether this occurrence is repeatable, and if so, whether it is specific to the product-field or to the measurement procedure used.

This example serves to illustrate two major principles about discrepancies in the context of the present theory. Firstly, it is of course always necessary to check any new kinds of data against the theoretical model, to see if perhaps some additional factor is at play. And this is particularly important when empirical conditions are on the border-line of what has so far been covered – examples being perhaps very *frequently* bought products like bread, cigarettes or milk (as mentioned in § 4.3) or rather *infrequently* bought ones like clothing (as discussed in § 5.3 in the next chapter).

* The first thought with any discrepancy must always be of a clerical or computing error, e.g. that the observed data in this case were recorded the wrong way round with the 1- and 10-week results and the 2- and 5-week ones inter-changed (the fit would be good!). But checks showed that this was ***not*** the case here.

Secondly, the important thing with a discrepancy is not that the observed data differ from the theoretical model as such (let alone that the theory itself could be "disproved" by one such happening), but that the new data differs from the patterns observed in all the other cases studied so far where the model does give a good fit. Thus the conclusion about the buyer behaviour for the product in Table 4.14 is that it is different from the kind of behaviour generally observed in other product-fields.

4.10. Summary: The Nature of the NBD/LSD Theory

To summarise the results outlined in this chapter, there is a *single* theory – the NBD model, together with the simplifying LSD approximations – which applies to a very wide range of circumstances, i.e. different kinds of product-fields and brands, different lengths of time-period, and so on.

Table *4.15.* Numerical Values of the LSD and NBD Formulae for the Percentage of Buyers of an Item in One Period Who Buy it Again in the Next Equal Period

(The values for $100b_R/b$ as given by Table 4.6, for various values of b and w)*

$100b_R/b$	Average Purchase Frequency per Buyer, w							
	1.1	1.3	1.5	2	3	5	10	20
NBD, for Proportions of Buyers, b =								
.8					85	89	92	93
.6				70	78	84	88	90
.4			52	64	74	80	85	88
.2**		37	47	60	70	77	83	86
.1		35	4.5	58	69	76	82	85
.01	16	34	44	57	68	75	81	85
LSD: Exact	16	34	44	57	68	75	81	85
Approximate	13	30	41	56	68	76	82	84

* Data for which $w < -\ln(1-b)/b$ cannot be fitted by an NBD.
** The percentage of repeat-buyers varies little for values of b less than .2.

The theory therefore is essentially a simple one, and the required input is also very simple – two observed quantities, primarily w, the average frequency of purchase of the item per buyer in some given period, and secondly b, the proportion of the population buying the item in that period. The way in which w is the dominant factor is illustrated in Table 4.15 in terms of the percentage of repeat-buyers from one period to the next, as given by the NBD and LSD formulae in Table 4.6 for various values of *b* and w.

One conclusion is that the observed patterns of repeat-buying loyalty are generally not intrinsic to the individual brand or product-field – it is not a case of "My brand (or my product) is different". Instead, repeat-buying patterns turn largely on the average purchase frequency of the item. Any two items with the same average purchase frequency will have the same repeat-purchasing patterns. Furthermore, as far as any two brands in a product-field tend to have similar w's, they will have repeat-buying characteristics which are similar not only in their general structure but also numerically *.

The underlying NBD/LSD model, and the empirically verified deductions from it, imply that there is generally no erosion of repeat-buying over time. Thus if there is no general trend in sales, the incidence of repeat-buyers in non-successive time-periods is virtually no lower than it is in successive periods. People who buy in one period but not in the other – so-called "lapsed" and "new" buyers – are not leaving the market altogether or coming in for the first time, but are simply *infrequent* (not necessarily *irregular*) buyers. The theory turns on the basic notion that different consumers have different long-run frequencies of purchase, and that this differing propensity-to-purchase manifests itself over time in a more or less random (i.e. stochastically regular) manner for each consumer.

Exceptions to this formulation can occur when there is a real turnover of buyers (the "leaky-bucket" theory) or when people have a special short- or medium-term enthusiasm for an item (the "jag buying" theory), but these *are* exceptions. (The "jag" type of exception seems for example to occur for the product-field analysed in Table 3.8 of Chapter 3, but – and this is the point – *not* in most other product-fields analysed, as is illustrated by Table 2.1 in Chapter 2 and Table 5.1 in Chapter 5.)

The simplicity of the repeat-buying model arises largely because the question of whether or not buyers of the item in question also buy any *other* item does not enter into any of the equations. This is essentially a matter of empirical fact – a good fit can be obtained by the NBD/LSD

* In contrast, the penetration *b* of different brands varies widely (and largely determines their different sales levels), but does not greatly affect their repeat-buying characteristics. The relation between *b* and *w* is discussed further in §§ 10.2, 11.4 and 11.5.

model as such, without making allowance for other brands. The reason behind all this is not that there is no multi-brand buying, as the reverse is generally true, but lies in the nature of multi-brand buying, as is discussed in Chapter 11.

All this is not to say that the NBD model is fundamentally "true". As has been pointed out already, the model breaks down at the boundaries, e.g. for very short time-periods and for very heavy buyers, and probably for very frequently-bought items and possibly for very infrequently bought ones. In particular, neither the Poisson- nor the Gamma-distribution assumption of the underlying stochastic model set out in § 4.5 can be altogether right. Some quite fundamental reformulation is required, as is discussed a little further in Chapter 11 and has in fact occurred in terms of the Dirichlet Model in Chapter 13, since the first edition of this book. But this must essentially give the same results and largely the same insights as the NBD theory in the vast majority of situations covered so far which the theory "models" rather successfully. The justification of the theory is therefore not the absolute truth of the theory in itself but that it works in practice and helps us to know and understand a good deal about empirical buyer behaviour and its essential simplicities.