

Chapter 15: Stochastic Choice

Prerequisites: Chapter 5, Sections 3.9, 3.10

15.1 Key Terminology

The topic of this chapter is a set of choice models that deal with consumer behavior over time. We will begin by looking at data that tabulates what consumers do on two sequential purchase occasions. Do they buy the same brand twice, or do they switch from one brand to another? Later in the chapter we will look at the number of times a particular brand has been purchased, a type of data often called *purchase-incidence* data.

In some cases, we will assume that the population being studied is homogeneous. This is tantamount to the Gauss-Markov assumption [presented in Equation (5.16)] that we typically make in the general linear model, that is, that each observation can be described by the same parameter. In other cases, we may assume that the population being studied is heterogeneous with that parameter taking on different values. The parameter may itself follow some sort of distribution, often called a *mixing distribution*.

There is a different sort of homogeneity-heterogeneity distinction that comes up in models dealing with data collected over time. Regardless as to whether each unit, browser, consumer or household in the population can be described by the same parameter, is it possible that the parameter can change over time? A parameter that remains invariant across time periods is generally referred to as being *stationary* rather than homogeneous. More formally, we would define stationarity for a parameter θ such that

$$\theta_t = \theta_{t'} = \theta \text{ for all } t, t' = 1, 2, \dots, T. \tag{15.1}$$

That terminology out of the way, let us now turn to the brand switching matrix which contains the key raw data for the models of the next few sections.

15.2 The Brand Switching Matrix

In what follows we will assume that we have three brands; call them A, B and C. Of course this terminology should not obscure the generality of the type of data we will be discussing. The three brands might actually be three Web sites, for example. In any case, in this section for each household we will be looking at a series of observations across T time periods: $y_1, y_2, \dots, y_t, \dots, y_T$. We might admit here that the y_t values should also have a subscript for household, but that is dropped for notational convenience. You can think of the value y_t as being randomly selected from some population of households. For now we will look at $T = 2$ purchase occasions and organize the data from these two occasions in a two way contingency table that might look a lot like the one below:

		Purchase Occasion Two				
		A	B	C		
Purchase Occasion One	A	10	5	10	25	
	B	8	12	5	25	
	C	10	10	30	50	

The table tells us that, for example, 10 households bought brand A on week one and then bought it again on week 2. On the other hand, of the 25 households who bought brand A on week one, 5 of

them switched to brand B on the second purchase occasion. It will be useful to be clear on different sorts of probabilities that can be formed from raw data such as these. An example of a *joint probability* would be the probability that a household in the sample bought A on week (occasion) one *and* then B on week 2, in other words $\Pr(y_1 = A \text{ and } y_2 = B)$. We can also write this as $\Pr(A, B)$. Making the notation a bit more general, let us define $\Pr(j, k)$ as the joint probability that brand j is chosen on the first occasion and k on the second. From the table we can see that $\Pr(A, B) = 5/100$ since 5 families from the sample of 100 families did just that.

A *marginal probability* gives the summary of a row or a column. For example, what is the probability of buying brand A on week one? The answer is $25/100$, as 25 out of 100 families did that, and that figure also happens to be the market share for brand A on week one. As such we might use the letter m and notate that value $m_A^{(1)}$. Alternatively we could also use an expression like $\Pr(A)$, where it is understood we are talking about week one.

Finally, a *conditional probability* involves subsetting the table in some way. A conditional probability looks at the odds of something happening within that subset of the table. We might ask, given that a family bought A on week one, what is the conditional probability that they would turn around and buy B on week two? In other words, what is $\Pr(y_2 = B \mid y_1 = A)$? A vertical bar is traditionally used to indicate a conditional probability. Here the numerator differs from the joint probability. You can think of this as the probability of B conditional on A, or *given* A. In either case, $\Pr(B \mid A) = 5/25$, as there are 25 families who bought brand A on week one, and of these, 5 bought B on the next occasion. Again we could make the notation a bit more general by referring to $\Pr(k \mid j)$, or p_{jk} , as the conditional probability that brand k is chosen on the next occasion given that j was chosen on the previous occasion. While the notation p_{jk} will be used to refer to $\Pr(k \mid j)$, this probability is actually in position j, k of the *transition matrix*, illustrated below.

We might note that

$$m_j^{(1)} = \sum_k \Pr(j, k), \quad (15.2)$$

$$\Pr(k \mid j) = \frac{\Pr(j, k)}{\Pr(j)} \text{ and that} \quad (15.3)$$

$$\sum_k \Pr(k \mid j) = 1. \quad (15.4)$$

In all three cases above the summation over the index k is taken to mean over all J brands in the study that appear in the switching matrix. Here the value $m_j^{(1)}$ is the share for brand j on week 1.

15.3 The Zero-Order Homogeneous Bernoulli Model

In this section we will once again be looking exactly two purchase occasions, i. e. $T = 2$. We begin by contemplating exactly two brands, A and B, and we will look at this situation with a particularly simple model. The zero-order homogeneous Bernoulli model assumes that on any purchase occasion the probability that A is bought is p . Here are the joint probabilities:

		Occasion Two	
		A	B
Occasion One	A	p^2	$p(1-p)$
	B	$(1-p)p$	$(1-p)^2$

For example, looking at the joint probability $\Pr(A, A)$, according to the model we have two independent events, each one occurring with a probability of p . The probability of two independent events can be calculated by multiplication. That the two events are independent is one of the strongest assumptions of the model. In effect, it assumes no feedback from one purchase event to the next. In other words, this model is zero-order just like a series of coin flips. Recall that with a fair coin, regardless of how many heads in a row have come up, the probability of a head on the next toss is still exactly .5.

The joint probability of any string of purchases can be calculated from multiplication as in

$$\Pr(A, B, A, A, B, \dots) = p \cdot (1-p) \cdot p \cdot p(1-p) \cdot \dots$$

The probability of r purchases of A out of T occasions would be

$$\binom{T}{r} p^r (1-p)^{T-r} \tag{15.5}$$

where the notation $\binom{T}{r}$ refers to the number of combinations of T things taken r at a time and is given by

$$\binom{T}{r} = \frac{T!}{r!(T-r)!}$$

and $T! = T \cdot (T-1) \cdot (T-2) \cdots 1$. The conditional probabilities can also be displayed in the same occasion-by-occasion format. When displayed as below, the table is called a transition matrix.

		Occasion Two	
		A	B
Occasion	A	p	$(1-p)$
One	B	p	$(1-p)$

The elements of the transition matrix, for example $\Pr(k | j)$, the probability that k is chosen given that j was chosen previously, are notated p_{jk} since that conditional probability arises from row j and column k .

15.4 Population Heterogeneity and The Zero-Order Bernoulli Model

Lets say that the value of p is itself a random variable, rather than a fixed parameter that describes the population of households, but there is still no feedback from one occasion to the next. On the surface it seems that this should imply, just as in a series of coin flips, that the next flip should not depend on what happens in any previous flips, right? It turns out the population heterogeneity and the lack of stationarity over time have similar implications in switching data. To get a handle on

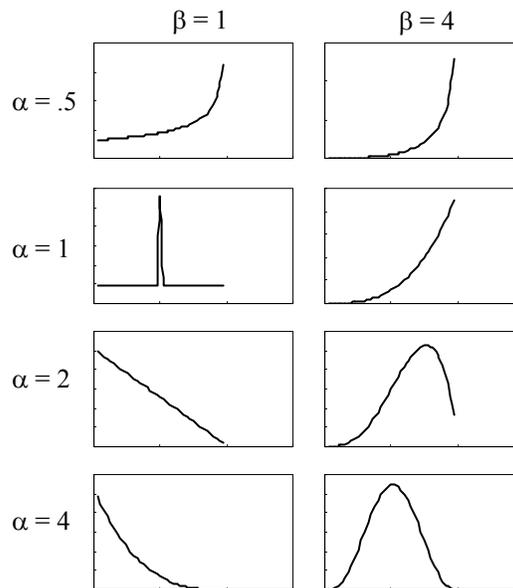
the nature of the heterogeneity of the value of p , we typically use the *Beta distribution* (Lilien and Kotler 1983), where

$$\Pr(p) = c_1 p^{\alpha-1} (1-p)^{\beta-1} \quad (15.6)$$

The constant c_1 is a place holder that needs to be there to make sure that the distribution integrates to 1, i. e. it must be the case that

$$\int_{-\infty}^{\infty} \Pr(p) dp = 1$$

because $\Pr(p)$ is a density function (see Section 4.2). The two parameters of this distribution, α and β , control the shape of it. As compared to the normal, a wide variety of shapes are possible! Some idealized examples are pictured below in a graph that shows the $\Pr(p)$ on each of the y-axes:



Given some value of p , the likelihood of r purchases out of T occasions (Lilien and Kotler 1983) is

$$\Pr(r, T | p) = c_2 p^r (1-p)^{T-r}. \quad (15.7)$$

The constant c_2 is a place holder for $\binom{T}{r}$, which does not figure into the derivation that follows.

At this time it is appropriate to invoke the name of the Reverend Thomas Bayes, given that his name is attached to a simple theorem that connects two different sorts of conditional probabilities. For any two events, a and b , we know that by definition

$$\Pr(a | b) = \frac{\Pr(a, b)}{\Pr(b)}$$

but also that

$$\Pr(b | a) = \frac{\Pr(a, b)}{\Pr(a)}.$$

This suggests that there are two ways to write $\Pr(a, b)$,

$$\Pr(a, b) = \Pr(a | b) \cdot \Pr(b) = \Pr(b | a) \cdot \Pr(a),$$

which, when set equal to each other yields

$$\Pr(a | b) \cdot \Pr(b) = \Pr(b | a) \cdot \Pr(a)$$

$$\Pr(a | b) = \frac{\Pr(b | a) \cdot \Pr(a)}{\Pr(b)}.$$

From his theorem we can deduce that

$$\Pr(p | r, T) = \frac{\Pr(r, T | p) \Pr(p)}{\Pr(r, T)}. \quad (15.8)$$

In the numerator of the right hand side we see the likelihood of the data given the model from Equation (15.7), i.e. $\Pr(r, T | p)$. The density for p , assumed to be beta distributed, is also in the numerator. This is usually called the *prior distribution*, or sometimes just the *priors*. The left hand side also has a name, the *posterior probability*. It is the posterior probability of choice on the next occasion given a history of r purchases out of T occasions. If we define c_3 as $1 / \Pr(r, n)$, then the posterior probability can be rewritten as

$$\begin{aligned} \Pr(p | r, T) &= c_3 \cdot \Pr(r, T | p) \cdot \Pr(p) \\ &= c_4 \cdot p^r (1-p)^{T-r} \cdot p^{\alpha-1} (1-p)^{\beta-1} \end{aligned} \quad (15.9)$$

which means that the posterior probability looks like a beta distribution with parameters $\alpha^* = \alpha + r$ and $\beta^* = \beta + T - r$. The upshot is that even though there is no memory or purchase feedback in this model, the posterior probability makes it look like there is. But the reason for this is that the population is not homogeneous. If we collect up all the households for which no one bought A, we probably have a group for whom p is lower than average. Dividing the sample of households in this way makes it look like there is contagion - a bunch of B's in a row lead to a higher probability of another B, not another flip of the coin.

We can estimate the choice parameter, p , using (Lilien and Kotler 1983)

$$\hat{p} = E(p | r, T) = \frac{\alpha + r}{\alpha + \beta + T}. \quad (15.10)$$

For example, for $T = 3$ we could look at eight possible triples that could occur with two brands and three weeks; AAA, AAB, ABA, ABB, BAA, BAB, BBA, BBB. The value of r is 0 for triple BBB. According to the model, the prediction for all those with three purchases of Brand B in a row would be

$$\hat{p} = E(p | 0, 3) = \frac{\alpha}{\alpha + \beta + 3}.$$

For $r = 1$ and $T = 3$ we could have ABB, BAB and BBA. All three sequences lead to the same estimate on trial 4,

$$\hat{p} = E(p | 1, 3) = \frac{\alpha + 1}{\alpha + \beta + 3}$$

As you can see, we can derive values for the choice probabilities, that is, values of p , on week 4. These probabilities arise from the more fundamental parameters that underlie the distribution of p , namely α and β , which are the unknowns and as such must be estimated from the sample. We could certainly minimize Pearson Chi Square across the eight data points from the triples. According to Minimum Pearson Chi Square, we pick values of α and β in such a way as to make

$$\hat{\chi}^2 = \sum_j \frac{(p_j - \hat{p}_j)^2}{\hat{p}_j} \quad (15.11)$$

as small as possible. We could also use modified minimum Chi Square or Maximum Likelihood. To do any of these we would need to determine the derivatives of the objective function and drive them to zero,

$$\frac{\partial \hat{\chi}^2}{\partial \alpha} = \frac{\partial \hat{\chi}^2}{\partial \beta} = 0,$$

using the methods described in Section 3.9. As there are eight triplets from three weeks worth of purchases, and two unknowns, the model can be tested against Chi Square on 6 degrees of freedom.

15.5 Markov Chains

Now we will look at models that assume homogeneity across consumers or households, but not zero memory. In fact, a defining aspect of a Markov chain is that the system has memory that goes back one time period. If we define y_t as the brand chosen on occasion t , this memory can be described as

$$\Pr(y_t = j | y_{t-1}, y_{t-2}, \dots, y_0) = \Pr(y_t = j | y_{t-1}). \quad (15.12)$$

We also assume stationarity which can be interpreted as the statement below:

$$\Pr(y_t = j | y_{t-1}) = \Pr(y_{t'} = j | y_{t'-1})$$

for all t, t' and j .

A Markov chain is characterized by a transition matrix and an initial state vector. The transition matrix consists of the conditional probabilities $\Pr(k | j)$ such that $\sum_k \Pr(k | j) = 1$. A sample transition matrix is presented below:

		Occasion t + 1	
		A	B
Occasion t	A	.7	.3
	B	.5	.5

For example, in the lower left hand corner we see $\Pr(A | B)$ which is element 2,1 (p_{21}) in the table and is equal to .5. The second characterizing feature of a Markov chain is the initial vector which represents the market shares at time zero. A typical element would be $\{m_j^{(0)}\}$ which is the market share for brand j at time 0. In that case we can define the J by 1 vector of shares as

$$\mathbf{m}^{(0)} = [m_1^{(0)} \quad m_2^{(0)} \quad \cdots \quad m_J^{(0)}]$$

Given a transition matrix and an initial state, we can now predict the market shares for any time period. For example, looking at brand k , we might ask what will the share of brand k be after one week. To do this, we can use the *Law of Total Probability*. After time 0 there are J things that could have happened, that is to say there are J ways for k to be picked at time 1. A purchaser of brand 1 could have switched to k , a purchaser of brand 2 could have switched to k , and so forth until we reach the last brand, brand J . This is illustrated below:

$$m_k^{(1)} = \underbrace{\Pr(k | 1)}_{\substack{\uparrow \\ \text{Pr (buy k given} \\ \text{previous purchase} \\ \text{of brand 1)}}} \cdot m_1^{(0)} + \Pr(k | 2) \cdot m_2^{(0)} + \cdots + \Pr(k | J) \cdot m_J^{(0)}$$

\uparrow Pr (bought 1 previously)

We can use a slightly more elegant notation to say the same thing as

$$m_k^{(1)} = \sum_j^J p_{jk} m_j^{(0)}$$

Here note that the law of total probability has us running down the rows of the \mathbf{P} matrix, that is, running through all the ways that event k can happen at time $t + 1$. We can also express all of the market shares at one time using linear algebra,

$$\begin{aligned} [\mathbf{m}^{(1)}]' &= [\mathbf{m}^{(0)}]' \mathbf{P} \\ [\mathbf{m}^{(2)}]' &= [\mathbf{m}^{(1)}]' \mathbf{P} = [\mathbf{m}^{(0)}]' \mathbf{P} \mathbf{P} \\ [\mathbf{m}^{(3)}]' &= [\mathbf{m}^{(2)}]' \mathbf{P} = [\mathbf{m}^{(1)}]' \mathbf{P} = [\mathbf{m}^{(0)}]' \mathbf{P} \mathbf{P} \mathbf{P} \\ \dots &= \dots = \dots = \dots \\ [\mathbf{m}^{(t)}]' &= [\mathbf{m}^{(0)}]' \mathbf{P}^t \end{aligned} \tag{15.13}$$

We frequently assume an equilibrium such that the share vector no longer changes and estimate the elements of \mathbf{P} from panel data. These elements themselves may be modeled with a smaller number of parameters that reflect the fundamental marketing concepts that are driving the data.

Recall that in the zero-order homogeneous Bernoulli model the transition matrix took on the appearance:

$$\begin{bmatrix} p & 1-p \\ p & 1-p \end{bmatrix}.$$

Here remember that the rows represent the state of the market at time t while the columns are the states at time $t + 1$. Element p_{jk} is the conditional probability, $\Pr(k | j)$.

Something we might call the *Superior-Inferior model* has a transition matrix

$$\begin{bmatrix} 1 & 0 \\ p & 1-p \end{bmatrix}.$$

No one who ever tries the first brand goes back to the second. One of the two states is an absorbing state - eventually the whole market will end up there.

In the *Variety-Seeking model* the propensity to buy a brand again is reduced by some fraction v :

$$\begin{bmatrix} p-vp & - \\ - & (1-p)-v(1-p) \end{bmatrix}$$

You will note that since $\sum_k^J p_{jk} = 1$, we can figure out one column by subtraction. Also note what happens as v goes from 0 to 1. The closer v gets to 0, the closer the model resembles the Bernoulli.

How would we estimate the parameters v and p ? We could look at the 8 triples that are possible, AAA, AAB, ..., BBB. Each one has a prediction from the model. For example, for AAA we would have

$$\Pr(\text{AAA}) = (p - vp)^3.$$

We would have 8 data points, and two unknowns, and we could just use Minimum Pearson Chi Square, Maximum Likelihood, or other methods as described in Section 12.4 as well as for the logit model in Sections 13.3 and 13.4.

15.6 Learning Models

The models of this section are part of *Linear Operator Theory*, which was originally applied in Marketing to learning about frozen orange juice by Kuehn. Here we are going to assume that we have but one brand of interest, that is either purchased or not:

$$y_t = \begin{cases} 1 & \text{if the brand of interest is bought at time } t \\ 0 & \text{if the brand of interest is not bought at time } t \end{cases}$$

If our brand is purchased at time $t - 1$, we apply the acceptance operator and presumably learning occurs:

$$p_t = \alpha_1 + \beta_1 p_{t-1}$$

while on the other hand, if the brand is rejected, we apply the rejection operator:

$$p_t = \alpha_2 + \beta_2 p_{t-1}.$$

Consider what happens when a loyal consumer repeatedly buys our brand,

$$p_1 = \alpha_1 + \beta_1 p_0$$

$$p_2 = \alpha_1 + \beta_1 [\alpha_1 + \beta_1 p_0]$$

$$\dots = \dots$$

or working recursively backwards from time t

$$p_t = \alpha_1 + \beta_1 p_{t-1}$$

$$p_t = \alpha_1 + \beta_1 [\alpha_1 + \beta_1 p_{t-2}]$$

$$p_t = \alpha_1 + \beta_1 [\alpha_1 + \beta_1 (\alpha_1 + \beta_1 p_{t-3})]$$

Now, multiplying out this last version we have

$$p_t = \alpha_1 + \beta_1 \alpha_1 + \beta_1^2 \alpha_1 + \beta_1^3 p_{t-3}.$$

Eventually, we note that a pattern emerges so that we have

$$\begin{aligned} p_t &= \alpha_1 + \beta_1 \alpha_1 + \beta_1^2 \alpha_1 + \beta_1^3 \alpha_1 + \beta_1^4 \alpha_1 + \dots \\ &= \alpha_1 [1 + \beta_1 + \beta_1^2 + \beta_1^3 + \dots]. \end{aligned} \tag{15.14}$$

The term in the brackets is an infinite series, but that does not mean that it is equal to infinity. Call it b

$$b = 1 + \beta_1 + \beta_1^2 + \beta_1^3 + \dots \tag{15.15}$$

$$= \sum_{i=0}^{\infty} \beta_1^i.$$

For Equation (15.15) to work requires that $0 \leq \beta_1 < 1$. If we multiply Equation (15.15) by β_1 we get

$$\beta_1 b = \beta_1 + \beta_1^2 + \beta_1^3 + \dots \tag{15.16}$$

Subtract Equation (15.16) from Equation (15.15) above and the difference is 1:

$$b - \beta_1 \cdot b = 1$$

$$b(1 - \beta_1) = 1$$

$$b = \frac{1}{1 - \beta_1}. \quad (15.17)$$

Combining Equation (15.14) and Equation (15.17), we can conclude that if the brand is always purchased, the probability will approach

$$p_t = \frac{\alpha_1}{1 - \beta_1}, \quad (15.18)$$

a phenomenon known as incomplete habit formation. In this Linear Operator Theory, if $\beta_1 = \beta_2 = 0$ then we end up with a transition matrix just like the one shown below:

$$\begin{bmatrix} \alpha_1 & 1 - \alpha_1 \\ \alpha_2 & 1 - \alpha_2 \end{bmatrix}$$

which is a zero-order Bernoulli model!

15.7 Purchase Incidence

The main exemplar of a purchase-incidence model uses the Negative Binomial Distribution or NBD. In order to lead into that, however, we will start with two simpler models, the first of which is the binomial, named after the terms in the expansion of

$$(q + p)^T$$

with $q = 1 - p$. Term number $r + 1$ is $q^{T-r} p^r$ which we have already seen used in the expression for the probability of T things taken r at a time in Equations (15.5) and (15.7):

$$\binom{T}{r} p^r (1-p)^{T-r}$$

The term $\binom{T}{r}$ gives the number of ways out T to have r "successes" while $p^r (1-p)^{T-r}$ is the probability of each one of those ways. Now, consider a table from a panel of n households, with each household being categorized in terms of how many purchases of our brand that they have executed during the T week study period:

r	Number of Households
0	f_0
1	f_1

$$\begin{array}{r} 2 \\ \dots \\ \hline \text{Total} \end{array} \qquad \begin{array}{r} f_2 \\ \dots \\ f_T \\ \hline n \end{array}$$

with a typical entry being f_r , which gives the number of households with r purchases during the study period. These are the data that we will attempt to account for with the model. The binomial model states simply that the probability of a purchase by any household on any week is p . We can estimate p using a particularly simple method called the *method of moments*. It is the case that

$$\bar{x} = E(r) = pT \tag{15.19}$$

gives the average number of purchases across households, or in other words, the average number of purchases per household. Solving for p we have simply

$$p = \frac{\bar{x}}{T}.$$

For example, if the average household purchase 2 items out of 4 occasions, then $p = 2/4 = .5$. According to the binomial model, we could substitute .5 for p in the formula

$$\hat{f}_r = T \cdot \binom{T}{r} (1-p)^{T-r} p^r. \tag{15.20}$$

We could test the model with a Chi Square that compares the predicted and observed frequencies of households with 1, 2, ..., T purchases.

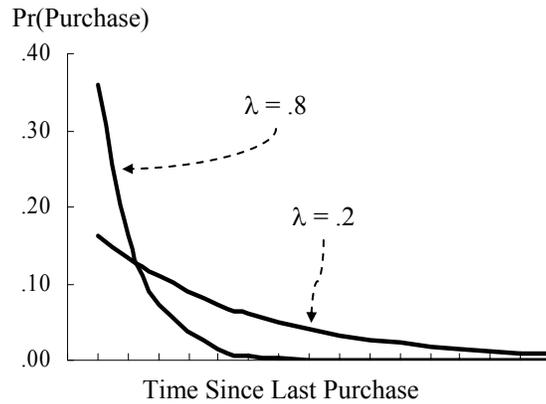
The Poisson model arises from the Binomial by letting $T \rightarrow \infty$ and $p \rightarrow 0$ but holding $Tp = \lambda$. The model originated from studies of the deaths of Prussian soldiers from kicks to the head by horses, apparently a worrisome occupational hazard. The number of Army corps with one death, two deaths, and so forth, was tabulated. The Poisson model asserts that

$$\hat{f}_r = n \cdot e^{-\lambda} \frac{\lambda^r}{r!}. \tag{15.21}$$

Fortunately for Prussian soldiers, the Poisson, which means fish in French but is actually named after it's inventor, is considered a distribution for "rare" events. The model assumes that there are a large number of small time periods with a small, but constant purchase probability in any time period. This is no doubt more realistic than the Binomial model, but unfortunately the Poisson makes an odd prediction about the probability that t time periods pass between one purchase occasion and the next

$$\text{Pr}(t) = \lambda e^{-\lambda t} \tag{15.22}$$

which is a special case of the *exponential distribution*. That this assumption is not in keeping with the reality of shopping can be seen in the graph below that looks at the relationship between time elapsed and the probability of a purchase:



The Poisson has the interesting property that its mean is equal to its variance,

$$E(r) = \bar{x} = \lambda$$

$$V(r) = s^2 = \lambda.$$

We could easily use the Method of Moments to estimate λ , and of course we also have at our disposal Minimum χ^2 , which would require that we compare the f_r and \hat{f}_r values, Maximum Likelihood, and so forth.

15.8 The Negative Binomial Distribution Model

The NBD model is named from the terms in the expansion of $(q - p)^{-r}$. The distribution can arise in a number of ways. For example, it could represent the probability that T trials will be needed for r successes. In effect, it is a binomial where the number of coin tosses is itself the random outcome. It could also represent a Poisson distribution with a contagion process such that the Poisson parameter λ changes over time. Another possible mechanism that leads to the NBD is where we have a Poisson model but the λ values is distributed across households according to the gamma distribution. The gamma is part of the general family of distributions that includes the Chi Square as a special case. According to the NBD model, the number of households purchasing the brand under study is

$$\hat{f}_r = n \cdot \left(\frac{k}{k+m} \right)^k \left(\frac{m}{k+m} \right)^m \frac{\Gamma(k+r)}{\Gamma(k)r!}. \quad (15.23)$$

The gamma function, $\Gamma(\cdot)$, not to be confused with the gamma distribution, acts like a factorial operator (the ! symbol) for non-integral arguments. For integral q , $\Gamma(q) = (q - 1)!$. In general,

$$\Gamma(q) = \int_0^{\infty} x^{q-1} e^{-x} dx. \quad (15.24)$$

Here we might note certain similarities between the Binomial model in Equation (15.20) and the Negative Binomial in Equation (15.23). In the latter, the role of p is played by $k/(k + m)$ while $1 -$

p is analogous to $m/(k + m)$. As before, we will be estimating k and m according to the method of moments, or using ML or Minimum Chi Square.

Here we might note certain similarities between the Binomial model in Equation (15.20) and the Negative Binomial in Equation (15.23). In the latter, the role of p is played by $k/(k + m)$ while $1 - p$ is analogous to $m/(k + m)$. As before, we will be estimating k and m according to the method of moments, or using ML or Minimum Chi Square.

References

Kahn, Barbara E., Manohar U. Kalwani and Donald G. Morrison (1988) "Nicheing Versus Change-of-Pace Brands: Purchase Frequencies and Penetration Rates to Infer Brand Positionings," *Journal of Marketing Research*, 25 (November), 384-90.

Kahn, Barbara E., Manohar U. Kalwani and Donald G. Morrison (1984) "Measuring Variety-Seeking and Reinforcement Behaviors Using Panel Data," *Journal of Marketing Research* 23 (May), 89-100.

Lilien, Gary L. and Philip Kotler (1983) *Marketing Decision Making*. New York: Harper and Row.

Schmittlein, David C., Albert C. Bemmaor and Donald G. Morrison (1985) "Why Does the NBD Model Work? Robustness in Representing Product Purchases and Imperfectly Recorded Purchases," *Marketing Science* 4 (Summer), 255-56.

Wheat, Rita D. and Donald G. Morrison (1990) "Assessing Purchase Timing Models: Whether or Not is Preferable to When," *Marketing Science*, 9 (Spring), 162-70.