Modelling buyer behaviour - 2
Rate-frequency models
By Chuck Chakrapani

As discussed in the previous article, sometimes it is useful to understand the underlying patterns of buyer behaviour that are not apparent from the survey data. Understanding such patterns has several practical applications, which will become evident as we continue our discussion of buyer behaviour models.

Predictable patterns of buyer behaviour
Let us begin with a research study that indicates that, on average, a household buys 4 litres of milk a week. We would expect that some households buy more than this amount and others less. Survey cross-tabs would also tell us how the buying frequencies are distributed. This distribution will vary form survey to survey due to sampling and other related and unrelated factors.

If we assume that there is an underlying regularity in the way consumers behave, our objective now becomes identifying this pattern, given the result of the survey data.

Basic assumptions
We begin with a few assumptions.

• Each household has a typical (or characteristic) rate of purchase;

• This rate may vary from household to household;

• Given the same rate of purchase, different households will exhibit different patterns of purchase. (Two households may buy an item at a mean rate of 2 per week. Yet in any given week, either of these households can buy any number of items).

The gamma distribution
When we make these assumptions and observe buying frequency, behaviour appears to follow a well-known mathematical curve known as the **Gamma distribution** (see graph). This distribution has several important properties as noted below.

• The distribution starts with zero and continues to infinity. In other words, while a household can never buy less than 0 number of times, in theory it can buy any number of times, in a given period.

• As the rate of buying increases, the number of consumers buying decreases.

• The majority of consumers have a low frequency of buying. Very few have a high purchase frequency rate. The pattern shows a peak on the left and a very long tail on the right.

In using the Gamma distribution, we assume that the behaviour represented by the model is not 'step-wise'. The model would also assume that at least 90% of consumers exhibit the behaviour under consideration.

When more than 10% of consumers *never* buy the product under question, one needs to exclude 'never' buyers from the modelling process. The Gamma distribution models the buying rather than the non-buying behaviour.

What does the distribution model? **The distribution actually models the underlying rates of purchase that give rise to the distribution observed in survey research or panel data.**

The Poisson distribution

We still need to address the assumption that different households have different underlying (characteristic) rates of purchase. Any household at any given time can deviate from its characteristic rate of purchase. Usually this happens for some identifiable reason for any given household. (For example, a household that eats out twice a month may have done so five times during the month under observation, because guests were visiting from abroad during that period). Yet such knowledge is not readily available to the researcher. Neither can the researcher do much with such knowledge, even if it were available to him/her.

Consequently, no matter how logical the deviations are for the household under question, we can treat these deviations as random fluctuations in our model. This random or stochastic component itself can be incorporated in the modelling process.

For instance, a household which has an underlying or characteristic purchase frequency of 1 purchase per month can have any of the following frequency during a given month: 0, 1, 2, 3,... this frequency can be approximated by another distribution known as the **Poisson distribution.**

To identify the value of a Poisson distribution all we need is the average purchase frequency of a household. Suppose a household eats out 2.3 times per month. What are the chances that it will not eat out at all in a given month? What are the chances that it will eat our more than 5 times? All this can be modelled using the Poisson distribution. (See Table 1). Such distributions can be verified through panel data or other records that may be available. (More on how to calculate the actual figures later).

Table 1
The probability of a household eating out given that the household eats out 2.3 times per month

# times	*%HHs*
0	9.9
1	22.8
2	27.0
3	19.9
4	12.0
5	5.4
6	2.1
7	0.7
8+	0.2

We can combine the Gamma and the Poisson models. This will facilitate our understanding of both rates or purchase and frequencies of purchase.

Understanding the implications of the models

It is important to understand that the models described above assume no causal relationships. There is no *reason* why a family with an average frequency of 2.3 should refrain from eating out 10% of the time. Neither can we claim predictive accuracy for any given household. A given household with a frequency of 2.3 can refrain from eating out 20% of the time and another household with the same frequency may do so only 5% of the time. But the models appear to work at the aggregate level when we apply them to actual empirical data. The observed regularity at the aggregate level is the justification for using them.

Using the model

At this stage we may want to know the practical application of a model like this. In many cases, we know only the average rate of purchase but not the underlying distribution that gives rise to the observed rate of purchase. Knowing the distribution may enable the manufacturer to market the product more efficiently. It may also point out any out-of-the-ordinary purchase patterns.

Assumptions of the model

Because the model assumes that there is a *relatively constant rate of purchase*, it follows that buying a product should not change the consumer's probability of buying it again. This is likely to be the case for new products, the buying of which may positively or negatively influence repurchase. One can think of several instances where this might be the case. For instance, when the product is introduced into the market for the first time or when an advertising or promotional campaign is undertaken (to attract new buyers) we may expect to find more instability. This will influence purchase-repurchase probabilities. One way to handle this situation might be to exclude first time buyers. We can then calculate the rates for those who have bought the product at least once before.
The second assumption is that the product usage should not be cyclical. For example, the average monthly rate of a product that is purchased mostly during Christmas or Easter is unlikely to produce a distribution that corresponds to the buyer's actual buying pattern.
A related observation is that the product should not be one that is bought at fixed intervals. Consider a car that is mostly used for driving to work. In this case, gasoline will be bought at regular intervals.
In short, for the model to work, there must be random variations around the characteristic rate for each household. Whenever this is not the case (eg. products bought at specified intervals or cyclical products) the model will not work satisfactorily. Once these conditions are satisfied, we can use the Gamma-Poisson model to estimate the frequency of purchase.

Calculating the parameters needed

The parameters of the model can be calculated fairly easily using any standard statistical text that deals with distributions. Such information is widely available. Therefore, I don't propose to discuss the details of calculations here. Instead, we will see how these parameters are calculated, using the gamma distribution as an example.
To use the Gamma model, all we need are two parameters, *c* and *a*. These are obtained as

Table 2
Actual and expected frequency of buying

Actual		Model	
# times	%HHs	%HHs	Diff.
0	59.8	60.0	-.2
1	21.5	21.0	.5
2	9.9	9.5	.5
3	4.3	4.6	-.3
4	1.8	2.3	-.5
5	1.1	1.2	-.1
6	0.6	.6	0
7	0.4	.3	.1
8	0.2	.2	0
9+	0.4	.3	.1

follows:
c (mean rate per unit of time) = survey mean frequency ÷ length of survey reporting period

a = square of the survey mean ÷ (survey variance - survey mean)
These parameters can easily be calculated from survey data. (In fact, practically all standard cross-tabulations provide - or can provide - mean rate and standard deviations. Once we have the two parameters *c* and *a*, the total distribution (such as the one shown in Table 2, column 3) can be calculated with any micro-computer or even a pocket calculator.
Note that the value of *a* is always positive. Should you obtain a negative value, it simply means (unless there is a calculation error) that this model is not appropriate for the data you have.

An example
These steps can be illustrated by a set of hypothetical data that resembles most purchase patterns. Suppose you collect data on the number of times beer was bought by 2,000 households during a given week. The data may look something like what are shown in columns 1 and 2 of Table 2. (The fact that such patterns are typical can be confirmed by any consumer panel that collects data on purchase behaviour.) If we calculate the mean of this purchase frequency, we arrive at 0.785 i.e. an ‘average’ household buys beer less than once a month - (.785 times to be exact) with a variance of .202.
According to our formula:

$c = .785 \div 1 = .785$
$a = (.196)^2 \div (.202 - .196) = .633$

Once we have these two parameters, the frequency of purchase can then be calculated by using simple algebraic formulas, as shown below:

% consumers who did not buy at all:
$F_0 = ((a/a+TC))^a *100$

% consumers who bought *m* times:
$F_m = f_{m-1} * ((a + m - 1) \div m)) * (Tc/(a+Tc))$

Where
F = Frequency of purchase (%)
T = Time period

For example, if you would like to know the percent of consumers who are likely to buy 3 times during a given week, then your m=3. We have already calculated the value of *c* and *a*. Because our calculated figures are for a week, T = 1. (If you want to use a different length of time e.g., 4 weeks instead of 1, set T = 4). We can simply substitute these values in the formula above and arrive at the percentage of consumers who would buy the product three times during the week.
Applying these formulas to our data we get the information shown in Table 3. You will note that the actual figures are fairly close. This is not an unusual result. For many established consumer products this model holds fairly well.

Table 3. How the Gamma and the Poisson distributions combine together to decompose the survey data into distribution of purchase frequency and rates

Rate	Mean rate	Gamma HH rate	Poisson HH Frequency					
		0	1	2	3	4	5+	
0.0-0.2	0.7	51%	87	12	.9	.1		
0.2-0.4	0.3	18	56	33	9	2	.2	
0.4-0.6	0.5	11	37	16	6	2	1	.3
……	…	…	…	…	…	…	…	…
……	…	…	…	…	…	…	…	…
……	…	…	…	…	…	…	…	…
Survey result			*62*	*20*	*10*	*4*	*2*	*2*

The Gamma and the Poisson distribution

Essentially this is the result of putting two separate (the Gamma and the Poisson) distributions together.

The Gamma distribution models the frequency of household buying at a given rate while the Poisson distribution accounts for the distribution of purchases within a household. These are described below.

The Gamma model

Going back to the purchase distribution curve (the Gamma distribution), we note that it is a continuous curve which means that any purchase frequency is possible. The x-axis of this curve represents the rate of purchase and the y-axis the proportion of buyers. The area under this curve between any two buying rates is the proportion of consumers who would buy that product at the rates defined by the rates under consideration.

A main characteristic of this model is that, as the rate of purchase of buyers decrease, the proportion of buyers between a given rate *r* and *2r* will always be higher than the proportion of buyers between *2r* and *3r*.

Reporting periods

In our example, we used a one week buying rate for our calculations. But why one week? Why not monthly or quarterly? The answer is that while survey or panel data may exist in any form, for each product, a characteristic purchase cycle may exist. For instance, the characteristic cycle for beer may be every week or every month rather than every 10 days.

Reporting periods in surveys should take into account several factors such as the cost of data collection and the possibility of recall error. However, for modelling purchases, what determines the time period is the probable purchase cycle.

While using a logical time cycle is desirable, it should be noted that the distribution itself is not scale dependent. If one week's rate is .196, the four week rate is simply .196 times 4, no more, no less.

The Poisson model

As noted before, each family will have a different distribution of buying frequencies. This will, of course, depend on their buying rates.

If we review the rate of distribution curve (Figure 1) we note that 50.7% of families show a rate of buying that ranges between 0 and .2 times per week. However, any family falling into this part of the distribution may have bought 0, 1, 2,... number of times. Let us assume that households falling in this range have a mean purchase frequency of 0.7. If the mean purchase frequency is 0.7, how many can be expected to buy the product 0, 1, 2,... times during the time period under consideration? the Poisson distribution provides the answer to this question. (For details of frequency calculations please consult any book on statistics.)

Combining the Gamma and the Poisson distributions

Recall the frequency of beer purchase we obtained from the survey. All we know from the survey is the number of times a household has bought an item and the average rate of purchase. What the Gamma-Poisson distributions show is how the survey results come about.

Models behind the numbers

Survey data provides the basic material for the Gamma and the Poisson distributions. If the Gamma-Poisson model fits the data (it appears to, for many established products) then it explains *how* the survey results came about.

The models identify the patterns of purchase, and this alone can be a valuable aid to the marketer in understanding buyer behaviour.

Dr. Chuck Chakrapani is a consultant in advanced marketing research and modelling techniques.