Modeling Response to Repetitive Promotional Stimuli

Richard J. Fox
Srinivas K. Reddy
University of Georgia

Bharat Rao
Harvard Business School

Marketers frequently include promotional stimuli which elicit some form of response from the recipient among the tactics used to market products or services. Print ads, including 800 numbers which allow consumers to respond, and direct mail campaigns are examples of such activities. Promotions of this nature are often repeated a number of times, thus providing several opportunities to respond. Understanding consumer response to such campaigns is critical for more efficient design and use of these activities. A conceptual framework of response to repetitive stimuli is proposed, and stochastic models of alternative response patterns are developed. Alternative contexts in which such models are useful are also noted. Variations of the models are provided for those situations when only a fraction of the target population will ever respond. Estimation of model parameters is discussed, and data from actual campaigns are used to demonstrate how to apply the models.

Repetitive promotions can take a variety of forms. A direct response ad, which offers merchandise or service for sale and which appears in several issues of a periodical, either consecutively or pulsed, is an example of a repetitive promotion. An offer repeatedly placed in monthly bills sent to utility customers and an infomercial shown repeatedly to viewers on a local cable channel are also repetitive promotions. U.S. advertisers spent $31.2 billion on direct mail in 1995, more than twice what they spent in 1984 (Economic Impact 1995). Mail-order catalogs are used very frequently in a repetitive manner to introduce prospects to a direct marketer’s offering. In 1992, 13.5 billion such catalogs were mailed as compared with 8.7 billion in 1983 ("USA Snapshots" 1993). This growth coincides with an increasing tendency of American consumers to purchase by phone or mail. The Simmons Study of Media and Markets reported that 55.2 percent of U.S. households made a purchase by phone or mail in 1992 (DMA Statistical Factbook 1993-1994). Since 1990, consumer mail-order sales growth, 9.9 percent annually, has exceeded the growth of in-store retail sales, reaching $43 billion in 1995 and projected to reach $63 billion by 2001 (Ray and Reis 1996). With billions of dollars being spent on repetitive promotional efforts and on changing consumer purchase patterns, it is important to develop models of consumer response to repetitive stimuli to make most efficient use of such campaigns as well as maximize related profits.

In any repetitive campaign, marketers face a fundamental question: How many times should a promotional stimulus, designed to induce a response, be sent to an individual? Related to this overriding question are the following additional considerations:
What is the cumulative impact and marginal effect of repeating a promotional stimuli?

Is there heterogeneity in response probability across the population, and if so, can this heterogeneity be measured and described?

Despite the recognized importance of modeling response to repetitive marketing efforts, academic researchers have only recently begun to address the issue. The objectives of this research are to propose and examine stochastic models representing alternative patterns of response to repetitive promotional stimuli. Initially, literature on efforts to model responses to repetitive marketing efforts is reviewed. Then, a conceptual framework of consumer response to repetitive promotional stimuli is presented. Alternative stochastic model parameters are proposed to accommodate different potential response patterns. Estimation of the model parameters is discussed next. The models are then modified to accommodate a dichotomous target population—prospects and non-prospects, or those who will eventually respond versus those who will never respond. Applications involving actual marketing campaigns are presented to illustrate the use of the various models. The final section discusses the implications of the models and their relevance to repetitive marketing efforts.

CONCEPTUAL FOUNDATIONS

Repetitive Advertising Research

Ray and Sawyer (1971) in their review of repetitive advertising research classified the various approaches into field experiments and lab experiments. Typically, these studies concentrated on testing psychological theories of learning and forgetting. Zietske's (1959) pioneering field experiment focused on repetitive print advertising. Using advertising recall as a proxy for ad effectiveness, Zietske found that, due to decay in recall, pulsed advertising is most effective. Strong (1972, 1974) also used a field experiment to investigate the effects of repetitive advertising on learning and recall. He developed a model which provided estimates of advertising recall for a variety of schedules. Others have attempted to study the impact of repetitive advertising in a laboratory setting rather than a field setting. Ray and Sawyer varied the type of advertising stimuli in different product and brand situations and found that different repetition functions and media strategies are necessary depending on the product classification (convenience vs. shopping good), brand position, advertising format, and advertising goals (recall vs. attitude).

The work of Appel (1971) and Grass (1968) sheds some light on the shape of the response function to repetitive advertising stimuli. They found that the response initially increases to a maximum and then declines. Grass and subsequently Krugman (1972) have popularized the three-exposure rule based on their studies. Basically, they suggest that mailing more than three times to a customer is wasteful.

Recent research in direct response modeling has taken two distinct paths—the individual-level approach and the aggregate. For example, researchers have used individual-level data to investigate who are the active and inactive customers in a mailing list (Schmittlein, Morrison, and Colombo 1987), how to select prospects in a mailing list (Rao and Steckel 1993), or segmenting customers (De Sarbo and Ramaswamy 1994). On the other hand, Buchanan and Morrison (1988) developed an aggregate model, using the beta-binomial model, to investigate the phenomenon of list falloff, which is a common problem in direct marketing. If their model applies, the researcher can determine the number of profitable solicitations based on just two test mailings. We also use an aggregate approach to investigate response to repetitive marketing.

Conceptual Framework

The behavioral process of response to repetitive marketing stimuli which forms the basis for the stochastic models is outlined in Figure 1. Potential customers (n) receive the stimulus in the form of an ad, a telephone contact, or a mail solicitation such as a catalog, request for donation, or an offer to provide product or service information such as a free auto insurance quote. The objective of the stimulus is to elicit a behavioral response, such as a purchase, mail-in subscription, or mail-in inquiry. A total of n potential customers respond, leaving a base of (n - n1) potential customers subjected to the next (identical) stimulus, and so on.

The two consumer response patterns to repetitive stimuli depicted in Figure 2 are suggested in the literature. The first pattern in which the response frequency is monotonically decreasing is quite common and occurs, for example, when consecutive stimuli are considered independent and the probability of a positive response is assumed to be the same across stimuli. In this case, the number of the stimulus which elicits a positive response follows the geometric distribution (see, e.g., Dwass 1970, p. 156). Clearly, such a model has intuitive appeal when considering response at some point to repetitive stimuli. The response pattern is decreasing whether the underlying population is assumed to be heterogeneous with respect to the response probability. Buchanan and Morrison (1988) investigate this response pattern, assuming heterogeneity, in the context of repeated direct mailings.

Learning theory provides a basis for the second response pattern in which the response frequency to an advertisement or promotion may be low initially but increases with repetition to a maximum level and diminishes with subsequent repetitions. In this case, several exposures may be required for consumers to fully evaluate the promotion or advertising. The probability of responding positively increases to a maximum value with repeated exposures to the same stimulus and decreases thereafter. Such a pattern was observed by Appel (1971) and Grass.
(1968) in their studies of attention response to repetitive television commercial exposures. In a lab study, Ray and Sawyer (1971) found that recall of an ad and purchase intent for the advertised product also produced this pattern. Calder and Sternthal (1980), investigating the wear-out effects of television commercials, found that mean product evaluations initially increased with repetition before dropping. Calder and Sternthal's (1980) information processing approach and Berlyne's (1970) two-factor theory provide
strong support for the two response patterns presented in Figure 2. Calder and Sterntahl draw heavily on the work of Cacioppo and Petty (1979), who argue that consumers, when confronted with advertising messages, have two kinds of thoughts—message-related thoughts and own thoughts. Message-related thoughts are directly stimulated by the message and to a large extent reflect message content. Own thoughts, on the other hand, are based on associations and reflecting on previous experiences. Calder and Sterntahl suggest that for initial exposures, the consumers' thoughts tend to be message related. After repeated exposures, the thoughts tend mainly to be based on associations only indirectly linked to the message (own thoughts). They argue that, in general, these thoughts are less positive than the initial message-related thoughts. Such a decrease in message-related thoughts and an increase in own thoughts will eventually produce a wear-out effect and an eventual decrease in response rates. This explanation can accommodate both the response patterns presented in Figure 2. The first pattern, which illustrates a constantly decaying response, suggests that the campaign elicits most of the message thoughts during the first exposure and that subsequent exposures elicit more and more own thoughts. The second response pattern is appropriate when increasing message thoughts occur during the initial few exposures and when subsequent exposures elicit more own thoughts.

Berlyne (1970) posits that two opposing factors come into play when individuals are exposed to repetitive stimuli. The first factor, which Berlyne refers to as positive habituation, reduces arousal from uncertainty and conflict and increases pleasure. Tedium or satiation, the second factor, increases with exposure, resulting in less pleasantness. During the early stages of repetition, positive habituation will dominate, and with repeated exposure to the same stimulus, tedium will eventually dominate, reducing the overall positive response (Sawyer 1981).  

In the next section, a variety of models consistent with the phenomenon of first response to promotional stimuli is developed. The models range in complexity from a simple constant probability of response model to more complex models allowing for a different pattern of response, as well as individual differences within the target population and a dichotomy in the target population—prospects, or those who will eventually respond, and non-prospects, or those who will never respond. Obtaining an estimate of the proportion of the target population who are prospects after only several stimuli is, of course, valuable information for managing a direct marketing campaign.

The models can be applied to actual campaign data "on the fly" to assess the merits of continuing, as well as projecting how much longer to continue and the associated anticipated return. Basu, Basu, and Batra (1995) discuss such an application. They use early returns to a direct marketing mailing to estimate the parameters of alternative models of cumulative response as a function of time. The models we develop can also be applied to data obtained in a test situation in which a stimulus or several variations of a stimulus are being evaluated for full-scale use. For example, a company might use the initial results from several repetitive campaigns conducted among relatively small samples from the target audience to estimate key parameters and the optimal number of repetitions for each, and ultimately to compare potential payoffs.

**ALTERNATIVE RESPONSE MODELS**

It is first assumed that a promotional stimulus is repeated $K \geq 1$ times and that each response can be traced to a particular one of the $K \geq 1$ identical stimuli. It is also assumed that individuals who have not yet responded after $K$ stimuli would do so eventually if the stimuli were continued indefinitely. The random variable $X$ represents the stimulus to which an individual responds and is only observed when $X \leq K$. Otherwise, it is known only that $X$ exceeds $K$.

**Geometric Response (Model A)**

The probability that an individual responds to the stimulus, denoted by $0 \leq r \leq 1$, is the same for all individuals in the target population. The stimulus is presented a total of $K \geq 1$ times, and each exposure is an independent trial. The random variable $X$, which again represents the stimulus to which an individual responds, follows the familiar geometric distribution (see, e.g., Dwass 1970, p. 156), and

$$P(X = i) = (1 - r)^{i-1}r, \text{ for } i = 1, 2, \ldots$$

and

$$P(X > K) = (1 - r)^K.$$  

(1)  

(2)

An obvious variation of this basic model is to assume that each stimulus is a Bernoulli trial in which the probability of exposure to the stimulus is $0 < q < 1$. As intuition suggests, the distribution of $X$ in this case is the same as given by Equations 1 and 2 with the product of $r$ and $q$ replacing $r$ so that $q$ and $r$ cannot be estimated separately. A formal proof is given in Appendix 1.

**Geometric Response With Population Heterogeneity (Model B)**

The logical extension of Model A is to allow $r$, the probability of responding to the stimulus, to vary across individuals. The beta distribution is assumed to represent the population heterogeneity in $r$. Thus, the density function describing the distribution of $r$ is $f(r) = \frac{\beta r^{\alpha-1}(1 - r)^{\beta-1}}{B(\alpha, \beta)}$, for $0 < r < 1$, where $B$ represents the beta function. If $\alpha > 0$ and $\beta > 0$ are the parameters of the associated beta distribution, then it is easily shown that
\[ P(X = 1) = \frac{\alpha}{\alpha + \beta} \]  

(3)

\[ P(X = i) = \left( \frac{\alpha}{\alpha + \beta} \right) \left( \frac{\beta}{\alpha + \beta + 1} \right)^{\beta + 1} \left( \frac{\alpha + \beta + 2}{\alpha + \beta} \right) \]

\[ \cdots \left( \frac{\beta + (i - 2)}{\alpha + \beta + (i - 1)} \right), \text{ for } i = 2, 3, \ldots, \]

(4)

and

\[ P(X > K) = \left( \frac{\beta}{\alpha + \beta} \right) \left( \frac{\beta + 1}{\alpha + \beta + 1} \right) \]

\[ \cdots \left( \frac{\beta + (K - 1)}{\alpha + \beta + (K - 1)} \right). \]

(5)

The mean of the beta distribution is \( \alpha / (\alpha + \beta) \), and the variance is \( \alpha \beta (\alpha + \beta + 1) (\alpha + \beta)^{-1} \). The polarization index, \( \phi = (\alpha + \beta + 1)^{-1} \) (Buchanan and Morrison 1985; Kalwani 1980; Sabavala and Morrison 1977), is typically used to characterize the amount of heterogeneity in the beta distribution of \( r \) across individuals. As \( \alpha \) and \( \beta \) both approach 0, the distribution of \( r \) approaches a discrete distribution with mass at \( r = 0 \) and \( r = 1 \), and \( \phi \) approaches 1. This represents extreme heterogeneity, in which individuals respond to the very first stimulus or do not respond at all. If \( \alpha \) and/or \( \beta \) become very large, then \( \phi \) approaches 0, and the distribution of \( r \) approaches a discrete distribution with mass concentrated at \( \alpha / (\alpha + \beta) \), which is Model A (no heterogeneity) with \( r = \alpha / (\alpha + \beta) \). Therefore, any distribution of Model A can be approximated arbitrarily closely with a Model B distribution by choosing \( \alpha \) and \( \beta \) sufficiently large, and in such a way that \( \alpha / (\alpha + \beta) \) is equal to the specified \( r \) of Model A. However, one would still consider and use Model A, if appropriate, because of its simplicity relative to Model B.

Model B for \( K = 2 \) is the model used by Buchanan and Morrison (1988) to analyze list tailoff—that is, the decline in response rate to a direct mail offer made several times in succession. They showed that, assuming this model, response rates decline systematically as functions of \( \alpha \) and \( \beta \) and that, having observed only the first two response rates, one can estimate \( \alpha \) and \( \beta \), and hence all future response rates. However, just two response rates are insufficient information to establish that Model B is appropriate, and estimation of \( \alpha \) and \( \beta \) is considerably more complicated when \( K > 2 \), which is the situation we consider. Also, their approach precludes the possibility of a segment of nonprospects who will never respond regardless of how many promotional stimuli they experience. We allow for this possibility in Model B’ discussed later in the article.

**Poisson Response (Model C)**

In both Models A and B, the response probability declines monotonically with the number of exposures. This is Pattern 1 depicted in Figure 2. Model C can, however, be used to fit responses which follow Pattern 2 of Figure 2. The number of exposures preceding the exposure to which a consumer responds is assumed to follow the Poisson distribution. That is, \( X = Y + 1 \), where \( Y \) follows the Poisson distribution with parameter \( \lambda > 0 \). Hence, for \( i = 1, 2, \ldots \), and \( \lambda > 0 \),

\[ P(X = i) = e^{-\lambda} \lambda^{i-1} / (i-1)! \]

(6)

and

\[ P(X > K) = \sum_{i=K}^{\infty} e^{-\lambda} \lambda^{i-1} / i!. \]

(7)

**Poisson Response With Population Heterogeneity—NBD (Model D)**

Suppose Model C is assumed, but it is also assumed that the population is heterogeneous with respect to \( \lambda \). To account for this heterogeneity, \( \lambda \) is assumed to have a Gamma distribution with parameters \( \alpha > 0 \) and \( s > 0 \). That is, \( \lambda \) is a positive random variable having density function \( f(t) = \alpha t^{\alpha-1} e^{-\alpha t / \Gamma(s)} \) for \( t > 0 \), where \( \Gamma \) denotes the Gamma function. It follows that for \( i = 1, 2, \ldots \),

\[ P(X = i) = \left( \frac{(i - 1) + s - 1}{i - 1} \right) \left( \frac{\alpha}{\alpha + 1} \right)^{i-1} \left( \frac{1}{\alpha + 1} \right)^{\alpha + 1}, \]

or equivalently

\[ P(X = 1) = \left( \frac{\alpha}{\alpha + 1} \right)^{1}, \]

(8)

and for \( i \geq 2, \)

\[ P(X = i) = \left( \frac{s + i - 2}{(i - 1)!} \right) \left( \frac{\alpha}{\alpha + 1} \right)^{s} \left( \frac{1}{\alpha + 1} \right)^{\alpha + 1}. \]

(9)

This, of course, is the familiar NBD model (see, e.g., Morrison and Schmittlein 1988). (By our definition, \( X \) represents the stimulus to which an individual responds, so \( X = 1 \), the number of stimuli preceding a response, actually follows the NBD distribution.) If \( s \) is a positive integer, then the number of exposures preceding a response has a distribution equivalent to the number of failures.
preceding the \( i \)th success in a sequence of Bernoulli trials in which the probability of success is \( \alpha/(\alpha + 1) \) and the probability of a failure is \( 1 - (\alpha/(\alpha + 1)) = 1/(\alpha + 1) \). In this case, the distribution of the number of exposures preceding a response is referred to as the binomial waiting time, Pascal, or Polya distribution (see, e.g., Cohen 1991, p. 208).

The mean of the Gamma distribution is \( s/\alpha \), and the variance is \( s/\alpha^2 \). Any distribution of Model C can be approximated arbitrarily closely with a Model D distribution by choosing \( s \) and \( \alpha \) so that the ratio \( s/\alpha \) is equal to the specified \( \lambda \) of Model C, and \( s/\alpha^2 \) is close to 0. It follows that, as with Models A and B, the class of distributions defined by Model D essentially contains the class defined by Model C, but not vice versa. Again, one would still consider and use Model C, if appropriate, because of its simplicity relative to Model D.

**DISCUSSION**

In this section, we derive distinguishing features associated with the probability distributions defined by Models A, B, C, and D. These conditions, summarized in Table 1, are useful for empirically determining which models, and associated parameter constraints, apply in given circumstances. To facilitate the discussion, for \( i = 1, 2, \ldots \), let \( f_i = P(X = i) \) and \( c_i = f_{i+1}/f_i \).

For Models A and B, \( f_i \) decreases as \( i \) increases. For Model A, \( c_i = (1 - r) \) by Equation 1, so that the ratio of consecutive probabilities is a constant—that is, the probability of responding on the \((i + 1)\)st stimulus is \((1 - r)\) times the probability of responding on the \(i\)th stimulus. For Model B, from Equations 3 and 4, \( c_i = (\beta + i - 1)/(\alpha + \beta + i) \), which increases in \( i \) to a limit of 1.

Under Model C, \( f_i \) increases to its maximum at \( i = [\lambda + 1] \), where \([\lambda]\) is the largest integer less than or equal to \( \lambda \), and decreases thereafter. (If \( \lambda \) is an integer, \( f_{\lambda+1} = f_{\lambda} \), and \( \lambda + 1 \) are both modal values of the distribution [see, e.g., Johnson and Kotz 1969, p. 92]). If \( \lambda < 1 \), then \( f_i \) decreases as \( i \) increases, so that in this special case of Model C, the response pattern is the same as the pattern for Models A and B—that is, Pattern 1 of Figure 2. Further, for Model C, \( c_i = \lambda \), so \( c_i \) decreases to zero.

For Model D, if \( s \geq \alpha + 1 \), then \( f_i \) increases to its maximum as \( i \) increases from 1 to \([(s + \alpha - 1)/\alpha] \), the mode, and decreases thereafter. (If \( s \geq \alpha + 1 \) and \((s + \alpha - 1)/\alpha\) is an integer, say \( i^* \), then \( f_{i^*} \) is equal to its maximum value at both \( i^* \) and \( i^* - 1 \).) If \( s < \alpha + 1 \), then \( f_i \) is a decreasing function of \( i \), which is the response pattern of Models A and B and the special case of Model C noted above. In the special case of \( s = 1 \), the distribution of \( \lambda \) is actually the negative exponential with parameter \( \alpha > 0 \), and Model D becomes Model A with \( \alpha/(\alpha + 1) \) and \( 1/(\alpha + 1) \) replacing \( r \) and \( 1 - r \), respectively. Finally, for Model D,

\[
c_i = (s + i - 1)/(s(\alpha + 1)) \text{ decreases in } i \text{ to the limit } (\alpha + 1)^{-1} \text{ when } s > 1, \text{ and } c_i \text{ is constant and equal to } (\alpha + 1)^{-1} \text{ (Model A)} \text{ when } s = 1, \text{ and } c_i \text{ increases in } i \text{ to } (\alpha + 1)^{-1} \text{ when } s < 1.
\]

In practice the conditions of Table 1 would be examined in the context of observed frequencies, \( n_i, i = 1, 2, \ldots, K \), and \( \tilde{c}_i = n_{i+1}/n_i, i = 1, 2, \ldots, K - 1 \). Models A and B and the special cases of Model C (\( \lambda < 1 \)) and D (\( s < \alpha + 1 \)) should be considered if the \( n_i \) follow Pattern 1 of Figure 2. The \( \tilde{c}_i \) sheds light on which of these options to choose. If the values of \( \tilde{c}_i \) are approximately constant, then Model A is appropriate. If \( \tilde{c}_i \) increases in \( i \) to a limit less than 1, then Model D, with \( s < 1 \) and \( (\alpha + 1)^{-1} \) equal to this limit, is appropriate. If this limit appears to be 1, then Model B is appropriate. If \( \tilde{c}_i \) decreases to a limit greater than 0, then Model D (\( 1 < s < \alpha + 1 \)) is appropriate. Further, in this case, \( 1/(\alpha + 1) \) is the lower limit of \( c_i \). If the limit of the decreasing \( \tilde{c}_i \) appears to be 0, then Model C (\( \lambda < 1 \)) is preferred.

If \( n_i \) behaves according to Pattern 2 of Figure 2, then only Model C (\( \lambda > 1 \)) or Model D (\( s > \alpha + 1 \)) is appropriate. The choice between these two models would depend on \( \tilde{c}_i \), which is decreasing in \( i \) in either case. If \( \tilde{c}_i \) appears to be decreasing to a limit less than 1 but greater than 0, then Model D (\( (\alpha + 1)^{-1} \) being the limit of \( c_i \)) is appropriate. If \( \tilde{c}_i \) appears to be decreasing to 0, then Model C (\( \lambda > 1 \)) is preferred.

**TABLE 1**

<table>
<thead>
<tr>
<th>Model</th>
<th>( f_i = P(X = i) )</th>
<th>( c_i = f_{i+1}/f_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Decreasing</td>
<td>Constant</td>
</tr>
<tr>
<td></td>
<td>( c_i = 1 - r )</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Decreasing</td>
<td>Increasing to 1.0</td>
</tr>
<tr>
<td></td>
<td>( c_i = (\beta + i - 1)/(\alpha + \beta + i) )</td>
<td>Decreasing</td>
</tr>
</tbody>
</table>
| C     | \( \lambda \geq 1 \) | Increasing to maximal  
|       | value at \( i = [\lambda] + 1 \) and decreasing thereafter | |
| D     | \( \lambda < 1 \)   | Decreasing           |
|       | \( d_i = (s + i - 1)/(i(\alpha + 1)) \) | Decreasing |
|       |                      |                      |
| s \geq \alpha + 1 | Increasing to maximal  
|       | value at \( \{(s + \alpha - 1)/\alpha \) and decreasing thereafter | |
| s = 1 | Model A              | Constant (1/\( (\alpha + 1) \) |
|       | \((r = \alpha/(\alpha + 1)) \) | |
| s < 1 | Decreasing           | Increasing to 1/(\( (\alpha + 1) \) |

**NOTE:** Model A and Model B are likely to be appropriate when response is more an impulse than an action resulting from careful consideration over an extended period of time. In the latter situation, Models C and D are likely to be appropriate.
ESTIMATION

Maximum Likelihood Estimates

The portion of the sample who respond (to one of the \( K \) stimuli) comprises a right-censored sample. For those who do respond, we observe the stimulus to which they respond; for those who do not respond, we know only that these observations exceed \( K \)—that is, these observations are censored. The likelihood function takes the form

\[
L = \left( \prod_{i=1}^{K} (P(X = i))^n \right) (P(X > K)^n - m),
\]

(10)

where \( n \) is the sample size, \( n_i \) is the number of individuals who respond to the \( i \)th stimulus, \( m = \sum_{i=1}^{K} n_i \) is the number in the sample who respond, and \( X \) is the generic random variable corresponding to the basic distribution specified by the model.

Generally speaking, closed-form solutions for the maximum likelihood estimators do not exist, and a computer optimization algorithm must be used to obtain numerical estimates for a particular set of results. One exception is the case of Model B with \( K = 2 \) (Buchanan and Morrison 1988). Also for Model C, the maximum likelihood estimator of \( \lambda \), say \( \hat{\lambda} \), based on the censored sample data, is the solution to the following equation (Cohen 1991, pp. 199-203):

\[
\bar{x} - 1 = \lambda \left[ 1 - \frac{n - m}{m} \left( \frac{\lambda^{k-1} e^{-\lambda} / (K-1)!}{\sum_{i=K}^{\infty} \lambda e^{-\lambda} / i!} \right) \right],
\]

where \( \bar{x} \) is the sample mean of the uncensored observations. An iterative trial-and-error procedure can be used to solve for \( \lambda \), as demonstrated in the following hypothetical example.

Example 1. A regional telephone company tests a promotion for a new service among a sample of 5,000 customers. The offer consists of a 20 percent discount on a future monthly bill of the customer’s choice as an incentive to subscribe to the service for at least a minimum period of time. Suppose 1,000 of the 5,000 customers in the sample subscribe and use their discounts as follows (\( n_i \) represents the month after subscribing that the discount is used): \( n_1 = 18 \), \( n_2 = 78 \), \( n_3 = 150 \), and \( n_4 = 192 \), so that \( m = 438 \) and \( n - m = 562 \).

The observed frequencies are increasing with the number of the stimulus. Hence, Models A and B are eliminated, and either Model C (\( \lambda \geq 1 \)) or Model D (\( s \geq \alpha + 1 \)) is feasible. Notice that under both Models C (\( \lambda \geq 1 \)) and D (\( s \geq \alpha + 1 \)), \( c \) are decreasing, but that under Model C, \( i_c \)

is constant and equal to \( \lambda \). The three values of \( \hat{\lambda} \) for this example are 4.33, 3.88, and 3.84, respectively. So, Model C seems reasonable. Since \( \bar{x} = 3.178 \), \( m = 438 \), and \( n - m = 562 \) for these data, the above equation, which is to be solved for \( \lambda \), becomes

\[
2.178 = \lambda \left[ 1 - \frac{562}{438} \left( \frac{\lambda^3 e^{-\lambda} / 3!}{1 - \sum_{i=0}^{3} \lambda^i e^{-\lambda} / i!} \right) \right].
\]

Solving by trial and error yields \( \hat{\lambda} = 3.95 \). Note that the modal value of the distribution of \( X \) corresponding to \( \lambda = 3.95 \) is 4 (\( \lfloor \hat{\lambda} \rfloor + 1 \)), which is consistent with the observed frequencies. The expected frequencies corresponding to \( \lambda = 3.95 \) are 19, 76, 150, and 198, respectively, and future response numbers are projected to decline. Expected frequencies for subsequent discounted are 195, 154, and so forth. The company now has a sense of not only the rate of acceptance and its cost but also how the cost will be distributed over time.

Another exception in which the maximum likelihood estimator can be calculated without the use of a computer optimization code is Model A. The maximum likelihood estimator for this case is derived in Appendix 2, and its use is demonstrated in the third application presented later in the article.

The following hypothetical example demonstrates the more typical scenario in which an optimization algorithm must be used to numerically obtain maximum likelihood estimators. GAUSS, a product of Aptech Systems, is a general purpose optimization PC software package, which contains a module designed specifically for maximum likelihood estimation. This subroutine was used in the applications appearing later in this article.

Example 2. A print ad including a direct response mail-in coupon for a free catalog of stereo equipment, available at discount prices, is included in four consecutive mailings to a sample of 1,000 members of a compact disc music club (\( K = 4 \) and \( n = 1,000 \)). The response pattern is as follows: \( n_0 = 200 \), \( n_1 = 270 \), \( n_2 = 210 \), and \( n_3 = 150 \), so that \( m = 830 \) and \( n - m = 170 \).

The frequencies increase to a maximum at \( i = 2 \) and decrease thereafter. Again, Model C (\( 1 < \lambda < 2 \)) or Model D (\( s > \alpha + 1 \)) is consistent with this response pattern. However, calculation of \( i_c = (m + i)/n \) yields 1.35, 1.54, and 2.13, so that the condition of \( i_c = \lambda \) for all \( i \) does not appear to be met, and Model D appears to be the appropriate choice. From Equations 8, 9, and 10, the likelihood function for Model D is

\[
L = \left[ \left( \frac{\alpha}{\alpha + 1} \right)^n \right]^{i_c} \left[ \left( \frac{\alpha}{\alpha + 1} \right)^s \left( \frac{1}{\alpha + 1} \right) \right]^{n-i_c}.
\]
\[
\left( \frac{s + K - 2}{(K - 1)!} \right) \left( \frac{1}{\alpha + 1} \right)^{s+1} \left( \frac{1}{\alpha + 1} \right)^{K-1} \left[ \sum_{i=1}^{K} \frac{n_i}{(i-1)!} \log(i-1)! \right]^{n-m} 
\]

The log of \( L \) of Equation 11 is given by

\[
\log(L) = sm \left( \log(\alpha) - \log(\alpha + 1) \right) - \log(\alpha + 1) 
\]

\[
\left[ \sum_{i=1}^{K} \frac{n_i}{(i-1)!} \right] + \sum_{i=2}^{s} \log(s) + \sum_{i=3}^{s} \log(s+1) 
\]

\[
+ \ldots + n_s \log(s+K-2) - \sum_{i=2}^{K} n_i \log((i-1)!) 
\]

(12)

\[
+ (n-m) \log \left[ 1 - \left( \frac{1}{\alpha + 1} \right)^{s+1} \left( \frac{1}{\alpha + 1} \right) \right]^{K-1} 
\]

Applying GAUSS to \( \log(L) \) of Equation 12 and using the above frequencies, the maximum likelihood estimates are \( \hat{\alpha} = 1.98 \) and \( \hat{s} = 3.92 \). Notice that \( \hat{s} > \hat{\alpha} + 1 \) and \( \lfloor (\hat{s} + \hat{\alpha} - 1)/\hat{\alpha} \rfloor = 2 \), as suggested by the pattern of observed frequencies. The expected frequencies corresponding to these parameter values are 201, 265, 219, and 145, respectively, and subsequent mailings would be expected to yield 84 responses, 45 responses, and so on.

**THE CASE OF PROSPECTS AND NONPROSPECTS**

Suppose it is not assumed that all individuals will eventually respond to a stimulus but, rather, that the audience consists of two mutually exclusive groups—those who will never respond and those who will eventually respond. Let \( 0 < p < 1 \) represent the fraction of the audience who are prospects—that is, who will eventually respond. Thus, \( 0 < 1 - p < 1 \) represents the fraction of the audience who are nonprospects. For instance, a more reasonable scenario for Example 2 is that a fraction of the customer base, \( p \), will ever respond, rather than assuming that all customers will eventually respond.

Morrison (1969) addressed a similar problem in the context of estimating future sales among nonbuyers in a prior period, assuming that the group of nonbuyers is composed of "hard-core" nonbuyers and consumers who just happened not to buy in the prior period. He assumed the NBD distribution (Model D) for the number of purchases by customers in the prior period and used an iterative method of moments approach to estimate the parameters. We make use of a computer optimization algorithm to obtain the more rigorous maximum likelihood estimates assuming a variety of possible distributions including the NBD (Model D).

The probability distributions of Models A, B, C, and D can all be modified to allow for a nonprospect segment. For example, Model A becomes Model A' defined by

\[
P(X = i) = p(1 - r)^{i-1} r, \; i = 1, 2, \ldots, K, 
\]

and

\[
P(X > K) = (1 - p) + p(1 - r)^K. 
\]

Likewise, Model B becomes Model B' defined by

\[
P(X = 1) = p(\alpha/(\alpha + \beta)) 
\]

\[
P(X = i) = \frac{p}{\alpha + \beta} \left( \frac{\beta}{\alpha + \beta + 1} \right) \left( \frac{\beta + 1}{\alpha + \beta + 2} \right) \ldots \left( \frac{\beta + (i-2)}{\alpha + \beta + (i-1)} \right) 
\]

for \( i = 2, 3, \ldots, K, \)

and

\[
P(X > K) = (1 - p) + \frac{p}{\alpha + \beta} \left( \frac{\beta}{\alpha + \beta + 1} \right) \left( \frac{\beta}{\alpha + \beta + 2} \right) \ldots \left( \frac{\beta}{\alpha + \beta + (K-1)} \right) 
\]

The analogous expressions for Models C' and D' are included in the following equations which demonstrate the use of Equation 10 in the contexts of these two models respectively:

\[
L = \prod_{i=1}^{K} \left( \frac{pe^{\lambda} \lambda^{i-1}}{(i-1)!} \right)^{n_i} \left( 1 - p \right) + p \left( 1 - e^{\lambda} \right)^{s+1} \left( \frac{\lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + \ldots + \frac{\lambda^{K-1}}{(K-1)!}}{(\alpha + 1)^{s+1}} \right)^{n-m} 
\]

(13)
TABLE 2
Direct Response Ad Data (Application 1) (n = 10,000)

<table>
<thead>
<tr>
<th>Number of Stimulus</th>
<th>Number Responding (ni)</th>
<th>( \hat{c}_i )</th>
<th>Expected Frequency (Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>236</td>
<td>966</td>
<td>235.8</td>
</tr>
<tr>
<td>2</td>
<td>228</td>
<td>719</td>
<td>224.3</td>
</tr>
<tr>
<td>3</td>
<td>164</td>
<td>701</td>
<td>167.6</td>
</tr>
<tr>
<td>4</td>
<td>115</td>
<td>741.2</td>
<td>113.5</td>
</tr>
<tr>
<td>( m = 743 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n - m = 9,257 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \chi^2 = 16^6 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Critical value of \( \chi^2 \) for \( \alpha = 0.05 \) and (5-1-3) = 1 degree of freedom is 3.84.

and

\[
L = p^m \left( \frac{\alpha}{\alpha + 1} \right)^m \prod_{i=2}^{K} \left( \frac{(s + i - 2) \ldots s}{(i - 1)!} \right) \left( \frac{1}{\alpha + 1} \right)^{i-1} \]

\[
\times \left[ 1 - p + p \left( \frac{\alpha}{\alpha + 1} \right)^s \right] \left[ 1 + \sum_{i=2}^{K} \left( \frac{(s + i - 2) \ldots s}{(i - 1)!} \right) \left( \frac{1}{\alpha + 1} \right)^{i-1} \right]. \tag{14}
\]

Once again, maximum likelihood estimation can be performed using an optimization algorithm to estimate model parameters. Kalwani and Silkt (1980) used a proprietary computer program developed by Kalwani (1975) to obtain maximum likelihood estimates for Model \( \hat{A} \) in the context of estimating depth of repeat purchasing of new products.

APPLICATIONS

The following three applications involve data from actual repetitive campaigns and demonstrate how the models developed in this article can be used in practice. The first application concerns a direct response ad which is inserted in a magazine targeted to industrial buyers with a readership of 10,000. The same ad is repeated for four consecutive issues. Table 2 shows the numbers responding to the first, second, third, and fourth ads, respectively. In this scenario, a prospect/prospect model, in which \( 1 - p \) represents the fraction of the audience who will never respond, appears most reasonable. Notice that in Table 2, the frequencies, \( n_i, i = 1, 2, 3, 4 \), are decreasing and that the \( \hat{c}_i \) are also decreasing, suggesting that Models \( C' \) (\( \lambda < 1 \)) and \( D' \) (\( s < \alpha + 1 \)) are appropriate. For Model \( C' \), \( \hat{c}_i = \lambda \) (see Table 1), which is less than 1.0 in the case of decreasing frequencies. However, for the data of Table 2, \( \hat{c}_i \) are .97, 1.44, and 2.1 for \( i = 1, 2, \) and 3, respectively, so Model \( C' \) is eliminated from consideration.

TABLE 3
Direct Response Ad Data (Application 2) (n = 5,000)

<table>
<thead>
<tr>
<th>Number of Stimulus</th>
<th>Number Responding (ni)</th>
<th>( \hat{c}_i )</th>
<th>Expected Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45</td>
<td>1.09</td>
<td>30.2</td>
</tr>
<tr>
<td>2</td>
<td>49</td>
<td>1.22</td>
<td>66.7</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>1.00</td>
<td>73.7</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>0.60</td>
<td>54.3</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td></td>
<td>30.0</td>
</tr>
<tr>
<td>( m = 250 )</td>
<td></td>
<td></td>
<td>254.9</td>
</tr>
<tr>
<td>( n - m = 4,750 )</td>
<td></td>
<td></td>
<td>4745.1</td>
</tr>
<tr>
<td>( \chi^2 = 16.3^6 )</td>
<td></td>
<td></td>
<td>5.96^6</td>
</tr>
</tbody>
</table>

a. Critical value of \( \chi^2 \) for \( \alpha = 0.05 \) and (6-1-2) = 3 degrees of freedom is 7.81.

b. Critical value of \( \chi^2 \) for \( \alpha = 0.05 \) and (6-1-3) = 2 degrees of freedom is 5.99.

Applying Model \( D' \)—that is, using GAUSS to obtain the values of \( \alpha, s, \) and \( p \) which maximize the log \( (L) \), where \( L \) is defined by Equation 14—yields

\( \hat{\alpha} = .84, \hat{s} = 1.75, \) and \( \hat{p} = .093. \)

Hence, ultimately 9.3 percent of the population are projected to respond to the ad. The number of ads preceding the one which generates a response follows a Poisson distribution with parameter \( \lambda > 0 \), and \( \lambda \) varies across the population of respondents according to a Gamma distribution with mean 2.08 (\( \hat{\lambda}/\hat{\alpha} \)) and variance 2.48 (\( \hat{\lambda}^2/\hat{\alpha}^2 \)). The \( \chi^2 \) goodness-of-fit test was applied, and the results, shown in the bottom half of Table 3, indicate that Model \( D' \) fits extremely well. Notice that \( 1 < \hat{s} = 1.75 < \hat{\alpha} + 1 = 1.84 \), which is consistent with \( n_i \) and \( \hat{c}_i \) decreasing. Subsequent stimuli are expected to produce frequencies of response given by

\[
10,000(.093) \left( \frac{(1.75 + i - 2) \ldots (1.75)}{(i - 1)!} \right) \left( \frac{.84}{1.84} \right)^{i-1}, i = 5, 6, \ldots
\]

For \( i = 5 \), we obtain 73.5 as the projected frequency of response, and so forth.

The second application involves another data set obtained in the same way as the one displayed in Table 2, but the response pattern is different. In this case, the stimulus was repeated five times to an audience of \( n = 5,000 \) buyers. Table 3 shows the frequencies of response to the stimuli as well as the values of \( \hat{c}_i, i = 1, 2, 3, 4 \). Note that \( n_i \) increases to a maximum value and decreases thereafter. The nature of these data suggests that Model \( C' \) (\( \lambda > 1 \)) and Model \( D' \) (\( s > \alpha + 1 \)) and \( [(s + \alpha - 1)/\alpha] < K = 5 \) are feasible, although \( c_s \) are strictly decreasing in both cases, and the \( \hat{c}_i \) for these data are not strictly decreasing. Once again, a case could be made for eliminating Model \( C' \) from consideration.
TABLE 4  
Book Club Attrition (Application 3) \( (n = 4,993) \)

<table>
<thead>
<tr>
<th>Number of Books</th>
<th>Number Dropping Out ( (n_0) )</th>
<th>Number Remaining ( \hat{N} )</th>
<th>Expected Frequencies</th>
<th>Projected Number Remaining</th>
<th>Expected Frequencies</th>
<th>Projected Number Remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,007</td>
<td>3,986</td>
<td>1.09</td>
<td>1,188</td>
<td>1,113</td>
<td>865</td>
</tr>
<tr>
<td>2</td>
<td>1,099</td>
<td>2,887</td>
<td>0.69</td>
<td>864</td>
<td>865</td>
<td>672</td>
</tr>
<tr>
<td>3</td>
<td>753</td>
<td>2,134</td>
<td>0.51</td>
<td>639</td>
<td>672</td>
<td>522</td>
</tr>
<tr>
<td>4</td>
<td>384</td>
<td>1,750</td>
<td>0.68</td>
<td>479</td>
<td>522</td>
<td>406</td>
</tr>
<tr>
<td>5</td>
<td>264</td>
<td>1,486</td>
<td></td>
<td>364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>221</td>
<td>1,265</td>
<td></td>
<td>280</td>
<td>315</td>
<td>1,100</td>
</tr>
<tr>
<td>7</td>
<td>185</td>
<td>1,080</td>
<td></td>
<td>210</td>
<td>244</td>
<td>856</td>
</tr>
<tr>
<td>8</td>
<td>162</td>
<td>918</td>
<td></td>
<td>160</td>
<td>190</td>
<td>666</td>
</tr>
<tr>
<td>9</td>
<td>159</td>
<td>759</td>
<td></td>
<td>120</td>
<td>148</td>
<td>518</td>
</tr>
<tr>
<td>10</td>
<td>84</td>
<td>675</td>
<td></td>
<td>95</td>
<td>115</td>
<td>403</td>
</tr>
<tr>
<td>11</td>
<td>73</td>
<td>602</td>
<td></td>
<td>75</td>
<td>89</td>
<td>314</td>
</tr>
</tbody>
</table>

because of nonconstant \( \hat{\alpha} \), but Model C' was retained to demonstrate the difference in fit to the data between the two models.

The maximum likelihood estimates for the two models (Equations 13 and 14, respectively), are

Model C': \( \hat{\lambda} = 2.24 \)

Model D': \( \hat{\alpha} = .58 \)

Hence, allowing for heterogeneity increases the estimate of the fraction who will ultimately respond ranges from 5.5 percent to 7.4 percent, and the heterogeneity in \( \lambda \) across responders is represented with a Gamma distribution having mean \( 3.72 (\hat{\lambda}/\hat{\alpha}) \) and variance \( 6.42 (\hat{\lambda}/\hat{\alpha})^2 \).

Notice that \( \hat{\lambda} = 2.2 > 1 \), or \( \hat{s} \geq \hat{\alpha} + 1 \) and \((\hat{\gamma} + \hat{\alpha} - 1)/\hat{\alpha}\) is an integer—namely 3—the values of \( f_i = P(X = i) \) corresponding to the estimated distribution increase to a maximal value which occurs at \( i = 2 \) and \( i = 3 \) and decrease thereafter—the pattern of the expected frequencies for Model D' (see Table 3). The expected frequencies and the associated \( \chi^2 \) goodness-of-fit test statistics (statistically significant at the 5% risk level for Model C' but not for Model D') indicate that Model D' fits the data better than Model C'. Model D' suggests that slightly more than 7 percent of the buyers will respond. Further, of these, the estimated percentage responding to the \( i \)th stimulus, \( i = 1, 2, \ldots, \), is

\[
\left( (i - 1) + 1.16 \right) \left( \frac{0.58}{1.58} \right)^{2.16} \left( \frac{1}{1.58} \right)^{i - 1}.
\]

In particular, about 32 responses are anticipated for the next (sixth) stimulus.

Models of response to repetitive promotional stimuli can also be used to investigate a negative phenomenon—how long before a customer terminates a buying relationship. For example, Starbucks' coffee club members receive a gift upon joining and agree to periodically purchase a minimal amount of coffee indefinitely. The event of interest in such a situation is when a response occurs but, rather, when an existing relationship is terminated. The third application demonstrates this frequently encountered context, which is analogous to the mailing list fallout phenomenon studied by Buchanan and Morrison (1988).

An offer of a free first book is used as an incentive for enrollment in a book club. The original offer is typically made by a direct response mailing, TV commercial/infomercial, or print advertisement. Subsequently, members periodically receive books for which they pay a predetermined amount plus shipping and handling fees. Membership can be terminated at any time, after buying one book, by returning the current offering. (Members refusing the first offering after the free book must return both books.) The total membership then consists of all who make the first purchase. Over time, attrition will fully deplete the base membership. Of interest, of course, is how rapidly this decay in membership takes place.

Table 4 contains actual data for the type of book club just described. In this case, 4,993 individuals enroll by making the first purchase; 1,007 fail to make another purchase, and so on. Conceptually, viewing each book sent to a member as an independent identical trial with two outcomes, purchase or not, argues for Model A (no heterogeneity) or Model B (heterogeneity). The data are somewhat inconsistent with these models in that the second frequency is greater than the first. However, thereafter, the frequencies are decreasing. Assuming \( K = 5 \)—that is, attrition for the first five book offers in the series (following the first purchased book) is observed—and applying Model D (Equation 11), which allows for response probabilities which increase to a maximum value and then decrease (\( s > \alpha + 1 \)), the maximum likelihood estimates are \( \hat{\alpha} = .283 \) and \( \hat{s} = .991 \). This is basically Model A with \( r = .22(.283/(1 + .283)) \) because \( \hat{s} \) is practically 1. Hence, attempting to accommodate the observed pattern of responses in Table 4 by applying Model D yields essentially

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
an estimated distribution with decreasing probabilities (Model A with \( r = .22 \)). Hence, we proceed by considering the more conceptually appealing Models A and B.

Assuming Model A with \( K = 5 \) and applying the method for calculating the maximum likelihood estimate of \( r \) described in Appendix 2, we have

\[
\hat{r} = \frac{m}{\sum_{i=1}^{m} n_i + 5(n - m)} = \frac{3.507}{8,320 + 5(1,486)} = \frac{3.507}{15,800} = .223.
\]

The maximum likelihood estimates for Model B (Equations 4 and 5 used in conjunction with Equation 10) are \( \alpha = 4.77 \) and \( \beta = 15.27 \). (The Buchanan and Morrison [1988] approach, which uses only the first two frequencies, cannot be applied to these data because the second observed frequency exceeds the first.) The mean for the estimated distribution of \( r \) under Model B is .238 \((\hat{\alpha}/(\hat{\alpha} + \hat{\beta}))\), and the polarization index is .05 \(((\hat{\alpha} + \hat{\beta} + 1)^{-1})\), which indicates relatively low heterogeneity.

Table 4 contains the expected frequencies and the projected number of remaining customers for each mailing beyond the fifth. Although both models obviously fail a fit test, the differences between observed and expected are considerably smaller for Model B (heterogeneity) than for Model A (no heterogeneity). Further, the projections based on Model B are superior to those based on Model A and are close enough to have managerial value in the context of prediction. Using Model B to project future sales suggests that the subsequent mailings will produce sales of 1,206 units (versus actual sales of 1,265 units), 996 units (versus actual sales of 1,080 units), and so forth. Similarly, it is estimated that 621 members will make 10 purchases in addition to the initial purchase (versus an actual figure of 675), and so on.

**SUMMARY**

Four models of response to repetitive promotional stimuli are presented. The models allow for two basic patterns of response—decreasing response rates versus response rates which increase to a maximum value and decrease thereafter—and heterogeneity versus homogeneity within the target population. A conceptual model of the process of response forms the basis for the development of these stochastic models. Heuristics are provided to make a preliminary determination regarding which of the four models is feasible based on the pattern of response to several consecutive stimuli. Also, the models can be extended to the likely situation of a dichotomy within the population—those who will never respond versus those who will eventually respond.

Aggregate response data for a consecutive stimuli are treated as a right-censored sample, and the method of maximum likelihood is used to estimate the parameters of the selected model(s). Having a model, one can then estimate the levels of response to additional stimuli as well as the fraction of the target population that will ultimately respond. Hence, judgements regarding the number of additional times the stimulus should be used can be made based on only response data for several initial stimuli.

An important application of the proposed models is pretesting. One could conduct a campaign on a random sample from the target population. From the observed pattern of initial responses, a response model could be estimated and then used to determine the desired number of mailings and to project the campaign's potential. Alternative campaigns could be compared on this basis.

An important question concerns the number of times a promotional stimuli should be directed to an individual to induce a response. The idea popularized by Krugman (1972) that more than three exposures may be wasteful was not based on any solid empirical evidence. The models proposed here take into account potentially differing response patterns and heterogeneity in response probabilities and provide a sound basis for examining this question.

We have not investigated the effects of factors such as type of stimulus, advertising medium, type of market, and so forth on response patterns. Future work should incorporate information from various types of campaigns so that the associations between response functions, as well as the parameters, and campaign characteristics can be explored.

The time interval between repetitions is an important issue in the development of models of repetitive stimuli. Scant literature exists concerning the timing of advertising pulses. Strong (1977) provides a computer heuristic model to schedule a limited number of advertising exposures under different scenarios of demand seasonality. He found synergistic effects of scheduling ads close together. He also found that grouping several ad exposures is beneficial early in an ad campaign. Simon (1982) presents an advertising pulsing model (ADPULS) which takes wear out into account. Wolff and Subramanian (1977) presented some stochastic models which provided insight into the optimal timing of advertising pulsing under various types of consumer response. Models allowing for response probabilities that vary according to the time between stimuli are considerably more complex than the ones presented in this article and provide an area of opportunity for future research.

**ACKNOWLEDGMENTS**

We acknowledge financial support provided to the authors by the Direct Marketing Educational Foundation and the Marketing Science Institute.
APPENDIX 1
Model A With Exposures Following Bernoulli Process

We adopt the following notation:
0 < q < 1 represents the probability that an individual is exposed to a stimulus,
0 < r < 1 represents the probability of responding to a stimulus once it is seen, and
X = stimulus to which an individual responds (a positive integer).

Let Y be the binomially distributed random variable corresponding to the number of stimuli to which an individual is exposed. It follows that for any K ≥ 1,

\[
P(X > K) = P(Y = 0) + \sum_{j=1}^{K} P(\text{No Response} / Y = j) \cdot P(Y = j)
\]

\[
= (1 - q)^K + \sum_{j=1}^{K} P(\text{No Response} / Y = j) \cdot P(Y = j)
\]

\[
= (1 - q)^K + \sum_{j=1}^{K} \binom{K}{j} q^j (1 - q)^{K-j}
\]

\[
= (1 - q)^K + \sum_{j=1}^{K} \binom{K}{j} (q(1 - r))^j (1 - q)^{K-j}. \tag{15}
\]

Applying the binomial expansion formula to the sum in the right-hand side of Equation 15, we have

\[
P(X > K) = (1 - q)^K + [(q(1 - r) + (1 - q))^K - (1 - q)^K] = (1 - qr)^K. \tag{16}
\]

Noting that P(X > 0) = 1, it follows from Equation 16 that for any j ≥ 1,

\[
P(X = j) = P(X > j - 1) - P(X > j)
\]

\[
= (1 - qr)^{j-1} - (1 - qr)^j = (1 - qr)^{j-1}qr. \tag{17}
\]

Hence, from Equation 17, allowing exposures to follow a Bernoulli process with parameter 0 < q < 1 in Model A results in the same model, but the parameter of Model A is now the product of q and r. Separate estimation of q and r from basic response data—that is, data which does not include information regarding the number of stimuli to which a respondent was exposed—is not possible.

APPENDIX 2
Maximum Likelihood Estimator for Model A

Let K be the number of stimuli, n = sample size, n_i, i = 1, 2, . . . , K represent the number in the sample who respond to the ith stimulus, and m = n_1 + n_2 + . . . + n_K. The log likelihood function is given by

\[
\log(L) = \sum_{i=1}^{K} n_i [\log(r) + (i - 1)\log(1 - r)]
\]

\[
+ K(n - m)\log(1 - r) = m\log(r) + \log(1 - r) \tag{18}
\]

\[
\cdot \left[ \sum_{i=1}^{K} in_i + nK - (K + 1)m \right].
\]

Taking the derivative of \log(L) of Equation 18 and setting it equal to 0 yields

\[
\frac{m}{r} - \frac{C}{1 - r} = 0,
\]

where C = \sum_{i=1}^{K} in_i + nK - (K + 1)m.

Solving Equation 19 for r yields

\[
\hat{r} = m/(C + m)
\]

or

\[
\hat{r} = \frac{m}{\sum_{i=1}^{K} in_i + K(n - m)}. \tag{20}
\]

The second derivative of \log(L), given by the derivative of the left side of Equation 19, with respect to r is

\[
\frac{m}{r^2} - \frac{C}{(1 - r)^2},
\]

which is negative, indicating that \hat{r} represents the value which maximizes the log likelihood function, or, equivalently, the likelihood function L.

NOTES

1. The conceptual framework developed here largely draws on assumptions made about individual consumers. An argument may be made about the relevance of such work on modeling aggregate responses. However, the conceptual basis for most observed aggregate (macro)
phenomenon is at the disaggregate, individual (micro) level. Calder and Sternthal (1980) and Berlyne (1970) use aggregated response patterns in their analysis and presentation of results.

2. All these models, of course, assume that one can trace the response to a particular ad or promotion. We find much evidence in the industry in which, with careful planning, this can be accomplished. In the case of print ads, which require sending in a coupon or a postcard, the response vehicle is coded so that the response can be traced to a specific stimulus. People using the telephone to order from a catalog are often asked for the code number that is attached to the mailing label of the catalog for tracking purposes. It is also possible to provide different toll-free telephone numbers for different repetitions of a campaign.

REFERENCES


“USA Snapshots-Mass Mailing.” 1993. USA Today, August 19, 1A.


ABOUT THE AUTHORS

Richard J. Fox is an associate professor of marketing at the University of Georgia’s Terry College of Business. He received his Ph.D. in mathematical statistics from Michigan State University. He joined the Terry College of Business after a 15-year career in industry, including 10 years in consumer research at Procter & Gamble. His research interests include discrete choice models, market response to promotional stimuli, forecasting marketing potential from initial sales results, and, in general, applications of quantitative methods to business problems. He has published articles in such journals as the Annals of Mathematical Statistics, Annals of Statistics, Journal of the Academy of Marketing Science, Journal of Retailing, and the Journal of Advertising.

Srinivas K. Reddy is an associate professor of marketing at the University of Georgia’s Terry College of Business. He earned his Ph.D. at Columbia University and has served as assistant and associate professor of marketing at New York University, as visiting assistant professor at Columbia University, and as visiting associate professor at the University of California, Los Angeles. He is a member of the editorial review boards of
Bharat Rao is a research associate at the Harvard Business School. He received his Ph.D. in marketing from the University of Georgia. His current research interests include marketing strategy, strategic alliances, relationship marketing, and the implications of new technologies on business research and practice. He has published in various conference proceedings and the *International Business Review*.