THE BEST THINKING IN BUSINESS ANALYTICS FROM THE DECISION SCIENCES INSTITUTE

DECISION SCIENCES INSTITUTE

EDITED BY MERRILL WARKENTIN

The Best Thinking in Business Analytics from the Decision Sciences Institute

This page intentionally left blank

The Best Thinking in Business Analytics from the Decision Sciences Institute

Decision Sciences Institute
Edited by Merrill Warkentin

Publisher: Paul Boger
Editor-in-Chief: Amy Neidlinger
Executive Editor: Jeanne Glasser Levine
Operations Specialist: Kristen Watterson
Cover Designer: Alan Clements
Managing Editor: Kristy Hart
Project Editor: Elaine Wiley
Copy Editor: Kitty Wilson
Proofreader: Sarah Kearns
Indexer: Tim Wright
Compositor: Nonie Ratcliff
Manufacturing Buyer: Dan Uhrig

Upper Saddle River, New Jersey 07458

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact international@pearsoned.com.

Printed in the United States of America

First Printing September 2015

ISBN-10: 0-13-407295-2
ISBN-13: 978-0-13-407295-1

Pearson Education LTD.
Pearson Education Australia PTY, Limited.
Pearson Education Singapore, Pte. Ltd.
Pearson Education Asia, Ltd.
Pearson Education Canada, Ltd.
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan
Pearson Education Malaysia, Pte. Ltd.

Library of Congress Control Number: 2015942756

I dedicate this volume of research to all my current and former students, and especially to my doctoral students, who have filled my heart with pride and joy as I have watched them develop intellectually and grow to pursue their own academic dreams. Well done!

Contents

Foreword

What is analytics? It's everywhere and, depending on who you ask, it's everything. But taking a moment to stop and think operationally about what analytics construes to us as scholars and decision analysts is a useful step.

At the SAS Institute (www.SAS.com), analytics is envisioned as an interdisciplinary field combining mathematics, statistics, predictive modeling, and machine learning to identify meaningful patterns and develop knowledge from large collections of data. At Teradata (www.teradata.com), the belief is that the exponential growth in data stores drives the demand for methods to manage and parse large data stores to generate intelligence to inform strategic business decisions. The business dictionary (www.businessdictionary.com) suggests that the goal of analytics is to improve business by gaining knowledge that can be used to make improvements or changes.

At the Decision Sciences Institute, the sponsor of this fine book on the emerging field of analytics, we have always been interested in interdisciplinary approaches to the gathering and analyzing data in support of improving business decisions. The Decision Sciences Institute advances the science and practice of decision making, and in view of the recent emergence of vastly more powerful data storage and statistical analysis tools, the practice and science of decision making is informed by more sophisticated mathematical and computation tools and more extensive data stores and systems. This is the reason for books such as this: to explicate the current state of the art in business decision making as supported by such emergent techniques.

The Decision Sciences Institute is dedicated to excellence in fostering and disseminating knowledge pertinent to decision making. The *Decision Sciences Journal* is dedicated to the interdisciplinary investigation of leading-edge techniques in support of business decision making. As such, analytics is at the heart of what we do, is at the core of our scholarly mission, and is the focus of some of our most interesting recent research. Such work is chronicled in our *Journal* and in books such as this, and we hope that your interest in the increasingly data-intensive, computationally sophisticated nature of business decision making will be fueled by these publications.

Read on! And consider returning to the spaces and pages of the Decision Sciences Institute, as well as our books and *Journal*, to demonstrate your own discoveries in this increasingly important aspect of business decision making.

Tom Stafford, Editor
Decision Sciences Journal

Acknowledgments

First, I want to acknowledge the authors who have contributed to this volume of research on business data analytics in the decision sciences. Their research efforts are at the core of the Decision Science Institute's purpose and mission, and their hard work on preparing these manuscripts for this publication was essential and exemplary.

I also wish to thank Jeanne Glasser Levine, the Executive Editor at Pearson/FT Press, who collaborated with me to conceive of this book and worked with me to bring it to life.

I also wish to thank former DSI President Marc Schniederjans and current DSI President Morgan Swink, who have supported this volume of research articles.

I wish to thank Mississippi State University and the leadership of my department and college for providing me with an environment that encourages rigorous scholarship and intellectual excellence.

I want to thank my parents, Harold and Rosena Warkentin, who raised me to love knowledge, to work hard and with excellence, and to look for ways to make others' lives better. My father was my own school teacher as a boy, and I am always grateful for his immeasurable influence on my intellectual development throughout my childhood. Thank you, Dad.

Most importantly, I want to thank Kimberly, my college sweetheart, who has been my wife for over 30 years, whose steadfast support and love enable me to pursue my dreams. Thank you, Kim!

Merrill Warkentin, Volume Editor
Mississippi State University

About the Author

Merrill Warkentin, Volume Editor

Merrill Warkentin is Professor of MIS and the Drew Allen Endowed Fellow in the College of Business at Mississippi State University, where he is also a member of the research staff of the Center for Computer Security Research (CCSR) and the Distributed Analytics and Security Institute (DASI). He has published more than 250 manuscripts, including more than 55 peer-reviewed journal articles, plus several books. His work has been cited more than 8,400 times, and his H-index is 24, according to Google Scholar in 2015. He has been ranked among the top 100 IS scholars in the world based on rankings of authors publishing in the AIS basket of six leading MIS journals. His research, on the impacts of organizational, contextual, situational, and dispositional factors on individual user behaviors in the context of information security and privacy, addresses security policy compliance/violation, and social media use (and formerly also on electronic collaboration systems and e-commerce/e-government) has appeared in such journals as *MIS Quarterly*, *Decision Sciences*, *European Journal of Information Systems*, *Decision Support Systems*, *Information & Management*, *Information Systems Journal*, *Communications of the ACM*, *Communications of the AIS*, *The DATABASE for Advances in Information Systems*, *Computers & Security*, *Information Resources Management Journal*, *Journal of Organizational and End User Computing*, *Journal of Global Information Management*, and others. Professor Warkentin is also the author or editor of six books.

Dr. Warkentin is currently an Associate Editor (AE) of *MIS Quarterly*, *Information & Management*, *Information Resources Management Journal*, and *Journal of Information Systems Security*, and he has previously served as AE of *Decision Sciences*, *European Journal of Information Systems*, and other journals. He is the Eminent Area Editor for MIS for *Decision Sciences* and Senior Editor of *AIS Transactions on Replication Research*. He is Program Co-Chair for AMCIS2016 and has held leadership positions for numerous international IS conferences, including Track Chair for Security and Privacy at AMCIS2015 (Puerto Rico), ICIS2013 (Milan), ECIS2012 (Barcelona), and DSI2008; Program Chair for WISE2007 and WISP2009; Program Chair for the 2009 IFIP Workshop on IS Security Research; AE at ICIS four times (Security Track); Track Chair at DSI three times (Security Track in 2008); and Program Committee member of over a dozen international conferences (IFIP, WISP, WEB, WITS, ICEIS, etc.). Dr. Warkentin is the Chair of the UN-sponsored IFIP Working Group on Information Systems Security Research (WG8.11/11.13) and the AIS Security Coordinator. In 2014, he chaired the search committee to select the Editor of the *Decision Sciences Journal*. He has Guest Edited several journal special issues, including two issues of *EJIS*. He is AE for a special issue of *Information Systems Research* and a recent ad hoc SE for *MISQ*. He also currently serves on the board of the *Journal of Computer Information Systems* and the editorial advisory board of *Information Management & Computer Security*.

Dr. Warkentin has served as a consultant to numerous companies and organizations and has been a featured speaker at almost 200 industry association meetings, executive development seminars, and academic conferences. He has been a Lecturer at the Army Logistics Management College and was named a "National Distinguished Lecturer" by the Association for Computing Machinery (ACM). He has been a visiting professor or an invited speaker at more than 25 universities around the world, including Georgia State, Indiana, LSU, Florida State, Clemson, USF, Copenhagen Business School, McMaster, Fudan, Oulu, Jyväskylä, Zhejiang, Cape Town, and others. He has earned various recognitions for his teaching at every level, from intro courses to doctoral research seminars—his primary focus has been teaching Systems Analysis classes and Research Design seminars. His research has been funded by the UN, NSF, IBM, NSA, DoD, U.S. Navy, Homeland Security, and others. He was previously on the faculty at George Mason University and held the Reisman Research Professorship at Northeastern University in Boston, where he was also the Director of MIS and e-commerce programs at both the graduate and undergraduate levels. Professor Warkentin's Ph.D. in MIS is from the University of Nebraska–Lincoln. He can be reached at m.warkentin@msstate.edu.

The Decision Sciences Institute, Sponsor

The Decision Sciences Institute (DSI) is an independent nonprofit educational multidisciplinary professional organization of academicians and practitioners interested in the application of quantitative and behavioral approaches to all managerial decision making in business, government, and society.

Through national, international, and regional conferences; competitions; and publications, DSI provides an international forum for presenting and sharing research in the study of decision processes across disciplines. DSI also plays a vital role in the academic community by offering professional development activities and job placement services.

Five regional subdivisions in the United States, as well as regions representing Europe, Mexico, Asia-Pacific, and the Indian subcontinent, operate independently within DSI. Each region has its own elected officers and holds annual meetings.

DSI's members specialize in functional areas such as information systems, finance, marketing, management, accounting, manufacturing/service management, supply chain management, and decision support processes, as well as institutional areas such as healthcare, public administration, resource management, and higher education. They employ leading rigorous research techniques, including experimental designs, empirical quantitative analysis, optimization, simulation, surveys, and other scientific methods, while also valuing innovative methodological horizons.

DSI's goals are to:

1. Enrich the diverse disciplines of the decision sciences
2. Integrate these disciplines into bodies of knowledge that are effectively utilized for decision making
3. Develop theoretical bases for such fundamental processes as implementation, planning, and design of decision systems
4. Improve educational programs in the decision sciences

This page intentionally left blank

1

Predictive Modeling of Customer Response Behavior in Direct Marketing

—Young H. Chun, Louisiana State University

—Yoonhyuk Jung, Ulsan National Institute of Science and Technology, Korea

Abstract

Using the records of customers' responses over time in direct marketing, many authors have proposed various curve-fitting models to describe and predict the number of responses received after the launch of a direct marketing campaign. Some of those models are based on simplifying assumptions that are not realistic in many practical situations. In this paper, we first propose a probabilistic response model that has many desirable properties. Our geometric response model has three meaningful parameters: (1) an ultimate response rate of recipients, (2) a daily delay rate of respondents, and (3) a total delivery time of the request and responses. We then show that these parameters can be estimated by the maximum likelihood method. Finally, we test our response model by using mail survey data to show its superior performance. One of the advantages of our response model is attributed to the Poisson delivery time that adequately describes the delivery and processing time of customer responses.

Introduction

Direct marketing is a type of advertising campaign that allows businesses and nonprofit organizations to communicate directly to a selected group of consumers. The communication methods include postal mail, telemarketing, email marketing, cell phone text messaging, interactive consumer websites, fliers, catalog distribution, and promotional letters. Direct marketing is practiced by businesses of all sizes and types—from the smallest startup companies to the leading Fortune 500 companies.

A key factor in direct marketing is a "call to action." Each customer is asked to take a specific action, such as returning a questionnaire, placing a catalog order, mailing a prepaid postcard, calling a toll-free telephone number, clicking a link to a specific website, redeeming a discount coupon, or ordering a product online with a promotional code (Bose and Chen 2009). With a call to action, the customers' responses are directly traceable and easily measured by the direct marketer. Using the data of customer responses over time, we can predict the customer response rate and speed, and we can use that information in making important marketing decisions.

Suppose, for example, that a direct marketer mailed a catalog simultaneously to all customers in a target population. After the launch of a direct marketing campaign, the marketer has recorded the number of orders that have been placed each day. Based on the daily sales record, the marketer needs to estimate the total number of catalog items that will eventually be ordered. If the marketer underestimates the total demand, the catalog item in stock will run out, and the marketer may suffer the loss of customer good will or extra ordering and expedite shipping costs. On the other hand, overstocking the catalog item may result in higher inventory, maintenance, and salvage costs.

A similar prediction problem was evident when we mailed out a questionnaire to individuals in a target population and recorded the number of individuals who responded to the questionnaire each day. The same type of prediction problem is applicable with solicitation letters for fundraising, credit card applications, discount coupons in the Sunday newspaper, and email advertisements with promotional codes.

In this paper, we propose a geometric response model with three parameters to predict the customers' response patterns in a direct marketing campaign. One of the key parameters is a delivery time that describes the delivery time of a direct marketer's request and the delivery time of customers' responses. With the use of mail survey data, we demonstrate the superior performance of our response model over other conventional curve-fitting models.

The remainder of the paper is structured as follows. The following section is a brief review of various response methods that have been proposed in marketing literature. We then develop a geometric response model with three parameters and demonstrate how to estimate these parameters via the maximum likelihood method. We consider three types of probability distributions of the delivery time. We use the weekly response data collected by Huxley (1980) to demonstrate how to estimate the parameter values and compare three different delivery time models. Some concluding remarks are given in the last section.

Preliminaries

Suppose that a survey form, catalog, or solicitation letter is sent to N customers in the selected group, and their responses are recorded over time. Let $\mathbf{y}=\{y_1, y_2, ..., y_k\}$ denote the number of responses received during each of the past k days (or weeks) after the launch of the direct marketing campaign. For notational convenience, let $s_i=y_1+y_2+...+y_i$ be the total number of responses accumulated by the end of the ith day. The cumulative number of responses s_i is usually a monotonically increasing function of time i.

Many researchers have proposed various types of growth curves and considered different methods of estimating the model parameters. For example, Huxley (1980) made the first formal attempt to model the response pattern of a mail survey by using the following equation:

$$E[s_i] = N - \alpha\beta^i , \tag{1-1}$$

where α (>0) and β (<1) are unknown parameters to be estimated empirically and N is the number of questionnaires mailed initially. The growth curve of the response rate is similar to the cumulative distribution of an exponential probability distribution:

$$E[s_i] = N(1 - \alpha' e^{-\beta' i}) , \tag{1-2}$$

where α=α/N and β=−ln(β). After a log-transformation, the growth curve in (1-1) can be written as a simple linear regression model,

$$\ln[N - s_i] = \ln\alpha + i\ln\beta , \tag{1-3}$$

from which he found the least square estimators of α and β for given data.

Huxley (1980) mailed out N=4,314 questionnaires initially and recorded the cumulative number of questionnaires $\{s_1, s_2, ..., s_{17}\}$ received during the 17-week period. Note that, in his response model in (1), he implicitly assumed that s_i approaches N as i increases to infinity, which implies that all questionnaires will be returned eventually. When i=0, on the other hand, the cumulative number of responses s_0 has a nonzero value.

Since Huxley's pioneering work, numerous researchers have modified his original model or proposed alternative ones (e.g., Hill 1981; McGowan 1986; Bauer 1987, 1991; Wilson and Singer 1991; Basu, Basu, and Batra 1995; Pan 2010; Chun 2012). Most response models have two or three parameters, whereas McGowan (1986) proposed a logistics curve that has five unknown parameters that have no meaningful interpretations.

In general, the customer response models are classified into (1) the growth curve model and (2) the probabilistic response model. First, most of the earlier research focused on how to find the best growth curve that fits a given response data (Huxley 1980; Hill 1981; McGowan 1986; Bauer 1987, 1991). The method of least squares is usually used to estimate the parameter values. Second, in the probabilistic response model, the daily response of each respondent is modeled as a Bernoulli process so that the total responses in each day can be a random variable from a geometric distribution. In such a case, the model parameters are estimated by the maximum likelihood method (Wilson and Singer 1991; Chun 2012).

The need for an accurate response model is significant in direct marketing. Based on the customer response rate and speed, a direct marketer can adjust the marketing campaign, the message, or the target population to identify the most likely responders and improve the return on investment. Finn (1983) concluded that "more research into the nature of response functions in mail surveys is needed. If a consistently accurate predictive technique can be found, it will be invaluable to users of mail surveys."

In the following sections, we propose a new probabilistic response model that has many desirable properties. First, the cumulative number of responses is s_i=0 when i=0, and has an asymptote $s_i < N$ when $i=\infty$. Second, the response model is flexible enough to represent various types of response patterns with different shapes and locations. Third, the response model is parsimonious, with a smaller number of parameters. Fourth, each of the model's parameters has a meaningful interpretation. Few researchers have proposed response models that have all four of these desirable properties.

Delivery Time

In most practical situations, the number of daily responses y_i is initially increasing, reaching a peak, and then showing a longer tail dwindling over time, as shown in Figure 1.1(a). However, many researchers have assumed that the daily number of responses y_i is a monotonically decreasing function over time, as shown in Figure 1.1(b). They have also considered growth curves that look like a banana-shaped concave function. The growth curves do not fit very well, particularly in postal mail surveys, and Bauer (1991) proposed to arbitrarily exclude the first one or two days (or weeks) to get a better fit. Alternatively, other researchers have assumed that the frequency distribution of y_i is symmetrical, as shown in Figure 1.1(c), and have proposed S-shaped logistics or Gompertz curves (Fildes et al. 2008).

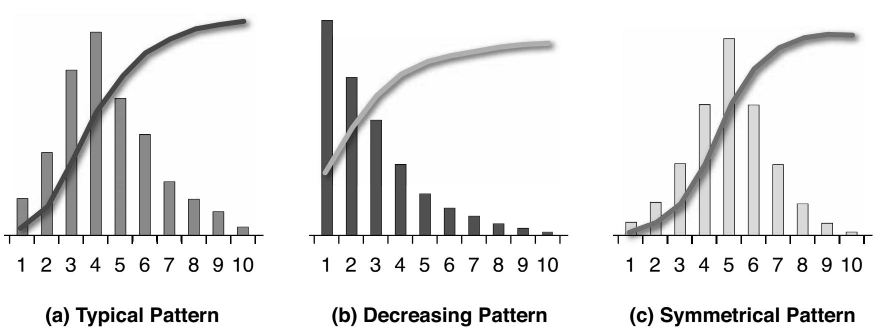

Figure 1.1 Frequency distribution of the number of daily responses over time.

Recently, Chun (2012) proposed a geometric response model with two meaningful parameters: (1) an ultimate response rate of the recipients and (2) a delay rate of respondents. His response model with the two parameters has many desirable properties but still has a limitation. The geometric response model is only appropriate for the cases in which the daily number of responses is geometrically decreasing in time, as shown in Figure 1.1(b). In this paper, we extend his model by adding a delivery time to effectively represent the typical *S*-shaped response pattern in Figure 1.1(a). If the delivery time is negligible, then the response pattern of our model is reduced to the banana-shaped concave function in Figure 1.1(b).

We can imagine many cases in which the processing and delivery time is non-negligible. For example, in postal mail surveys or catalog sales, it takes a longer time to deliver the request to a customer and receive his or her response. In such a case, the delivery time includes the time the postal service takes to deliver a questionnaire (or catalog) to the recipient, the time for a respondent to review and fill out the questionnaire, and the time it takes for responses to get back to the direct marketer.

The response model with a delivery time is called a "heterogeneous starting point" model in Basu, Basu, and Batra (1995), who assume that the delivery time is a uniform (a.k.a., rectangular) distribution. In addition to the uniform distribution, we consider two more probability distributions of delivery time and compare their performances. In the next section, we propose a geometric response model in which the delivery time is expressed in a general form. For a given set of response data, the three parameters in the model can be estimated via the method of maximum likelihood.

Customer Response Model

Suppose that we send out a request to N individuals simultaneously in a direct marketing campaign. Among the N individuals, the proportion of the "respondents" who will eventually respond to the request is π. We call π the "ultimate response rate," which is an unknown constant that should be estimated empirically.

Due to procrastination, even those respondents do not reply immediately. For each respondent, let p be the probability that he or she replies during a given day, and $q = 1-p$ denote the daily "delay rate" of a respondent. Thus, the number of Bernoulli trials for each respondent to react is a geometric distribution with a parameter q.

Chun (2012) considered the geometric response model with the two parameters, π and q, in which the expected number of daily responses is decreasing over time, as shown in Figure 1.1(b). Now, we assume that each reply will be delivered d days later ($0 \le d < \infty$), and the "delivery time" d is a discrete random variable. At the cost of introducing the additional variable d, we can represent various types of response patterns with different locations and shapes. Figure 1.2 illustrates the flowchart of responses during the first three days.

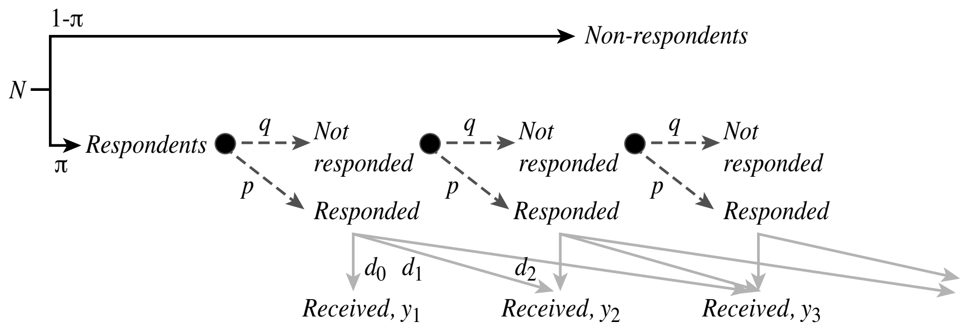

Figure 1.2 Flowchart of response patterns during the first three days.

For a respondent, let P_i be the probability that the reply of a respondent will be received i days after the launch of a direct marketing campaign. As shown in Figure 1.2, P_i does not depend on π, but it is a function of the unknown q and d. (Various types of functional forms of P_i will be considered in the next section.) The probability of receiving a series of responses, $\mathbf{y}=\{y_1, y_2, ..., y_k\}$, during the first k days can be described as a multinomial distribution with $(k+1)$ classes:

$$P[\mathbf{y} \mid \pi, q, d] = \frac{N!}{(N-s_k)! \prod_{i=1}^{k} y_k!} \left[1 - \pi \sum_{i=1}^{k} P_i\right]^{N-s_k} \prod_{i=1}^{k} (\pi P_i)^{y_i}. \tag{1-4}$$

from which we can find the expected values of y_i and s_i as follows:

$$E[y_i] = N\pi P_i \text{ and} \tag{1-5}$$

$$E[s_i] = N\pi \sum_{j=1}^{i} P_j \text{, for } i=1, 2, ..., k. \tag{1-6}$$

If we have the estimates of the parameters π, q, and d, we can predict the expected number of responses by a certain time and anticipate the time period needed to achieve a certain level of responses. Thus, our primary goal is to estimate π, q, and d empirically based on the sample observations $\mathbf{y}=\{y_1, y_2,..., y_k\}$.

Suppose that response data $\mathbf{y}=\{y_1, y_2, ..., y_k\}$ is available at time k. It follows from the multinomial distribution in (1-5) that the "likelihood function" of π is

$$L_{\mathbf{y}}(\pi) = \left[1-\pi\sum_{i=1}^{k} P_i\right]^{N-s_k} \prod_{i=1}^{k} (\pi P_i)^{y_i}. \tag{1-7}$$

The maximum likelihood estimator of π maximizes this likelihood function in (1-7). It is well known that the optimal value that maximizes the likelihood function $L_{\mathbf{y}}(\pi)$ also maximizes its log-likelihood function, $\ln L_{\mathbf{y}}(\pi)$. Therefore, it is more convenient to find the maximum likelihood estimator of π from the following log-likelihood function:

$$\ln L_{\mathbf{y}}(\pi) = (N - s_k)\ln\left[1-\pi\sum_{i=1}^{k} P_i\right] + s_k \ln \pi + \sum_{i=1}^{k} y_i \ln P_i . \tag{1-8}$$

If we take the first-order derivative with respect to π and set the equation equal to 0, we have

$$\frac{d}{d\pi}\ln L_{\mathbf{y}}(\pi) = -\frac{(N-s_k)\sum_{i=1}^{k} P_i}{1-\pi\sum_{i=1}^{k} P_i} + \frac{1}{\pi}s_k = 0 . \tag{1-9}$$

Solving this equation gives us the maximum likelihood estimator of the response rate π, as follows:

$$\hat{\pi} = \frac{s_k}{n}\left(\sum_{i=1}^{k} P_i\right)^{-1} . \tag{1-10}$$

If we plug $\hat{\pi}$ in (1-10) into the log-likelihood function in (1-8) and rearrange the expression, we have

$$\ln L_{\mathbf{y}}(q,d) \propto \sum_{i=1}^{k} y_i \ln P_i - s_k \ln \sum_{i=1}^{k} P_i , \tag{1-11}$$

where $\propto$ denotes "is proportional to."

The maximum likelihood estimates $\hat{q}$ and $\hat{d}$ are the ones that maximize this log-likelihood function in (1-11). Any optimization software, such as Microsoft Excel Solver, can be used to find the maximum likelihood estimates of q and d. With $\hat{q}$ and $\hat{d}$, we then find the maximum likelihood estimate of π from (1-10).

Note that P_i is a function of q and d, where the delay rate q is an unknown constant, and the delivery time d is a random variable. If a specific distribution of the delivery time d is given, then we can specify the probability P_i in the log-likelihood function in (1-11). In the next section, we consider three different types of probability distribution function of the delivery time d.

Delivery Time Models

The reply of a respondent is delivered i days after the launch of a direct marketing campaign due to the delay rate q and the delivery time d. Thus, in the geometric response model, the probability P_i that a respondent's reply will be received on day i is

$$P_i = \sum_{j=1}^{i} q^{i-j}(1-q)\, P[d_{j-1}] = \sum_{j=1}^{i} q^{j-1}(1-q)\, P[d_{i-j}], \tag{1-12}$$

where $P[d_j]$ is the probability mass function of the delivery time. In a special case in which the delivery is instant, the probability distribution becomes

$$P[d_j] = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \geq 1. \end{cases} \tag{1-13}$$

In such a case, we can simply have

$$P_i = q^{i-1}(1-q)\,. \tag{1-14}$$

Let us consider three different probability mass functions of d with a single parameter. First, suppose that the delivery time d has a discrete *uniform* distribution as in Basu, Basu, and Batra (1995);

$$P[d \mid u] = \frac{1}{u+1} \quad \text{where } d = 0, 1, 2, \ldots u, \tag{1-15}$$

where u is the upper limit of the uniform random variable. The delivery is instant if u=0. The expected value of the uniform delivery time is

$$E[d \mid u] = \frac{u}{2} \tag{1-16}$$

It follows from (1-12) and (1-15) that

$$P_i = \frac{1-q}{u+1} \sum_{j=\max\{1,\, i-u\}}^{i} q^{j-1}, \tag{1-17}$$

which can be simplified further as

$$P_i = \frac{1}{u+1} \min\left\{ (1-q^i), (q^{i-u-1} - q^i) \right\}. \tag{1-18}$$

Second, suppose that the delivery time d has a *geometric* distribution:

$$P[d \mid r] = r^d (1-r),\ d = 0, 1, 2, ..., \infty, \tag{1-19}$$

where r is a parameter, $0<r<1$, to be estimated empirically. If r is close to zero, then the delivery time is negligible. The expected value of the geometric random variable is

$$E[d \mid r] = \frac{r}{1-r}. \tag{1-20}$$

It follows from (1-12) and (1-19) that

$$P_i = (1-q)(1-r) \sum_{j=1}^{i} q^{j-1} r^{i-j}. \tag{1-21}$$

Third, suppose that the delivery time d has a *Poisson* distribution:

$$P[d \mid s] = \frac{s^d e^{-s}}{d!},\ d = 0, 1, 2, ..., \infty, \tag{1-22}$$

where s is a parameter, $s>0$, to be estimated empirically. The delivery time is negligible if s is close to 0. The average delivery time in (1-22) is

$$E[d \mid s] = s. \tag{1-23}$$

With the Poisson delivery time, it follows from (1-12) and (1-22) that

$$P_i = (1-q)e^{-s}\sum_{j=1}^{i} q^{i-j}\,\frac{s^{j-1}}{(j-1)!}\,. \tag{1-24}$$

Figure 1.3 illustrates the three probability distributions in which the average delivery time is $E[d] = 2$ days. Among the three distributions, the Poisson delivery in Figure 1.3(c) appears to be the most realistic in most practical situations.

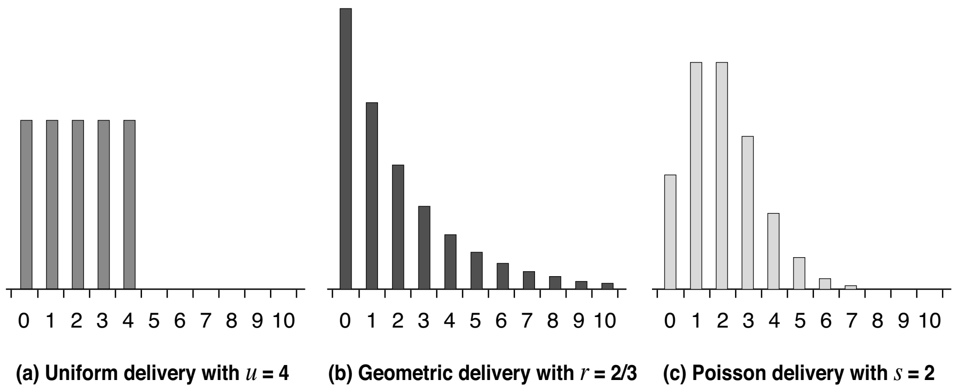

Figure 1.3 Various delivery time distributions with $E[d]$=2 days.

Note that we may consider other discrete probability distributions with more than one parameter. For example, the negative binomial distribution has been widely used in various consumer behavior models (Wagner and Taudes 1987) and in product inspection models (Chun and Sumichrast 2007). However, we restrict our attention to the single-parameter delivery time to have a parsimonious response model. Thus, our geometric response model has only three parameters: response rate, delay rate, and delivery time. All of the parameters have meaningful interpretations. In the next section, we compare the performance of the three delivery time models using weekly response data and propose the best one.

Numerical Example

To illustrate our response model with a delivery time, we use the response data collected by Huxley (1980) as a part of his dissertation research. He mailed out questionnaires to N=4,314 manufacturing firms, and he recorded the number of responses received by the end of each week during the 17-week period. Huxley's response data has been extensively used as a benchmark in subsequent studies by Hill (1981), Parasuraman (1982), McGowan (1986), Bauer (1991), and others.

Table 1.1 Various Delivery Time Models with Estimates of π, q, and d

Delivery Time Model	Response Rate, π	Delay Rate, q	Delivery Time, d	SSE	ML
No delivery time	0.5897	0.8836	—	428,948	–5959.4
Uniform distribution	0.5498	0.8365	u=2.000	129,894	–5628.6
Geometric distribution	0.5357	0.7414	r=0.742	135,515	–5700.5
Poisson distribution	0.5308	0.7746	s=2.163	91,876	–5578.3

Using Huxley's response data, we estimate the parameter values of our geometric response model. The results are given in Table 1.1. As a performance measure, we consider the sum of squared errors (SSE) of the cumulative number of responses s_i. The maximum value of the likelihood function in (1-11) is also considered as a performance measure.

Without the delivery time, the ultimate response rate is estimated as π=0.58972. The maximum likelihood estimate of the weekly delay rate is q=0.88355. The SSE of our geometric response model with an instant delivery time is 428,948, which is much better than the SSE=649,503 of Huxley's (1980) classical regression model in (1-3). If we include a delivery time, the geometric response model performs even better, as shown in Table 1.1.

Among the three probability distributions of the delivery time, the Poisson distribution appears to be the best, followed by the uniform distribution. The Poisson delivery time model has the smallest SSE and the largest value of the likelihood function. The superior performance of the Poisson distribution is anticipated from Figure 1.3, where the Poisson delivery time looks more realistic than the uniform or geometric delivery distribution. By changing the parameter value of the Poisson distribution, we can represent a wide variety of delivery time distribution with different shapes and locations. In practice, we strongly suggest using the geometric response model with the Poisson delivery time.

Figure 1.4 illustrates Huxley's (1980) original response data, along with the cumulative number of responses s_i predicted by our geometric response model with the Poisson delivery. The dotted curve in Figure 1.4 represents the predictions of Huxley's (1980) classical response model. As contrasted in the figure, our S-shaped response curve with a delivery time is clearly a better choice than Huxley's banana-shaped concave curve for the 17-week mail survey data.

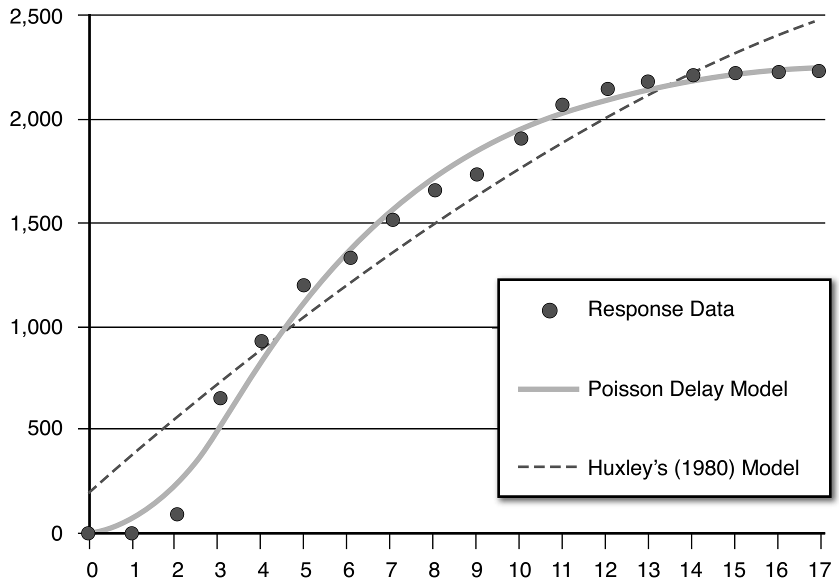

Figure 1.4 Actual and fitted values of the cumulative number of responses at time k=17.

Figure 1.5 displays the cumulative number of responses s_i, up to k=25, predicted by Huxley's model and by our geometric response model with the Poisson delivery. When the first k=10 week data is available, the Huxley's growth curve has a negative value at k=0, as shown in Figure 1.5(a), and it significantly overestimates the actual values from k=11 to 25. Furthermore, Huxley's model approaches N=4,314 as k approaches infinity. On the other hand, our geometric response model with a Poisson delivery slightly underestimates the actual values from k=11 to 17, but it fits much better than Huxley's response model.

The predicted values based on the first 15 weeks' worth of data are shown in Figure 1.5(b). The S-shaped growth curve of our geometric response model predicts the cumulative number of responses by the end of the 25th week much better than Huxley's banana-shaped concave curve.

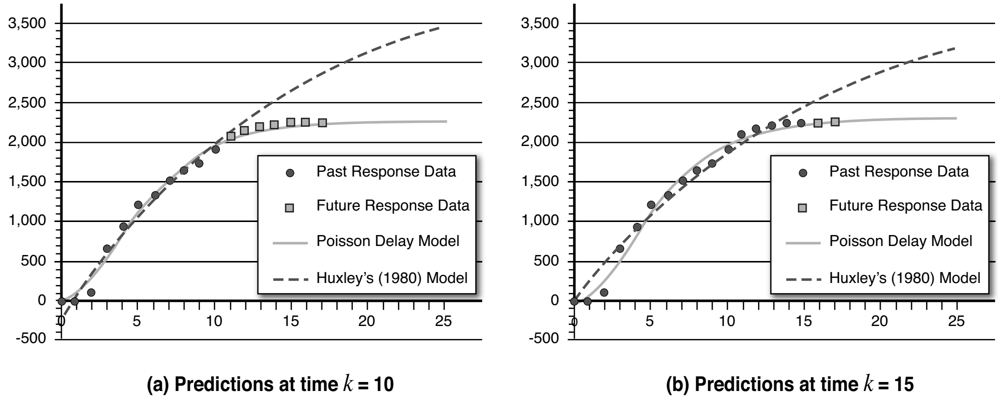

Figure 1.5 Predictions of the cumulative number of responses.

Concluding Remarks

One of the most important issues for direct marketers is how to sample targets from a population for a direct marketing campaign. Many authors have proposed various customer response models, in which the response variable is the probability of whether a customer with various characteristics will respond to a direct marketing campaign. Unlike those response models, the objective of this paper is to analyze the customers' response patterns and speed over time.

For observational data on the number of responses over time, an S-shaped sigmoid function can be used to describe and predict the growth pattern of customer responses (Freeland and Weinberg 1980). For example, McGowan (1986) proposed a logistics curve with five unknown parameters, which have no meaningful interpretations. In this paper, we proposed a probabilistic model with three parameters that can be interpreted as the ultimate response rate, daily delay rate, and total delivery time. Furthermore, we showed that the geometric response model with a Poisson delivery time has many desirable properties.

Our response model was fitted to Huxley's (1980) empirical data to show its superior performance over conventional models. However, Huxley's response data has the following anomalies: The first week is only two days, while other weeks each have five days. In addition, follow-up mails were sent in weeks 4 and 7. To compare the performance of our proposed response model with that of conventional models, we may need more empirical data or extensive simulation studies. In any case, we believe that our response model with the Poisson delivery is clearly an improvement over the traditional growth curve models.

Certainly, it is possible to construct richer and more complex response models with more model parameters. For example, we assume that the delay rate q is constant throughout the entire process, but it could be a function of time or could be changed by some form of follow-up or reminder mailings. Although we only considered a discrete-time case in this paper, our response model could be extended to a continuous-time case, in which each time period is not necessarily the same. This can be achieved by making appropriate modifications to our geometric response model with varying degrees of difficulty.

Another potentially fruitful area of research lies in a Bayesian response model that could incorporate our prior knowledge from similar direct marketing campaigns or expert opinions (Rossi and Allenby 2003). Unlike other conventional response models that only give point estimates of unknown parameters, the Bayesian model can construct confidence intervals of parameters and test various hypotheses under different loss functions. The geometric response model in this paper has three unknown parameters; however, the computational difficulties with the three prior distributions can be overcome with an appropriate Monte Carlo Markov chain method or a Gibbs sampler (Chun 2008).

With the increasing popularity of personal computers and the Internet, many researchers have analyzed the differences in shopping behavior of online customers (Van den Poel and Buckinx 2005). Thus, it would be interesting to compare the ultimate response rate, daily delay rate, and total delivery time between a traditional mail survey and a web-based survey (Cobanoglu, Warde, and Moreo 2001; Kwak and Radler 2002). We can also analyze the effects on the parameter values based on various response stimulants such as providing advance notice to respondents, utilizing different forms of postage, giving a variety of monetary and non-monetary premiums, and so on (Cobanoglu and Cobanoglu 2003).

Our response model can be applied to other areas as well. Meade and Islam (1998) reviewed various "diffusion models" for the spread of technological innovation or the penetration of a new product into the market. The response rate in a direct marketing campaign can be represented as a growth curve over time. Thus, it would be possible to use our geometric response model with a delivery time for diffusion models that describe the process of how new products get adopted over time (Tapiero 1983; Shore and Benson-Karhi 2007).

References

Basu, A. K., Basu, A., and Batra, R. (1995). Modeling the response pattern to direct marketing campaigns. *Journal of Marketing Research* 32, 204–212.

Bauer, C. L. (1987). Direct response advertising: Forecasting responses over time. *Journal of Direct Marketing* 1, 38–50.

Bauer, C. L. (1991). Logistic versus decaying exponential equations for describing mail survey response curves: A conceptual rationale and reanalysis. *Journal of Direct Marketing* 5, 15–26.

Bose, I., and Chen, X. (2009). Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research* 195, 1–16.

Chun, Y. H. (2008). Bayesian analysis of the sequential inspection plan via the Gibbs sampler. *Operations Research* 56, 235–246.

Chun, Y. H. (2012). Monte Carlo analysis of estimation methods for the prediction of customer response patterns in direct marketing. *European Journal of Operational Research* 217, 673–678.

Chun, Y. H., and Sumichrast, R. T. (2007). Bayesian inspection model with the negative binomial prior in the presence of inspection errors. *European Journal of Operational Research* 182, 1188–1202.

Cobanoglu, C., and Cobanoglu, N. (2003). The effect of incentives in web surveys: Application and ethical considerations. *International Journal of Market Research* 45, 475–488.

Cobanoglu, C., Warde, B., and Moreo, P. J. (2001). A comparison of mail, fax, and web-based survey methods. *International Journal of Market Research* 43, 441–452.

Fildes, R., Nikolopoulos, K., Crone, S. F., and Syntetos, A. A. (2008). Forecasting and operational research: A review. *Journal of the Operational Research Society* 59, 1150–1172.

Finn, D. W. (1983). Response speeds, functions, and predictability. *Journal of the Academy of Marketing Science* 11, 61–70.

Freeland, J. R., and Weinberg, C. B. (1980). S-shaped response functions: Implications for decision models. *Journal of the Operational Research Society* 31, 1001–1007.

Hill, R. W. (1981). Using S-shaped curves to predict response rates. *Journal of Marketing Research* 18, 240–242.

Huxley, S. J. (1980). Predicting response speed in mail surveys. *Journal of Marketing Research* 17, 63–68.

Kwak, N., and Radler, B. (2002). A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of Official Statistics* 18, 257–273.

McGowan, I. (1986). Fitting S-shaped curves to mail response data. *Journal of the Royal Statistical Society, Series D*, 35, 73–77.

Meade, N., and Islam, T. (1998). Technological forecasting—Model selection, model stability and combining models. *Management Science* 44, 1115–1130.

Pan, B. (2010). Online travel surveys and response patterns. *Journal of Travel Research* 49, 121–135.

Parasuraman, A. (1982). More on the prediction of mail survey response rates. *Journal of Marketing Research* 19, 261–268.

Rossi, P. E., and Allenby, G. (2003). Bayesian statistics and marketing. *Marketing Science* 22, 304–328.

Shore, H., and Benson-Karhi, D. (2007). Forecasting *S*-shaped diffusion processes via response modelling methodology. *Journal of the Operational Research Society* 58, 720–728.

Tapiero, C. S. (1983). Stochastic diffusion models with advertising and word-of-mouth effects. *European Journal of Operational Research* 12, 348–356.

Van den Poel, D., and Buckinx, W. (2005). Predicting online-purchasing behavior. *European Journal of Operational Research* 166, 557–575.

Wagner, U., and Taudes, A. (1987). Stochastic models of consumer behaviour. *European Journal of Operational Research* 29, 1–23.

Wilson, J. G., and Singer, E. M. (1991). Modelling survey response data and determining the optimal number of questionnaires. *Decision Sciences* 22, 898–910.

About the Authors

Young H. Chun is Professor of Decision Science and Cherie H. Flores Endowed Chair of MBA Studies at E. J. Ourso College, Louisiana State University. He received a Ph.D. in quantitative methods from Krannert Graduate School of Management, Purdue University, in 1990. His current research interests are in business analytics, quality control and reliability, statistical decision analysis, and warranty analysis. His work has been published in various journals, including *Operations Research*, *Decision Sciences*, *IIE Transactions*, *IEEE Transactions on Engineering Management*, *Naval Research Logistics*, *European Journal of Operational Research*, *Journal of the Operational Research Society*, *Journal of Quality Technology*, and *American Statistician*. Dr. Chun is on the Editorial Review Board of Production and Operations Management (POMS). Contact Dr. Chun at chun@lsu.edu.

Yoonhyuk Jung is Associate Professor at Ulsan National Institute of Science of Technology, South Korea. He received a Ph.D. in Management Information System from Louisiana State University. His main research interest is users' sense-making and adoption of emerging information technologies, with a special emphasis on social technologies, wireless technology applications, and health information systems. Contact Dr. Jung at yjung@unist.ac.kr.

Index

A

B

C

D

E

K

L

M

Q-R

S

T

U-V

W-X-Y-Z