European Journal of Operational Research 217 (2012) 673–678

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

Monte Carlo analysis of estimation methods for the prediction of customer response patterns in direct marketing

Young H. Chun

Department of Information Systems and Decision Sciences, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803, USA

ARTICLE INFO

Article history:
Received 20 October 2010
Accepted 11 October 2011
Available online 21 October 2011

Keywords:
Response pattern
Direct marketing
Probabilistic modeling

ABSTRACT

In direct marketing, customers are usually asked to take a specific action, and their responses are recorded over time and stored in a database. Based on the response data, we can estimate the number of customers who will ultimately respond, the number of responses anticipated to receive by a certain period of time, and the like. The goal of this article is to derive and propose several estimation methods and compare their performances in a Monte Carlo simulation. The response patterns can be described by a simple geometric function, which relates the number of responses to elapsed time. The "maximum likelihood" estimator appears to be the most effective method of estimating the parameters of this function. As we have more sample observations, the maximum likelihood estimates also converge to the true parameter values rapidly.

1. Introduction

"Direct marketing" is an interactive marketing system that uses various channels to target potential customers – one-on-one. It attempts to send its message directly to the customers without the use of *intervening* media. The most popular form of communication is direct mail, which is also referred to as "ad-mail" and may involve bulk mail. Other direct marketers also use e-mail marketing, telemarketing, broadcast faxing, door-hangers, and coupons.

In direct marketing, customers are usually asked to take a specific action which is quantifiable, such as calling a toll-free telephone number, clicking a link to a website, ordering a product online with a promotional code, redeeming a discount coupon, or returning a prepaid postcard. Due to this aspect of direct marketing, the customers' responses are tractable and measurable, and their activities are usually stored in a database. Using the customers' response records currently available, we may estimate the expected number of responses or the overall response rate, and use such information in making important managerial decisions.

Suppose, for example, that a store manager mailed out discount coupons for a digital camera as a promotional tool to n customers k weeks ago. Since then, the manager has recorded the number of coupons that have been redeemed in each week. Based on the weekly response records $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$ for the past k weeks, the manager wants to estimate the total number of coupons that will be redeemed ultimately or by a certain point in time. If the manager underestimates the total demand, the promotional items in stock could be run out and the store may suffer the loss of customer good will or extra ordering and shipping costs (Bijvank and Vis, 2011). On the other hand, over-stocking the promotional items in the first place may result in higher inventory, maintenance, and salvage costs.

Consider another example in which, as a part of its annual fund-raising campaign, a university sent out n donation request letters with prepaid post cards to its n alumni. Based on the number of daily or weekly replies $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$ for the past k time periods, the alumni office wants to estimate the number of alumni who will participate in the fund-raising campaign by a certain point in time. The admissions office at the same university is in the similar situation when they mailed out informational brochures with reply cards to n high school seniors. We have the same type of estimation problem with mail or online survey forms, discount coupons in Sunday newspaper, solicitation letters for credit card or mortgage applications, membership renewal requests, or e-mail advertisements with online coupons.

Recently, Bose and Chen (2009) reviewed various quantitative models for direct marketing, including the problem of estimating consumer response patterns, from a systems perspective. The problem of predicting the number of respondents can be also treated as a forecasting problem as in Fildes et al. (2008). The response rates in a specific field have been analyzed by various researchers – in operations management by Frohlich (2002) and in human resource management and organizational behavior by Roth and BeVier (1998). Various factors that could increase the response rate in mail survey were analyzed by Bruvold and Comer (1990).

In the paper, we (i) develop a geometric response model, (ii) consider various methods of estimating its parameter values, and then (iii) compare their performances in a Monte Carlo simulation. Amemiya (1976) previously compared the efficiencies of the maximum likelihood estimator, minimum Chi-square estimator,

E-mail address: chun@lsu.edu

0377-2217/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2011.10.008

and nonlinear weighted least squares method when the response variable is a dichotomous or polytomous variable. In the Monte Carlo simulation, we observe that the "maximum likelihood" method performs very well, displaying very accurate estimates with smaller variances. We also observe that the estimate of the response rate converges rapidly to the true parameter value as more response data becomes available.

In the next section, we formally define several terms, introduce the notation that will be used throughout the paper, and develop a geometric response model. In Section 3, we propose the maximum likelihood estimator of the customer response rate and discuss necessary conditions for its existence. We also consider additional estimation methods such as the minimum Chi-square method and the non-linear regression method in Section 4. In Section 5, we evaluate their performances in Monte Carlo simulations, followed by concluding remarks in Section 7.

2. Preliminaries

In direct marketing, there are many reasons why people in the target population choose not to respond. Many customers may see it as a nuisance. They find spam, junk mail, and telemarketing intrusive and irritating. In some cases, time is a factor. People may feel they cannot spare the time to participate in a survey or respond to a solicitation letter. They also do not respond to a discount coupon for a promotional item if the monetary incentive is not big enough to entice their participations or they simply do not need the item (Yin and Dubinsky, 2004).

Thus, the target population of n customers can be divided into two groups: Z "respondents" who will ultimately respond to the survey form, and $(n - Z)$ "non-respondents" who have no intention of responding to it at all. Let π be the "ultimate response rate," which is the proportion of the respondents in the target population who will ultimately respond.

Due to procrastination, however, not every respondent reacts immediately. For each respondent, let p be the "daily reaction rate," which is the probability that he or she reacts on a given day. Likewise, let $q = 1 - p$ be the "daily delay rate." As in most practical situations, we assume that the parameters π and p are unknown constants, which should be estimated.

Let $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$ be the daily response data during the past k days. For notational convenience, let $s_i = x_1 + x_2 + \cdots + x_i$ be the "cumulative" number of responses received by and including the ith day. By definition, we assume that $s_0 = 0$ and $s_\infty = z$. In some response models proposed in early days (e.g., Huxley, 1980), they assume that $s_\infty = n$ which I believe is not valid. The tree diagram in Fig. 1 illustrates the daily numbers of responses $\mathbf{x} = \{x_1, x_2, x_3\}$ over the three-day period. In most practical situations, the daily number of responses x_i is skewed to the right, showing a longer tail dwindling over time.

The primary goal in many managerial decision problems is to estimate the unknown parameter values π and p based on the sample observations $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$. Once we obtain the accurate estimates of π and p, we can find the point and interval estimates of the number of respondents Z, the total number of responses that would be received by a certain date, and the like. We may also be able to determine how many survey forms n should be mailed out initially in order to receive a specific number of responses by a certain date.

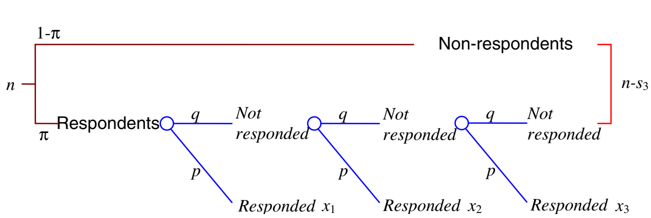

Fig. 1. Number of daily responses during the three-day period.

First, suppose that π and p are *known* constants. Then the cumulative number of responses S_k that will be received during the first k days is a binomial random variable:

$$P[S_k = i] = \frac{n!}{i!(n-i)!}\left[\pi(1-q^k)\right]^i\left[1-\pi(1-q^k)\right]^{n-i}. \tag{1}$$

The expected number of responses received during the first k days is

$$E[S_k] = n\pi(1-q^k), \tag{2}$$

and its variance is

$$Var[S_k] = n\pi(1-q^k)\left[1-\pi(1-q^k)\right]. \tag{3}$$

To find the confidence interval of S_k, we may use the binomial distribution in (1) or use a normal approximation if n is relatively large.

Second, since s_k is the cumulative number of responses we have received for the first k days, the number of future responses $Z - s_k$ is also a binomial distribution:

$$P[Z - s_k = i]$$
$$= \frac{n - s_k!}{i!(n - s_k - i)!}\left(\frac{\pi q^k}{1-\pi+\pi q^k}\right)^i\left(\frac{1-\pi}{1-\pi+\pi q^k}\right)^{n-s_k-i}. \tag{4}$$

When we have received s_k responses so far, the expected number of remaining responses is

$$E[Z - s_k] = \frac{(n-s_k)\pi q^k}{1-\pi(1-q^k)}, \tag{5}$$

and its variance is

$$Var[Z - s_k] = \frac{(n-s_k)\pi q^k(1-\pi)}{\left[1-\pi(1-q^k)\right]^2}. \tag{6}$$

With the expected value in (5) and the variance in (6), we can easily find the point and interval estimate of the number of remaining responses at time k. If $k = 0$, then the expected value and the variance of Z are reduced to $n\pi$ and $n\pi(1 - \pi)$, as expected from the binomial distribution in (4).

For many practical situations in which the population parameters π and p are *unknown* constants, we first propose the method of maximum likelihood in the next section.

3. Maximum likelihood estimator

In statistics, the most popular method of estimating the parameters of a statistical model is the maximum likelihood estimator. In the context of quality control, for example, Sim (1988) used the method of maximum likelihood to obtain asymptotically consistent, minimum variance estimates of the inspector's detection probability and the number of defective items still remaining in the lot.

3.1. Likelihood function

Since we assume that the ultimate response rate π is an unknown constant, the total number of respondents Z among n individuals is a binomial random variable with n and π:

$$P[Z = z] = \binom{n}{z}\pi^z(1-\pi)^{n-z}, \quad \text{for } z = 0, 1, 2, \ldots, n. \tag{7}$$

At the beginning of day i, the number of respondents who have not replied yet is $z - s_{i-1}$. Thus, the number of respondents who will reply on day i is expressed as a binomial distribution: