A Bernoulli–Gaussian mixture model of donation likelihood and monetary value: An application to alumni segmentation in a university setting

Pablo L. Durango-Cohen a, Elizabeth J. Durango-Cohen b,⇑, Ramón L. Torres c

a Civil & Environmental Engineering and Transportation Center, Northwestern University, 2145 Sheridan Road, Room A332, Evanston, IL 60208, USA
b Stuart School of Business, Illinois Institute of Technology, 565 W. Adams Street, Room 447, Chicago, IL 60616, USA
c Civil & Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

A R T I C L E   I N F O

Article history:
Received 15 December 2012
Received in revised form 30 May 2013
Accepted 9 August 2013
Available online 20 August 2013

Keywords:
Market segmentation
Finite mixture models
Alumni giving
Fundraising
Bernoulli–Gaussian Distribution
Expectation–Maximization Algorithm

A B S T R A C T

Advances in computational power and enterprise technology, e.g., Customer Relationship Management (CRM) software and data warehouses, allow many businesses to collect a wealth of information on large numbers of consumers. This includes information on past purchasing behavior, demographic characteristics, as well as how consumers interact with the organization, e.g., in events, on the web. The ability to mine such data sets is crucial to an organization’s ability to deliver better customer service, as well as manage its resource allocation decisions. To this end, we formulate a Bernoulli–Gaussian mixture model that jointly describes the likelihood and monetary value of repeat transactions. In addition to presenting the model, we derive an instance of the Expectation–Maximization Algorithm to estimate the associated parameters, and to segment the consumer population.

We apply the model to an extensive dataset of donations received at a private, Ph.D.-granting university in the Midwestern United States. We use the model to assess the effect of individual traits on their contribution likelihood and monetary value, discuss insights stemming from the results, and how the model can be used to support resource allocation decisions. For example, we find that participation in alumni-oriented activities, i.e., reunions or travel programs, is associated with increased donation likelihood and value, and that fraternity/sorority membership magnifies this effect. The presence/characterization of unobserved, cross-sectional heterogeneity in the data set, i.e., unobserved/unexplained systematic differences among individuals, is, perhaps, our most important finding. Finally, we argue that the proposed segmentation approach is more appealing than alternatives appearing in the literature that consider donation likelihood and monetary value separately. Among them and as a benchmark, we compare the proposed model to a segmentation that builds on a multivariate Normal mixture model, and conclude that the Bernoulli–Gaussian mixture model provides a more coherent approach to generate segments.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

As the world economy steers through a global recession, the “once-booming nonprofit sector [in the United States] is in the midst of a shakeout” (Banjo & Kalita, 2010), a consequence of government funding cuts and declines in donations from individuals. As a case in point, institutions of higher education experienced declines in contributions by individuals in 2009 and 2010 of 6.2% and 11.9%, respectively (Hall & Joslyn, 2011) – record drops in 50 years of recordkeeping. Overall impacts are discussed in studies such as Nicas and McWhirter, 2012, who argue that the severity of the cuts is adversely affecting educational quality, and even the solvency of institutions. This landscape has, in part, motivated universities to increase the volume and sophistication of their fundraising efforts. Central to university fundraisers’ efforts are two primary goals: meeting fundraising targets set by university administrators, and increasing participation rates in fundraising campaigns among alumni. Solicitation strategies are often intended to foster “loyalty/commitment” with younger alumni by encouraging nominal contributions to annual funds, e.g., $50/year. The importance of increasing participation is magnified by the role it plays in university ranking systems. US News & World Report, the most widely recognized college ranking service, for example, assigns a 5% weight to an “Alumni Giving” category, which the magazine
defines as the percentage of undergraduate alumni of record who donate money to the college or university in a given year. The percentage of alumni donors serves as a proxy for student satisfaction. At the same time, fundraisers seek to increase the monetary value of contributions as alumni advance professionally and financially. This goal is increasingly critical because a substantial portion of the funds raised by universities come from individual donors; in 2011, for instance, $13.45 billion (44%) of the $30.30 billion raised by colleges came from donations made by individuals (Council for Aid to Education, 2012). The development of effective solicitation strategies aimed at increasing the value and frequency of donations relies on understanding the contribution behavior of alumni, and how it evolves over time. In this paper, we present a model that advances this objective.

Specifically, we formulate a finite mixture model that jointly describes the likelihood and monetary value of donations in a university setting. The proposed model, a Bernoulli–Gaussian (BG) mixture model, relies on the assumption that the population is comprised of a finite set of latent classes/segments in unknown proportions. Each segment is characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in any given year. In addition to presenting the model, we derive an instance of the Expectation–Maximization (EM) Algorithm to estimate the associated parameters, and to assign donors to the segments in the population. For a given individual, segment membership is established based on the probabilities that her contribution sequence is consistent with the distribution functions that characterize the segments. In terms of validation, we use the model to analyze donations at a private, Ph.D.-granting university in the Midwestern United States. The dataset consists of 282,888 annual contributions from 75,922 individuals taking place between fiscal years 2000 and 2010. The results show that the proposed model adequately fits the contribution data. Further, we use the model to assess the effect of individual traits on their contribution likelihood and monetary value, discuss insights stemming from the results, and how the model can be used to support resource allocation decisions. The presence/characterization of unobserved, cross-sectional heterogeneity in the data set, i.e., unobserved/unexplained systematic differences among individuals, also constitutes one of the significant findings/results of our analysis. Finally, we argue that the proposed segmentation approach is more appealing than alternatives that consider donation likelihood and monetary value separately. Among them and as a benchmark, we compare the proposed model to a segmentation that builds on a multivariate Normal mixture model, and discusses the managerial implications of the results.

This paper contributes a new model that can be used to mine extensive panel data sets in situations where longitudinal data, i.e., transactions, are intermittent, and where there is significant cross-sectional (unobserved) heterogeneity among individuals. As such, the model has applicability to any setting where individuals are associated with large numbers of transactions, and where it is of interest to focus on/allocate resources based on a small number of segments instead of a large number of individuals. Advances in technology and computational power are, of course, making these settings increasingly common in research and in practice. Example application areas in Industrial Engineering include (dynamic) catalog mailing decisions (Simester, Sun, & Tsitsiklis, 2006), devising marketing strategies for (online) retailing (Ha, Bae, & Park, 2002; Jonker, Persina, & Van den Poel, 2004), monitoring usage rates of (subscription-based) services (Samimi & Aghaie, 2011), and others. The remainder of the paper is organized as follows: Section 2 positions our work with respect to the literature. In Section 3 we describe the data used in this study. In Section 4, we introduce notation and assumptions to formulate BG mixture models. We also present an instance of the EM Algorithm to estimate the model’s parameters. Results from an extensive empirical study, highlighting the insights gained by processing transaction sequences, and by segmenting the alumni population on the basis of donation likelihood and monetary value, are presented in Section 5. We conclude with a summary of the findings in Section 6.

2. Related work

Rather than providing comprehensive overviews of the fundraising or segmentation literatures, our objective is to contrast the approach implied by the proposed segmentation model to others appearing in the literature. In this case, the university fundraising context provides an interesting and relevant application – one that has been the subject of significant recent research. For broader reviews of fundraising and segmentation readers are referred to Bekkers and Wiepking, 2011a; Bekkers and Wiepking, 2011b; Wiepking and Bekkers, 2012 and Wedel and Kamakura, 2000.

To position the proposed model, we rely on the taxonomy of market segmentation models presented in Wedel and Kamakura, 2000. Among others, they rely on the following attributes: segmentation approach, segmentation bases, and type of statistical model used to explain responses. With respect to segmentation approach, models are classified as either a priori or post hoc. In a priori segmentation models, the number and types of segments are determined in advance of the analysis, whereas in post hoc segmentation they are determined as a result of the analysis based on goodness-of-fit or other criteria. The conditions used to assign individuals to segments are referred to as segmentation bases. Segmentation bases are either observed when they rely on observed/measured trait, response, or institutional variables; or unobserved, frequently attributed either to unobserved/missing data or to latent/unobservable (psychographic) variables. Finally, with respect to the capability to explain responses, models are classified as descriptive or predictive. Predictive models relate explanatory variables to the outcomes of a set of dependent variables. In contrast, descriptive models represent the (joint) distribution of the variables without distinction between outcome and explanatory variables (Green, 1977). Based on this taxonomy, the proposed model can be categorized as a post hoc segmentation procedure where the underlying statistical models are descriptive, and where the segmentation basis is unobserved. We note that this is consistent with the assumption that cross-sectional heterogeneity can be, at least partly, attributed to unobserved, and potentially unobservable factors.

The prevailing approaches in segmentation/econometric studies of (higher education) fundraising involve assigning individuals to a set of pre-established segments based on observed variables and criteria. Some examples in the literature include: undergraduate vs. graduate-degreed alumni segments, and segments of annual fund contributors vs. major donors who contribute at least $X per year or over a lifetime. These approaches are, therefore, a priori segmentation methods relying on observed segmentation bases. Segmentation bases generally consist of either trait variables (e.g.: demographic or socioeconomic characteristics), institutional variables (e.g.: size, private vs. public institutions, etc.), or response variables (e.g.: Recency, Frequency and Monetary Value of donations – RFM data). Numerous studies combine segmentation and econometric analyses, generally, to explain and reveal insights
about the underlying giving behavior of each segment. Segmentation bases consisting of trait or institutional variables lead to estimation of segment-level differential effects and predictions; whereas bases consisting of response variables lead to identification/inference of the factors that determine/influence or are correlated with the likelihood and monetary value of donations. Baade and Sundberg, 1996 is an example of a study where individuals in a large, multi-institution database of alumni donors are classified based on the type of institution they attended for their undergraduate studies: public university, private university, or liberal arts college. The authors then estimate differential effects of variables, including gender and age, on the monetary value of annual contributions for the three segments. Lindahl and Winship, 1992 is a seminal example of segmentation bases consisting of response variables. They estimate logit models for both major donors and annual fund contributors that include both trait and response factors as explanatory variables. In their analysis, past giving turns out to be the most important factor to predict the likelihood of future donations for both segments.

In terms of the underlying approach, the BG mixture model presented herein differs from others appearing in the literature in that:

- The model provides a joint description of the likelihood and monetary value of donations across a population. Most other studies formulate separate, (segment-level) econometric models, or focus on either variable (but not both). In addition to Lindahl and Winship, 1992, discussed above, Holmes, 2009, a recent and representative example, uses probit and tobit models, respectively, to predict donation likelihood and monetary value in any given year. The empirical study in Section 5 shows that the proposed model provides a rigorous and managerially-appealing basis for segmentation based on both variables.

- The model provides a description of the contribution sequences in the data set, and thus differs from studies/models that describe the distribution of aggregate, and possibly biased statistics, e.g., average annual contributions, derived from these sequences. Thus, the proposed model yields a segmentation that is sensitive to the (variation in the) contribution events/decisions and amounts, which we argue, in the context of the empirical study, are better indicators of donor behavior and potential.

- The assumption of a population comprised of latent classes, and the development of a post hoc segmentation methodology is appealing to describe populations with unobserved heterogeneity. With only a handful of exceptions, e.g., Lindahl and Winship, 1994; Weerts and Ronca, 2009; Le Blanc and Rucks, 2009, segmentation studies in the alumni giving/university fundraising literature rely on observed variables and subjective judgments to establish segmentation bases, e.g., monetary value thresholds to identify major donors. While intuitive and simple, it is clear that such approaches can introduce bias, particularly in situations where the cross-sectional heterogeneity is caused/influenced by possibly unobservable factors. We present evidence of unobserved heterogeneity in the data set used in the empirical study, and discuss how objective segmentation criteria improve within-segment homogeneity and between-segment heterogeneity, which can result in better resource allocation decisions. It is relevant to mention that instead of exploring heterogeneity at the individual level as is done in fixed or random effects specifications of predictive models (see, e.g., Clogg, 2003; Meer & Rosen, 2009; Gottfried, 2010; Meer, 2011), in the proposed model heterogeneity is captured at the segment level.

The use of aggregate statistics in segmentation models can, in part, be traced to computational and technical difficulties in the analysis of longitudinal data, i.e., dealing with unbalanced or intermittent sequences, accounting for serial dependence, etc. Wedel and Kamakura, 2000 review seminal studies, and elaborate on the difficulties. Notwithstanding, a small number of dynamic segmentation models have appeared recently in the segmentation and (higher education) fundraising literature. As stated above, these models focus on explaining the progression of either donation likelihood or monetary value (but not both). Fader, Hardie, and Shang, 2010 and Netzer, Lattin, and Srinivasan, 2008 are representative examples of models describing the likelihood of donations across populations. In Fader et al., 2010 contribution sequences are represented as binary strings: 1 in periods where donors are active, and 0 otherwise. The authors present a discrete-time, Beta-Geometric/Beta-Bernoulli model of the likelihood of donations. Netzer et al., 2008 segment a population of alumni donors based on their level of interaction with the institution. Further, they present Hidden Markov Models representing transitions between states of donor inactivity, occasional and high frequency activity. Van Diepen, Donkers, and Frances, 2009 is worthy of note among models that study how the monetary value of donations evolves (in the presence of competitive solicitations). They propose a random effects specification of a tobit model to capture unobserved heterogeneity. They also provide a recent and detailed overview of other similar dynamic models.

To the best of our knowledge, the proposed model is the first to represent the distribution of likelihood and monetary value of donations/transactions across a population. The distribution is constructed by assuming that, for each individual, donation events/decisions correspond to a set of independent Bernoulli trials, and that the monetary value is drawn from a Normal Distribution. Including the probability of a donation in a given period provides a simple and appealing explanation of gaps arising in longitudinal data. Further, we assume that the probability distribution functions describing donations are drawn from a finite set. That is, the donation sequences for individuals in a given segment follow a Bernoulli–Gaussian Distribution with parameters that correspond to the donation likelihood (in a given year), mean and standard deviation. Methodologically, the model presented in Section 4 extends the multivariate Normal mixture modeling framework, which is widely used to describe and support segmentation of heterogeneous populations in various disciplines. More broadly, our model follows in a long line of works estimating finite mixture models using the EM Algorithm, dating back to the seminal work of Dempster, Lair, and Rubin, 1977. McLachlan and Peel, 2000 and McLachlan and Thriyambakam, 2001 are seminal references that include numerous examples and cite hundreds of applications. Our work also extends the use of the BG Distribution, which is used in signal processing, i.e., approximate message passing (cf., Vila & Schniter, 2012 and the references therein), to model random errors.

3. Data

The data used in this study were provided by a major private university in the Midwestern United States. The dataset consists of the contribution sequences of 75,922 active donors, who contributed at least $10 in any year but less than $25,000 in any 3 consecutive years, over the period from fiscal year 2000 through 2010. Fiscal years begin September 1, i.e., fiscal year 2000 begins September 1, 1999. The dataset consists of 342,847 transactions, which we aggregate into 282,888 annual contributions.

Statistics describing the dataset appear in Table 1. We observe that, in each year, the mean annual contribution is larger than the contribution amount corresponding to the 75th-percentile, which indicates that the distributions display skewness toward smaller monetary values. Another important feature of the data
is the abundance of zeros stemming from intermittent alumni giving behavior, and that, in turn, motivates the proposed BG mixture model. Specifically, we note that of the individuals in the dataset, 32.5% contribute only once, 16.7% contribute twice, 11.4%, 8.3%, 6.3%, 5.1%, 4.3%, 3.8%, 3.6%, 3.7%, 4.3% contribute 3, 4, 5, 6, 7, 8, 9, 10, and 11 times, respectively. Overall, the dataset includes a total of 558,831 observations when $0 contributions are included, i.e., in periods where an individual does not make a contribution. We also note that, not surprisingly, approximately 20% of the donor population contribute 80% of the total amount raised.

In addition to the response data summarized above, the university provided trait data including age, zip-code and state of preferred address, class year(s), degree(s) earned, marital status, fraternity/sorority affiliation, participation in the alumni travel program, participation in alumni reunion events, and information about family members, i.e., spouse or children, who attended the university.

4. Model formulation and estimation

In this section, we present a BG mixture model to describe the contribution behavior of the alumni population, as well as to support segmentation based on the likelihood and monetary value of their donations. The relevant information is contained in the contribution sequences and thus, technically, we consider the problem of describing and classifying contribution sequences. First, we introduce notation and state assumptions to formulate BG mixture models, and explain how such models support segmentation. We then present an instance of the Expectation–Maximization (EM) algorithm to estimate the associated parameters.

Notation and assumptions to formulate a finite mixture model are as follows:

- The sequence of contributions for individual $m$ is denoted $y_m = \{y_m^1, y_m^2, \ldots, y_m^T\} = \{y_{m,t}\}_{t=1}^T$, where $y_m^t$ is the contribution of $m$ in period (year) $t$, $m = 1, \ldots, M$, $t = 1, \ldots, T$. $M$ represents the total number of individuals in the population, and $T$ corresponds to the total number of periods in the dataset. In terms of specifying the contribution sequences (for the data used in the empirical analysis), no entries exist in the periods preceding the first donation, and thus the sequences can be unbalanced. However, to simplify the exposition, we ignore the fact that each contribution sequence is specified as $y_m = \{y_{m,t}\}_{t=1}^T$, where $y_{m,t}$ corresponds to the period of individual $m$’s first contribution.
- We use $y$ to represent the set of sequences, i.e., $y = \{y_m\}_{m=1}^N$.
- We assume the population is comprised of $S$ segments in proportions $\lambda_1, \lambda_2, \ldots, \lambda_S$. $\lambda$ represents the set of proportions, i.e., $\lambda = \{\lambda_s\}_{s=1}^S$. Collectively, the mixture proportions correspond to the probability mass function describing an individual’s segment membership. That is, $\lambda_s$ corresponds to the (a priori) probability that a randomly selected individual from the population belongs to segment $s$. Mathematically, letting $z_m = 1$ if individual $m$ belongs to segment $s$, and $z_m = 0$ otherwise, and assuming every individual belongs to exactly one segment, i.e., $\sum_{s=1}^S z_m = 1$, $m = 1, 2, \ldots, M$, we have $P(z_m = 1 | s) = \lambda_s$, $s = 1, 2, \ldots, S$. Thus, $\lambda_s \geq 0$, $s = 1, \ldots, S$, and $\sum_{s=1}^S \lambda_s = 1$, i.e., $\lambda$ is in the $-S$-dimensional unit simplex, $\mathcal{S}$. Also, we define $z_m = \{z_{m,s}\}_{s=1}^S$.
- Each segment is characterized by a BG probability function, $f_s(y_m; \theta_s)$, representing the probability that an individual belonging to $s$ contributes sequence $y_m$. $\theta_s = \{q, \mu, \sigma\}$ represents the set of parameters that define the function $f_s(\cdot)$. More specifically, $q_s$ represents the probability of a contribution by an individual in segment $s$ in a given period, and $\mu$ and $\sigma$ are the parameters of the Gaussian distribution describing the monetary value of contributions of individuals in segment $s$. The $f_s(\cdot)$ functions are given as follows:

$$f_s(y_m; \theta_s) = \prod_{t=1}^T h_s(y_{m,t}|q_s, \mu, \sigma_s) \tag{1}$$

where

$$h_s(y_{m,t}|q_s, \mu, \sigma_s) = \begin{cases} q_s(y_{m,t}|\mu, \sigma_s) & \text{with probability } q_s \\ 0 & \text{otherwise} \end{cases}$$

From the above, it follows that the total probability of observing sequence $y_m$ is:

$$f(y_m; \theta) = \sum_{s=1}^S \lambda_s f_s(y_m; \theta_s) \tag{3}$$

where $\theta = \{\theta_s\}_{s=1}^S$. The total probability is a weighted sum, i.e., a mixture, of the probabilities associated with each segment. Eq. (3) is referred to as a finite mixture model.

In addition to describing the distribution of sequences across the population, the specification of a finite mixture model provides a segmentation framework based on updating each individual’s membership probabilities in response to her contribution sequence. That is, given $y_m$, Bayes Law can be applied to update the probability that individual $m$ belongs to segment $s$, $p_{ms}$, as follows:

$$p_{ms} = P(z_m = 1 | y_m) = \frac{\lambda_s f_s(y_m; \theta_s)}{\sum_{s=1}^S \lambda_s f_s(y_m; \theta_s)} \tag{4}$$

The segmentation is overlapping/stochastic in that (4) yields conditional membership probabilities, as opposed to deterministic assignments of individuals to segments.

4.1. Maximum (log)likelihood estimation problem

In the previous subsection, we used a finite mixture model. Eq. (3), to describe the distribution of contribution sequences. Here, we discuss the approach to estimate the associated parameters, i.e., the mixture proportions, $\lambda$, and the segment-specific parameters, $\theta$. As is commonly done in the estimation of finite mixture models, we present an implementation of the EM Algorithm. The data for the estimation problem consist of the set of contribution sequences for the population, $y$. Having specified (3), the data likelihood for $\lambda$ and $\theta$ is given by
$L(y; \lambda, \theta) = \prod_{m=1}^{M} p(y_m|\lambda, \theta)$

The objective is to find parameters, $\lambda, \theta$, that maximize (5). As applied in the estimation of finite mixture models, the EM Algorithm relies on the fact that if individual memberships, $z \equiv \{z_m\}_{m=1}$, were known, the ensuing estimation problem would be simplified. We can write the complete data likelihood for each instance were $K$ randomly generated initial parameters: proportions, donation algorithm was implemented in MATLAB, with results shown below.

Eq. (7). A detailed derivation of an EM Algorithm to obtain the complete data log-likelihood function, Eq.(6). The algorithm consists of two steps: An expectation step, E-Step, to evaluate the expectation of $\ln (L_c)$ over the possible realizations of $z$, given the observed data, $y$, and estimates of $\lambda$ and $\theta$; and a maximization step, M-Step, to update $\lambda$ and $\theta$ with arguments that maximize the expectation of $\ln (L_c)$. The EM Algorithm alternates between the E and M steps (until a convergence criterion is met). Mathematically, the EM Algorithm can be written as follows:

**E-Step:**

$$Q(\lambda, \theta) = \mathbb{E}_z[\ln (L_c(y; \lambda, \theta))]$$

$$= \mathbb{E}_z[\mathbb{E}_{z_m} \left( \ln \left( \sum_{s=1}^{S} z_m \ln (f_s(y_m|\theta_s)) \right) \right)]$$

**M-Step:**

$$\lambda, \theta = \arg \max_{\lambda, \theta} Q(\lambda, \theta) : \lambda \in \mathbb{R}; \theta \in \text{set of valid parameters of } f_s(y_m|\theta_s)$$

Once estimates, $\hat{\lambda}$ and $\hat{\theta}$, are obtained, the membership probabilities, $\hat{p}_{ms}$, $m = 1, \ldots, M; s = 1, \ldots, S$, are updated in order to evaluate Eq. (7). A detailed derivation of an EM Algorithm to obtain the parameters of the BG mixture model appears in Appendix A. The algorithm was implemented in MATLAB, with results shown below. For a given number of BG segments, we ran 150 instances with randomly generated initial parameters: proportions, donation likelihoods, means and standard deviations. The convergence criteria for each instance were $K = 100$ iterations or a tolerance $\epsilon = 5 \times 10^{-5}$. The results reported the subsequent section are for instances with the maximum log-likelihood.

Algorithm 1 (BG Mixture Models).

**Initialize**

$k = 0$

$\lambda^k, \hat{\theta}^k$

**While** $k < K$ and $Q(\lambda^{k+1}, \hat{\theta}^{k+1}) - Q(\lambda^k, \hat{\theta}^k) > \epsilon$

Given $y$

For $s = 1, \ldots, S$

$$\hat{p}_{ms} = \frac{\lambda^k \left(1 - \hat{q}_s^k\right) \prod_{t=1}^{T} \prod_{h=1}^{H(t)} [g_s(y_m|\mu_{sh}^k, \sigma_{sh}^k)]}{\sum_{s'=1}^{S} \lambda^{k_s} \left(1 - \hat{q}_{s'}^k\right) \prod_{t=1}^{T} \prod_{h=1}^{H(t)} [g_{s'}(y_m|\mu_{sh}^k, \sigma_{sh}^k)]}$$

where $\hat{z}_m$ is the number of zeros in the sequence $y_m$.

$$\lambda_{s+1} = \frac{1}{M} \sum_{m=1}^{M} \hat{p}_{ms}$

Updating Distribution Parameters:

For $s = 1, \ldots, S$,

$$\mu_{s+1} = \frac{1}{M \hat{z}_m} \sum_{m=1}^{M} \hat{p}_{ms} [T - \hat{z}_m] \sum_{t=1}^{T} y_{mt}$$

$$\sigma_{s+1}^2 = \sqrt{\frac{1}{M \hat{z}_m} \sum_{m=1}^{M} \hat{p}_{ms} [T - \hat{z}_m] \sum_{t=1}^{T} (y_{mt} - \mu_{s+1})^2}$$

5. Empirical study: Alumni segmentation based on donation likelihood and monetary value

5.1. Results: Parameter estimates

As other post hoc segmentation methods, the number of segments in finite-mixture models is established based on goodness-of-fit and other criteria, such as obtaining segments that represent sufficiently large proportions of the population, and/or obtaining segments that align with suggested contribution levels. In our analysis, we considered mixtures of $S = 1, \ldots, 12$ segments and, as shown in Fig. 1, evaluated them with the Consistent Akaike Information Criterion (CAIC). The CAIC is a widely-used statistic that trades off goodness-of-fit and overfitting. The CAIC is given as $-2 \cdot LL + (p \cdot S + S - 1) \ln (n + 1)$, where, respectively, $LL, p, S$ and $n$ represent the log-likelihood, number of parameters per segment, the number of segments, and the number of observations. For BG mixture models, $p = 3$, which corresponds to the cardinality of the set $\theta$. For a given dataset, $n$ is constant. In the dataset used herein, $n = 558,831$.

The parameter estimates for models with 7–12 segments appear in Table 2. The table also includes the scaled LLs, given by $LL/n$, and CAICs. Fig. 1 and the statistics presented in Table 2
suggest that there is little value in considering mixtures with more than 12 segments. The effect of \( S \) on CAIC is small relative to the effect of \( L \), which motivates the use of alternate criteria, including managerial appeal, for model selection. Our analysis is presented below.

Based on the parameter estimates presented in Table 2, we observe that when the number of segments is small, the dominant criterion characterizing the segments appears to be the mean annual value of contributions. Contribution likelihood becomes relevant when the number of segments increases. A comparison of the models with 8 and 9 segments illustrates this effect. In particular, consider the changes to the contribution likelihoods and monetary values of segments 6 and 7 in the 8-segment mixture model. Specifically, approximately 21% of the alumni, with similar monetary value distributions, but very different donation likelihoods. In particular, segments 1 and 2 with \( q_1 = 50\% \) versus \( q_2 = 73\% \), segments 3 and 4 with \( q_3 = 34\% \) and \( q_4 = 87\% \), segments 5, 6, and 7 with \( q_5 = 82\% \), \( q_6 = 40\% \) and \( q_7 = 82\% \), and segments 8 and 9 with \( q_8 = 89\% \) and \( q_9 = 37\% \). These groups may be interpreted as reflecting individuals with similar contribution potential, but different commitment levels. The differences between the groups, and with respect to the unmatched segments (10, and 11), are driven by monetary value.

5.2. Results: Segmentation

The results presented in the previous subsection document the estimation and selection of a finite mixture model to explain the contribution behavior across a population. Here, we shift focus to describe how individuals in the population are assigned to segments. For illustration purposes, we begin by considering the assignment of a small set of alumni donors. We then consider the assignment of individuals sharing traits to identify factors related to contribution behavior, and discuss how the results might lead to tailored marketing strategies.

Individual assignment relies on the (estimated) posterior membership probabilities, \( p_{ms} \). Recall that \( p_{ms} \) represents the conditional probability of observing the sequence of contributions by member \( m \), given the mixture proportion and parameter estimates, \( \lambda \) and \( \theta \). To illustrate the assignment process, Tables 3 and 4 show the contribution sequences and membership probabilities for a small sample of alumni.

The contribution sequences and posterior membership probabilities for individuals 1 and 2 highlight the relationship between data availability, i.e., length of the contribution sequences, and the confidence with which individuals are assigned to segments.
Individual 1’s contribution sequence is shortest among the members in the example. The ensuing membership probabilities reflect high uncertainty with a relatively small mode and positive probabilities for 8 of 11 segments. Qualitatively, this observation is opposite that for individual 2, who has the longest contribution sequence in the example. She is assigned to segment 4 with certainty despite appreciable within-sequence variation, and an average annual contribution about 1 standard deviation below the mean. The sequences and probabilities for individuals 3, 4 and 5 highlight an appealing feature of the BG model. In these cases, the average annual contributions of the three individuals are identical, $140. Their contribution frequencies are also similar. We observe, however, that the probabilities that each belongs to segments 5, 6, or 7, with $\mu_6 = 122.60$, $\mu_7 = 176.04$, and $\mu_8 = 204.15$, are vastly different. While individual 3’s contribution frequency appears to play a role in her assignment, it seems that, overall, the result is driven by the variation in the contribution amounts. The result highlights the fact that the correct interpretation of the membership probabilities is that they apply to the event that the entire sequence is generated by one of the probability functions, $f_i(\cdot)$.

The sequences and probabilities for individuals 3, 4 and 5 highlight an appealing feature of the BG model. In these cases, the average annual contributions of the three individuals are identical, $140. Their contribution frequencies are also similar. We observe, however, that the probabilities that each belongs to segments 5, 6, or 7, with $\mu_6 = 122.60$, $\mu_7 = 176.04$, and $\mu_8 = 204.15$, are vastly different. While individual 3’s contribution frequency appears to play a role in her assignment, it seems that, overall, the result is driven by the variation in the contribution amounts. The result highlights the fact that the correct interpretation of the membership probabilities is that they apply to the event that the entire sequence is generated by one of the probability functions, $f_i(\cdot)$.

The sequences and probabilities for individuals 3, 4 and 5 highlight an appealing feature of the BG model. In these cases, the average annual contributions of the three individuals are identical, $140. Their contribution frequencies are also similar. We observe, however, that the probabilities that each belongs to segments 5, 6, or 7, with $\mu_6 = 122.60$, $\mu_7 = 176.04$, and $\mu_8 = 204.15$, are vastly different. While individual 3’s contribution frequency appears to play a role in her assignment, it seems that, overall, the result is driven by the variation in the contribution amounts. The result highlights the fact that the correct interpretation of the membership probabilities is that they apply to the event that the entire sequence is generated by one of the probability functions, $f_i(\cdot)$.

The sequences and probabilities for individuals 3, 4 and 5 highlight an appealing feature of the BG model. In these cases, the average annual contributions of the three individuals are identical, $140. Their contribution frequencies are also similar. We observe, however, that the probabilities that each belongs to segments 5, 6, or 7, with $\mu_6 = 122.60$, $\mu_7 = 176.04$, and $\mu_8 = 204.15$, are vastly different. While individual 3’s contribution frequency appears to play a role in her assignment, it seems that, overall, the result is driven by the variation in the contribution amounts. The result highlights the fact that the correct interpretation of the membership probabilities is that they apply to the event that the entire sequence is generated by one of the probability functions, $f_i(\cdot)$.

The sequences and probabilities for individuals 3, 4 and 5 highlight an appealing feature of the BG model. In these cases, the average annual contributions of the three individuals are identical, $140. Their contribution frequencies are also similar. We observe, however, that the probabilities that each belongs to segments 5, 6, or 7, with $\mu_6 = 122.60$, $\mu_7 = 176.04$, and $\mu_8 = 204.15$, are vastly different. While individual 3’s contribution frequency appears to play a role in her assignment, it seems that, overall, the result is driven by the variation in the contribution amounts. The result highlights the fact that the correct interpretation of the membership probabilities is that they apply to the event that the entire sequence is generated by one of the probability functions, $f_i(\cdot)$.

The sequences and probabilities for individuals 3, 4 and 5 highlight an appealing feature of the BG model. In these cases, the average annual contributions of the three individuals are identical, $140. Their contribution frequencies are also similar. We observe, however, that the probabilities that each belongs to segments 5, 6, or 7, with $\mu_6 = 122.60$, $\mu_7 = 176.04$, and $\mu_8 = 204.15$, are vastly different. While individual 3’s contribution frequency appears to play a role in her assignment, it seems that, overall, the result is driven by the variation in the contribution amounts. The result highlights the fact that the correct interpretation of the membership probabilities is that they apply to the event that the entire sequence is generated by one of the probability functions, $f_i(\cdot)$.

The sequences and probabilities for individuals 3, 4 and 5 highlight an appealing feature of the BG model. In these cases, the average annual contributions of the three individuals are identical, $140. Their contribution frequencies are also similar. We observe, however, that the probabilities that each belongs to segments 5, 6, or 7, with $\mu_6 = 122.60$, $\mu_7 = 176.04$, and $\mu_8 = 204.15$, are vastly different. While individual 3’s contribution frequency appears to play a role in her assignment, it seems that, overall, the result is driven by the variation in the contribution amounts. The result highlights the fact that the correct interpretation of the membership probabilities is that they apply to the event that the entire sequence is generated by one of the probability functions, $f_i(\cdot)$.

The sequences and probabilities for individuals 3, 4 and 5 highlight an appealing feature of the BG model. In these cases, the average annual contributions of the three individuals are identical, $140. Their contribution frequencies are also similar. We observe, however, that the probabilities that each belongs to segments 5, 6, or 7, with $\mu_6 = 122.60$, $\mu_7 = 176.04$, and $\mu_8 = 204.15$, are vastly different. While individual 3’s contribution frequency appears to play a role in her assignment, it seems that, overall, the result is driven by the variation in the contribution amounts. The result highlights the fact that the correct interpretation of the membership probabilities is that they apply to the event that the entire sequence is generated by one of the probability functions, $f_i(\cdot)$.
Alumni with affiliation in fraternities/sororities (11% of the population, or 8398 individuals) are also more likely to be assigned to high-dollar value segments (22.9% are assigned to segments 8–11) than non-affiliated alumni (only 16.5% are assigned to segments 8–11).

Alumni who participated in travel organized by the alumni association (4% of the population, or 3828 individuals) are also more likely to make larger contributions (50.6% are assigned to segments 7–11 whereas only 34.2% of alumni who did not participate in alumni travel are assigned to these segments), and these alumni are also more likely (23.6% versus 11.3%) to belong to high-likelihood segments (segments 7, 8, and 10).

Building on the last point above, Table 5 shows the distribution of alumni across segments based on whether they have attended a reunion or participated in alumni travel, done both, or neither activity. The last row of the table shows the distribution of alumni who participated in a fraternity/sorority during their tenure at the university, and who have attended reunions or participated in alumni travel activities.

Engagement in alumni-centric activities affects the distribution of alumni across the segments in important ways. To highlight the effects, in Table 5, the labels for low likelihood segments are underlined. Participation in one or both activities shifts alumni from low dollar-value segments (segments 1–4). It also increases the proportion of alumni in all high-likelihood segments with moderate to high contribution levels (5, 7, 8, and 10). In addition, we note that while fraternity/sorority affiliation increases the proportion of alumni in higher dollar values (third bullet above), there appears to be a substantive combined effect tied to fraternities/sororities and participation in reunion or travel activities, as show in the last row Table 5—this level of engagement translates into a further reductions in the proportion of alumni in low dollar-value segments, and leads to a clear increase in the percentage of alumni into the highest value segments (segment 10 and segment 11).

The above descriptive results have prescriptive implications for fundraisers, both in terms of tailoring (alumni) fundraising strategies, but also in terms of developing programs that increase student engagement (in fraternities/sororities as undergraduates), and encourage participation, possibly through subsidizing, in alumni-centric activities. It is also important to mention that the results reveal significant unobserved heterogeneity in the population, i.e., the distribution of the subgroups considered is fairly disperse. Nevertheless, the BG model can be an important tool to tailor fundraising strategies; for example, by trying to engage alumni who display high potential, but low commitment.

5.3. Results: Comparison of BG mixture model to a multivariate Normal mixture model

In this section, we discuss the advantages of the proposed model vis-à-vis a segmentation approach that builds on a multivariate Normal mixture model of annual contributions. The comparison shows that the BG mixture model provides a more coherent approach to generate segments, which has the potential to lead to more effective and efficient marketing strategies. In addition, we find it instructive to compare the proposed model to a well-established benchmark, multivariate Normal mixture models. In terms of background, a rigorous treatment of the formulation and estimation of Normal mixture models can be found in McLachlan and Peel, 2000. The estimation problem consists of finding the population proportions, \( \hat{\pi} \), and the first two moments of the segment density functions. That is, the set \( \theta \) is given by \( \{ \mu, \sigma \} \). The parameter estimates and mixture proportions for Normal mixture models...
with five to eight segments, as well as the scaled LL and CAIC associated with each model are presented in Table 6.

As with the BG mixture model, we note that the scaled LL and CAIC improve at a decreasing rate as the number of segments increases. The parameter estimates suggest that the improvement arises from fitting a small percentage of donors at the higher contribution levels. In addition to these considerations, our selection of the 7-segment model for additional analysis was also related to a loose correspondence with the 11-segment BG mixture model. This observation, in some sense, mirrors the discussion at the end of Section 5.1 on the presence of groups of segments in the university setting. The proposed model relies on the assumption that the population is comprised of a finite set of latent segments/classes in unknown proportions, with each segment characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in a given year. In terms of messaging, how should appeals, targeted to specific segments, be crafted when alumni have such disparate contribution frequencies? In terms of effective use of mailing dollars, how can the university differentiate between alumni who are most likely to give (whether or not they receive a solicitation) from those who need to be prompted as a way to increase the likelihood that they will give in an efficient way?

The above discussion is intended to illustrate the difficulty of using a Normal mixture model as a starting point, which can be used to independently assess loyalty and generate a set of coherent segments that can be exploited via direct marketing. The BG mixture model, in contrast, does achieve this objective in an integrated, statistically-rigorous, and managerially-appealing fashion.

6. Summary and conclusions

We formulate a Bernoulli–Gaussian mixture model that jointly describes the likelihood and monetary value of donations in a university setting. The proposed model relies on the assumption that the population is comprised of a finite set of latent segments/classes in unknown proportions, with each segment characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in a given year. In addition to presenting the model, we derive an instance of the

with five to eight segments, as well as the scaled LL and CAIC associated with each model are presented in Table 6.

As with the BG mixture model, we note that the scaled LL and CAIC improve at a decreasing rate as the number of segments increases. The parameter estimates suggest that the improvement arises from fitting a small percentage of donors at the higher contribution levels. In addition to these considerations, our selection of the 7-segment model for additional analysis was also related to a loose correspondence with the 11-segment BG mixture model. This observation, in some sense, mirrors the discussion at the end of Section 5.1 on the presence of groups of segments in the university setting. The proposed model relies on the assumption that the population is comprised of a finite set of latent segments/classes in unknown proportions, with each segment characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in a given year. In terms of messaging, how should appeals, targeted to specific segments, be crafted when alumni have such disparate contribution frequencies? In terms of effective use of mailing dollars, how can the university differentiate between alumni who are most likely to give (whether or not they receive a solicitation) from those who need to be prompted as a way to increase the likelihood that they will give in an efficient way?

The above discussion is intended to illustrate the difficulty of using a Normal mixture model as a starting point, which can be used to independently assess loyalty and generate a set of coherent segments that can be exploited via direct marketing. The BG mixture model, in contrast, does achieve this objective in an integrated, statistically-rigorous, and managerially-appealing fashion.

6. Summary and conclusions

We formulate a Bernoulli–Gaussian mixture model that jointly describes the likelihood and monetary value of donations in a university setting. The proposed model relies on the assumption that the population is comprised of a finite set of latent segments/classes in unknown proportions, with each segment characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in a given year. In addition to presenting the model, we derive an instance of the

with five to eight segments, as well as the scaled LL and CAIC associated with each model are presented in Table 6.

As with the BG mixture model, we note that the scaled LL and CAIC improve at a decreasing rate as the number of segments increases. The parameter estimates suggest that the improvement arises from fitting a small percentage of donors at the higher contribution levels. In addition to these considerations, our selection of the 7-segment model for additional analysis was also related to a loose correspondence with the 11-segment BG mixture model. This observation, in some sense, mirrors the discussion at the end of Section 5.1 on the presence of groups of segments in the university setting. The proposed model relies on the assumption that the population is comprised of a finite set of latent segments/classes in unknown proportions, with each segment characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in a given year. In terms of messaging, how should appeals, targeted to specific segments, be crafted when alumni have such disparate contribution frequencies? In terms of effective use of mailing dollars, how can the university differentiate between alumni who are most likely to give (whether or not they receive a solicitation) from those who need to be prompted as a way to increase the likelihood that they will give in an efficient way?

The above discussion is intended to illustrate the difficulty of using a Normal mixture model as a starting point, which can be used to independently assess loyalty and generate a set of coherent segments that can be exploited via direct marketing. The BG mixture model, in contrast, does achieve this objective in an integrated, statistically-rigorous, and managerially-appealing fashion.

6. Summary and conclusions

We formulate a Bernoulli–Gaussian mixture model that jointly describes the likelihood and monetary value of donations in a university setting. The proposed model relies on the assumption that the population is comprised of a finite set of latent segments/classes in unknown proportions, with each segment characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in a given year. In addition to presenting the model, we derive an instance of the

Table 6
Parameter estimates: Normal mixture model

<table>
<thead>
<tr>
<th>Segment</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S = 5$ segments, $LL_{scaled} = -5.982$, CAIC = 3.385E+06</td>
<td>$\mu_A$</td>
<td>$\sigma_A$</td>
<td>$\lambda_A$</td>
<td>$q_A$ (%)</td>
<td>$LL_{scaled}$</td>
<td>CAIC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_A$</td>
<td>$\sigma_A$</td>
<td>$\lambda_A$</td>
<td>$q_A$ (%)</td>
<td>$LL_{scaled}$</td>
<td>CAIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_A$</td>
<td>$\sigma_A$</td>
<td>$\lambda_A$</td>
<td>$q_A$ (%)</td>
<td>$LL_{scaled}$</td>
<td>CAIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_A$</td>
<td>$\sigma_A$</td>
<td>$\lambda_A$</td>
<td>$q_A$ (%)</td>
<td>$LL_{scaled}$</td>
<td>CAIC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7
Parameter estimates for 7-segment multivariate normal mixture models; as well as percentage of alumni with different contribution frequencies, by segment.

| Segment | $\mu_A$, $\mu_B$, $\mu_C$, $\mu_D$, $\mu_E$, $\mu_F$, $\mu_G$ | $\sigma_A$, $\sigma_B$, $\sigma_C$, $\sigma_D$, $\sigma_E$, $\sigma_F$, $\sigma_G$ | $\lambda_A$, $\lambda_B$, $\lambda_C$, $\lambda_D$, $\lambda_E$, $\lambda_F$, $\lambda_G$ | Contribution frequency |
|---------|--|--|--|--|--|--|--|--|
| $[0-25\%)$ | $25-50\%)$ | $50-75\%)$ | $75-100\%)$ |
| $A$ | $31.99$ | $34.09$ | $75.98$ | $278.55$ | $601.37$ | $1277.30$ | $3137.50$ | $30.4\%$ | $28.6\%$ | $19.2\%$ | $11.6\%$ |
| $B$ | $78.61$ | $34.09$ | $75.98$ | $278.55$ | $601.37$ | $1277.30$ | $3137.50$ | $30.4\%$ | $28.6\%$ | $19.2\%$ | $11.6\%$ |
| $C$ | $277.55$ | $163.57$ | $149.13$ | $385.96$ | $877.48$ | $2960.30$ | $3484.00$ | $31.0\%$ | $31.1\%$ | $21.8\%$ | $12.5\%$ |
| $D$ | $601.37$ | $163.57$ | $149.13$ | $385.96$ | $877.48$ | $2960.30$ | $3484.00$ | $31.0\%$ | $31.1\%$ | $21.8\%$ | $12.5\%$ |
| $E$ | $31.99$ | $78.61$ | $149.13$ | $278.55$ | $601.37$ | $1277.30$ | $3137.50$ | $30.4\%$ | $28.6\%$ | $19.2\%$ | $11.6\%$ |
| $F$ | $1277.30$ | $3137.50$ | $31.99$ | $78.61$ | $149.13$ | $278.55$ | $601.37$ | $1277.30$ | $3137.50$ | $30.4\%$ | $28.6\%$ | $19.2\%$ | $11.6\%$ |
| $G$ | $3484.00$ | $3137.50$ | $1277.30$ | $601.37$ | $149.13$ | $278.55$ | $78.61$ | $149.13$ | $278.55$ | $601.37$ | $1277.30$ | $3137.50$ | $30.4\%$ | $28.6\%$ | $19.2\%$ | $11.6\%$ |

3 This observation, in some sense, mirrors the discussion at the end of Section 5.1 on the presence of groups of segments in the university setting. The proposed model relies on the assumption that the population is comprised of a finite set of latent segments/classes in unknown proportions, with each segment characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in a given year. In terms of messaging, how should appeals, targeted to specific segments, be crafted when alumni have such disparate contribution frequencies? In terms of effective use of mailing dollars, how can the university differentiate between alumni who are most likely to give (whether or not they receive a solicitation) from those who need to be prompted as a way to increase the likelihood that they will give in an efficient way?

The above discussion is intended to illustrate the difficulty of using a Normal mixture model as a starting point, which can be used to independently assess loyalty and generate a set of coherent segments that can be exploited via direct marketing. The BG mixture model, in contrast, does achieve this objective in an integrated, statistically-rigorous, and managerially-appealing fashion.

6. Summary and conclusions

We formulate a Bernoulli–Gaussian mixture model that jointly describes the likelihood and monetary value of donations in a university setting. The proposed model relies on the assumption that the population is comprised of a finite set of latent segments/classes in unknown proportions, with each segment characterized by a BG probability distribution function that explains both the likelihood and the distribution of donations in a given year. In addition to presenting the model, we derive an instance of the
Expectation–Maximization Algorithm to estimate the associated parameters, and to segment the donor population.

From a marketing perspective, the model provides an appealing and rigorous framework to describe transaction sequences, i.e., panel data, and to segment individuals based on the likelihood and distribution of donations. Segmentation/econometric models in the literature typically consider either dimension, but not both. Also, models in the literature tend to rely on aggregate statistics to describe transaction sequences. We show that such an approach can introduce bias, but more importantly, that the variation within the transaction sequences has behavioral implications (that can be exploited for fundraising). From a methodological perspective, the BG mixture model extends multivariate Normal mixture models, widely used in segmentation/cluster analysis to describe heterogeneous populations with latent classes in engineering, social and physical sciences, as well as the use of Bernoulli–Gaussian probability distribution functions, which have appeared in signal processing applications.

In terms of segmentation, we observe that approximately two-thirds of the donors are assigned specific segments with high confidence (with probability of 75% or greater). This is both appealing to fundraisers, and provides a basis to support tailored marketing strategies aimed at increasing donation frequency and value. The other third of the population can be approached with ‘exploratory’ solicitations partially aimed at refining the confidence with which individuals are assigned to segments. Our analysis of the factors that influence contribution behavior reveals that graduate and professional alumni are more likely to donate larger amounts; though, the commitment levels across the population are similar. The latter differs from other studies that suggest that undergraduate alumni tend to be more committed/loyal donors. In terms of activities, our analysis shows that alumni who attended class reunions or participated in university travel programs are significantly over-represented in segments with high donation likelihoods and monetary value. We also noticed a significant combined effect of fraternity/sorority membership and participation in the aforementioned activities. This reinforces the importance of developing programs that increase student engagement/satisfaction, e.g., membership social, academic, professional, athletic clubs/events, and encourage participation, possibly by subsidizing, in alumni-oriented activities. The result also suggests that social interaction may influence donations. Perhaps most importantly, the results reveal significant unobserved heterogeneity in the population, i.e., the distribution of the subgroups considered is fairly disperse. The BG model captures such heterogeneity and, therefore, can be an important tool to tailor fundraising strategies; for example, by trying to engage alumni who display high potential, but low commitment. Finally, we argue that the proposed segmentation approach is more appealing than alternatives appearing in the literature that consider donation likelihood and value separately. Among them and as a benchmark, we compare the proposed model to a segmentation that builds on a multivariate Normal mixture model, and conclude that the Bernoulli–Gaussian mixture model provides a more coherent approach to generate segments.

The results presented here provide actionable information that can be used by alumni development offices to better focus their efforts (on alumni who are more likely to donate, and/or those who donate more generously), and improve the efficiency of resource allocation (e.g., telemarketing dollars). Some caution is needed, however, in interpreting the results due to some limitations stemming from available data. In particular, the dataset provided by the university excluded major donors, and alumni who did not donate over the analysis period. This, in turn, introduces biases (in the interpretation of) the results because the individuals who were included, annual fund contributors, may not be representative of the overall alumni population. As such, results and insights need to be qualified accordingly. It is also relevant to mention that the transactions in the dataset constitute the responses to the university’s fundraising strategy. As such, interpretation of the results cannot be attributed exclusively to donor behavior. For example, it is possible that variation in the transaction sequences can be (at least partly) attributed to (changes in) the university’s solicitation strategy. To conclude, we note that the BG mixture model is formulated under the assumption that transactions are serially-independent (and identically distributed). Because many of the results and insights are related to the variation within the transaction sequences, extending the BG mixture model to account for systematic variation, i.e., to represent serial dependence, appears to be an interesting direction for additional research.

Acknowledgement

The research was partially supported by grants from the National Science Foundation

Appendix A. Derivation of the EM Algorithm for BG mixture models

In this section, we present the derivation of the instance of the EM Algorithm appearing in Section 4.1. We begin by establishing two intermediate results. First, we observe that for estimates \( \hat{\theta}, \theta \), and a contribution sequence, \( y_m = \{x_m , y_m, \hat{\theta}, \theta \} \), which may be rewritten as

\[
\mathbb{P}(y_m | \hat{\theta}) = \mathbb{P}(y_m | \theta) = \prod_{t=1}^{n} \mathbb{P}(y_t | \hat{\theta}, \theta) = \prod_{t=1}^{n} \mathbb{P}(y_t | \hat{\theta}, \theta) = \prod_{t=1}^{n} \mathbb{P}(y_t | \hat{\theta}, \theta)
\]

(A.1)

where \( z_m \) is the number of zeros in the sequence \( y_m \).

The second result is obtained by evaluating \( \ln(g(y_m | \hat{\theta}, \theta)) \), recalling that, in the BG Distribution, \( g(\cdot) \) corresponds to the Normal Density Function. That is, \( g(y_m | \hat{\theta}, \theta) = \frac{1}{\sqrt{2\pi} \sigma_{\theta}} e^{-\frac{(y_m - \theta)^2}{2\sigma_{\theta}^2}} \), and thus,

\[
\ln(g(y_m | \hat{\theta}, \theta)) = -\frac{1}{2} \left( \frac{y_m - \theta}{\sigma_{\theta}} \right)^2
\]

(A.2)

Now, we evaluate (7) by considering the expectation (with respect to \( z \)) of the complete data log-likelihood function, Eq. (6):

\[
\mathbb{E}(\ln(y_m | \hat{\theta}, \theta)) = \mathbb{E}_x(z, \cdots, z) \left[ \sum_{m=1}^{M} \sum_{i=1}^{T} \mathbb{E}_y(z_m, \cdots, z_m) \ln(Q_i(y_m | \hat{\theta}, \theta)) + \ln(z_m) \right]
\]

where \( \mathbb{E} \) is the expectation operator.

\[
= \sum_{m=1}^{M} \sum_{i=1}^{T} \mathbb{E}_x(z_m, \cdots, z_m) \left[ \sum_{i=1}^{T} \mathbb{E}_y(z_m, \cdots, z_m) \ln(Q_i(y_m | \hat{\theta}, \theta)) + \ln(z_m) \right]
\]

which may be rewritten as

\[
= \sum_{m=1}^{M} \sum_{i=1}^{T} \mathbb{E}_x(z_m, \cdots, z_m) \left[ \sum_{i=1}^{T} \mathbb{E}_y(z_m, \cdots, z_m) \ln(Q_i(y_m | \hat{\theta}, \theta)) + \ln(z_m) \right]
\]

(A.3)

where \( z_m \) is the number of zeros in the sequence \( y_m \).

(continued...)

Following (8), the optimization problem to update the parameter estimates \( \hat{\lambda}, \hat{\theta} \) is as follows:

\[
\max_{\hat{\lambda}, \hat{\theta}} Q(\hat{\lambda}, \hat{\theta}) \quad (A.3)
\]

Subject to:

\[
\sum_{i=1}^{s} \hat{\lambda}_i = 1 \quad (A.4)
\]

\[
\tilde{q}_i \leq 1, \ s = 1, \ldots, S \quad (A.5)
\]

\[
\hat{\lambda}_i, \tilde{q}_i \geq 0, \ s = 1, \ldots, S \quad (A.6)
\]

Under the assumption that (A.4) is the only binding constraint, we may write the Lagrangean of the problem as follows:

\[
L(\hat{\lambda}, \hat{\theta}) = Q(\hat{\lambda}, \hat{\theta}) - v \left[ \sum_{i=1}^{s} \hat{\lambda}_i - 1 \right]
\]

The first-order optimality conditions are:

\[
\frac{\partial L(\cdot)}{\partial \hat{\lambda}_i} \frac{1}{\hat{\lambda}_i} \sum_{m=1}^{M} \tilde{p}_m - v = 0, \ s = 1, \ldots, S
\]

\[
\frac{\partial L(\cdot)}{\partial p} = 1 - \sum_{i=1}^{s} \hat{\lambda}_i = 0
\]

\[
\frac{\partial L(\cdot)}{\partial \mu_s} = \sum_{m=1}^{M} \tilde{p}_m \left( \sum_{t=1}^{T} \tilde{y}_m^t - \hat{\mu}_s \right) = 0, \ s = 1, \ldots, S
\]

\[
\frac{\partial L(\cdot)}{\partial \sigma_i^2} = \sum_{m=1}^{M} \left( \sum_{t=1}^{T} (\tilde{y}_m^t - \hat{\mu}_s)^2 - \sum_{t=1}^{T} \sigma_i^2 \right) = 0, \ s = 1, \ldots, S
\]

\[
\frac{\partial L(\cdot)}{\partial \tilde{q}_i} = \sum_{m=1}^{M} \left[ \frac{-\tilde{c}_m}{1 - \tilde{q}_i} + \sum_{t=1}^{T} \frac{1}{\tilde{q}_i} \right] = 0, \ s = 1, \ldots, S
\]

Now we find a set of values \( \hat{\lambda}, \hat{\theta}, \) and \( v \) that satisfy conditions (A.7)–(A.11). From conditions (A.7) and (A.8), we have

\[
\hat{\lambda}_s = \frac{1}{v} \sum_{m=1}^{M} \tilde{p}_m, \ s = 1, \ldots, S
\]

\[
\sum_{i=1}^{s} \hat{\lambda}_i = \frac{1}{v} \sum_{m=1}^{M} \tilde{p}_m = 1
\]

\[
\Rightarrow v = M
\]

\[
\hat{\lambda}_s = \frac{1}{M} \sum_{m=1}^{M} \tilde{p}_m, \ s = 1, \ldots, S
\]

From conditions (A.9) and (A.10) we find

\[
\mu_s = \frac{1}{M} \sum_{m=1}^{M} \tilde{p}_m \left( \frac{1}{T - \tilde{c}_m} \right) \sum_{t=1}^{T} \tilde{y}_m^t
\]

\[
\sigma_i^2 = \frac{1}{M} \sum_{m=1}^{M} \tilde{p}_m \left( \frac{1}{T - \tilde{c}_m} \right) \sum_{t=1}^{T} (\tilde{y}_m^t - \hat{\mu}_s)^2
\]

Finally, from (A.11)

\[
\tilde{q}_i = \frac{1}{M} \sum_{m=1}^{M} \tilde{p}_m \left( 1 - \frac{\tilde{c}_m}{T} \right)
\]

References


Council for Aid to Education (2012). Voluntary support of education report.


