J. R. Statist. Soc. A (2006)
169, *Part* 2, *pp.* 361–379

Analysis of Irish third-level college applications data

Isobel Claire Gormley and Thomas Brendan Murphy

Trinity College Dublin, Republic of Ireland

[Received March 2005. Revised October 2005]

Summary. The Irish college admissions system involves prospective students listing up to 10 courses in order of preference on their application. Places in third-level educational institutions are subsequently offered to the applicants on the basis of both their preferences and their final second-level examination results. The college applications system is a large area of public debate in Ireland. Detractors suggest that the process creates artificial demand for 'high profile' courses, causing applicants to ignore their vocational callings. Supporters argue that the system is impartial and transparent. The Irish college degree applications data from the year 2000 are analysed by using mixture models based on ranked data models to investigate the types of application behaviour that are exhibited by college applicants. The results of this analysis show that applicants form groups according to both the discipline and the geographical location of their course choices. In addition, there is evidence of the suggested 'points race' for high profile courses. Finally, gender emerges as an influential factor when studying course choice behaviour.

Keywords: College course choice; EM algorithm; Higher education; Mixture models; Ranked data

1. Introduction

The Irish college applications system involves prospective college students ranking up to 10 degree courses in order of preference before they sit their final second-level examinations (Leaving Certificate). Applications are processed by the Central Applications Office (CAO), who deal with applications for all third-level degree programmes in Ireland.

Typically, seven or eight subjects are taken for the Leaving Certificate examination. Once graded the best six examination results are used to produce a 'points' score; each grade A1, A2, B1, ..., F has an associated number of points.

Subsequently to examination grading, the CAO fixes a universal points requirement for each degree programme. The points requirements are set so that applicants are offered a place in the highest preference course for which they have achieved the points requirement; in the case of applicants being tied for the last positions in a course, random allocation is used to choose which applicant is offered a place.

It is worth emphasizing that applicants do not know the course points requirement before they complete their application or take their examinations. The points requirement is influenced by the examination results of applicants who applied for the course and by the number of available positions on the course.

Some courses have minimum entry standards; for example, a sufficient standard of mathematics may be required for an engineering degree. However, the actual subjects that are taken at Leaving Certificate level do not have an effect on the applicant's points score; nor do interviews

Address for correspondence: Isobel Claire Gormley, Department of Statistics, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Republic of Ireland.
E-mail: gormleyi@tcd.ie

 0964–1998/06/169361

or previous examination performance. A few courses have interviews, but these are not common. The subjects Irish, English and mathematics taken at Leaving Certificate level are entry requirements for Irish applicants for many courses but the remaining subjects are the student's choice. In addition, the Leaving Certificate can be taken several times without having any effect on an application.

International applicants are dealt with in a similar manner. For example, the UK A-level results are converted into points—these are totalled and subsequently such applicants are allocated to a course by the same method as Irish Leaving Certificate students.

A similar college applications system is used in Australia where applicants rank up to nine courses which are processed by the Universities Admission Centre. Both the Irish CAO and the Australian Universities Admission Centre systems can be likened to the 'Tote' in horse-racing betting. In the Tote, no punter knows the odds on any horse before all bets have been placed. Similarly under the CAO system no applicant knows the points requirement for any course before the publication of all examination results. Extensive details of the college applications system are available on the CAO Web page (`http://www.cao.ie`).

The method of gaining entry to third-level education, as managed by the CAO, is a much debated subject among the Irish media, students, parents and education circles. Many aspects of the CAO system appear annually as headlines in the Irish media—national front pages carry stories of fluctuating points requirements and volatile applicant numbers, particularly for the weeks surrounding the announcement of who is admitted to each course.

Detractors suggest that applicants are influenced by the annual media excitement and rank courses according to points requirements, ensuring that they study a current 'high profile' course and therefore they may ignore their vocational callings. They claim that artificial demand is created for courses that are deemed to be of high social standing. Supporters insist that it is a fair system where each applicant is dealt with in a consistent and transparent manner. The supporters claim that the so-called 'points race' for courses is media generated and has no significant effect on applications.

If students are actually selecting courses according to their prestige rather than by vocational callings, then there should be groups of applicants where the discipline of their ranked courses are quite different, but the common feature of the courses is that they have high points requirements. Therefore, if the points race drives applicants' choices, then groups of applicants ranking high points requirement courses together (such as law, medicine, pharmacy, dentistry and actuarial science) should be present, but where the courses are from different disciplines. However, if the system does work in its intended manner, then applicants should belong to groups where the discipline of their ranked courses is consistent.

In 1997, the Minister for Education and Science set up the Commission on the Points System to review the current college applications system. This led to the publication of a report (Hyland, 1999) which reviewed the system and made a series of recommendations concerning its future. A series of four research reports were also published in conjunction with the Commission's final report. Of particular interest is the report of Tuohy (1998) who studied the college application data by using exploratory techniques; this work is the closest to our analysis. Of some interest is the report of Lynch *et al.* (1999) who investigated the predictive performance of the points that are awarded to applicants in determining overall performance in higher education. These reports received an enormous amount of coverage in the Irish media and were discussed at length by the public. The general conclusion of the exercise was that, although the current system is not perfect, it works very well in practice. Clancy (1995) studied the admissions (rather than applications) data for students in Irish third-level institutions, but this work is closely related to our analysis.

Our analysis focuses on analysing the set of degree course applications that were made through the CAO in the year 2000; there is a separate applications system for diploma and nursing courses. We use mixture models to investigate the presence of groupings in the set of applications (Section 3.1). The resulting mixture model can be used to complete a clustering of the applicants by using a model-based approach. The idea of using mixture models to cluster data has been exploited with much success by Banfield and Raftery (1993) and Fraley and Raftery (2002) among others. A motivation for the use of mixture models when clustering data was given by Aitkin *et al.* (1981) in the discussion of their paper where they say that

> 'when clustering samples from a population, no cluster method is *a priori* believable without a statistical model'.

The mixture models that are employed use the Plackett–Luce model for ranked data (Section 3.2). This model describes the ranking process as a sequence of choosing the next most favoured course.

The model is fitted by maximum likelihood using the EM algorithm (Dempster *et al.*, 1977). The M-step of the EM algorithm is completed by using the MM algorithm (Lange *et al.*, 2000; Hunter, 2004) as an optimization technique (Section 3.4).

To increase the modelling flexibility of our method, we investigate using a noise component in the mixture model to allow for applications that are very different from the majority of the remaining applications.

The resulting groupings of applicants that are suggested by the mixture model (Section 4) reveal that applicants generally appear to be driven by their vocational interests as discipline emerges as the defining characteristic of applicant groups. The geographical position of the institution to which an applicant applies also transpires to have a significant influence on choice of course (Section 4.1). Crucially, in Section 4.2, some weight is added to the CAO system detractors' arguments. A deeper analysis of the groups revealed highlights a subtle influence of the points on the applicants' choices.

A separate analysis of the data for males and females suggests that applicants of different gender have different course choice behaviours (Section 4.3).

2. Central Applications Office data

The data set that is used in this analysis was collected in the year 2000 and it consists of the course choices of 53 757 applicants to degree courses that are offered in Irish third-level institutions. A total of 533 degree courses were selected by the applicants.

A college application that is made through the CAO allows an applicant to rank up to 10 degree courses in order of their preference. Course places are subsequently offered by using these ordered choices.

The gender of the applicants is also known and this information is only used in the analysis that is described in Section 4.3. There were 29338 female and 24419 male applicants in the year 2000.

Characteristic features of these data include the large number of applicants giving preferences for a large number of courses and the constraint that applicants are restricted in the number of courses that they may rank.

3. Statistical methodology

The data that were collected from the Irish college applications contain students from many different backgrounds and with many different interests. We model the choices of course of these students by using a mixture model, so that we can discover groups of students with different choice behaviour.

The finite mixture model provides a sound model-based basis for making rigorous statements about the presence of groups and the structures of these groups. Statements can be based on sound statistical theory rather than being of a descriptive nature.

3.1. Mixture models

We assume that the course choices that were made by the CAO applicants form a sample from a heterogeneous population. This assumption is justified because of the differing backgrounds and interests of the college applicants. Mixture models appropriately model situations where data are collected from heterogeneous populations. Therefore, we propose to use a mixture model to model the college applications data.

A finite mixture model assumes that the population consists of a finite collection of components (or groups). We assume that the (unknown) probability of belonging to component k is π_k. In addition, observations within component k have a probability density $f(\mathbf{x}_i|\underline{p}_k)$, where p_k are unknown parameters. Hence, the resulting model for a single observation is

$$f(\mathbf{x}_i)=\sum_{k=1}^{K}\pi_k\, f(\mathbf{x}_i|\underline{p}_k).$$

Thus given our data $\mathbf{x}=(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_M)$ and the assumption of a mixture model the likelihood is

$$L(\pi_1,\pi_2,\ldots,\pi_K;\underline{p}_1,\underline{p}_2,\ldots,\underline{p}_K|\mathbf{x})=\prod_{i=1}^{M}\sum_{k=1}^{K}\pi_k\, f(\mathbf{x}_i|\underline{p}_k). \tag{1}$$

Extensive reviews of mixture modelling are given by McLachlan and Peel (2000) and Titterington *et al.* (1985); in addition, an excellent overview of using mixture models to produce model-based methods for clustering has been given by Fraley and Raftery (2002). Previous applications of mixture models for analysing ranked data are given in Marden (1995) and Murphy and Martin (2003) among others.

3.2. The Plackett–Luce model

We need to specify an appropriate density for each component of the mixture model. Each applicant's data consist of a ranking of up to 10 courses. Hence, a model that is appropriate for modelling ranked data is required. Many possible models for ranked data are described in Marden (1995), Diaconis (1988) and Critchlow (1985). Multistage ranking models (Marden (1995), section 5.6) have a nice interpretation in terms of sequentially choosing items in order of preference. One parsimonious multistage ranking model that is easy to interpret is the Plackett–Luce model (Plackett, 1975). We propose to use this model for each component of the mixture model.

Plackett (1975) motivated the Plackett–Luce model in terms of modelling horse-races where a vector of probabilities for each horse winning is used to construct a probability model for the finishing order. Similar characteristics can be identified between horse-races and the process of ranking courses; once a course has been chosen it cannot be selected again, and following a choice being made the probability of any remaining course being selected at the next stage is altered.

In the Plackett–Luce model with parameter $\underline{p}=(p_1,p_2,\ldots,p_N)$, the probability that course c_1 is ranked in first position is p_{c_1}. The probability that course c_2 is ranked second, given that course c_1 is ranked first, is $p_{c_2}/\Sigma_{c\neq c_1}\, p_c$, i.e. it is equal to the probability that c_2 is ranked first when all courses except course c_1 are available for selection. The probability that course c_3 is ranked third, given that courses c_1 and c_2 are selected first and second, is $p_{c_3}/\Sigma_{c\notin\{c_1,c_2\}}\, p_c$. The process continues to give the other placing probabilities, i.e. the choices of course are modelled as the product of the probabilities of each chosen course being ranked first where, at each level of preference, the probabilities are appropriately normalized.

In the mixture model that is used in this study, we let $p_{kc(i,t)}$ denote the probability that the course that is chosen in tth position by applicant i is selected first, given that the applicant belongs to the kth component. The rank t of a selected course must be less than or equal to n_i, where n_i is the number of choices expressed by applicant i. The Plackett–Luce model then suggests that the probability of applicant i's ranking conditional on coming from component k is

$$\begin{aligned}\mathbf{P}(\mathbf{x}_i|\underline{p}_k) &= \frac{p_{kc(i,1)}}{\sum_{s=1}^{N} p_{kc(i,s)}}\,\frac{p_{kc(i,2)}}{\sum_{s=2}^{N} p_{kc(i,s)}}\cdots\frac{p_{kc(i,n_i)}}{\sum_{s=n_i}^{N} p_{kc(i,s)}}\\ &= \prod_{t=1}^{n_i}\frac{p_{kc_{(i,t)}}}{\sum_{s=t}^{N} p_{kc(i,s)}}\end{aligned}$$

where, for $s \geqslant t$, the values of $c(i,t)$ are any arbitrary ordering of the unselected courses.

The Plackett–Luce model assumes a weak dependence between the objects that are ranked at different levels. In particular, $\mathbf{P}\{c(i,t)=j\}$ depends on $\{c(i,1),\ldots,c\{i,(t-1)\}\}$ but is independent of their order. This appears to be a reasonable assumption as the ranking process involves deciding whether an object should be placed higher or lower than other alternatives and not on the specific level at which the other alternatives are ranked. Thus at each choice level in the Plackett–Luce model the probabilities are adjusted such that they account for the objects that have already been ranked.

3.3. The EM algorithm

The EM algorithm (Dempster *et al.*, 1977) is a widely used tool for obtaining maximum likelihood estimates in missing data problems; mixture models can be formulated as having the component membership of each observation as missing data. Maximization of the likelihood function is simplified by augmenting the data to include the missing membership variables. Furthermore, the EM algorithm provides estimates not only of the model parameters but also of the unknown component memberships of the observations.

We denote the complete data by $(\mathbf{x},\mathbf{z})=((\mathbf{x}_1,\mathbf{z}_1),\ldots,(\mathbf{x}_M,\mathbf{z}_M))$, where $\mathbf{x}_i$ is applicant i's application and

$$\mathbf{z}_i=(z_{i1},z_{i2},\ldots,z_{iK}) \qquad \forall i=1,\ldots,M$$

with

$$z_{ik}=\begin{cases}1 & \text{if applicant } i \text{ belongs to component } k,\\ 0 & \text{otherwise.}\end{cases}$$

The missing data $\mathbf{z}$ can be interpreted as an indicator of component membership. On convergence of the EM algorithm, the estimated values of z_{ik} are the conditional probabilities that applicant i belongs to component k.

Hence, the complete-data log-likelihood is given as

$$\sum_{i=1}^{M}\sum_{k=1}^{K} z_{ik}\left\{\log(\pi_k)+\sum_{t=1}^{n_i}\log(p_{kc(i,t)})-\sum_{t=1}^{n_i}\log\left(\sum_{s=t}^{N} p_{kc(i,s)}\right)\right\} \tag{2}$$

(see Appendix A).

The EM algorithm involves two steps: an expectation step (E-step) followed by a maximization step (M-step). In the context of finite mixture models, the expectation step estimates the unknown values of z_{ik}. The maximization step then proceeds to maximize the complete-data log-likelihood (2) to estimate the model parameters.

The EM algorithm is an iterative technique and continually repeats the E- and M-steps until convergence to stable estimates and/or a predetermined criterion is achieved. Aitken's acceleration criterion (McLachlan and Peel (2000), section 2.11) was employed in this application as a convergence criterion.

Specifically, the EM algorithm proceeds as follows.

Step 1: initialize—choose starting values for $\underline{p}_1^{(0)}, \underline{p}_2^{(0)}, \ldots, \underline{p}_K^{(0)}$ and $\pi_1^{(0)}, \pi_2^{(0)}, \ldots, \pi_K^{(0)}$. Let $l=0$.

Step 2: E-step—compute the values

$$\hat{z}_{ik}=\frac{\pi_k^{(l)} f(\mathbf{x}_i|\underline{p}_k^{(l)})}{\sum_{k'=1}^{K}\pi_{k'}^{(l)} f(\mathbf{x}_i|\underline{p}_{k'}^{(l)})},$$

where the value $\hat{z}_{ik}$ is the estimated posterior probability that observation i belongs to group k.

Step 3: M-step—maximize the function

$$\sum_{i=1}^{M}\sum_{k=1}^{K} \hat{z}_{ik}\left\{\log(\pi_k)+\sum_{t=1}^{n_i}\log(p_{kc(i,t)})-\sum_{t=1}^{n_i}\log\left(\sum_{s=t}^{N} p_{kc(i,s)}\right)\right\}$$

to yield new parameter estimates $\underline{p}_1^{(l+1)}, \underline{p}_2^{(l+1)}, \ldots, \underline{p}_K^{(l+1)}$ and $\pi_1^{(l+1)}, \pi_2^{(l+1)}, \ldots, \pi_K^{(l+1)}$. Increment l by 1.

Step 4: convergence—repeat the E-step and M-step until convergence. The final parameter values are the maximum likelihood estimates $\hat{\underline{p}}_1, \hat{\underline{p}}_2, \ldots, \hat{\underline{p}}_K$ and $\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_K$.

The E-step is relatively straightforward when fitting a mixture of Plackett–Luce models. The optimization with respect to $\pi_1, \pi_2, \ldots, \pi_K$ in the M-step is also straightforward. However, optimization with respect to $\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_K$ in the M-step is problematic; this optimization is discussed in Section 3.4.

3.4. *MM algorithm*

The M-step of the EM algorithm aims to maximize

$$Q(\mathbf{p})=\sum_{i=1}^{M}\sum_{k=1}^{K} \hat{z}_{ik}\left\{\log(\pi_k)+\sum_{t=1}^{n_i}\log(p_{kc(i,t)})-\sum_{t=1}^{n_i}\log\left(\sum_{s=t}^{N} p_{kc(i,s)}\right)\right\}, \tag{3}$$

where $\mathbf{p}=(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_K)$.

The $\Sigma_{t=1}^{n_i}\log(\Sigma_{s=t}^{N} p_{kc(i,s)})$ term makes maximization of equation (3) difficult. However, Lange *et al.* (2000) provided a summary of a method called optimization transfer using surrogate objective functions which was later renamed the MM algorithm. The MM algorithm is a prescription for constructing optimization algorithms more so than a directly implementable algorithm.

To maximize the function $Q(\mathbf{p})$, the MM algorithm forms a surrogate function that minorizes the objective function. A particular minorizing function for $Q(\mathbf{p})$ is given by $q(\mathbf{p})$, which is of the form (up to a constant)

$$Q(\mathbf{p}) \geqslant q(\mathbf{p}) = \sum_{k=1}^{K} \sum_{i=1}^{M} \hat{z}_{ik} \log(\pi_k) + \sum_{k=1}^{K} \sum_{i=1}^{M} \sum_{t=1}^{n_i} \hat{z}_{ik} \left\{ \log(p_{kc(i,t)}) - \sum_{s=t}^{N} p_{kc(i,s)} \Big/ \sum_{s=t}^{N} p_{kc(i,s)}^{(l)} \right\}, \qquad (4)$$

where the $p_{kj}^{(l)}$-values are parameter estimates.

Optimizing the surrogate function $q(\mathbf{p})$ yields new parameter values $\mathbf{p}^{(l+1)}$ which give a higher value for $Q(\mathbf{p})$, i.e. $Q(\mathbf{p}^{(l+1)}) \geqslant Q(\mathbf{p}^{(l)})$. Explicit details of the steps that are involved in deriving and maximizing expression (4) are given in Appendix B.

Note that the EM algorithm itself is an MM algorithm where the maximization of the log-likelihood function (2) is transferred to a surrogate function (3); the relationship between the EM and MM algorithms is discussed in Lange *et al.* (2000) and Hunter and Lange (2004).

3.5. Noise component

We investigate the inclusion of a 'noise' component in the mixture with

$$(p_{k1}, p_{k2}, \ldots, p_{kN}) = (1/N, 1/N, \ldots, 1/N).$$

This component 'soaks up' observations that have low probability of belonging to the other components. The net result is that outlying observations have less of an effect on the overall results. This component is analogous to the Poisson noise component that is introduced in model-based clustering (Fraley and Raftery, 2002).

When choosing the appropriate number of components in the mixture model, we also investigate whether a noise component should be included in the model.

3.6. Choice of model

A fundamental issue in mixture modelling is the choice of an appropriate number of components. We use the Bayesian information criterion BIC (Schwarz, 1978; Kass and Raftery, 1995) to choose the number of components in the mixture model.

BIC is based on an approximation to the logarithm of the Bayes factor for choosing the number of components (Kass and Raftery, 1995). In the case of mixture models the approximation is not valid. However, BIC has been shown to give good results (Fraley and Raftery, 2002) and has been theoretically supported by Leroux (1992) and Keribin (1998, 2000).

The value of the BIC is calculated as

$$\text{BIC} = 2(\text{maximized log-likelihood}) - r \log(M)$$

where r is the number of parameters that are estimated in the model.

The log-likelihood will increase with the number of fitted components, as will the number of estimated parameters. Thus BIC strikes a compromise between model fit and model complexity by penalizing for large r.

BIC was also used to determine whether a noise component (Section 3.5) should be included in the mixture model. Interestingly, we found that, for the CAO data, the use of BIC suggested including a noise component in the mixture.

4. Results

Mixtures of Plackett–Luce models were fitted by maximum likelihood to the CAO data with the number of components ranging from $K=1$ to $K=30$. Random starting values for both the **p**- and the $\underline{\pi}$-parameters were employed in the EM algorithm with good results. In addition, the option for allowing one of the components to be a noise component (Section 3.5) was also investigated. The mixture model with the highest BIC-value was chosen and the resulting model was carefully examined.

The maximum likelihood estimates of the choice probabilities $\underline{\hat{p}}_k=(\hat{p}_{k1},\hat{p}_{k2},\ldots,\hat{p}_{kN})$ for each component were examined and sorted into decreasing order. From these probabilities it is possible to determine which types of course have the highest probability of being selected by applicants from the component. By examining these probabilities, the component was given a summarizing label. Clearly, we would expect that the courses with high probability values will occur with high frequency in the rankings of applicants that belong to this component.

In addition, for each component, the estimated probability that applicants come from each component $\hat{\pi}_k$ was recorded and examined.

When the full set of all CAO applicants is examined, the BIC-values suggest that a 22-component mixture model should be used. The model selected had a noise component as one of the components. The mixing proportions π_k describe the percentage of the population that is assigned to each component. Table 1 gives the resulting 22 components in decreasing order of their mixing proportions.

An evaluation of Table 1 verifies the argument of supporters of the CAO system—the defining characteristic of the mixture components is the common discipline of the courses with high probabilities, as opposed to courses' common entry requirements. For example, the mixture model contains a component reflecting applicants who chose engineering courses, a component describing applicants who chose educational courses and a component for applicants who chose health science courses. There is no evidence, from the examination of the probabilities, for a component representing applicants who appear to apply for high status (usually high points standard) courses. The resulting components suggest that CAO applicants do follow their vocational interests when applying to Irish institutions of third-level education.

Interestingly, science-based applicants are very distinctly partitioned. Applicants to biological sciences, engineering, mathematical sciences and health sciences are well segregated rather than constituting a single science component.

Table 1. Names and proportions of the 22 components that were detected when the set of all applicants was analysed

Component	*Proportion*	*Component*	*Proportion*
Business and marketing	0.08	Engineering	0.04
Hospitality management	0.08	Cork-based courses	0.04
Arts and humanities	0.07	Galway- and Limerick-based courses	0.04
Biological sciences	0.06	Education	0.03
Business and commerce	0.06	Health sciences	0.03
Communications and media	0.06	Art and design	0.03
Construction studies	0.06	Law	0.03
Computer science (ex-Dublin)	0.05	Mathematical and physical sciences	0.03
Social science	0.05	Business and languages	0.03
Munster-based courses	0.05	Music	0.02
Computer science (Dublin)	0.05	Noise component	0.002

Also of note are the mixing proportion values. Ranking the components in order of mixing proportions indicates that more applicants have a tendency to apply for humanities and business degrees than for more science-based programmes.

However, the results require further examination and discussion; this is done in Sections 4.1–4.3.

4.1. *The geographical effect*

The components that are reported in Table 1 reveal important traits within the population of applicants. Most obvious are the presence of components that highlight a geographical effect on applications.

Interestingly, five of the 22 components that were identified have a geographical basis. The Munster-based courses and Cork-based courses components are epitomized by applicants who predominantly apply to institutions that are within the province of Munster or to institutions in County Cork respectively. The Galway- and Limerick-based component emerges from similarly motivated applicants. Although it is possibly surprising that a geographical effect would be so well defined in such a relatively small island, readers who are acquainted with Irish society will be familiar with such a phenomenon. Firstly, many Irish students opt to live at home during their college studies; this differs from the situation in many other countries. Also, Irish people are very parochial and show strong affinity to their home region. People from Munster, and Cork in particular, have a very strong affinity to their region and tend to avoid travelling for their studies unless the course that they wish to study is not available in the region. Galway and Limerick are the main cities on the west coast of Ireland and a similar impetus is revealed by this component.

Also of note with regard to the geographical effect is the frequent distinction between sets of applicants who apply for degrees of a similar discipline but are deemed separate on the basis of whether or not the institutions to which they apply are in Dublin (the capital of Ireland). Of Ireland's 3.92 million population, 1.12 million reside in County Dublin, and 2.11 million in the province of Leinster (the area around Dublin). Dublin is the centre of Irish governmental, financial and business dealings. Therefore some applicants are drawn to living there, whereas others prefer to stay away to avoid living in a large city. This goes some way in explaining why applicants view courses of a similar type as different, on the basis of whether the institution is in Dublin or not. This effect is clear on the groups of applicants applying for computer science courses, and to a lesser extent on the applicants for business, marketing and commerce degrees.

4.2. *The points race*

On the surface the components that were determined by our model-based clustering verify the arguments of the supporters of the CAO system. Detractors insist that applicants are influenced by media excitement and by the perceived social standing of some courses (which is revealed through their high points requirements). An examination of the reported components and their associated parameters provides deeper insight into the behaviour of the CAO applicants.

We take two approaches to examining this phenomenon. We examine courses according to the probability that the course is chosen within a component, i.e. by using the $\mathbf{P}(\text{course } j|\text{component } k) = p_{kj}$ values (which are estimated by $\hat{p}_{kj}$). We also examine the posterior probability of belonging to a component given that a particular course is chosen, i.e. by using the $\mathbf{P}(\text{component } k|\text{course } j) \propto \pi_k p_{kj}$ values (which are estimated by $\hat{\pi}_k \hat{p}_{kj}$).

To demonstrate that there may be a points race, we examine the results for the health sciences component by using the two approaches that were described above. The results of this deeper analysis are given in Sections 4.2.1–4.2.2.

Table 2. 30 most probable courses to be ranked on an application form, given that an applicant belongs to the health sciences component†

Institution	*Course*	*Probability*
University College Dublin	Medicine	0.4723
Trinity College Dublin	Medicine	0.2413
University College Galway	Medicine	0.2004
University College Cork	Medicine	0.1219
Royal College of Surgeons in Ireland	Medicine	0.0610
University College Dublin	Science	0.0351
Trinity College Dublin	Science	0.0297
Trinity College Dublin	Pharmacy	0.0280
Trinity College Dublin	Dentistry	0.0280
University College Dublin	Physiotherapy	0.0260
Trinity College Dublin	Physiotherapy	0.0241
University College Cork	Dentistry	0.0233
University College Dublin	Veterinary medicine	0.0163
Royal College of Surgeons in Ireland	Medicine with Leaving Certificate Scholarship	0.0153
University College Galway	Science	0.0140
University College Cork	Biological and chemical sciences	0.0125
Trinity College Dublin	Human genetics	0.0121
University College Galway	Biomedical science	0.0116
Dublin Institute of Technology	Optometry	0.0104
University College Dublin	Radiography	0.0101
Trinity College Dublin	Medicinal chemistry	0.0099
University College Dublin	*Arts*	*0.0092*
University College Dublin	*Law*	*0.0091*
Trinity College Dublin	*Law*	*0.0085*
University College Dublin	*Engineering*	*0.0083*
Trinity College Dublin	Therapeutic radiography	0.0081
Trinity College Dublin	Psychology	0.0074
Royal College of Surgeons in Ireland	Physiotherapy	0.0069
Royal College of Surgeons in Ireland	Medicine with Royal College of Surgeons in Ireland scholarships	0.0065
University College Dublin	Psychology	0.0059

†Clearly health science degrees dominate, but the presence of high status law degrees adds some weight to the argument that applicants are influenced by the prestige of some courses' points requirements.

4.2.1. *Examination of component parameters*

Table 2 shows the 30 courses with the highest probability of selection (listed in decreasing order) given that an applicant belongs to the health sciences component (see Table 1). Table 2 illustrates how components were assigned a summarizing label—from a glance it is clear that applicants belonging to this component have high probability of choosing courses leading to a degree in the health sciences sector. Many health science degree programmes have high entry requirements, due to demand, a limited supply of places and the fact that these courses attract highly achieving second-level students. Medicine, pharmacy, dentistry and veterinary medicine are annually reported as degree programmes with higher points requirements than other courses and the resulting careers are highly esteemed within Irish society. They also are vocationally driven careers, and thus we would expect applicants to have a tendency to apply for many courses within a discipline for which they feel that they have a vocation.

Within the top 30 courses in Table 2 four have been highlighted. Arts as offered by University College Dublin, law as offered by University College Dublin and Trinity College Dublin and engineering as offered by University College Dublin. Although the probabilities of ranking these courses given that an applicant belongs to the health sciences component are small, in relative terms applicants are almost equally likely to rank medicinal chemistry, law or therapeutic radiography. Although some would, perhaps correctly, argue that a career in law is also a vocation, it could also be argued that equally so are careers such as those in the education sector. The difference between law and education degrees, in Ireland at least, is their points requirements. Law would be considered a consistently high requirement degree, whereas an education degree would have lower points requirements. There is little evidence that health science applicants choose education programmes with high probability. Therefore, some weight has been added to the assertions of CAO detractors that the CAO system influences applicants to apply for courses that are prestigious (in terms of points). Another explanation is that the applicants are attracted to courses that tend to lead to high salaried professions. In any case, this implies that courses are being chosen by their status in society rather than by the discipline. How otherwise would health science applicants be as likely to choose law as therapeutic radiography?

The four courses that are highlighted in Table 2 also have the common characteristic of being based in institutions in Dublin. It is clear from other components that the geographical location of a course influences applicants. Within the health sciences component it appears that both geography and course status affect the way in which applicants rank courses. However, it is difficult to distinguish fully between these influences.

Also of note is the high probability of choosing the arts degree (in University College Dublin) and engineering (in University College Dublin). As the name partially suggests, in an arts degree students study one or two subjects from a range of arts and humanities subjects. Thus the arts degree is a very general degree that provides a broad basis from which many different career paths can emerge. In fact, it is the most frequently ranked degree programme among all CAO applicants and has relatively achievable points requirements. Its popularity or perhaps its reputation as a 'fail-safe' third-level choice are possible explanations of its high choice probability within the health science component.

The inclusion of engineering as a high probability course could also be because applicants include a fail-safe alternative. The points required for engineering in University College Dublin were much lower than those for the health science degrees in Table 2, so the points status would not appear to be a contributing factor. It is clear that health science applicants select a general science degree with high probability and perhaps are then also attracted to the general scientific aspects of an engineering degree. More of note perhaps is that in 2000 engineering as offered by University College Dublin was a general entry degree where students did not choose a specific vein of engineering until later in their degree. Although both Trinity College Dublin and University College Galway ran a similar style of programme, the points required that year were considerably higher than those required for entry to University College Dublin's degree. Thus, similarly to an arts degree, engineering may have been viewed as the fail-safe science-based option.

4.2.2. Examination of posterior component membership probabilities

An alternative approach can be taken in the analysis of the parameter estimates by examining the posterior probability of belonging to component k given that course j was selected, i.e. $\mathbf{P}(\text{component } k|\text{course } j)$. Table 3 shows the 25 courses whose selection gives highest posterior probability of belonging to the health sciences component.

Table 3. 25 courses whose selection on a CAO application form gives the highest probability of belonging to the health sciences component

Institution	*Course*	*Probability*
University College Dublin	Medicine	0.9440
Trinity College Dublin	Medicine	0.9392
Royal College of Surgeons in Ireland	Medicine	0.8886
Royal College of Surgeons in Ireland	Medicine with Leaving Certificate scholarship	0.8813
University College Galway	Medicine	0.8724
Royal College of Surgeons in Ireland	Medicine with Royal College of Surgeons in Ireland scholarships	0.8010
University College Cork	Medicine	0.7633
Trinity College Dublin	Dentistry	0.5391
University College Cork	Dentistry	0.3674
Trinity College Dublin	Pharmacy	0.3110
Trinity College Dublin	Medicinal chemistry	0.2725
Trinity College Dublin	Therapeutic radiography	0.2662
Trinity College Dublin	Human genetics	0.2579
University College Dublin	Radiography	0.2230
Trinity College Dublin	Physiotherapy	0.2141
Royal College of Surgeons in Ireland	Physiotherapy	0.2035
Trinity College Dublin	*Mathematics and Latin*	*0.2013*
Dublin Institute of Technology	Optometry	0.1989
University College Dublin	Physiotherapy	0.1986
University College Dublin	Veterinary medicine	0.1796
Trinity College Dublin	*Mathematics and psychology*	*0.1642*
University College Galway	Biomedical science	0.1517
University College Galway	Biomedical engineering	0.1093
Trinity College Dublin	Science	0.1053
Trinity College Dublin	Occupational therapy	0.0992

An examination of the mixture model in this way further highlights the subtle effect that the points race may have on some applicants' choices. Within the top 25 courses that suggest high probability of belonging to the health sciences component are mathematics and Latin and mathematics and psychology, which are both offered by Trinity College Dublin. It appears strange to have high probability of belonging to a component that is dominated by health sciences courses due to the selection of either of these courses. Both are part of Trinity College's version of the general arts degree—the two-subject moderatorship (TSM) programme. In the TSM programme, students choose two modules from a range of arts and humanities subjects and study them simultaneously. However, each combination is viewed as a separate course by the CAO and owing to the wide range of subjects, and therefore combinations, their choice is usually quite rare, leading to sparse data. Some TSM courses lead to some strange results when analysing the CAO data owing to the rarity of some course selections within the TSM programme. However, the inclusion here of only two of the wide range of TSM courses suggests a contributing factor other than sparsity of data. These two TSM courses both include mathematics; in that particular year points requirements for TSM courses involving mathematics were at a similar level to that of many of the health science programmes listed in Table 3. Therefore, a deeper investigation of the posterior probabilities highlights again the possibility of a subtle effect that a course's points requirements may affect CAO applicants.

Why is there a focus on law programmes and mathematics programmes as examples of the points race? Other high points courses such as actuarial and financial studies (in University College Dublin) also appear within the top 50 programmes in both views of the model; again, this course seems to be a strange course to appear among a component that is dominated by the ‘vocational’ health science sector.

Why is there a focus on the health sciences component only? It seems natural also to consider the law component that is also deemed as high points and high status. The points race effect is also apparent here—psychology in both University College Dublin and Trinity College Dublin, which had high entry requirements that year, have a high probability of being selected given that an applicant belongs to the law component. Although law and psychology have some similarities, they would not be deemed to be members of the same discipline, suggesting that some element of the points race is present. However, an examination of the posterior probabilities for the law component gives less of an indication of the presence of a points race. It seems that the points status of courses has more of an effect in the health science component than in the other components in the mixture model.

4.3. The gender effect

The gender of each CAO applicant in 2000 was available in addition to their choices of course; of the 53757 applicants, 24419 were male. The data were partitioned according to applicants’ gender and mixtures of Plackett–Luce models were fitted to the two resulting data sets. An examination of the resulting parameter estimates $\hat{p}_{kj}$ led to the summarizing component labels as outlined in Table 4.

The resulting mixtures fitted to the partitioned data provide good insight into the different choice behaviour of the male and female applicants. The predominant aspect of the component labels is subject discipline, thus enhancing the supporting view of the CAO that applicants are

Table 4. Resulting 16 components from the analysis for the female applicants, and the resulting 17 from the male applicants

Results for females		*Results for males*	
Component	*Proportion*	*Component*	*Proportion*
Hospitality management	0.11	Construction studies	0.09
Social science	0.11	Communications and journalism	0.09
Business and marketing (Dublin)	0.09	Business and marketing (Dublin)	0.09
Biological sciences	0.08	Computer science (ex-Dublin)	0.08
Cork-based courses	0.08	Hospitality management	0.07
Applied computing (ex-Dublin)	0.07	Computer science (Dublin)	0.07
Communications and journalism	0.07	Arts and humanities	0.06
Business and commerce (ex-Dublin)	0.07	Engineering (ex-Dublin)	0.06
Law and psychology	0.06	Business and commerce	0.06
Galway-and Limerick-based courses	0.06	Cork-based courses	0.06
Education	0.05	Law and business	0.06
Engineering and computer science	0.04	Engineering (Dublin)	0.05
Art and design; music	0.04	Sports science and education	0.05
Business and languages	0.04	Science	0.04
Health sciences	0.04	Limerick-based courses	0.04
Noise component	0.003	Health sciences	0.03
		Noise component	0.004

inclined to follow their vocational interests. The geographical effect that was discussed in Section 4.1 is again apparent, but it is more apparent in the results for males. In particular, some components for males reveal a common discipline but at different geographical locations; this occurs more so than in the components for females. For example, the male engineering applicants are partitioned by the location of the institution in Dublin, as are the computer science applicants.

Stereotypical differences between the two genders are very apparent in the resulting components—there appear to be distinct components for females in social science, art and design, music and education, whereas female applicants with an interest in engineering and computer science are grouped together. Not only are the engineering and computer science components for males separate; they are also further divided within these disciplines by geography. Further, the largest component (with probability 0.09) in the results for males involves construction studies courses whereas this does not appear as a distinct component in the results for females.

Other results of interest are the popularity of biological sciences among females whereas males have a general science component in their results; both genders have education components but the education component for males also has a sports aspect.

In addition, in close similarity to the results for all applicants (see Section 4.2.1), the health sciences component for males contains three law degrees in the top 30 most probable courses. Similar results are revealed for the females, but the probability of selection of the law courses is lower within the health sciences component.

4.4. Clustering of applicants

A major advantage of fitting mixture models via the EM algorithm, as detailed by Fraley and Raftery (1998), is that the value $\hat{z}_{ik}$ at convergence is an estimate of the conditional probability that observation i belongs to component k; these values can be used to cluster observations into groups. A clustering of the set of applicants is simply achieved by examining $\max_k\{\mathbf{P}(\text{component } k|\text{application } i)\}, \forall i$, and assigning applicants to the group for which the maximum is achieved.

The clustering of applicants can be scrutinized in different ways. As suggested by Bensmail *et al.* (1997), the uncertainty that is associated with an applicant's component membership can be measured by

$$U_i = \min_{k=1,\ldots,K}\{1 - \mathbf{P}(\text{component } k|\text{application } i)\}.$$

When i is very strongly associated with group k then $\mathbf{P}(\text{component } k|\text{application } i)$ will be large and so U_i will be small. Fig. 1 illustrates the uncertainty that is associated with the clustering of the male and female applicants.

Clearly, the clustering uncertainty values tend to be very small, with 61% of females and 59% of males classified with an uncertainty of less than 0.05. Summary statistics for the uncertainty values further demonstrate how well the model allocates applicants to components; these are given in Table 5.

5. Discussion

This paper presents a model-based statistical analysis of degree level applicants to Irish institutions of third-level education. The methods seek to find groups of similar applicants, and to draw conclusions about the merits and failures of the centralized applications system from the defining characteristics of these groups.

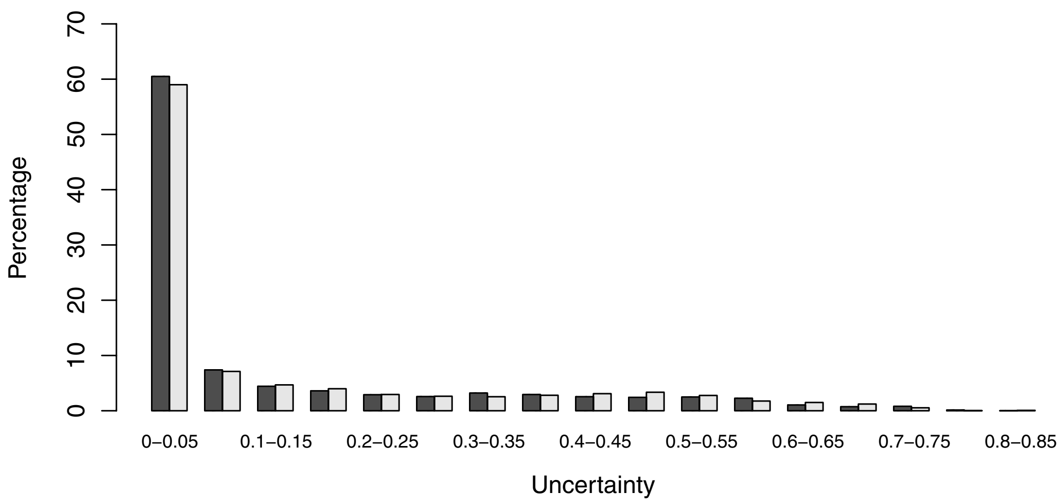

Fig. 1. Uncertainty in the clustering of the female (■) and male (■) applicants

Table 5. Summary statistics associated with the clustering uncertainty of male and female applicants

Applicants	*1st quartile*	*Mean*	*3rd quartile*
Females	0.0002	0.1228	0.1866
Males	0.0002	0.1301	0.2043

A top level view of the groups of applicants that is suggested by the analysis verifies a supporting view of the CAO system—applicants appear to follow their vocational interests and rank their third-level course choices in a manner which reflects this. The analysis suggests that the majority of CAO applicants use the system as it is intended and rank courses in view of their genuine preferences and/or choice of career. However, it is apparent that more subtle influences also contribute to the choice of course and a detailed examination of the mixture components indicates the faint presence of the reported points race. It appears that there are those who choose courses on the points levels of previous years and therefore on the prestige that is attached to some of these courses.

Whereas most discussions of the CAO system in Irish education circles focus on the influence of the points race this work highlights other factors which have an influence on an applicant's choice of course. The geographical location of the institution to which an applicant applies has a clear effect on the process of choice. Whether this is due to a vocational desire to study a particular course in a specific institution, the desire to live in a certain area or because of financial viability, it is a striking feature of the groups of applicants. A course's geographical location appears to be almost as important as vocational interest in an applicant's process of choice. Whether this feature is a benefit of the CAO system or not remains to be researched.

Further to the effects of vocation, geography and the points race, the gender of the applicant also affects the choice of course. Geography and the points race may have a larger effect on male applicants than on females. Stereotypical gender differences are also apparent—only 4% of female applicants are 'classified' as engineering and computer science students compared with 26% of the population of male applicants. Further differences (see Section 4.3) indicate that males and females need to be targeted in different ways, with regard to third-level education, and this should be of interest to third-level institutions and to governmental education departments.

In terms of the model that is employed within components, the Plackett–Luce model performs well when modelling the rankings of the preferred third-level choices of the CAO applicants. The model does suffer from independence from irrelevant alternatives (see Train (2003)). The Plackett–Luce model is said to exhibit irrelevant alternatives as the ratio of the probabilities of choosing one alternative over another is independent of all other available alternatives and independent of the level of choice. Although it can be argued that such models are unrealistic in some situations, in this application the model appears to provide a realistic representation of the course choice process.

The only covariate that is available for this analysis is the gender of the applicant—relationships between choice of course and other covariates are very likely to be present. Expanding the analysis to include other covariates would also be desirable, but further covariates were not available for this study.

Acknowledgements

Both authors thank Adrian Raftery and other members of the Working Group on Model-based Clustering at the University of Washington, Seattle, for important inputs into this research. The authors also extend thanks to the Joint Editor and Associate Editor of the journal and to the referees for constructive comments and improvements. These have greatly improved the paper.

Isobel Claire Gormley was supported by a Government of Ireland Research Scholarship in Science, Engineering and Technology provided by the Irish Research Council for Science, Engineering and Technology, funded by the National Development Plan.

Appendix A: Formation of the complete-data log-likelihood

Under the Plackett–Luce model the probability of applicant i's ranking is

$$\mathbf{P}(\mathbf{x}_i) = \prod_{t=1}^{n_i} \frac{p_{kc(i,t)}}{\sum_{s=t}^{N} p_{kc(i,s)}}.$$

Accounting for the missing component membership indicator $\mathbf{z}_i$, it follows that the complete-data log-likelihood for applicant i is

$$\begin{aligned}\mathbf{P}(\mathbf{x}_i, \mathbf{z}_i) &= \mathbf{P}(\mathbf{x}_i|\mathbf{z}_i)\,\mathbf{P}(\mathbf{z}_i)\\ &= \left\{ \prod_{k=1}^{K} \left(\prod_{t=1}^{n_i} \frac{p_{kc(i,t)}}{\sum_{s=t}^{N} p_{kc(i,s)}} \right)^{z_{ik}} \right\} \prod_{k=1}^{K} \pi_k^{z_{ik}}.\end{aligned}$$

Hence, the complete-data log-likelihood for all applicants is

$$\begin{aligned} l &= \log\left\{ \prod_{i=1}^{M} \prod_{k=1}^{K} \left(\pi_k \prod_{t=1}^{n_i} \frac{p_{kc(i,t)}}{\sum_{s=t}^{N} p_{kc(i,s)}} \right)^{z_{ik}} \right\} \\ &= \sum_{i=1}^{M} \sum_{k=1}^{K} z_{ik} \left\{ \log(\pi_k) + \sum_{t=1}^{n_i} \log(p_{kc(i,t)}) - \sum_{t=1}^{n_i} \log\left(\sum_{s=t}^{N} p_{kc(i,s)} \right) \right\}. \end{aligned}$$

Appendix B: MM algorithm calculations

We detail the steps that are involved in the formation of the required surrogate function to which maximization is transferred in the M-step of the EM algorithm and from which the maximum likelihood estimates $\hat{p}_{kj}$ are derived. This derivation is closely related to calculations that were given in Hunter (2004). The general reviews of the MM algorithm that have been given by Lange *et al.* (2000) and Hunter and Lange (2004) are also of interest.

The construction of a surrogate function relies on the exploitation of properties of convex functions. For convex function $f(x)$ with differential $\mathrm{d}f(u)$, the *supporting hyperplane property of a convex function* f,

$$f(x) \geqslant f(y) + \mathrm{d}f(y)(x-y) \qquad x, y \geqslant 0, \tag{5}$$

provides a linear minorizing function that can be utilized as a surrogate function in an optimization transfer algorithm. Sometimes it is preferable to form a quadratic or higher order surrogate function. Expanding expression (5) by using higher order expansions can yield such higher order functions.

In this application, we wish to construct a minimizing surrogate function to be iteratively maximized. This iterative maximization gives a sequence of parameter estimates with increasing values for the expected complete-data log-likelihood function (3). The strict convexity of the $-\log(x)$ function implies that

$$-\log(x) \geqslant -\log(y) + 1 - x/y.$$

We let $f(x) = -\log(\Sigma_{s=t}^{N} p_{kc(i,s)})$. Thus,

$$-\log\left(\sum_{s=t}^{N} p_{kc(i,s)}\right) \geqslant -\log\left(\sum_{s=t}^{N} p_{kc(i,s)}^{(l)}\right) + 1 - \sum_{s=t}^{N} p_{kc(i,s)} \Big/ \sum_{s=t}^{N} p_{kc(i,s)}^{(l)}.$$

It follows that, up to a constant,

$$\begin{aligned} Q(p_{kj}) &\geqslant q \\ &= \sum_{k=1}^{K}\sum_{i=1}^{M} \hat{z}_{ik} \log(\pi_k) + \sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{t=1}^{n_i} \hat{z}_{ik} \left\{ \log(p_{kc(i,t)}) - \sum_{s=t}^{N} p_{kc(i,s)} \Big/ \sum_{s=t}^{N} p_{kc(i,s)}^{(l)} \right\}. \end{aligned}$$

By iterative maximization of the surrogate function q we produce a sequence of p_{kj}- (and of the mixing proportions π_k) values which have monotonically increasing Q-value. The values converge to the maximum of Q with respect to p_{kj} and π_k.

Differentiation of q with respect to p_{kj} gives

$$\frac{\partial q}{\partial p_{kj}} = \sum_{i=1}^{M} \hat{z}_{ik} \left(\sum_{t=1}^{n_i} \frac{1}{p_{kc(i,t)}} \mathbf{1}_{\{j=c(i,t)\}} - \sum_{t=1}^{n_i} \frac{1}{\sum_{s=t}^{N} p_{kc(i,s)}^{(l)}} \mathbf{1}_{\{j \in \{c(i,t),\ldots,c(i,N)\}\}} \right). \tag{6}$$

We denote

$$\omega_{kj} = \sum_{i=1}^{M} \hat{z}_{ik} \sum_{t=1}^{n_i} \mathbf{1}_{\{j=c(i,t)\}}$$

and

$$\delta_{ijt} = \begin{cases} 1 & \text{if } j \in \{c(i,t),\ldots,c(i,N)\}, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, equating expression (6) to 0 and substituting in ω_{kj} and δ_{ijt} as detailed above, we obtain

$$\frac{\omega_{kj}}{p_{kj}} = \sum_{i=1}^{M} \hat{z}_{ik} \sum_{t=1}^{n_i} \left(\sum_{s=t}^{N} p_{kc(i,s)}^{(l)} \right)^{-1} \delta_{ijt}$$

which implies that

$$p_{kj}^{(l+1)} = \omega_{kj} \Big/ \sum_{i=1}^{M}\sum_{t=1}^{n_i} \hat{z}_{ik} \delta_{ijt} \left(\sum_{s=t}^{N} p_{kc(i,s)}^{(l)} \right)^{-1},$$

for $k = 1, \ldots, K$ and $j = 1, \ldots, N$.

Similarly, to obtain π_k the expected complete-data log-likelihood function

$$Q(\mathbf{p}) = \sum_{i=1}^{M} \sum_{k=1}^{K} \hat{z}_{ik} \left\{ \log(\pi_k) + \sum_{t=1}^{n_i} \log(p_{kc(i,t)}) - \sum_{t=1}^{n_i} \log\left(\sum_{s=t}^{N} p_{kc(i,s)} \right) \right\}$$

is maximized with respect to π_k, subject to the constraint $\Sigma_{k=1}^{K} \pi_k = 1$. Thus, denoting a Lagrange multiplier by λ, we have

$$\frac{\partial}{\partial \pi_k} \left\{ Q - \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) \right\} = 0$$

$$\Rightarrow \sum_{i=1}^{M} \hat{z}_{ik} / \pi_k - \lambda = 0.$$

Since

$$\sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} \sum_{i=1}^{M} \hat{z}_{ik} / \lambda = 1$$

we have $M/\lambda = 1$ and hence

$$\pi_k = \sum_{i=1}^{M} \hat{z}_{ik} / M.$$

References

Aitkin, M., Anderson, D. and Hinde, J. (1981) Statistical modelling of data on teaching styles (with discussion). *J. R. Statist. Soc.* A, **144**, 419–461.

Banfield, J. D. and Raftery, A. E. (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics*, **49**, 803–821.

Bensmail, H., Celeux, G., Raftery, A. E. and Robert, C. P. (1997) Inference in model-based cluster analysis. *Statist. Comput.*, **7**, 1–10.

Clancy, P. (1995) *College Entry in Focus: a Fourth National Survey of Access to Higher Education*. Dublin: Higher Education Authority.

Critchlow, D. E. (1985) Metric methods for analyzing partially ranked data. *Lect. Notes Statist.*, **34**.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Diaconis, P. (1988) *Group Representations in Probability and Statistics*. Hayward: Institute of Mathematical Statistics.

Fraley, C. and Raftery, A. E. (1998) How many clusters?: which clustering method?—answers via model-based cluster analysis. *Comput. J.*, **41**, 578–588.

Fraley, C. and Raftery, A. E. (2002) Model-based clustering, discriminant analysis, and density estimation. *J. Am. Statist. Ass.*, **97**, 611–612.

Hunter, D. R. (2004) MM algorithms for generalized Bradley-Terry models. *Ann. Statist.*, **32**, 384–406.

Hunter, D. R. and Lange, K. (2004) A tutorial on MM algorithms. *Am. Statistn*, **58**, 30–37.

Hyland, A. (1999) *Commission on the Points System: Final Report and Recommendations*. Dublin: Stationery Office.

Kass, R. E. and Raftery, A. E. (1995) Bayes factors. *J. Am. Statist. Ass.*, **90**, 773–795.

Keribin, C. (1998) Estimation consistante de l'ordre de modèles de mélange. *C. R. Acad. Sci. Par.* I, **326**, 243–248.

Keribin, C. (2000) Consistent estimation of the order of mixture models. *Sankhya* A, **62**, 49–66.

Lange, K., Hunter, D. R. and Yang, I. (2000) Optimization transfer using surrogate objective functions (with discussion). *J. Comput. Graph. Statist.*, **9**, 1–59.

Leroux, B. G. (1992) Consistent estimation of a mixing distribution. *Ann. Statist.*, **20**, 1350–1360.

Lynch, K., Brannick, T., Clancy, P. and Drudy, S. (1999) *Points and Performance in Higher Education: a Study of the Predictive Validity of the Points System*. Dublin: Stationery Office.

Marden, J. I. (1995) *Analyzing and Modeling Rank Data*. London: Chapman and Hall.

McLachlan, G. J. and Peel, D. (2000) *Finite Mixture Models*. New York: Wiley.

Murphy, T. B. and Martin, D. (2003) Mixtures of distance-based models for ranking data. *Computnl Statist. Data Anal.*, **41**, 645–655.

Plackett, R. L. (1975) The analysis of permutations. *Appl. Statist.*, **24**, 193–202.

Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.
Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985) *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.
Train, K. E. (2003) *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.
Tuohy, D. (1998) *Demand for Third-level Places*. Dublin: Stationery Office.