Epitope profiling via mixture modeling of ranked data

Cristina Mollica*† and Luca Tardella

We propose the use of probability models for ranked data as a useful alternative to a quantitative data analysis to investigate the outcome of bioassay experiments when the preliminary choice of an appropriate normalization method for the raw numerical responses is difficult or subject to criticism. We review standard distance-based and multistage ranking models and propose an original generalization of the Plackett–Luce model to account for the order of the ranking elicitation process. The usefulness of the novel model is illustrated with its maximum likelihood estimation for a real data set. Specifically, we address the heterogeneous nature of the experimental units via model-based clustering and detail the necessary steps for a successful likelihood maximization through a hybrid version of the expectation–maximization algorithm. The performance of the mixture model using the new distribution as mixture components is then compared with alternative mixture models for random rankings. A discussion on the interpretation of the identified clusters and a comparison with more standard quantitative approaches are finally provided. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: ranking data; Plackett–Luce model; multistage ranking models; mixture models; EM algorithm; epitope mapping

1. Introduction

Ranked data arise in several contexts, especially when objective and precise measurements of the phenomena of interest can be impossible or deemed unreliable and the observer gathers ordinal information in terms of orderings, preferences, judgments, relative, or absolute rankings among competitors. Research fields where ranked data analysis is frequently required are the social and behavioral sciences, whose studies often ask a sample of \( N \) people to rank a finite set of \( K \) items according to certain criteria, typically their personal preferences or attitudes. In marketing, political surveys, or psychological experiments, items to rank could be consumer goods, political candidates or goals, words or topics considered to be more or less associated to a reference one according to the individual perception. Another typical context is sport, where teams, horses, or cars compete, and the final outcome is a ranking among competitors. A detailed and well-structured reference monograph concerning ranking data analysis and modeling is [1].

Statistical analysis of observed rankings is less usual in experimental research, where the availability of (sometimes sophisticated) measuring devices allows to express phenomena of interest in terms of precise quantitative information. This work is aimed at verifying the usefulness of probability models for ranked data in an experimental study, where quantitative outcomes are indeed available, but for reasons due to numerical instability of the measurements and absence of universally accepted ways of rescaling the original data, one could instead investigate the ranking evidence. Our interest has been motivated by a new technology for epitope mapping of the binding between the antibodies present in a biological tissue and a target protein. Specifically, we used the real data set from the large fragment phage display (LFPD) bioassay experiment described in [2]. Researchers set up a new promising technology in order to get further insights into the understanding of molecular recognition of the immune system via epitope mapping of the HER2 oncoprotein. They employed a sample of patients and recorded for each subject the binding level, expressed on quantitative scale, between antibodies and specific partially overlapping fragments of the HER2 oncoprotein. The sample was actually composed of three different groups according to the
known breast cancer state. A preliminary exploratory analysis of the LFPD data showed differential outcome profiles for the three cancer-specific groups [2]. Hence, we assumed the sample as drawn from a heterogeneous, multimodal population and opted to describe it through a mixture modeling approach for the individual ranked binding sequences. Well-studied parametric distributions for ranked data and a new extension of the Plackett–Luce (PL) model have been employed as mixture components, and the resulting performances have been compared. Maximum likelihood estimates (MLE) have been obtained with the implementation of the expectation–maximization (EM) algorithm or with hybrid versions thereof.

This article is organized as follows. In Section 2, we define the notation and review the distance-based (DB) and the PL model, with some generalizations already proposed in literature. The presentation of our extended PL (EPL) model follows in Section 2.3, whereas the MLE for a finite mixture is discussed in Section 3. The application to the LFPD data and the comparison of the novel model with alternative ranking probability distributions are detailed in Section 4, along with an interpretation of the inferential findings. The article ends with conclusions and proposals for future developments in Section 5.

2. Statistical models for ranked data

2.1. Notation and basic definitions

Before reviewing some of the approaches for the probabilistic modeling of ranked data, it is convenient to fix some notation. Formally, a full (or complete) ranking \( \pi \) is a bijective mapping of a finite set \( I = \{i_1, \ldots, i_K\} \) of labeled items into a set of ranks \( R = \{1, \ldots, K\} \), that is,

\[
\pi : I \to R.
\]

With some abuse of notation, each item label will be identified with its subscript: instead of writing a forward ranking process \( \pi(i) \), and entries give the corresponding assigned ranks, which means that \( \pi(i) \) must be read as the rank attributed to the \( i \)-th item. The underlying convention is that if \( \pi(i) < \pi(i') \), then item \( i \) is ranked higher than item \( i' \) and hence preferred to it.

In the literature, it is common to distinguish between a full and a partial (or incomplete) ranking where, in the latter case, the rank assignment process is not completely carried out. This happens, for instance, when a judge expresses only her first \( t \) preferences out of \( K \) items (\( t < K \)), producing the so-called top-\( t \) partial ranking. In the present context, partial rankings are not contemplated because of surjectivity of the mapping \( \pi \), and ties are not allowed because of injectivity.

The inverse \( \pi^{-1} = (\pi^{-1}(1), \ldots, \pi^{-1}(K)) \) of a ranking \( \pi \) is called ordering. Positions of the components in \( \pi^{-1} \) refer to ranks, and elements correspond to the items. Hence, \( \pi^{-1}(j) \) is the item ranked in the \( j \)-th position. In order to avoid confusion with \( \pi \), we will henceforth make explicit use of the inverse function notation to denote the corresponding ordering \( \pi^{-1} : R \to I \).

We denote with \( S_K \) the set of all \( K! \) possible permutations, conceived as a symmetric group. This special finite subset of \( \mathbb{R}^K \) is endowed with a composition operation \( \circ \) such that two elements \( \pi \) and \( \sigma \) in \( S_K \) may yield a permutation of either \( R \) or \( I \). For instance, \( \pi \sigma^{-1} = \pi \circ \sigma^{-1} = (\pi(\sigma^{-1}(1)), \ldots, \pi(\sigma^{-1}(K))) \) indicates ranks under \( \pi \) relative to the items ranked \( 1, \ldots, K \) by \( \sigma \). In the symmetric group, this is formally defined as the mapping \( \sigma^{-1} \), which acts on the right of \( \pi \), to stress that the composition is not in general commutative. Reversing the order in the composition, in fact, gives \( \sigma^{-1} \pi = \sigma^{-1} \circ \pi = (\sigma^{-1}(\pi(1)), \ldots, \sigma^{-1}(\pi(K))) \), which lists items receiving from \( \sigma \) those ranks that \( \pi \) has attributed to items \( 1, \ldots, K \).

When a judge proceeds to elicit from her best choice (rank 1) up to the worst one (rank \( K \)), one has the so-called forward ranking process; the inverse ranking procedure is named backward ranking process. This formal definition has been originally introduced in [3], but, to our knowledge, the rank assignment scheme has not received an explicit consideration in a model setup in the attempt to improve the description of random ranked data. Obviously, any other order for the rank assignment process is admissible and potentially leads to different models. This aspect has inspired us to expand an existing and well-known parametric ranking model and employ such a new class in the analysis of the LFPD data in order to verify whether and how the reference order can influence the inferential results and the final model-based clustering.

Copyright © 2014 John Wiley & Sons, Ltd. Statist. Med. 2014, 33 3738–3758
2.2. Probability models for random rankings

In this section, we give a brief account of ranked data modeling. For a more systematic review, see [1] and [4]. The collection of all discrete distributions for random rankings can be identified with the whole \((K! - 1)\)-dimensional simplex \(P(S_K)\). This is equivalent to saying that a random ranking and its distribution can be denoted with \(\pi \sim P\), where the set \(\{P \in P(S_K)\}\) represents the most general statistical model. We refer to this general form as saturated model (SM), parameterized by \(K! - 1\) free parameters, that is, the probabilities of each ordered sequence. Within this very general class, two particular cases play a special role: the uniform (or null) model (UM), represented by the single flat distribution which assigns equal probability to each ranking, and the degenerate models, which conversely concentrate all the probability mass on a single ranking. Although the SM allows for the maximum degree of flexibility, the fast-growing dimension of the ranking space makes it intractable and cumbersome to interpret even with a relatively small number \(K\) of items. These practical limitations have motivated the introduction of simplifying assumptions on the ranking process and justify the wide assortment of restricted parametric models developed in the rank data theory.

2.2.1. Distance-based models. A fundamental class of parametric distributions is the so-called DB model. Roughly speaking, the DB can be interpreted as the analog of the normal distribution on the finite discrete space \(S_K\). In fact, it is an exponential location-scale model indexed by a discrete parameter \(\sigma \in S_K\), called modal or central ranking, and a nonnegative real concentration parameter \(\lambda \in \mathbb{R}_0^+\). Each distribution in a DB model has the following form

\[
P(\pi | \sigma, \lambda) = \frac{1}{Z(\lambda)} e^{-\lambda d(\pi, \sigma)} \quad \pi \in S_K, \tag{2.1}
\]

where \(Z(\lambda) = \sum_{\pi \in S_K} e^{-\lambda d(\pi, \sigma)}\) is the normalization constant and \(d\) is a metric on \(S_K\). The probability mass function in (2.1) attains its maximum at \(\pi = \sigma\) and decreases as the distance from \(\sigma\) increases. Under (2.1) rankings at the same distance from the modal sequence \(\sigma\) are equally probable. The central ranking \(\sigma\) expresses the so-called consensus in the population, whereas the concentration/precision parameter \(\lambda\) calibrates the effect of \(d\) on the probability of the ranking: the higher the value of \(\lambda\), the more concentrated the distribution around its mode. Hence, when \(\lambda \to +\infty\), Equation 2.1 becomes the degenerate model at \(\pi = \sigma\); conversely, when \(\lambda = 0\), it turns out to be the UM.

Changing the distance measure \(d\) in (2.1), one can define different families of parametric distributions for ranked data. Formally, a function \(d : S_K \times S_K \to \mathbb{R}_0^+\) is a distance between rankings if it satisfies the usual three properties of a metric (identity, symmetry, and triangle inequality) and the additional fourth condition of right-invariance, that is,

\[
d(\pi, \sigma) = d\left(\pi \delta^{-1}, \sigma \delta^{-1}\right), \tag{2.2}
\]

for all \(\pi, \sigma, \delta \in S_K\). Condition (2.2) guarantees the desirable property of invariance of \(d\) with respect to (WRT) arbitrary relabeling of items. One of the most frequently used metrics is the Kendall distance defined as

\[
d_K(\pi, \sigma) = \sum_{1 \leq i < j \leq K} I_{((\pi(i) - \pi(j))(\sigma(i) - \sigma(j)) < 0)},
\]

which counts the number of pairwise disagreements, that is, the pairs of items with relative discordant order under \(\pi\) and \(\sigma\). It is also equal to the minimum number of adjacent transpositions needed to transform \(\pi^{-1}\) into \(\sigma^{-1}\). Relaxing the adjacency requirement, one has the Cayley distance, expressed also as

\[
d_C(\pi, \sigma) = K - C(\pi^{-1} \sigma),
\]

where \(C(\eta)\) is the number of cycles in \(\eta\). It can be shown that both \(d_K\) and \(d_C\) are decomposable into the sum of independent terms, and Fligner and Verducci [5] exploited such a property to formulate multi-parametric generalizations of the basic DB. The decomposition into sums is appealing also because it leads to a closed form expression for \(Z(\lambda)\). The normalizing constant, in fact, could be computationally demanding, as it requires the summation over all possible rankings. A convenient simplification relies on the relation of \(Z(\lambda)\) with the moment generating function of the random distance \(D(\pi, \sigma)\), as stressed in
Hence, when performing a statistical ranked data analysis, one should balance between interpretation purposes, choosing the distance which best accommodates the problem at hand and computational feasibility. In light of these motivations, in our application to the LFPD data, we chose the Kendall distance \( d_K \) for the specification of the DB, whereas both \( d_K \) and \( d_c \) were employed in the inferential procedure for the novel model. These aspects will be better clarified later in the paper. For a more complete and detailed description of the metrics on rankings, the reader is referred to [6] and [7].

2.2.2. The Plackett–Luce model and related extensions. The PL model is a very popular parametric family of ranking distributions. Its name arises from independent contributions supplied by both [8] and [9]. The monograph in [8] provides an in-depth theoretical description of the individual choice behavior based on a general axiom, whereas Plackett [9] derived the PL in the context of horse races. Its probabilistic expression moves from the decomposition of the ranking process in independent stages, one for each rank that has to be assigned, combined with the underlying assumption of standard forward procedure on the ranking elicitation. In fact, a ranking can be elicited through a series of sequential comparisons in which a single item is preferred to all the remaining alternatives and, after being selected, is removed from the next comparisons. For this reason, the PL is said to belong to the family of multistage ranking models. Specifically, the PL probability distribution is completely specified by the so-called support parameter vector \( p = (p_1, \ldots, p_K) \), where \( p_i > 0 \) for all \( i = 1, \ldots, K \) and \( \sum_{i=1}^{K} p_i = 1 \). Note that in the present PL formulation, the support parameters are constrained to add up to one; this restriction is introduced to avoid nonidentifiability due to possible multiplication with an arbitrary positive constant. The generic parameter \( p_i \) expresses the probability that item \( i \) is selected at the first stage of the ranking process and hence preferred among all other items. The probability of choosing item \( i \) at lower preference levels \( t > 1 \) is proportional to its support value \( p_i \). Because the set of available items in the sequence of random selections is reduced by one element after each step, the computation of the choice probabilities at each stage requires suitable normalization of the support probabilities WRT the set of remaining items. It follows that under the PL model, the probability of the random ordering \( \pi^{-1} \) is

\[
P(\pi^{-1} \mid p) = \frac{p_{\pi^{-1}(1)} \cdot \sum_{v=1}^{K} p_{\pi^{-1}(v)}}{\sum_{i=1}^{K} p_{\pi^{-1}(i)}} \quad \pi^{-1} \in S_K.
\] (2.3)

The vase model metaphor, originally introduced by Silverberg [10], is an alternative way to interpret the aforementioned random stage-wise item selections and a useful representation in order to understand PL extensions developed in the literature (see [1] for a review). Let us consider a vase containing item-labeled balls such that the vector \( p \) expresses the starting vase composition. The vase differs from an urn simply because the former contains an infinite number of balls in order to allow continuous values of the proportions. At the first stage, a ball is drawn, and the corresponding item is ranked first. At the second stage, another ball is drawn from the vase: if its label is different from \( \pi^{-1}(1) \), the rank 2 is then assigned to the corresponding item; otherwise, the ball is put back in the vase, and the drawing is repeated until a distinct item is chosen and then ranked in the second position. The multistage experiment ends when there is only one item not yet selected, and this is automatically ranked last. The probability of a generic sequence of drawings turns out to be (2.3). In such a scheme, the vase configuration is constant over all stages, and interactions among items are not contemplated. A first attempt to generalize this basic scheme consists in retaining the absence of item interactions but letting the vase composition vary among stages, as formalized in [10]. In this model setting, the support parameters become stage-dependent, that is, \( p_{ni} \) for \( t = 1, \ldots, (K-1) \) and \( i = 1, \ldots, K \). Setting the special form \( p_{ni} = p_i^{nt} \), one obtains the Benter model (BM) introduced by Benter [11], that is,

\[
P(\pi^{-1} \mid p, \alpha) = \frac{p_{\pi^{-1}(1)}^{nt} \cdot \sum_{v=1}^{K} p_{\pi^{-1}(v)}^{nt}}{\sum_{i=1}^{K} p_{\pi^{-1}(i)}^{nt}} \quad \pi^{-1} \in S_K.
\]

The BM is characterized by the additional vector \( \alpha = (a_1, \ldots, a_K) \) with \( 0 \leq a_t \leq 1 \) for all \( t = 1, \ldots, K \) containing the dampening parameters. These parameters accommodate for the possible different degree of accuracy of the choice at each selection stage. The PL corresponds to the BM with \( a_t = a = 1 \) for all \( t = 1, \ldots, K \), meaning that all ranks are attributed with the maximum degree of care. In order to overcome overparametrization problems, one assumes \( a_1 = 1 \) and \( a_K = 0 \) in any inferential analysis.
[12, 13]. Relaxing also the noninteraction hypothesis, so that the vase configuration at each stage relies on the previous selected items, Plackett [9] defined a hierarchy of further PL extensions. Marden [1] refers to such generalizations as Lag L models, where \( L = 0, \ldots, K - 2 \) indicates that the vase at stage \( t \) depends on the previous choices only through the last \( L \) selected items \( \{\pi^{-1}(t - L), \ldots, \pi^{-1}(t - 1)\} \). The Lag 0 model coincides with the ordinary PL; the Lag 1 model is such that at each choice step \( t \), the vase depends only on the item \( \pi^{-1}(t - 1) \). In general, the total number of parameters in the Lag \( L \) model is given by \( K(K - 1) \cdots (K - L) - 1 \); thus, the \( L = K - 2 \) model covers the whole SM.

2.3. Novel extension of the Plackett–Luce model

In this section, we introduce an original proposal, which generalizes the standard PL. Multistage ranking models previously reviewed implicitly suppose that preferences are expressed with the canonical forward order, proceeding from the assignment of the first rank up to the last one. This is just a preliminary assumption, and other reference orders can be contemplated, but, to our knowledge, this aspect has not been addressed in the literature. Indeed, even the individual experience in choice problems suggests the plausibility of alternative paths for the ranking elicitation. For example, one can think of situations where the judge has a clearer perception about her most-liked and least-liked items but only a vaguer idea relative to middle ranks; alternatively, the ranker could build up her best alternatives following an exclusion process starting with the final position, which would be better described by a backward model. Besides these motivations aimed at characterizing typical behaviors in real choice/selection problems, we are also interested in developing a more flexible tool in order to improve the description of phenomena observed in the form of ordered data. All these intuitive and practical instances make the forward hypothesis too restrictive when approaching a flexible inferential analysis of a ranking data set. Hence, we propose to extend the PL in the following way: rather than fixing \( a \ priori \) the stepwise order leading the judge to her final ranked sequence, we represent it with a specific free model parameter \( \rho \in S_K \) and let data guide inference about the reference order followed in the rank assignment scheme. Such an approach would also alleviate the asymmetry toward ranks assigned at the extreme (the first and the last) stages of the ranking procedure, which affects the PL (see also Appendix B). It turns out that the reference order \( \rho = (\rho(1), \ldots, \rho(K)) \) is the result of a bijection between the stage set \( S \) and the rank set \( R \), that is,

\[
\rho : S \to R,
\]

where the entry \( \rho(t) \) indicates the rank attributed at the \( t \)-th stage of the ranking process. Then, \( \rho \) identifies a discrete parameter taking values in \( S_K \). The composition of an ordering \( \pi^{-1} \) with a reference order \( \rho \) yields

\[
\eta^{-1} = \pi^{-1} \rho,
\]

representing the sequence which lists the items selected at each stage. This means that \( \eta^{-1}(t) = \pi^{-1}(\rho(t)) \) is the item chosen at step \( t \) and receiving rank \( \rho(t) \). The probability of a random ordering under the EPL model can be written as

\[
P_{EPL}\left(\pi^{-1} | \rho, p\right) = P_{PL}\left(\pi^{-1} \rho | p\right) = \prod_{t=1}^{K} \frac{P_{\pi^{-1}(\rho(t))}}{\sum_{v=t}^{K} P_{\pi^{-1}(\rho(v))}} \pi^{-1} \in S_K, \tag{2.4}
\]

where the additional discrete parameter \( \rho \) acts directly by composition on the right-hand side of the generated outcome of a standard PL. It follows that by construction, the EPL inherits from the PL the property of label-invariance formalized by Critchlow et al. [4]. Hereafter, we will shortly refer to (2.4) as EPL(\( \rho, p \)). The vector \( p \) still denotes the support parameters, with the probabilities for each item to be selected at the first stage and to be ranked in the position given by the first entry in \( \rho \). Obviously, the standard PL is a special case of the EPL, obtained setting \( \rho \) equal to the identity permutation \( \varepsilon = (1, 2, \ldots, K) \). Similarly, when \( \rho = (K + 1 - \varepsilon, \ldots, K) \), one has the backward PL. For a toy example elucidating the definition and the notation of the EPL given in (2.4), the reader is referred to Appendix A.

From a theoretical point of view, the novel EPL is a proper generalization of the (2.3) if and only if such a new class covers a wider portion of the SM, that is, it allows to describe additional probability functions that can not be derived from the PL with any parameter specification. In other words, one should give a formal proof concerning the existence of a ranking distribution, generated by the new EPL, which does
not belong to the standard PL family. Such a proof is given in Appendix B. In Section 3.1, we describe in detail the MLE for such a new model.

2.4. Finite mixture modeling for ranked data

One of the formal properties satisfied by the DB with \( d = d_K \) is strong unimodality, meaning that the probability of a ranking decreases as the distance from the modal sequence increases [1]. However, strong unimodality is expected to be violated in real data, especially when the sample composition is heterogeneous with respect to factors related to the ranking elicitation. A well-established statistical tool to address inference in the presence of unobserved heterogeneity is given by the finite mixture approach. A finite mixture model assumes that the population consists of a finite number \( G \) of subpopulations. In this setting, the probability of observing the ranking \( \pi_s \) for the \( s \)-th unit is

\[
f(\pi_s) = \sum_{g=1}^{G} \omega_g f_g(\pi_s), \quad \pi_s \in S_K,
\]

where \( f_g(\cdot) \) denotes the \( g \)-th component of the mixture, that is, the statistical distribution of data within the \( g \)-th group, and \( \omega_g \) is the probability for the \( s \)-th observation to belong to the \( g \)-th group. The membership probabilities \( \omega = (\omega_1, \ldots, \omega_G) \) are usually termed weights of the mixture. Mixture components are often modeled with members of the same parametric family, that is, \( f_g(\cdot) = f(\cdot | \theta_g) \in \{ f(\cdot | \theta) : \theta \in \Theta \} \) for all \( g = 1, \ldots, G \) and thus, they are identified by the group-specific parameter \( \theta_g \). For a more extensive introduction to finite mixture models, the reader can refer to [14]. After the early seminal papers [15, 16], several more recent mixture model applications account for unobserved heterogeneity and make the ranked data modeling more flexible. For example, Murphy and Martin [17] analyzed the popular 1980 American Psychological Association presidential election data set (specifically the subdata set of complete rankings) fitting a mixture of DB models. They aimed at inquiring voters’ orientation toward candidates within the electorate, assessing the possible adequacy to incorporate a noise component (UM) in the mixture. Such a component, in fact, could collect outliers and/or observations characterized by atypical preference profiles with a possible final improvement of model fitting. Similar methods have been adopted in other preference studies. Gormley and Murphy [18] fitted a mixture of PL to the 2000 Central Applicant Office data set to investigate what motivates Irish college applicants in their degree course choice. Gormley and Murphy [12] applied mixtures of both PL and BM to infer the structure of the Irish political electorate and characterize voting blocks. In more recent works, the same authors attempted to extend the mixture approach in different directions (for further details, see [13] and [19]). The reader can find further examples and references in [1].

In Section 4.2, we will present our mixture model application to data originated from the LFPD bioassay experiment. For the analysis of the LFPD data set, we will implement different mixture models, adopting as mixture components elements from the following parametric families:

- DB with \( d = d_K \);
- PL with known forward and backward reference order;
- BM;
- our novel EPL.

DB and PL represent two of the most frequently used distributions for inferring ranking data, and both parameterizations allow a clear interpretation. In the former case, the central ranking summarizes the overall profile in assessing the orderings of the items, whereas the concentration parameter expresses how representative the modal ranking is. In the latter case, a higher value of the item support parameter implies a greater probability for that item to be preferred at each selection level. The implementation of the BM mixture facilitates investigation of how the BM and the EPL compete in terms of flexibility, although they generalize the PL in totally different directions.

3. Inferring ranking models

3.1. Maximum likelihood estimates of the mixture of extended Plackett–Luce models

In this section, we only provide inferential details for the novel extended model, starting with the simpler case of homogenous population (\( G = 1 \)). As mentioned before, in fact, the conventional forward PL is a
of PL can be easily derived from the mixture of EPL with all known reference orders \( \rho_s = e \). However, explicit estimation formulas for this special case can be found in [18], whereas inference concerning the mixture of BM is detailed in [12]. The inferential procedure for MLE of mixture of DB with \( d = d_k \) is described in [17].

Postulating the EPL(\( \rho, p \)) as the underlying mechanism generating the observed orderings \( \pi^{-1} = (\pi_1^{-1}, \ldots, \pi_N^{-1}) \), the log-likelihood has the following expression

\[
l(\rho, p) = N \sum_{i=1}^{K} \log p_i - \sum_{j=1}^{N} \sum_{i=1}^{K} \log \left( \sum_{j=1}^{K} p_{\pi^{-1}(\rho(v))} \right).
\]

Note that in order to find MLE solutions, the direct maximization of the log-likelihood WRT the \( p \)'s is made arduous by the presence of the annoying term \( \log \left( \sum_{j=1}^{K} p_{\pi^{-1}(\rho(v))} \right) \). Therefore, we derived the estimation formula for the support parameters borrowing the approach detailed in [20] and based on the minorization/maximization (MM) algorithm. This iterative optimization method is reviewed in its general form by Lang et al. [21, 22], whereas Hunter [20] discusses the specific application of the MM algorithm for the PL estimation. The basic idea consists in performing the optimization step for the \( p \)'s on a surrogate objective function rather than on (3.1). The surrogate is obtained by exploiting the strict convexity of \(- \log \left( \sum_{j=1}^{K} p_{\pi^{-1}(\rho(v))} \right)\), and in particular, the supporting hyperplane property for convex functions. From Taylor’s linear expansion, in fact, one has

\[
- \log \left( \sum_{j=1}^{K} p_{\pi^{-1}(\rho(v))} \right) \geq 1 - \log \left( \sum_{j=1}^{K} l_{\pi^{-1}(\rho(v))} \right) - \sum_{j=1}^{K} \frac{1}{\sum_{j=1}^{K} p_{\pi^{-1}(\rho(v))}} \cdot
\]

and disregarding terms not depending on \( p \), the minorizing auxiliary objective function can be written as

\[
q = N \sum_{i=1}^{K} \log p_i - \sum_{j=1}^{N} \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{1}{\sum_{j=1}^{K} p_{\pi^{-1}(\rho(v))}}.
\]

As emphasized by Hunter [20], the convenience of optimizing the more tractable (3.2) in place of (3.1) is the possibility to estimate each support parameter \( p_i \) separately. Furthermore, the iterative maximization of \( q \) returns a sequence \( p^{(1)}, p^{(2)}, \ldots \) that is guaranteed to converge at least to a local maximum of the original objective function. Thus, differentiating WRT each \( p_i \) and equating the partial derivatives to zero, at the current iteration, one obtains the following updating rule for the support parameters

\[
p^{(i+1)}_i = \frac{N}{\sum_{j=1}^{N} \sum_{i=1}^{K} \delta^{(i)}_{ji}} \quad i = 1, \ldots, K,
\]

where

\[
\delta^{(i)}_{ji} = \begin{cases} 
1 & \text{if } i \in \{ \pi_1^{-1}(\rho^{(t)}(t)), \ldots, \pi_1^{-1}(\rho^{(t)}(K)) \} \\
0 & \text{otherwise}
\end{cases}
\]

corresponds to the binary indicator for the event that item \( i \) is still available at stage \( t \) or, equivalently, which is not selected by unit \( s \) before stage \( t \). Notice that the binary array has a superscript because of the dependence on the \( \rho = \rho^{(t)} \) available at the current iteration. Using the original log-likelihood, the update of the reference order parameter is derived as follows

\[
\rho^{(i+1)} = \arg \min_{\rho} \sum_{i=1}^{N} \sum_{j=1}^{K} \frac{1}{\sum_{j=1}^{K} p_{\pi^{-1}(\rho(v))}}.
\]

Solving (3.3) with a global search in \( S_K \) is prohibitive when \( K \) is large, as in our application to the LFPPD data. So, we implemented a local search similarly to the method suggested by Busse et al. [23] and Lee and
Yu [24], constraining the optimization within a fixed distance from the current estimate of the reference order \(\rho^{(0)}\). To evaluate the sensitivity of the algorithm WRT the choice of a particular distance in the local search step, we focused on both the Kendall and the Cayley distance and compared the corresponding estimation performances.

Now, we relax the hypothesis of homogeneous population and consider a more flexible mixture model with EPL components, discussing the related inferential issues. Augmenting data with the missing individual group membership indicator \(z_s = (z_{s1}, \ldots, z_{sG})\), such that

\[
z_{sg} = \begin{cases} 
1 & \text{if } s\text{-th unit belongs to the } g\text{-th mixture component} \\
0 & \text{otherwise,}
\end{cases}
\]

one obtains the following expression for the complete log-likelihood

\[
l_C(\rho, p, \omega, z) = \sum_{s=1}^N \sum_{g=1}^G z_{sg} \left[ \log \omega_g + \sum_{i=1}^K \log p_{gi} - \sum_{i=1}^K \log \left( \sum_{v=1}^K \sum_{g' = 1}^G p_{g' \pi_{v-1}^{-1}(\rho_{g'}(v))} \right) \right].
\]

In the EM algorithm, which represents one of the major schemes to address MLE in the presence of missing data [25], the maximization problem is transferred to the expectation of the \(l_C\) WRT the posterior distribution of the latent variables \(\tilde{z}\) represented by \(\hat{z}\), that is,

\[
Q = \mathbb{E} \left[ l_C^{\hat{z}^{-1}}; \rho, p, \omega \right] = \sum_{s=1}^N \sum_{g=1}^G \hat{z}_{sg} \left[ \log \omega_g + \sum_{i=1}^K \log p_{gi} - \sum_{i=1}^K \log \left( \sum_{v=1}^K \sum_{g' = 1}^G p_{g' \pi_{v-1}^{-1}(\rho_{g'}(v))} \right) \right],
\]

where

\[
\hat{z}_{sg}^{(l+1)} = \frac{\omega_g^{(l)} \mathbb{P}_{\text{EPL}} \left( \pi_{v-1}^{-1}(\rho_{g'}(v)) \mid \hat{z}_{sg}^{(l)} \right)}{\sum_{g'=1}^G \omega_{g'}^{(l)} \mathbb{P}_{\text{EPL}} \left( \pi_{v-1}^{-1}(\rho_{g'}(v)) \mid \hat{z}_{sg}^{(l)} \right)},
\]

for \(s = 1, \ldots, N\) and \(g = 1, \ldots, G\). Similarly to [18], we combined the EM with the MM algorithm into a hybrid version of the former, called EMM algorithm, using the minorizing surrogate function \(q\) in place of the original \(Q\) such that

\[
Q \geq q = \sum_{s=1}^N \sum_{g=1}^G \hat{z}_{sg} \sum_{i=1}^K \log p_{gi} - \sum_{s=1}^N \sum_{g=1}^G \hat{z}_{sg} \sum_{i=1}^K \sum_{v=1}^K \sum_{g' = 1}^G p_{g' \pi_{v-1}^{-1}(\rho_{g'}(v))},
\]

The maximization of \(q\) leads to the following updating rule for \(p_{gi}\) at the current iteration

\[
p_{gi}^{(l+1)} = \frac{\sum_{s=1}^N \hat{z}_{sg}^{(l+1)}}{\sum_{s=1}^N \hat{z}_{sg}^{(l+1)} \sum_{i=1}^K \sum_{v=1}^K \sum_{g' = 1}^G p_{g' \pi_{v-1}^{-1}(\rho_{g'}(v))}},
\]

for \(g = 1, \ldots, G\) and \(i = 1, \ldots, K\) with

\[
\hat{z}_{sg}^{(l)} = \begin{cases} 
1 & \text{if } i \in \pi_{v-1}^{-1}(\rho_{g'}(t)), \ldots, \pi_{v-1}^{-1}(\rho_{g'}(K)) \\
0 & \text{otherwise,}
\end{cases}
\]

indicating if, under the group-specific reference order \(\rho_g\), the unit \(s\) has not extracted the \(i\)-th item before stage \(t\), and hence if, at that step, it still belongs to the set of available alternatives or not. The estimate for the reference orders in each subgroup is the solution of the optimization problem given by

\[
\rho_{g}^{(l+1)} = \arg \min_{\rho} \sum_{s=1}^N \hat{z}_{sg}^{(l+1)} \sum_{i=1}^K \log \left( \sum_{v=1}^K \sum_{g' = 1}^G p_{g' \pi_{v-1}^{-1}(\rho_{g'}(v))} \right),
\]

\(g = 1, \ldots, G\).
which is solved locally as in the case $G = 1$. The M-step ends with the update of the mixture weights, computed as the posterior proportions of sample units belonging to each group

$$
\omega_g^{(l+1)} = \frac{\sum_{i=1}^{N} \hat{z}_i^{(l+1)} g}{N}, \quad g = 1, \ldots, G.
$$

Finally, to address the issue of local maxima, we run the algorithm with a suitably large number of different starting values.

3.2. Algorithm convergence and model selection

We conducted MLE inference relying on the EM algorithm and on a hybrid version thereof. We developed a suite of functions written in R language [26], which are available upon request from the first author. In these estimation procedures, the log-likelihood is iteratively maximized until convergence is achieved. As suggested by McLachlan and Peel [14], the Aitken acceleration criterion has been employed as stopping rule in place of the standard lack of progress criterion based on the absolute/relative increment of the log-likelihood. For a discussion on the relative merits of the Aitken acceleration criterion and other related proposals, see [27]. As far as convergence of the proposed algorithm is concerned, it is difficult to provide theoretical support. In fact, we can not exploit well-known sufficient conditions provided in [28] and [29] because of the discreteness of the $\rho$ parameter component. To our knowledge, in the presence of mixed-type parameter space, no previous positive result has been developed. However, we have always experienced a strictly monotone likelihood updating in our algorithm, with stopping rule achieved in a suitable number of iterations. A similarly convincing behavior of the EM algorithm in the presence of mixed-type parameter space is found also in the successful implementations for mixtures of DB models in [17] and [24].

Another crucial issue in a mixture model setting is the choice of the number of components. In the statistical literature, this problem is addressed with several criteria; we opted for the popular $BIC$

$$
BIC = -2 l \left(\hat{\theta}_{ML}\right) + \nu \log N,
$$

where $l \left(\hat{\theta}_{ML}\right)$ is the maximized log-likelihood and $\nu$ is the number of free parameters. The BIC, introduced by [30], is a measure which balances between two conflicting goals typically aimed at when fitting a statistical model: good fit and parameter parsimony, where the latter is modulated through the penalty term. In the presence of competing mixture models, the one associated with the lowest BIC value is preferred. In the next section, we detail MLE results derived from alternative mixture models fitted to the LFPD data set.

4. Statistical analysis of large fragment phage display data

4.1. The large fragment phage display data set

Our investigation is motivated by a real data set coming from a new technology for epitope mapping of the binding between the antibodies present in a biological tissue and a target protein. The biological foundation of the experiment is detailed in [2] and consists of repeated binding measurements of human blood exposed to $K = 11$ partially overlapping fragments of the HER2 oncoprotein, denoted sequentially by Hum 1, · · · , Hum 11 (Figure 1). Researchers were originally interested in testing the validity of their

![Figure 1](https://example.com/figure1.png)

**Figure 1.** 1-D scheme of the HER2 oncoprotein and its segmentation into 11 partially overlapping fragments (Hum) employed in the large fragment phage display bioassay experiment. Hum 12 and Hum 13 indicate, respectively, the whole HER2 oncoprotein (positive control) and the empty phage vector (negative control).
4.2. Mixture models for the large fragment phage display data

The original raw absorbance data derived from the LFPD experiment were somehow wildly fluctuating and looked indeed very heterogeneous as apparent in Figure 2. However, there were certainly some manifest peaks corresponding to recurrent fragments, especially high for some patients, most frequently those diagnosed with cancer. It is also apparent that the individual absorbance profiles are measured at different mean levels for different patients and with different variability. A simple logarithmic transformation and recentering WRT the individual mean were performed providing some more stable evidence of differential profiles among groups. However, there are some specific profiles which seem pretty much overlapped among different subgroups, although with some different overall pattern (Figure 3).

Since data emerged from the development of an innovative technology, miscalibrations or inaccuracies of the measuring device may occur, and/or subject-specific characteristics may alter somehow the observed numerical outcome. This makes it more difficult to adjust the statistical analysis on the basis of the original quantitative scale measurements.
Figure 3. Mean-centered log-absorbance profiles for the three groups of patients in the large fragment phage display study: HD = healthy (green), EBC = diagnosed with breast cancer in an early stage (red), and MBC = diagnosed with metastatic breast cancer (blue). Each continuous piecewise linear function represents the mean-centered log-absorbance levels in the HER2 oncoprotein fragments (Hum) of a single experimental unit. Hum 12 and Hum 13 indicate, respectively, the whole HER2 oncoprotein (positive control) and the empty phage vector (negative control).

of raw or ad hoc preprocessed data. Unfortunately, for this kind of data, a consolidated and fully shared normalization technique is lacking. For all these reasons, rather than basing our analysis on the quantitative output of the LFPD study, we verified the possible usefulness of the ranking profiles as a more robust and unambiguously defined evidence, capable to capture and characterize the sample heterogeneity. Hence, we first derived ordered sequences ranking the absorbance levels of the individual protein fragments taken in decreasing order (Rank 1 = highest value; Rank K = lowest value). We performed a simple exploratory analysis by cancer state computing both the $K \times K$ first-order marginal matrices $\hat{M}$ and the so-called Borda orderings $\pi^{-1}$. The generic entry $\hat{M}_{ij}$ in the marginal matrix denotes the observed relative frequency that item $i$ is ranked $j$-th, whereas the sequence $\pi^{-1}$ lists items taken in order from the highest to the lowest mean rank. These matrices are displayed as image plots in Figure 4, together with the Borda sequences at the bottom of each panel. The color intensity is an increasing function of the corresponding observed frequency. The analysis of the first-order marginal matrices suggests that some protein fragments are very often associated with lower ranks, as pointed out by the presence of darker rectangles in correspondence of bottom positions. This constantly occurs for all disease subgroups with Hum 10, but some interesting differential evidence is apparent for EBC subjects with Hum 5 and 6, for MBC with Hums 2 and 13, and also for HD patients with Hum 9 (Figure 4). Such a precious discriminant information could be better captured by our EPL. To validate this claim, we fitted both the PL and the new EPL to the three disease subgroups separately. For the former, we used two known orders, forward (PL-$\rho_1$) and backward (PL-$\rho_2$), whereas for the latter, the reference order is a parameter to be estimated. Estimation performances are shown in terms of BIC values in Table I and reveal that the EPL fit is better or at most comparable with those relative to the PL with fixed reference orders. The interest in relaxing the traditional forward assumption is supported also by the BIC values for the PL-$\rho_2$, showing that such a model constantly outperforms the PL-$\rho_1$ when fitted to HD and MBC subjects. These BIC results represent a strong evidence motivating the need of a PL extension.
In what follows, we considered a more comprehensive analysis in a mixture model setting. With this approach, we aimed at

- addressing the heterogenous nature of the LFPD data using the evidence provided by the orderings of the absorbance levels;
- assessing if and how the path in the sequential ranking process can impact the final model-based classification and select the most appropriate one;
- identifying possible characteristic subgroups related to the disease state;
- characterizing each subgroup with the estimates of the cluster-specific parameters.

4.3. Empirical findings from mixture models fitted to large fragment phage display data

Considering all the 67 available orderings, we fitted mixtures of DB with d = d_K (DBmix), mixtures of PL with both forward and backward reference order (PLmix-ρ_1 and PLmix-ρ_2), mixtures of BM (BMmix) with dampening parameters shared by all groups as in [12], and mixtures of EPL. In the most general mixture model (EPLmix), each EPL component has a group-specific parameter ρ_g to be inferred. We have also considered a constrained version with an unknown reference order ρ common to all components, and we have denoted it with PLmix-ρ. All mixtures have been implemented with a number of components...
Table I. BIC values and corresponding differences \( \Delta \text{BIC} \) with respect to the best fitting model.

| Model  | HD BIC | \( \Delta \text{BIC} \) | EBC BIC | \( \Delta \text{BIC} \) | MBC BIC | \( \Delta \text{BIC} \) |
|--------|--------|----------------|--------|----------------|--------|----------------|
| \( PL_\rho_1 \) | 694.04 | 17.11 | 776.13 | 2.96 | 499.46 | 25.56 |
| \( PL_\rho_2 \) | 685.85 | 8.92 | 804.61 | 31.44 | 498.67 | 24.77 |
| EPL  | **676.93** | 0 | **773.17** | 0 | **473.90** | 0 |

\( K = 13 \)

| Model  | HD BIC | \( \Delta \text{BIC} \) | EBC BIC | \( \Delta \text{BIC} \) | MBC BIC | \( \Delta \text{BIC} \) |
|--------|--------|----------------|--------|----------------|--------|----------------|
| \( PL_\rho_1 \) | 899.63 | 25.92 | **1025.71** | **0** | 658.44 | 28.39 |
| \( PL_\rho_2 \) | 894.44 | 20.73 | 1039.45 | 13.74 | 652.15 | 22.10 |
| EPL  | **873.71** | **0** | **1026.71** | **0** | **630.05** | **0** |

Maximum likelihood estimates of \( PL_\rho_1, PL_\rho_2 \), and EPL have been computed separately for the three disease groups (HD = healthy, EBC = diagnosed with early stage breast cancer, and MBC = diagnosed with metastatic breast cancer) and for a different number \( K \) of binding probes included in the rankings. The smallest BIC and \( \Delta \text{BIC} \) values indicating the best fitted models are highlighted in bold.

Table II. BIC values resulting from the maximum likelihood estimates of the DBmix on the large fragment phage display data with a varying number \( G \) of components, when either \( K = 11 \) or \( K = 13 \) binding probes are included in the rankings.

| \( G \) | \( K = 11 \) | \( K = 13 \) |
|--------|--------------|--------------|
| 1      | 2078.77      | 2700.02      |
| 2      | 2003.65      | 2617.66      |
| 3      | 1940.86      | 2551.38      |
| 4      | 1899.33      | 2512.19      |
| 5      | 1882.32      | 2483.25      |
| 6      | 1863.17      | 2451.38      |
| 7      | 1846.98      | 2421.71      |
| 8      | 1829.81      | 2392.10      |
| 9      | 1817.06      | 2366.78      |
| 10     | 1798.12      | 2342.60      |

varying from \( G = 1 \) to \( G = 7 \). Of course, the case \( G = 1 \) coincides with the assumption that observations come from a homogeneous population without an underlying group structure. We separately applied all mixture models to the ranked absorbance levels relative to the \( K = 11 \) partially overlapping protein fragments as well as to the \( K = 11 + 2 \) binding probes (spots), including also the whole HER2 oncoprotein (positive control) and the empty phage vector (negative control).

Focusing on the BIC for \( G = 1 \) compared with \( G > 1 \), the MLE of the DBmix provided an overall evidence in favor of heterogeneity when both \( K = 11 \) or \( K = 13 \) binding probes are considered. We highlighted a remarkable decreasing behavior for the associated BIC, which persists even when the fitting is carried out up to \( G = 10 \) components as shown in Table II.

Indeed, fitting DBmix with an increasing number of groups pointed out a particular feature of the DB, probably because of the sparse nature of LFPD data. We remind, in fact, that in the present application, the sample size is small WRT the cardinality (11! or 13!) of the discrete ranking space. As the value of \( G \) in the DBmix increases, some components start to represent just a single observation. This can be explained, perhaps, by the fact that, once the modal ranking \( \sigma \) has been fixed, DB has only one parameter left for fitting the amount of the uncertainty around \( \sigma \). It follows that for these components, the concentration parameter \( \lambda \) is typically estimated as a very high value. This behavior, of course, could make the DBmix model not sufficiently parsimonious and suitable in some sparse-data situations because it could lead to a more sparse clustering of the observations and to a less enlightening inferential findings.

When stage-wise models were fitted to the LFPD data, we found again evidence in favor of the heterogeneous structure, as indicated in Table III. In the case \( K = 11 \), all types of mixtures consistently identify four groups in the sample. When also the control probes are included in the ordered sequences, five groups are consistently selected with the only exception of PLmix-\( \rho_2 \). Bold BIC values in Table III
Table III. BIC values, corresponding differences ΔBIC WRT the best fitting model and number $G$ of components of the best PLmix-$\rho_1$, PLmix-$\rho_2$, BMmix, PLmix-$\rho$, and EPLmix fitted to large fragment phage display data for different number $K$ of binding probes included in the ranking.

| Mixture model | $K = 11$ | $K = 13$ |
|---------------|----------|----------|
|               | BIC      | ΔBIC     | $G$ | BIC      | ΔBIC     | $G$ |
| PLmix-$\rho_1$ | 1964.87  | 29.10    | 4   | 2589.93  | 71.37    | 5   |
| PLmix-$\rho_2$ | 1984.53  | 48.76    | 4   | 2619.60  | 101.04   | 4   |
| BMmix         | 1995.09  | 59.32    | 4   | 2615.45  | 96.89    | 5   |
| PLmix-$\rho$  | 1969.24  | 33.47    | 4   | 2597.42  | 78.86    | 5   |
| EPLmix        | **1935.77** | **0**   | 4   | **2518.56** | **0**   | 5   |

The smallest BIC and ΔBIC values indicating the best fitted models are highlighted in bold.

Figure 5. BIC trends resulting from the maximum likelihood estimates of the PLmix-$\rho_1$, PLmix-$\rho_2$, BMmix, PLmix-$\rho$, and EPLmix on the LFPD data with varying number $G$ of mixture components, when either $K = 11$ (left) or $K = 13$ (right) binding probes are included in the ranking. The symbol $\bullet$ indicates the minimum BIC values for the final selection of the number of groups.

point out the EPLmix as the best model. Optimal BIC values of the EPLmix are, in both cases, significantly smaller than those corresponding to the competing mixtures, as also stressed by ΔBIC values in the same table. Indeed, the outperformance of EPLmix for any fixed $G$ of the mixture is apparent in Figure 5. Hence, the introduction of the discrete parameter in the mixture component leads to a remarkable improvement of fit when it is allowed to be variable among groups. Note, in fact, the gap between the BIC trend associated to the more flexible EPLmix and the one relative to its restricted version PLmix-$\rho$.

Comparing also the EPLmix with the BMmix, BIC results apparently show that the larger flexibility provided by the BM does not lead to an improvement, because with these data, the penalization for a larger number of parameters exceeds the gain in terms of loglikelihood. As regards the comparison between the Kendall and the Cayley distance employed in the local search step, neither of the two metrics yielded a consistently better solution when fitting $G = 1, \ldots, 7$ groups. We decided to report in Table III only the BIC values obtained with the Kendall distance.

The selected EPLmix exhibits a good accuracy in the discrimination of sample units WRT the real disease state. The two resulting clusterings, in fact, agree with the most relevant distinction of the real disease state between healthy and unhealthy patients, as pointed out in Table IV(a) and (b). Specifically, collapsing the model-based group membership into the aforementioned basic bipartition, we recognize that healthy subjects are well isolated, with only one or two false positive cases; for diseased patients, instead, we have nine misclassifications in the $K = 11$ case but only four with the addition of the control spots (Table IV(a) and (b)). As expected, the inclusion of the positive and negative controls provided an additional discriminating power, measured by the increment of the adjusted rand index (ARI) from 0.52 to 0.77. Healthy
Table IV. Correspondence between the model-based clustering derived from the MLE of the EPLmix and the true disease state of the large fragment phage display experimental units: HD = healthy, EBC = diagnosed with early stage breast cancer, and MBC = diagnosed with metastatic breast cancer.

| Disease state | Group 1 | Group 2 | Group 3 | Group 4 |
|---------------|---------|---------|---------|---------|
| HD            | 0       | 2       | 10      | 8       |
| EBC           | 13      | 12      | 2       | 21      |
| MBC           | 0       | 15      | 3       | 1       |

Corresponding values for the adjusted rand index based on the basic healthy/unhealthy bipartition are shown in parentheses.

Table V. Modal orderings derived from the best PLmix-\(\rho_1\), PLmix-\(\rho_2\), and EPLmix fitted to large fragment phage display data for a different number \(K\) of binding probes included in the rankings.

| Mixture model | \(K = 11\) | \(K = 13\) |
|---------------|------------|------------|
|               | \(g\)     | D.C.       | \(\hat{\delta}_g^{-1}\) | \(g\)     | D.C.       | \(\hat{\delta}_g^{-1}\) |
| PLmix-\(\rho_1\) | 1         | (11,1,3)  | (6,1,5,4,7,11,3,8,10,9,2)* | 1         | (2,11,3)  | (12,1,11,7,8,9,2,13,3,4,6,5,10) |
|               | 2         | (0,10,0)  | (9,3,7,11,4,1,8,2,5,6,10)  | 2         | (6,0,0)   | (7,2,1,2,8,11,13,5,6,10,4,3,9) |
|               | 3         | (3,17,15) | (1,11,7,8,3,9,4,5,2,6,10)  | 3         | (0,7,14)  | (9,3,4,12,11,7,18,13,2,5,6,10) |
|               | 4         | (6,0,1)   | (7,2,1,8,11,5,6,4,10,3,9)  | 4         | (0,7,14)  | (1,12,11,7,8,3,5,9,4,6,2,13,10) |
|               |           |           |                   | 5         | (12,3,2)  | (1,5,6,7,3,4,11,12,8,9,10,2,13)* |
| PLmix-\(\rho_2\) | 1         | (14,5,3)  | (1,6,11,7,5,8,2,4,9,3,10)* | 1         | (12,2,2)  | (1,6,12,5,7,11,4,2,13,3,10,9,8)* |
|               | 2         | (6,0,1)   | (7,2,1,8,11,5,6,3,4,10,9)  | 2         | (0,15,0)  | (12,9,3,7,4,11,13,8,2,5,6,10) |
|               | 3         | (0,12,15) | (1,17,11,8,3,9,4,5,2,6,10) | 3         | (1,11,17) | (1,2,11,7,8,3,9,5,6,4,13,2,10) |
|               | 4         | (0,11,0)  | (9,3,4,7,11,1,8,2,5,6,10)  | 4         | (7,0,0)   | (7,2,1,2,8,13,11,5,6,3,4,10,9) |
| EPLmix        | 1         | (0,13,0)  | (9,8,1,3,11,7,2,4,5,6,10)  | 1         | (1,12,14) | (1,2,11,7,8,3,4,9,5,6,2,13,10) |
|               | 2         | (2,12,15) | (1,11,7,8,9,3,4,5,6,2,10)  | 2         | (10,0,2)  | (5,2,14,4,3,6,10,7,8,9,12,1,13) |
|               | 3         | (10,2,3)  | (5,4,11,1,6,3,10,2,9,7,8)  | 3         | (0,9,9)   | (9,12,11,3,4,7,2,13,8,5,6,10) |
|               | 4         | (8,1,1)   | (7,2,1,8,11,5,6,4,10,9,3)  | 4         | (1,7,3)   | (12,9,1,11,3,8,7,2,4,5,6,10) |
|               |           |           |                   | 5         | (8,0,0)   | (11,2,1,6,12,8,13,5,7,10,4,3,9) |

*D.C.* stands for ‘disease composition’ and lists sequentially the number of HD = healthy, EBC = diagnosed with early stage breast cancer, and MBC = diagnosed with metastatic breast cancer patients in each group. The symbol * indicates mixture components, which are very close to the UM.

Patients are always modeled by two components in all the fitted mixtures. This hints at possibly different subtypes of healthy profiles. In fact, we can easily verify that such subdivision reflects two different absorbance patterns in cancer-free units, made evident in Figure 2 by the green continuous piecewise linear functions: a first subgroup whose immune defenses essentially did not react at all to the exposition to the HER2 oncoprotein (lower panel) and a second one with some manifest and characterized binding profile (upper panel).

On the other hand, among the components representing diseased patients, the subclassification between EBC and MBC is only partially recovered, especially for the latter subgroup. This is proved by the presence of at least one model-based group entirely composed of EBC subjects in all the fitted mixtures, whereas MBC patients always belong to mixed-type components.

The varying correspondence between the real cancer state and the inferred clustering structure confirms the presumed dependence of the classification results on the adopted reference ranking process \(\rho\). Furthermore, the good agreement obtained with the EPLmix (Table IV(a) and (b)) suggests that researchers should not focus exclusively on differential epitope identification but could extend their analysis considering also a more general global understanding of differential bindings. Hence, in order to characterize disease groups WRT ranking profiles, it is interesting to interpret the component-specific modal orderings (Table V), derived by ordering the corresponding support parameter estimates (Figure 6). Results
Figure 6. Support parameter estimates represented via mosaic plots for the best PLmix-$\rho_1$, PLmix-$\rho_2$, and EPLmix fitted to the large fragment phage display data. Bar widths are proportional to group weights. Upper panel refers to the data with $K = 11$ protein fragments, whereas lower panel concerns the case with $K = 13$ binding probes, including the whole HER2 oncoprotein (Hum 12 = positive control) and the empty phage vector (Hum 13 = negative control).

Table VI. Mixture weights and reference order estimates of the best PLmix-$\rho_1$, PLmix-$\rho_2$, and EPLmix fitted to large fragment phage display data for a different number $K$ of binding probes included in the rankings.

| Mixture model | $K = 11$ | $K = 13$ |
|---------------|----------|----------|
|               | $\hat{\omega}_g$ | $\hat{\rho}_g$ | $\hat{\omega}_g$ | $\hat{\rho}_g$ |
| PLmix-$\rho_1$ | 1 | 0.22 | $\rho_1$ | 1 | 0.24 | $\rho_1$ |
|               | 2 | 0.15 | $\rho_1$ | 2 | 0.09 | $\rho_1$ |
|               | 3 | 0.53 | $\rho_1$ | 3 | 0.11 | $\rho_1$ |
|               | 4 | 0.10 | $\rho_1$ | 4 | 0.31 | $\rho_1$ |
|               |           |           | 5 | 0.25 | $\rho_1$ |
| PLmix-$\rho_2$ | 1 | 0.35 | $\rho_2$ | 1 | 0.25 | $\rho_2$ |
|               | 2 | 0.10 | $\rho_2$ | 2 | 0.22 | $\rho_2$ |
|               | 3 | 0.39 | $\rho_2$ | 3 | 0.43 | $\rho_2$ |
|               | 4 | 0.16 | $\rho_2$ | 4 | 0.10 | $\rho_2$ |
| EPLmix        | 1 | 0.19 | (11,10,9,7,8,4,2,3,6,5,1) | 1 | 0.39 | (2,1,3,4,5,6,8,9,7,10,11,12,13) |
|               | 2 | 0.44 | (1,2,3,4,6,5,7,8,9,10,11) | 2 | 0.18 | (6,9,2,12,13,4,8,1,3,7,11,5,10) |
|               | 3 | 0.22 | (6,9,7,10,4,5,8,2,1,11,3) | 3 | 0.14 | (12,11,8,10,9,5,7,6,3,4,2,1,13) |
|               | 4 | 0.15 | (3,1,2,4,5,9,11,10,8,7,6) | 4 | 0.17 | (1,4,3,7,8,2,9,5,6,10,12,13,11) |
|               |           |           | 5 | 0.12 | (8,13,12,10,11,1,6,7,4,5,9,2,3) |

refer to PLmix-$\rho_1$, PLmix-$\rho_2$, and EPLmix; inferential findings for the BMmix and EPL-$\rho$ were very similar to PLmix-$\rho_1$ and are not shown. Weights and reference order estimates of the identified clusters are shown in Table VI.
Focusing on the analysis based on 13 binding probes, we stress that in the best fitted models, the positive control probe (Hum 12) repeatedly occupies top positions in modal orderings of EBC and EBC + MBC mixture components. We remind that Hum 12 denotes the absorbance level corresponding to the entire HER2 oncoprotein. Thus, in theory, its level should reflect the total binding, and it is reasonably expected to be higher than absorbance level detected in limited portions of the oncoprotein. On the other hand, immunological response in healthy patients may either be unaffected by the exposition to the HER2 oncoprotein or yield a mild binding. This implies an exchangeability of binding probes in the ordering of absorbance levels, which is typical of the UM. These aspects reinforce the presence of Hum 12 in top positions as a signal that the immunological response actually occurred, and it can be reasonably interpreted as a distinguishing feature of the unhealthy patients. It turns out that with our wildly fluctuating LFPD data, it would not be possible to identify a simple threshold for the raw (or normalized) binding outcome to discriminate unhealthy patients. This objective is better achieved using binding profiles based on rankings. Moreover, the combination of Hum 12 with the pattern (Hum 1, Hum 11, and Hum 7) in top positions seems to characterize mixed (EBC + MBC) diseased groups, such as the first and the fourth on rankings. Moreover, the combination of Hum 12 with the pattern (Hum 1, Hum 11, and Hum 7) in top positions seems to characterize mixed (EBC + MBC) diseased groups, such as the first and the fourth components in PLmix-ρ₁, the third one in PLmix-ρ₂, and the first one in EPLmix. In fact, the protein fragments Hum 1, Hum 11, and Hum 7 were already recognized in [2] as the relevant epitopes. Referring to EBC-specific components, similar results are valid for the fragment pair (Hum 9 and Hum 3), which, together with the positive control, occupies the very first top positions (see, e.g., the third group in PLmix-ρ₁ and the second one in PLmix-ρ₂). This means that for some EBC patients, the binding reaction mainly occurs in a different section of the oncoprotein, improving the discrimination of this subgroup among diseased patients. Relevant findings can also be highlighted for healthy patients. The absent or negligible immunological response observed for some of them is well described in the estimated models by the presence of a component which is very close to the UM, as shown by the corresponding inferred values \( \hat{p}_g \). In this case, the modal orderings are poorly representative, so we marked them with the symbol ∗ in Table V. These UM-like components involve prevalently HD patients. They are also involved in another more characterized mixture component. The interpretation of the nonuniform component parameters suggests that some HD subjects share the epitopes Hum 1 and Hum 7 with other patients, but they also have a distinctive Hum 2 in top positions; Hum 11, instead, appears in middle positions. We can also look at low absorbance patterns if bottom ranks can be regarded as meaningful signatures for the problem at hand. Note, for example, that although Hum 10 appears consistently in last positions for almost all of the fitted components, Hum 9 seems to be a sort of ‘anti’-epitope signature for HD units. The same role is played by the Hum pattern (Hum 5 and Hum 6) for EBC subjects. Another interesting feature regards Hum 13; it corresponds to the empty phage vector and hence, theoretically, one would expect it to be associated with bottom ranks. This is true, instead, for those groups composed for the most part of MBC units (see, e.g., the fourth component in the PLmix-ρ₁, the third one in the PLmix-ρ₂, and the first one in the EPLmix). Therefore, a minimum absorbance level in Hum 13 could be an important feature to discriminate MBC patients, the subgroup which is only weakly characterized by the present analysis. Similar observations are valid for the case \( K = 11 \) omitting, naturally, Hum 12 and Hum 13. Finally, we remark that the EPLmix, selected as the best model in terms of the BIC, involves 69 parameters in the case of \( K = 13 \).

**4.4. Alternative quantitative data analysis**

In this section, we show that our analysis based on ranked data and the EPL mixture model compares favorably with a more conventional quantitative data approach. We implemented the flexible mixture of multivariate normal distributions (MNorm-mix) with the R package mclust described in [31].

As urged in Section 4.2, we must preliminarily decide whether there exists a more appropriate way of transforming and rescaling the original quantitative measures. Because a consolidated normalization method is lacking for this type of experiments, we worked with three alternative reasonable options: the original raw data, the log-transformed absorbances, and the rescaled log-transformed absorbances so that the individual average log-absorbance of all the considered spots is null for each patient. Results derived from the quantitative analysis are very different according to the measurement scale adopted in the input data. In fact, only with the raw data, the best fitting mixture model provides evidence in favor of a heterogeneous model, namely a mixture with \( G = 3 \) components. However, as shown in Table VII(a), the correspondence with the known disease state is poorer (ARI = 0.32) than the one obtained with the ranking-based analysis. In all other cases, the MNorm-mix model selected the single component homogeneous model as the best fitting. However, if the model is forced to be fitted as heterogeneous,
Table VII. Correspondence between the model-based clustering derived from the maximum likelihood estimates of the MNorm-mix and the true disease state of the large fragment phage display experimental units: HD = healthy, EBC = diagnosed with early stage breast cancer, and MBC = diagnosed with metastatic breast cancer.

(a) Raw LFPD data with 11 Hum (ARI = 0.32)  
(b) Rescaled log-transformed LFPD data with 13 Hum (ARI = 0.88)

| Group | Disease state | 1 | 2 | 3 | Disease state | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------|---------------|---|---|---|---------------|---|---|---|---|---|---|---|
| HD    |               | 7 | 9 | 4 |               | 7 | 11| 2 | 0 | 0 | 0 | 0 |
| EBC   |               | 3 | 2 | 23|               | 0 | 0 | 12| 1 | 10| 5 | 0 |
| MBC   |               | 9 | 0 | 10|               | 0 | 0 | 3 | 5 | 0 | 9 | 2 |

Corresponding values for the adjusted rand index based on the basic healthy/unhealthy bipartition are shown in parentheses.

then a variable number of groups is selected, ranging from 4 to 7. Indeed, the best classification is the one obtained with a MNorm-mix applied to the rescaled log-transformed absorbances of all the 13 Hum. This grouping has a very good agreement with the three disease subgroups (ARI = 0.88), as shown in Table VII(b). However, we stress that this model does not represent the best fitting in terms of BIC and yields a more scattered clustering. Moreover, this model requires 117 parameters; hence, it is less parsimonious and can be more difficult to interpret than the best fitting mixture for ranked data.

5. Concluding remarks and future developments

In the present work, we presented a novel extension of the popular and widely-used PL model relaxing the standard assumption of forward ranking elicitation and detailed its estimation in the MLE framework. We verified the usefulness of the EPL with a successful application to the real LFPD data set from a bioassay experiment, comparing its performance with alternative and more standard probability distributions for rankings. Specifically, taking into account the heterogeneous origin of the sample units, we considered several parametric models in a mixture model setting. Inferential results of our mixture modeling approach pointed out a good capability of the absorbance rankings to fit heterogeneous and wildly fluctuating binding data as well as a good accuracy in discriminating the actual disease state. Interestingly, an almost UM component has been estimated from the data. Differently from previous applications in the literature, where the UM component was introduced to fit outliers/untypical observations, for the LFPD data, such a component has a precise interpretation in characterizing healthy patients. The utility of the ranking-based analysis for epitope mapping experiments is reinforced by the possibility to partially overcome difficulties related to the choice of the preliminary normalization needed for the raw quantitative absorbance profiles. Additionally, the fitted model turns out to be more parsimonious than alternative quantitative analyses in the present multivariate setting and exhibits an interesting interpretation, unaffected by ad hoc monotone preprocessing transformations of the original raw data. Hence, even when quantitative data are available in a bioassay experiment, statistical analysis of the underlying ordinal information may provide a useful and more robust tool for the description of the outcomes. Cluster-specific parameter estimates, characterizing groups of patients, are very useful to construct epitope mapping profiles. These can identify protein fragments whose binding can be related to the disease development and help to detect spots relevant for possible classification/prediction purposes. Moreover, the significantly improved fit obtained with the more general EPL class could reflect that our proposal accounts for the absence of a natural and a priori known reference order of the binding mechanism and consequently allows to better capture the discriminant contribution of all positions. This suggests that the understanding of the binding outcomes should not be limited to the use of the standard forward PL.

A first natural way to further develop our work could be the implementation of the EPL mixture model in a Bayesian framework in order to allow the incorporation of pre-experimental information in the analysis. This extension could benefit from the conjugacy of the PL with the Gamma prior distribution, already exploited for the Bayesian inference in [32] and [33] but restricted to the homogenous population case. Moreover, a further contribution to ranking modeling could be based on the combination of the EPL with the BM. Such models, in fact, describe substantially different but compatible attributes of the ranking procedure. Hence, selection accuracy and reference order in the ranking process could be combined to
construct a more flexible parametric PL generalization, which incorporates both the BM and the EPL as specific parameter configurations. Another interesting direction could be the setting up of a flexible framework aimed at integrating the use of mixed-type (ordinal and quantitative) data with the possible inclusion of individual covariates.

Appendix A: A toy example of the EPL model

Here is a toy example which illustrates the idea and notation of the EPL model specification presented in (2.4). Suppose we have fixed a parameter \((\rho, p)\) so that \(p = (0.1, 0.2, 0.3, 0.4)\) and the entries of the reference order \(\rho = (1, 4, 2, 3)\) correspond to the following alternating selection scheme: at the first stage the judge expresses her best preference \((\rho(1) = 1)\), at the second stage she chooses her least-liked item \((\rho(2) = 4)\), and finally at the third and forth stage, she attributes, respectively, the second \((\rho(3) = 2)\) and the third \((\rho(4) = 3)\) position. Let \(\pi^{-1} = (4, 3, 1, 2)\) be the ordering of interest for which we want to compute the probability mass under the specified EPL. The EPL retains the same stage-wise ranking construction of the PL but allows the rank attribution order to be different than the ordinary best-to-worst path. The EPL postulates that the probability associated to \(\pi^{-1}\) is equivalent to the probability under the PL of sequentially selecting item 4 at the first stage, item 2 at the second stage, item 3 at the third stage and item 1 at the last stage, as indicated by the composition

\[
\pi^{-1} \rho = \left(\pi^{-1}(\rho(1)), \pi^{-1}(\rho(2)), \pi^{-1}(\rho(3)), \pi^{-1}(\rho(4))\right) = (4, 2, 3, 1),
\]

hence

\[
\mathbf{P}_{\text{EPL}}((4, 3, 1, 2) \mid \rho, p) = \mathbf{P}_{\text{PL}}((4, 2, 3, 1)(0.1, 0.2, 0.3, 0.4)) = \frac{0.4}{1} \cdot \frac{0.2}{0.1 + 0.2 + 0.3} \cdot \frac{0.3}{0.1 + 0.3} \cdot 1 = 0.1
\]

Note that, in order to read the order of preferences from \(\pi^{-1} \rho\), one needs to refer to the reference order \(\rho\).

Appendix B: Relation between the novel EPL and the PL

We prove here the presence of distributions on orderings in the novel EPL family, which are not members of the canonical PL family. For this purpose, let us remind that the PL implies the \textit{independence of irrelevant alternatives} (IIA), stating that the probability ratio of selecting an item over another is unaffected by the precedences toward the other alternatives in the choice set [8]. Equivalently, one can say that in a PL, the choice probability ratio between two items is constant over all stages as long as such alternatives are both still available. In the \(K = 3\) case, the IIA lemma translates into the following set of conditions on the probabilities \(q_{\pi^{-1}} = \mathbf{P}(\pi^{-1})\) of each possible ordering

\[
\begin{align*}
q_{(1,2,3)} &= q_{(2,1,3)} + q_{(2,3,1)} \\
q_{(1,3,2)} &= q_{(3,1,2)} + q_{(3,2,1)} \\
q_{(2,1,3)} &= q_{(1,2,3)} + q_{(1,3,2)} \\
q_{(2,3,1)} &= q_{(3,1,2)} + q_{(3,2,1)} \\
q_{(3,1,2)} &= q_{(1,2,3)} + q_{(1,3,2)} \\
q_{(3,2,1)} &= q_{(2,1,3)} + q_{(2,3,1)}
\end{align*}
\]  

(B.1)

and they have to be simultaneously satisfied for a generic ranking distribution to belong to the forward PL. Now, let us consider the EPL with parameter \(p = (2, 1, 3)\) and support parameter vector \(p\). The generic induced probability distribution on random orderings is given by

\[
\begin{pmatrix}
q_{(1,2,3)} & q_{(1,3,2)} & q_{(2,1,3)} & q_{(2,3,1)} & q_{(3,1,2)} & q_{(3,2,1)} \\
\frac{p_{1}p_{2}}{1-p_{1}} & \frac{p_{1}p_{3}}{1-p_{1}} & \frac{p_{1}p_{3}}{1-p_{1}} & \frac{p_{1}p_{3}}{1-p_{1}} & \frac{p_{1}p_{3}}{1-p_{1}} & \frac{p_{1}p_{3}}{1-p_{1}}
\end{pmatrix}
\]  

(B.2)
Substituting (B.2) in (B.1) and solving WRT \( p \), one obtains as unique solution \( p = (1/3, 1/3, 1/3) \), meaning that the two model classes can share only the UM. This formally shows what has been hinted at in [3] on the possibility to define new ranking models relaxing the forward hypothesis. To give an intuition about the types of ranking distributions that are not covered by the traditional PL, let us consider the EPL with parameter configuration \( \rho = (2, 1, 3) \) and \( p = (1 - 2\epsilon, \epsilon, \epsilon) \) where \( \epsilon \to 0 \). The corresponding probability function over the six possible orderings has two equally supported modes on the sequences with item 1 ranked second capturing almost the total mass, as shown in Figure B1(a). This represents a distribution that can not be obtained with any parameter specification from the forward PL. In fact, the suitable calibration of the support parameters can lead only to degenerate marginal choices of item 1 for the first and the last rank (Figure B1(b) and (c)). Therefore, the introduction of the extra parameter \( \rho \) running in the permutation space allows to overcome this asymmetry among ranks.

**Acknowledgements**

We are deeply grateful to Augusto Amici and his research group for providing LFPD data and many insightful discussions on the biological foundations of the experimental outcomes. We also thank both anonymous referees, whose comments allowed us to improve the article, and Thomas Brendan Murphy and Isobel Claire Gormley for providing their C code and helpful hints.

**References**

1. Marden JI. *Analyzing and Modeling Rank Data*, Monographs on Statistics and Applied Probability, vol. 64. Chapman & Hall: London, 1995.
2. Gabrielli F, Salvi R, Garulli C, Kalogris C, Arima S, Tardella L, Monaci P, Papa SM, Tagliabue E, Montani M, Quaglino E, Cucici C, Marchini C, Amici A. Identification of relevant conformational epitopes on the HER2 oncoprotein by using large fragment phage display (LFPD). *PloS ONE* 2013; 8(3):e58358.
3. Fligner MA, Verducci JS. Multistage ranking models. *Journal of the American Statistical Association* 1988; 83(403): 892–901.
4. Critchlow DE, Fligner MA, Verducci JS. Probability models on rankings. *Journal of Mathematical Psychology* 1991; 35(3):294–318.
5. Fligner MA, Verducci JS. Distance based ranking models. *Journal of the Royal Statistical Society: Series B* 1986; 48(3):359–369.
6. Critchlow DE. *Metric Methods for Analyzing Partially Ranked Data*, Lecture Notes in Statistics, vol. 34. Springer-Verlag: Berlin - Heidelberg - New York, 1985.
7. Diaconis PW. *Group Representations in Probability and Statistics*, IMS Lecture Notes Monogr. Ser., vol. 11. Inst. of Math. Stat.: Hayward, CA, 1988.
8. Luce RD. *Individual Choice Behavior: A Theoretical Analysis*. John Wiley & Sons Inc.: New York, 1959.
9. Plackett RL. The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 1975; 24(2):193–202.
10. Silverberg AR. Statistical models for q-permutations. *Ph.D. dissertation*, Princeton University, ProQuest LLC, Ann Arbor, MI, 1980.
11. Benter W. Computer based horse race handicapping and wagering systems: a report. In *Efficiency of Racetrack Betting Markets*, Hausch DB, Lo VSY, Ziema WT (eds). Academic Press: San Diego, CA, 1994; 183–198.
12. Gormley IC, Murphy TB. Exploring voting blocs within the Irish electorate: a mixture modeling approach. *Journal of the American Statistical Association* 2008; **103**(483):1014–1027.

13. Gormley IC, Murphy TB. A mixture of experts model for rank data with applications in election studies. *Annals of Applied Statistics* 2008; **2**(4):1452–1477.

14. McLachlan G, Peel D. *Finite Mixture Models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience: New York, 2000.

15. Croon MA. Latent class models for the analysis of rankings. In *New Developments in Psychological Choice Modeling*, De Soete G, Feger H, Klauer KC (eds). Elsevier Science Publisher B. V.: North Holland, 1989.

16. Croon MA, Luijkx R. Latent structure models for ranking data. In *Probability Models and Statistical Analyses for Ranking Data (Amherst, MA, 1990)*, Lecture Notes in Statist., vol. 80. Springer: New York, 1993; 53–74.

17. Murphy TB, Martin D. Mixtures of distance-based models for ranking data. *Comput. Statist. Data Anal.* 2003; **41**(3–4): 645–655.

18. Gormley IC, Murphy TB. Analysis of Irish third-level college applications data. *Journal of the Royal Statistical Society: Series A* 2006; **169**(2):361–379.

19. Gormley IC, Murphy TB. A grade of membership model for rank data. *Bayesian Analysis* 2009; **4**(2):265–295.

20. Hunter DR. MM algorithms for generalized Bradley–Terry models. *Annals of Statistics* 2004; **32**(1):384–406.

21. Lange K, Hunter DR, Yang I. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics* 2000; **9**(1):1–59. With discussion, and a rejoinder by Hunter and Lange.

22. Hunter DR, Lange K. A tutorial on MM algorithms. *The American Statistician* 2004; **58**(1):30–37.

23. Busse LM, Orbanz P, Buhmann JM. Cluster analysis of heterogeneous rank data. In *Proceedings of the 24th International Conference on Machine Learning—ICML 2007*, Ghahramani Z (ed.) Omnipress: Madison, WI, 2007; 113–120.

24. Lee PH, Yu PLH. Distance-based tree models for ranking data. *Computational Statistics & Data Analysis* 2010; **54**(6):1672–1682.

25. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B* 1977; **39**(1):1–38.

26. R Core Team. *R: a language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2012. ISBN 3-900051-07-0.

27. McNicholas PD, Murphy TB, McDaid AF, Frost D. Serial and parallel implementations of model-based clustering via parsimonious gaussian mixture models. *Computational Statistics & Data Analysis* 2010; **54**(3):711–723.

28. Vaida F. Parameter convergence for EM and MM algorithms. *Statistica Sinica* 2005; **15**(3):831.

29. McLachlan G, Krishnan T. *The EM Algorithm and Extensions*, Vol. 382. John Wiley & Sons: New York, 2007.

30. Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**(2):461–464.

31. Fraley C, Raftery AE. Enhanced model-based clustering, density estimation, and discriminant analysis software: Mclust. *Journal of Classification* 2003; **20**(2):263–286.

32. Guiver J, Snelson E. Bayesian inference for Plackett–Luce ranking models. In *Proceedings of the 26th International Conference on Machine Learning—ICML 2009*, Bottou L, Littman M (eds). Omnipress: Madison, WI, 2009; 377–384.

33. Caron F, Doucet A. Efficient Bayesian inference for generalized Bradley–Terry models. *Journal of Computational and Graphical Statistics* 2012; **21**(1):174–196.