FINITE SAMPLE PERFORMANCE OF THE E-M ALGORITHM FOR RANKS DATA MODELLING (*)

Angela D'Elia

1. INTRODUCTION

The analysis of preferences towards a set of "objects" is generally obtained by means of surveys conducted on the consumers/users (e.g. the choice among cars, marketing products, financial options, political opinions, etc.). The same approach is applied in the evaluation of services both in public and private institutions (e.g. teaching, sanitary services, public transportations, etc.).

In these cases, constrained by budget limitations, it is relevant to plan adequate sampling experiments in order to get significant results; thus, it seems useful to assess the minimum size of the sample scheme that allows meaningful statistical inference on the parameters of interest.

Usually, in descriptive approaches to preferences and evaluations analyses, people list average ranks and some measure of variability. However, a statistical approach to the choice mechanism seems more fruitful, since it conveys a parametric setting where the role of the subjects covariates could be analysed more usefully. In fact, by using a modelling structure we can estimate probabilities, expectations and variances, and test for relevant hypotheses; in this way, we gain experience and inference on the choice mechanism from the observed results, and we may be able to learn about future behaviours.

In this vein, many proposals appeared in the literature as the models discussed by Fligner and Verducci (1993), Marden (1995), D'Elia (1999, 2003) among the others, and some of them have been supported by interesting applications to different real datasets (ranging from Sport to Economics, Politics to Psychology, Economics to Marketing, etc.).

A more recent probabilistic model, which outperformed previous results in several contexts, has been proposed by D'Elia and Piccolo (2004): in their approach, the choice expressed by the rater is explained by a mixture random variable where a weighted measure of the uncertainty is designed to take into account the composite nature of the elicitation mechanism. The statistical application of such model requires a maximum likelihood estimation via the E-M algorithm. In

(*) We acknowledge the support received by Dipartimento di Scienze Statistiche, Università di Napoli Federico II (within Legge Regionale 5/2002 - Regione Campania) and the research structures of the Centro per la Formazione in Economia e Politica dello Sviluppo Rurale, Portici.

this way, the asymptotic results for the two parameters can be derived and effectively used in many empirical datasets. Thus, it seems necessary to establish the finite sample performance of such asymptotic estimators in order to assess their validity also for the sample size of the real experiments and/or to establish a required minimum sample size for deriving acceptable results.

The paper is organized as follows: in the next section we introduce the model, establish the main notation and discuss briefly the probabilistic mechanism for the random generation of the data. Then, in section 3 we describe a Monte Carlo experiment and discuss the main results we obtained. In section 4 we deal with the problem of the joint performance of the estimators of the model's parameters. Some concluding remarks end the paper.

2. A PROBABILISTIC MODEL FOR THE PREFERENCES ANALYSIS

The two problems of preference and evaluation analyses are conceptually different: in the first case, we rank different objects/items/opinions and we may be interested in studying the distribution of just one of them at a time; in the second, we express a satisfaction degree among a list of ordered alternatives following our liking/disliking feeling for something.

Also from a statistical point of view, the two issues are different. Indeed, let **R** be the ranks matrix where r_{ij} is the rank assigned by the i-th rater to the j-th item; then:

- in a preference analysis each row of **R** represents the preference ranking given by the i-th rater towards a set of items (and, in general, it is a permutation of the integers, since no ties are allowed among the items themselves);
- on the other hand, in an evaluation analysis, each row of **R** represents the i-th rater satisfaction degree towards different aspects of something (thus, it is not a permutation).

As a consequence, in a preference study each single column of the data matrix is not independent from the others (each row sum is strictly defined by the number of the items), while in an evaluation survey the vector of the selected item/aspect exhausts all the information in the data concerning the selected item.

However, when we study the preference feeling with respect a single object or the result of an evaluation study we are faced with probabilistic structures which are completely similar. In both cases:

- the subject is asked to select with reference to a specific question/object/item an integer value included in the set of the positive integer defined by the number of options;
- the choice is the result of a composite decision which is related to the liking/agreement or disliking/disagreement with the specific option;
- finally, the result of the decision derives from a comparison judgement among the alternatives.

Then, it seems reasonable to unify both the circumstances and study them in an encompassing manner; in the following formal setting, for simplicity, we discuss about objects also when we are in an evaluation context, but the conclusions hold for both cases.

2.1. *The probabilistic framework and main inferential developments*

In a set of m well defined objects, we let r be the rank assigned by a single rater to a given item (henceforth, we drop out the subscripts for shortness); throughout the paper, we assume that $r = 1$ means "most preferred" while $r = m$ means "least preferred". Then, we say that r is the observed value of a discrete random variable defined on the support $\{1, 2, \ldots, m\}$.

Within the class of all such random variables, D'Elia and Piccolo (2004) introduced the Mixture of a Uniform and a shifted Binomial (MUB) model, such that $R \sim \text{MUB}(m, \pi, \xi)$ if:

$$\Pr(R = r) = \pi \binom{m-1}{r-1} (1-\xi)^{r-1} \xi^{m-r} + (1-\pi)\frac{1}{m}, \qquad r = 1,\ 2,\ \ldots, m.$$

The expectation and the variance of R are, respectively:

$$\text{E}(R) = \frac{m+1}{2} + \pi(m-1)\left(\frac{1}{2} - \xi\right);$$

$$\text{Var}(R) = (m-1)\left\{\pi\xi(1-\xi) + \frac{(1-\pi)(m+1)}{12} + \pi(1-\pi)\frac{(m-1)(2\xi-1)^2}{4}\right\}.$$

Of course, the first of these expressions shows that $\text{E}(R) = (m+1)/2$ in the symmetric situation (that is, $\xi = 1/2$).

The parameter π weights for the first component (a shifted Binomial random variable where the ξ parameter increases with the liking towards the object) and, thus, $(1 - \pi)/m$ is a direct measure of the uncertainty share. In fact, the parameters are jointly related to the choice probability and it is not easy to distinguish the marginal contribution of each of them. However, it results that the joint increase of both π and ξ towards 1 implies a greater preference feeling, since this circumstance lowers the mean value of R.

For generating a sample of such random variable we refer to the interpretation of a mixture random variable as a two steps choice mechanism between a shifted Binomial and a discrete Uniform random variable. In this way:

i) we choose the first component (random variable) with probability π;

ii) then, given this random variable, we generate a pseudo-random number from it.

The code in the Appendix, written in the Gauss© language, seems to be an efficient method for generating a large number of pseudo-random numbers for simulation purposes.

The information contained in the sample of the observed ranks for n subjects $(r_1, r_2, \ldots, r_n)$ is strictly equivalent to that contained in the reduced vector $(n_1, n_2,$

…, n_m,) of the observed frequencies for the ordered ranks. Then, from a statistical viewpoint, efficient and consistent maximum likelihood (ML) estimators for both π and ξ can be derived by an implementation of the E-M algorithm (McLachlan and Krishnan, 1997; McLachlan and Peel, 2000). This algorithm is the most efficient way to compute the maximum of the log-likelihood function defined by:

$$\log L(\pi,\xi) = \sum_{r=1}^{m} n_r \log(\Pr(R = r \mid \pi,\xi)).$$

The asymptotic properties of the ML estimators ($\hat{\pi}$, $\hat{\xi}$) are well known and exploiting a result valid from grouped data (Rao, 1973, pp. 367-368) we derived asymptotic standard errors and testing criteria for the parameters. Specifically, the asymptotic confidence ellipse at the 100(1 – α)% level is given by:

$$\{(\pi,\xi) : d_{\pi\pi}(\hat{\pi}-\pi)^2 + d_{\pi\xi}(\hat{\pi}-\pi)(\hat{\xi}-\xi) + d_{\xi\xi}(\hat{\xi}-\xi)^2 \le -2\log(\alpha)/n\}.$$

The quantities $d_{\pi\pi}$, $d_{\pi\xi}$, $d_{\xi\xi}$ are n times the estimates of the elements of the inverse of the asymptotic variance-covariance matrix of the ML estimators.

Moreover, the correlation between these estimators, expressed by Corr($\hat{\pi},\hat{\xi}$) $= \dfrac{d_{\pi\xi}}{\sqrt{d_{\pi\pi}d_{\xi\xi}}}$, is a relevant issue except when $\left(\hat{\xi}-\dfrac{1}{2}\right) \cong 0$ (that is, a symmetric distribution for ranks), since in this case the estimators turn out to be uncorrelated.

3. THE MONTE CARLO EXPERIMENT

Generally, in the preference and/or evaluation surveys the number m of objects or values to be compared is fixed and known in advance. Thus, the ratio $k = n/m$, where n is the number of raters for the designed experiment, is a relevant issue for the analysis. In a sense, k measures the worst preference situation since it is the relative frequency of each rank for $r = 1, 2, \dots, m$ when we are in the case of *equi-preference*, that is when the choice among the items is absolutely random: of course, in this case $\pi = 0$, and R is a discrete Uniform random variable over the support $\{1, 2, \dots, m\}$.

We generated *nsimul* = 10000 experiments for varying ratios k = 10, 15, 20, 25, 30, 40, 50, 100, 200 when $R \sim$ MUB(m, π, ξ), and m = 5, 7, 12. Indeed, these values seem to be modal choices in the current literature on these issues and also in our experiences from several researches areas.

Although the experiment was conducted for all such models, for shortness we limit ourselves to discuss only the results for m = 7, reporting that in the other cases the same conclusions apply, since there is a substantial coincidence of patterns and performances.

As a preliminary choice, we selected a large number of models for varying parameters over the admissible parameters space, but the choice of the results we present here is dictated by the final discussion.

Table 1 synthesizes the main features of the parametric models we selected for our experiment and Figure 1 shows their probability distributions emphasizing their diversity in location, variability and skewness aspects.

TABLE 1

Models selected for the Monte Carlo experiment (m=7)

Model	π	ξ	E(*R*)	Var(*R*)	Mode(*R*)
A	0.1	0.1	4.24	4.17	7
B	0.1	0.4	4.06	3.78	5
C	0.2	0.2	4.36	3.91	6
D	0.4	0.1	4.96	4.00	7
E	0.4	0.5	4.00	3.00	4
F	0.7	0.3	4.84	2.38	5
G	0.3	0.8	3.46	3.77	2
H	0.7	0.6	3.58	2.28	3
I	0.9	0.9	1.84	1.40	1

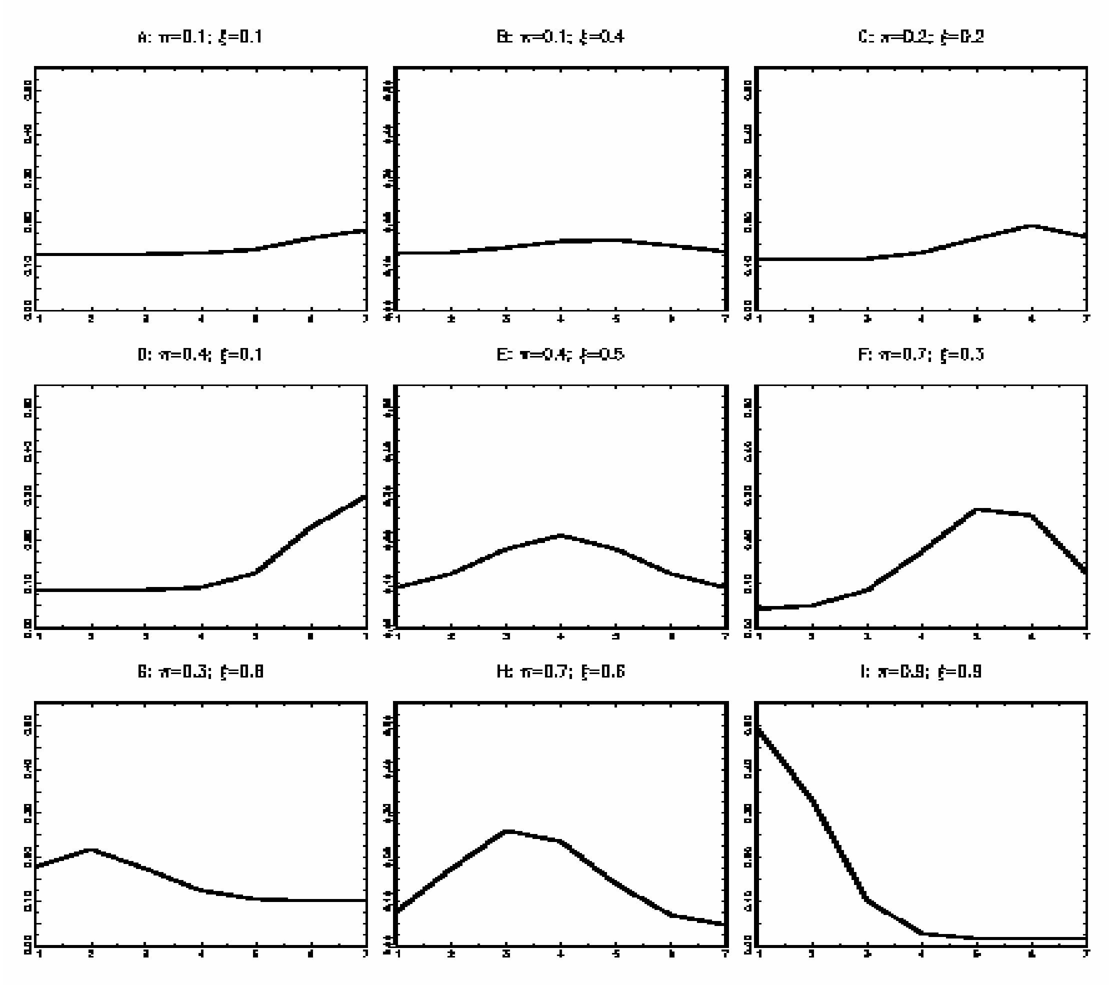

Figure 1 – Probability distributions for the models in Table 1.

The following tables show the main results concerning the finite bias (Table 2) and relative efficiency (Table 3) of the ML estimators, respectively.

TABLE 2

Bias of the ML estimators for π and ξ (m=7)

Model		k								
		10	15	20	25	30	40	50	100	200
A	π	0.05562	0.03696	0.02544	0.02038	0.01677	0.01173	0.00975	0.00511	0.00444
	ξ	0.21220	0.17429	0.15109	0.12750	0.10561	0.08483	0.06309	0.02349	0.00753
B	π	0.08857	0.06625	0.05332	0.04545	0.03857	0.03190	0.02659	0.01678	0.01165
	ξ	0.04559	0.03711	0.02930	0.02660	0.02022	0.01683	0.01416	0.00942	0.00510
C	π	0.03506	0.02267	0.01713	0.01387	0.01099	0.00889	0.00669	0.00303	0.00145
	ξ	0.04452	0.02371	0.01357	0.00972	0.00496	0.00348	0.00247	0.00111	0.00021
D	π	0.00727	0.00643	0.00448	0.00334	0.00228	0.00175	0.00149	0.00077	0.00028
	ξ	0.00162	0.00097	0.00038	0.00032	0.00013	-0.00023	0.00020	0.00003	0.00005
E	π	0.02439	0.01720	0.01225	0.00950	0.00822	0.00639	0.00468	0.00193	0.00139
	ξ	-0.00118	-0.00012	-0.00112	-0.00003	0.00033	-0.00050	-0.00016	0.00045	-0.00015
F	π	0.00799	0.00485	0.00521	0.00387	0.00261	0.00156	0.00171	0.00102	-0.00003
	ξ	0.00069	-0.00006	0.00004	0.00025	0.00025	0.00004	0.00004	-0.00012	0.00000
G	π	0.01936	0.01328	0.01009	0.00644	0.00540	0.00497	0.00451	0.00107	0.00098
	ξ	-0.00868	-0.00240	-0.00098	-0.00074	-0.00067	-0.00002	-0.00063	-0.00028	-0.00038
H	π	0.00898	0.00714	0.00584	0.00385	0.00331	0.00177	0.00228	0.00089	0.00118
	ξ	-0.00025	-0.00034	-0.00006	0.00011	0.00015	0.00002	-0.00015	-0.00003	-0.00005
I	π	0.00167	0.00137	0.00060	0.00026	0.00062	0.00073	0.00076	0.00023	0.00011
	ξ	-0.00008	-0.00025	0.00010	-0.00032	0.00011	-0.00015	-0.00010	0.00008	-0.00001

It seems evident that the bias is generally limited and the asymptotically unbiased nature of the ML estimators is confirmed. Indeed, it appears that the bias decreases when *k* becomes large, for both the parameters. Thus, as far as it concerns the unbiasedness, a ratio of *k* greater than 30 seems to be an acceptable bound for inferring on both the parameters. The worst performance of the ξ estimator happens for low values of *k* and of the parameter itself (e.g. model A, where $\xi = 0.1$). In finite sample we always find a positive bias for the parameter π, while for the parameter ξ a negative bias is recognisable in model G, and in minor extent in models E, H and I.

In Table 3 we compare the finite variability of the estimators with the asymptotic ML variances. The relative efficiency (which should converge to 1) is measured as the ratio between the asymptotic variance (as computed by D'Elia and Piccolo, 2004) and the observed MSE in the simulated data.

TABLE 3

Relative efficiency of the ML estimators for π and ξ (m=7)

Model		k								
		10	15	20	25	30	40	50	100	200
A	π	0.96672	1.00558	1.06324	1.01714	1.04553	1.01199	1.01101	0.99122	1.05841
	ξ	0.15209	0.12656	0.11208	0.10935	0.11453	0.11115	0.12905	0.20921	0.58245
B	π	1.20468	1.21728	1.22561	1.24161	1.28932	1.24280	1.25548	1.22042	1.17294
	ξ	1.42554	1.15715	1.04724	0.98225	0.95008	0.91831	0.90498	0.90974	1.00567
C	π	1.14987	1.10395	1.07305	1.07563	1.05331	1.04323	1.04296	1.07010	1.01510
	ξ	0.37656	0.45698	0.51425	0.63612	0.68234	0.77924	0.85671	0.93504	0.98342
D	π	0.98864	0.98611	0.98515	0.99237	1.00036	1.01417	1.01556	1.03373	1.01300
	ξ	0.81141	0.90268	0.92968	0.97213	0.96137	0.97388	0.99735	0.99088	1.02480
E	π	1.07762	1.05729	1.04023	1.01218	1.01432	1.02130	1.00211	0.99544	1.01150
	ξ	0.77584	0.86511	0.90835	0.91755	0.92588	0.95403	0.97962	0.98650	0.97664
F	π	0.97474	0.97952	0.99286	1.01200	1.01546	0.98419	1.00305	1.00055	0.98024
	ξ	0.94366	0.98436	1.01271	0.98617	0.99956	0.99480	1.00827	0.95721	0.98149
G	π	1.02959	1.03221	0.99432	1.01042	1.02864	0.98940	0.99479	0.98519	1.00874
	ξ	0.53174	0.74920	0.85016	0.87374	0.90555	0.92237	0.95039	0.98734	0.99204
H	π	0.99448	0.98501	0.99359	0.99671	0.99922	0.99069	1.00389	0.98868	0.99565
	ξ	0.93565	0.96613	0.98003	0.98317	1.00927	0.99318	0.99596	0.96965	1.01942
I	π	1.01631	0.98552	0.99144	0.98211	1.01215	0.99235	0.99478	1.00807	1.00179
	ξ	0.96659	0.99041	0.96727	0.97032	0.97135	0.99534	1.01565	0.99905	1.01519

The results show a different performance of the two estimators. As matter of fact, the π estimator exhibits a larger efficiency over the whole parameter space. On the contrary, for the ML estimator of ξ, we observe a very poor efficiency for the model A, and also a limited performance for the model C and G; however in these last two cases, increasing values of k affect positively the observed efficiency.

At this regard, it seems important to notice that the models A, C and G are all characterised by low values of the π parameter (that are: 0.1, 0.2, 0.3, respectively). This evidence may support the idea that the problem for the efficiency of the ξ estimator originates when the π parameter is small. In fact, when $\pi \to 0$, we are moving from a genuine mixture distribution to an almost Uniform discrete random variable: in this case (*equipreference* or *equiprobability*) for all the alternatives it is quite natural to expect that the inference drawn from a finite sample is more difficult, since the Uniform random variable maximises the entropy, among all the discrete distributions with finite support $\{1,2, \ldots, m\}$, for a fixed m (Papoulis, 1984, pp. 514-515).

Finally, we would stress the acceptable performance of the ML parameter for the E model which derives from a symmetric distribution; further studies should confirm that the uncorrelation among the estimators improves their finite performances.

For enhancing the univariate assessment of the parameters performance, we report also some of the results from the simulation experiments in a graphical format. In Figure 2, the box-plots for the estimators simulated distributions are shown only for the models A, E, and I, for increasing k values (from 10 to 200).

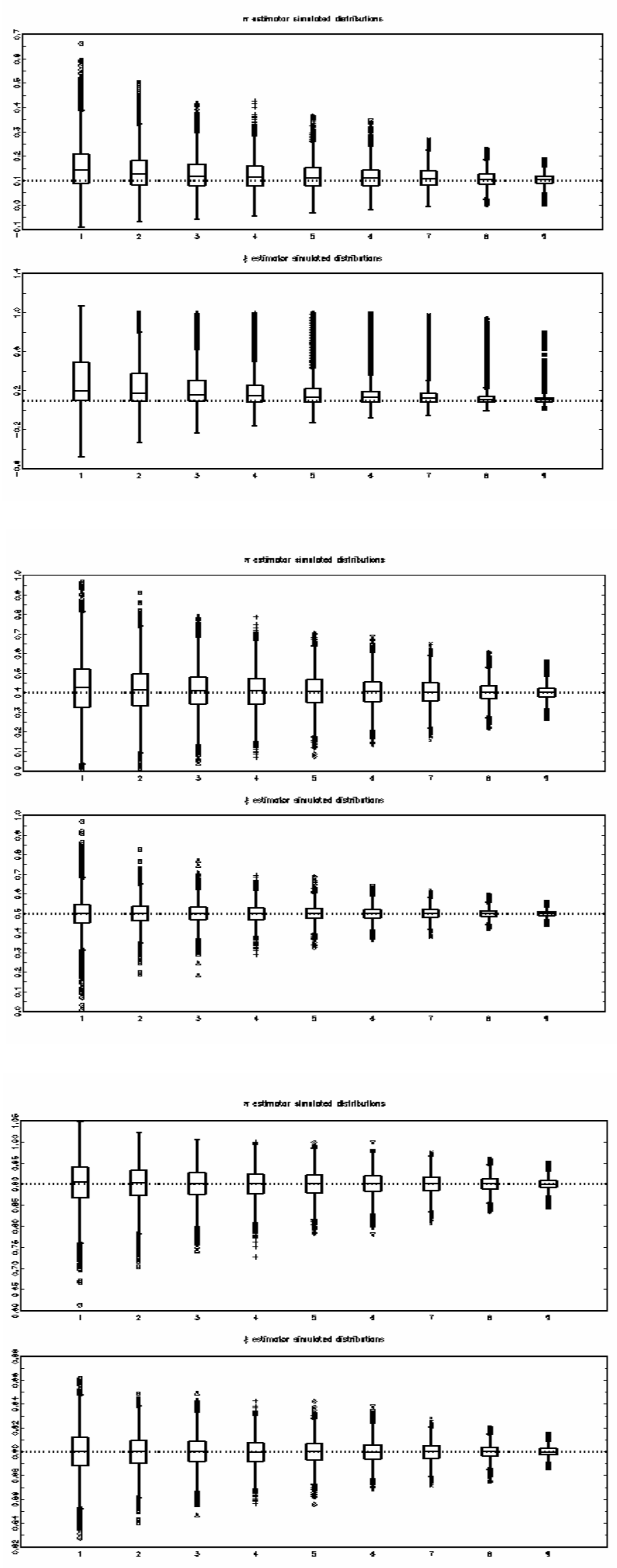

Figure 2 – Box-plots for the ML estimators (models A, E and I).

As a consequence of the previous evidence, we could conclude that, in order to achieve adequate results in finite samples, we should collect as many data as requested by at least k = 30. Of course, this bound may be lowered when we have *a priori* information that the object/item possesses a relatively high degree of preference or disliking; on the contrary, it should be increased if we suspect that the experiment concerns an object/item with an equipreference pattern.

4. THE JOINT PERFORMANCE OF THE ML ESTIMATORS

It is not strictly correct to assess the quality of the ML estimators in a finite sample without an adequate judgement of their bivariate distributions.

If we had unbiased estimators, we should compare the asympotic variance-covariance matrix $\mathbf{V}_\infty$ with the finite sample variance-covariance matrix $\mathbf{V}_{n,h}$ for a sample size of n and the h-th replication of the Monte Carlo experiment. Indeed, their inverse elements characterize the confidence ellipse previously defined, both in the asymptotic and finite sample cases.

In the literature, some joint measures are derived for this aim, as the trace $tr(\mathbf{V})$ or the determinant $|\mathbf{V}|$. In fact, the trace is just the sum of the variances, and thus it seems unable to take into account the correlations between the estimators, which is a relevant issue for judging the adequacy of their performance. On the contrary, the determinant is a synthesis of both the variances and covariance of the estimators and it is immediately related to the area of the confidence ellipse. But, although one should prefer in this kind of comparisons the determinant with respect to the trace, we think that a simple inspection of the area could be misleading since a similar measure could result also for ellipses displaced with respect to the true parameter values.

Thus, among the many approaches to this kind of comparisons, we choose to study the joint performance of the estimators with respect to the asymptotic ellipses which should include 95% of them (if we let α = 0.05). In this way, we control both the single placement of the parameters' estimators (bias and variability with respect to the true values) and their correlations (with respect to the ellipse orientation).

Table 4 shows the percentages of correct inclusions of the ML estimators $(\hat{\pi}, \hat{\xi})$ into the asymptotic ellipse, for each of the selected models and for varying k. Such results show a uniform and adequate convergence of the observed values towards the nominal one, starting for values of k as lows as $k \approx 15$. Noticeable exceptions are models C and, especially, model A where the same convergence is significantly slower. Thus, when both parameters are small, both joint and univariate performances are expected to be good only for very large sample size. Finally, we note that for all the models (except for B and I) the convergence towards the nominal confidence coefficient stems from values less than 95%.

TABLE 4

Percentages of correct inclusion (m=7)

Model	*k*								
	10	15	20	25	30	40	50	100	200
A	0.6683	0.7006	0.7244	0.7500	0.7732	0.8055	0.8355	0.8983	0.9539
B	0.9857	0.9683	0.9577	0.9547	0.9536	0.9501	0.9527	0.9537	0.9597
C	0.8904	0.9113	0.9216	0.9282	0.9358	0.9388	0.9409	0.9473	0.9494
D	0.9394	0.9407	0.9440	0.9490	0.9470	0.9470	0.9493	0.9515	0.9523
E	0.9355	0.9429	0.9440	0.9434	0.9452	0.9469	0.9505	0.9461	0.9487
F	0.9389	0.9449	0.9511	0.9476	0.9493	0.9511	0.9491	0.9471	0.9462
G	0.9217	0.9340	0.9385	0.9396	0.9434	0.9458	0.9471	0.9443	0.9483
H	0.9473	0.9474	0.9468	0.9480	0.9494	0.9482	0.9504	0.9472	0.9499
I	0.9518	0.9488	0.9492	0.9470	0.9515	0.9490	0.9515	0.9495	0.9503

5. CONCLUDING REMARKS

The paper has shown that the performances of the ML estimators derived by an E-M algorithm for a mixture model for preference/evaluation analyses are satisfactory when the sample size is at least 30 times the number of objects/items to be compared/evaluated. This bound may be lowered when there is *a priori* information about the presence of a strong liking/disliking feeling towards the considered item; viceversa, the bound should be increased if an equipreference pattern in expected.

Then, this kind of evidence may be usefully exploited for planning adequate sample sizes in surveys and researches, where the available budget and time are quite limited.

Dipartimento di Scienze Statistiche ANGELA D'ELIA
Università di Napoli Federico II

APPENDIX

Gauss© code for generating pseudo-random numbers from the random variable R~MUB(m, π, ξ).

```
PROC SIMULMUB(nsimul,m,pai,csi);
LOCAL  vett0,vett1,vett2,vett;
          vett0=(rndu(nsimul,1).<=pai);
          vett1=1+sumc((floor(1-csi+rndu(nsimul,m-1))'));
          vett2=floor(m*rndu(nsimul,1))+1;
          vett=vett0.*(vett1-vett2)+vett2;
RETP(vett);
ENDP;
```

ACKNOWLEDGEMENTS

The author is grateful to an anonymous referee for helpful comments on a first version of the paper.

REFERENCES

A. D'ELIA (1999), *A proposal for ranks statistical modelling*, in H. FRIEDL, A. BERGHOLD, G. KAUERMANN (eds.), *Statistical Modelling*, Graz-Austria, pp. 468-471.
A. D'ELIA (2003), *Modelling ranks using the Inverse Hypergeometric distribution*, "Statistical Modelling: an International Journal", 3, pp. 65-78.
A. D'ELIA, D. PICCOLO (2004), *A mixture model for preference data analysis*, "Computational Statistics & Data Analysis", forthcoming.
M.A. FLIGNER, J.S. VERDUCCI (1999), *Probability Models and Statistical Analyses of Ranking Data*, Springer-Verlag, New York.
J.I. MARDEN (1995), *Analyzing and Modelling Rank Data*, Chapman & Hall, London.
G. MCLACHLAN, T. KRISHNAN (1997), *The EM algorithm and extensions,* J. Wiley & Sons, New York.
G. MCLACHLAN, G.J. PEEL (2000), *Finite mixture models*, J. Wiley & Sons, New York.
A. PAPOULIS (1984), *Probability, Random Variables, and Stochastic Processes,* (2nd edition), McGraw-Hill Int. Book Co., Singapore.
C.R. RAO (1973), *Linear statistical inference and its applications*, (2nd edition), J. Wiley & Sons, New York.

RIASSUNTO

La performance in campioni finiti dell'algoritmo E-M per un modello per variabili rango

Nell'articolo si valuta la performance in piccoli campioni degli stimatori di massima verosimiglianza dei due parametri di un modello mistura, recentemente introdotto per variabili rango. Da un punto di vista computazionale, le stime sono ottenute mediante l'algoritmo E-M, ed il loro comportamento è considerato sia in termini univariati che congiuntamente. I risultati ottenuti, mediante un ampio esperimento Monte Carlo, evidenziano che le performance degli stimatori dei due parametri sono entrambe soddisfacenti per quanto concerne l'assenza di distorsione; viceversa, in termini di efficienza, si mette in luce un comportamento differenziato, che sembra dipendere dalla rispettiva posizione nello spazio parametrico. Alcuni considerazioni e suggerimenti operativi concludono il lavoro.

SUMMARY

Finite sample performance of the E-M algorithm for ranks data modelling

We check the finite sample performance of the maximum likelihood estimators of the parameters of a mixture distribution recently introduced for modelling ranks/preference data. The estimates are derived by the E-M algorithm and the performance is evaluated both from an univariate and bivariate points of view. While the results are generally acceptable as far as it concerns the bias, the Monte Carlo experiment shows a different behaviour of the estimators efficiency for the two parameters of the mixture, mainly depending upon their location in the admissible parametric space. Some operative suggestions conclude the paper.