Finite Mixture Analysis of Beauty-Contest Data from Multiple Samples *

Antoni Bosch-Domènech
José G. Montalvo
Rosemarie Nagel
and
Albert Satorra
Universitat Pompeu Fabra, Barcelona

January 27, 2004

Abstract

This paper develops a finite mixture distribution analysis of Beauty-Contest data obtained from diverse groups of experiments. ML estimation using the EM approach provides estimates for the means and variances of the component distributions, which are common to all the groups, and estimates of the mixing proportions, which are specific to each group. This estimation is performed without imposing constraints on the parameters of the composing distributions. The statistical analysis indicates that many individuals follow a common pattern of reasoning described as iterated best reply (degenerate), and shows that the proportions of people thinking at different levels of depth vary across groups.

Keywords: Beauty-Contest experiments, reasoning hierarchy, finite mixture distribution, EM algorithm.

Journal of Economic Literature Classification: C24, C44, C91.

1   Introduction

In recent years there has been an increasing interest in evaluating experimentally individuals' choices, decision processes and beliefs formation. From an econometric perspective, the potential multiplicity of decisions and beliefs favors clustering procedures to separate the different outcomes of each decision process. These procedures differ in the estimation techniques used and the amount of structure imposed on the econometric model.

In this paper we seek to interpret the choice data reported in A. Bosch-Domènech, J. G. Montalvo, R. Nagel and A. Satorra (2002), by constructing a finite mixture model. These data were obtained in seventeen different experiments involving the Beauty-Contest (BC) game. In a basic BC game, each player simultaneously chooses a decimal number in an interval. The winner is the person whose number is closest to p times the mean of all chosen numbers, where $p < 1$ is a predetermined and known number. The winner gains a fixed prize. In this game there exists only one (Nash) equilibrium in which all players choose the lowest possible number. In the seventeen experiments reported, $p = 2/3$ and the interval, in sixteen out of the seventeen, is $[0, 100]$. In one experiment the choice set is $[1, 100]$.

Several types of reasoning processes have been proposed to explain the individuals' decisions in the BC game (see references in Section 5). One such reasoning process, denoted as IBRd, for Iterated Best Reply with degenerate beliefs (i.e., the belief that the choices of all others are at, or around, one precise value),[1] classifies subjects according to the depth, or number of levels,

---

[1] See, e.g., Bosch-Domènech et al. (2002) or Stahl (1996).

of their reasoning. It assumes that, at each level, every player has the belief that she is exactly one level of reasoning deeper than all the rest. A Level-0 player chooses randomly in the given interval $[0, 100]$, with the mean being 50. Therefore, a Level-1 player gives best reply to the belief that everybody else is a Level-0 player and thus chooses $50p$. A Level-2 player chooses $50p^2$, a Level-k player chooses $50p^k$, and so on. A player who takes infinite steps of reasoning, and believes that all players take infinite steps, chooses zero, the equilibrium. This hypothesis of iterated best reply, together with $p = 2/3$, and an interval $[0, 100]$, predicts that choices (in addition to random and haphazard choices, corresponding to Level-0 players) will be on the values 33.33, 22.22, 14.81, 9.88, ... and, in the limit, 0.

The seventeen different experiments whose data we are analyzing take place in differents settings, and are classified in six groups as described in Table 1.[2]

Note that the experiments are performed in very different environments, involving different subject pools, sample sizes, payoffs, and settings: the data have been collected in classrooms, conferences, by e-mail, through newsgroups or among newspaper readers, as well as in laboratories with undergraduate students. The non-laboratory sessions typically allow more time to participants and use economists, game theorists, or the general public as subjects. We are, therefore, dealing with a rich and heterogeneous data set.

This paper presents a statistical analysis of these BC data allowing for two types of heterogeneity: one that is unobserved, namely the reasoning

---

[2]More details of these seventeen experiments and the IBRd hypothesis can be found in Bosch-Domènech et al. (2002).

Table 1: The data of the 6 different groups of experiments

Group	# of experiments	Description of subjects	Sample size n_g
1 (Lab)	5	Undergraduate students in labs (Bonn & Caltech)	86
2 (Class)	2	Undergraduate students, UPF	138
3 (Take-Home)	2	Undergraduate students in Take-Home tasks, UPF	119
4 (Theorists)	4	Game Theory students and experts in Game Theory in conferences and e-mail	92
5 (Internet)	1	Newsgroup in Internet	150
6 (Newspapers)	3	Readers of FT, E and S	7900
		Financial Times	*1476*
		Expansión	*3696*
		Spektrum der Wissenschaft	*2728*

level of each individuals in the sample; and another one that is manifest, the group membership. We specify a finite mixture distribution model, with all parameters of the composing distributions unconstrained (to be estimated) but equal across groups, and with mixture proportions that are group specific. This approach contrasts with the previous literature, where data sets were more homogeneous, and the models more restrictive.

The paper is organized as follows. The next section describes the data and the characteristics of each group of experiments. Section 3 proposes a finite mixture distribution model to interpret the unobserved heterogeneity associated with the reasoning processes of agents playing the BC game. Section 4 contains estimation results that give empirical support to the IBRd hypothesis. Section 5 compares our results with those using alternative statistical procedures applied to BC data. Section 6 concludes.

2   Data description

Inspecting the histogram for the whole distribution, when all the groups are pooled together (see Figure 1), we observe that the peaks closely correspond to the numbers that individuals would have chosen if they had reasoned according to the IBRd hypothesis, at reasoning levels one, two, three and infinity. If we take the histograms for the six groups of data separately (Figure 2), the peaks at level one, two and infinity are still discernible, but their frequency varies considerably across groups of experiments.

The first group, Lab-experiments with undergraduates, is clearly distinguished from the rest, because the Nash equilibrium was rarely selected. When subjects have some training in game theory, the proportion of subjects choosing the equilibrium seems to increase. The highest frequencies are attained when experimenting with theorists, in which case, the greater confidence that others will reach similar conclusions may be reinforcing the effect of training. In Newspapers, the frequency of equilibrium choices falls somewhere in between,[3] as should be expected from the heterogeneous level of training of their readers.

Yet, for some subgroups of data in particular, the regularity of choices can be striking. Take the responses from readers of *Financial Times (FT)* and *Spektrum (S)*. Despite catering to different types of readers (*S* to scientists and *FT* to businessmen) and the severe non-normality of the data, a comparison of the results of the experiment performed with *S* and *FT* readers yields a very similar distribution, as can be observed in the quantile-quantile

---

[3]In *Expansión* the choices were in $[1, 100]$. If we include choices at 1 as equilibrium choices, then the frequency would increase.
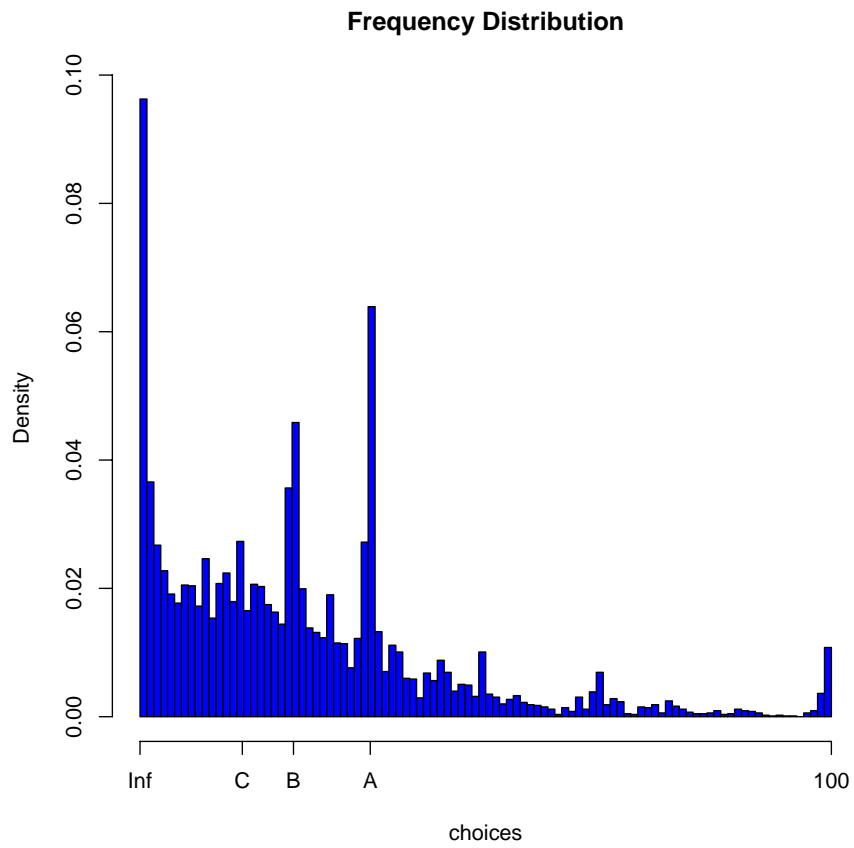
Figure 1: Histogram for the whole sample. The points A,B,C and Inf, correspond to the choices of subjects with first, second, third and infinite levels of reasoning.
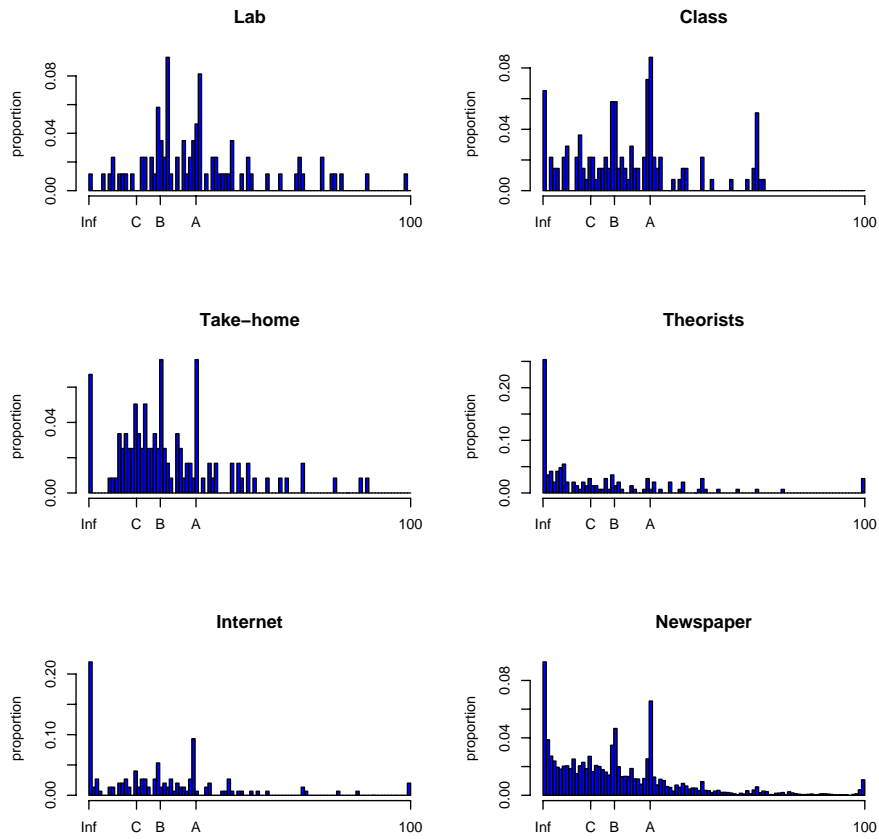
Figure 2: Histograms for the six different groups. As in Figure 1, the values A,B,C and Inf, correspond to first, second, third and infinite levels of reasoning.
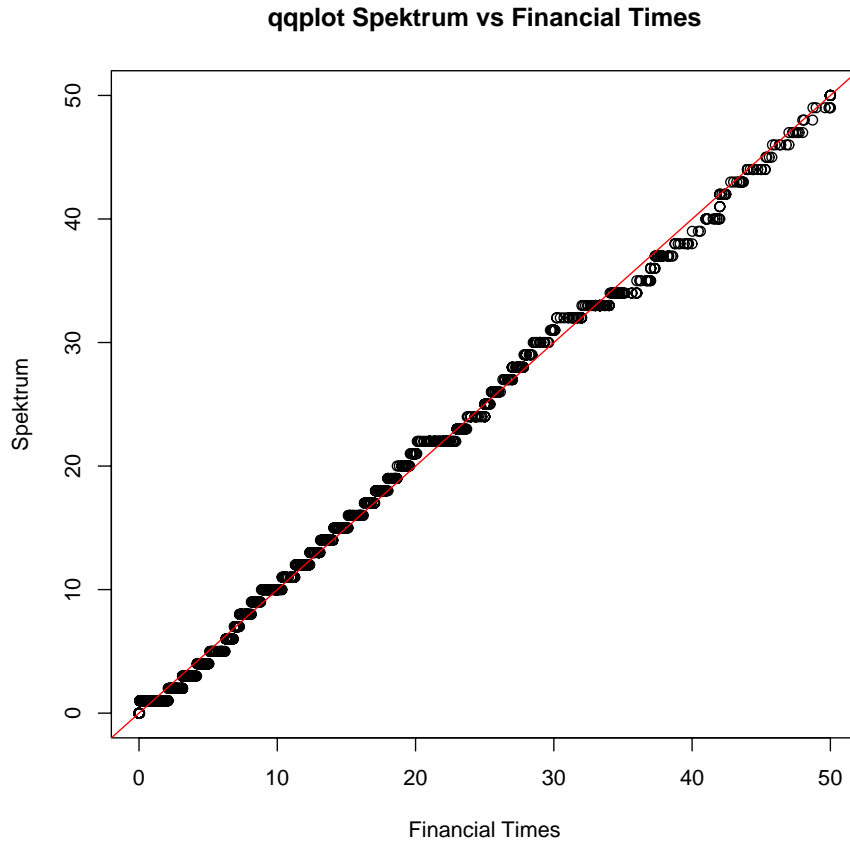
qqplot Spektrum vs Financial Times

Figure 3: Quantiles of *Spektrum* vs *Financial Times* for choices smaller than 50.

plot of Figure 3.[4] The Kruskal-Wallis chi-squared test statistic for the null hypothesis that the two distributions are the same is equal to 0.002 (p-value equal to 0.964), i.e., the two distributions cannot be distinguished.

_____

[4]In this type of graphs, equality of distributions corresponds to points lying on the diagonal.

3 The finite mixture model and estimation procedure

From our previous discussion it appears that the basic problem in fitting a statistical model to the BC data we consider is the existence of unobserved heterogeneity (the different levels of reasoning), in addition to a multiple group structure. Statisticians and, more recently, economists, have developed models of finite mixture distributions to deal with this type of problems.[5] This section proposes an interpretation of the BC data as a mixture of distributions and provides a statistical strategy to estimate such a model.

3.1 A multi-sample finite mixture model

Let us denote the multiple-sample data in Table 1 by $\{y_{ig}; \, i = 1, \ldots, n_g\}_{g=1}^{6}$, where y_{ig} is the number chosen by individual i in the group g of experiments, and n_g is the sample size of group g.

For each of the six different groups, we specify the following $(K+1)$-mixture probability density function for y,

$$f_y(y, \psi) = \pi_0 f_0(y) + \pi_1 f_1(y, \theta_1) + \ldots \pi_K f_K(y, \theta_K),$$

where the $f_0, f_1, \ldots, f_K$ are the components of the mixture distribution, θ_k denoting a mean and variance parameter vector of component k, and

- $f_0(y) = 1/100$, i.e., the density of the uniform distribution in $[0, 100]$.

- $f_k, k = 1, \ldots, K - 1$, are (truncated below 0 and above 100) normal distributions of means μ_k and variances σ_k^2.

---

[5]Titterington et al.(1992) covers many issues related to the statistical properties of finite mixture distributions models.

- f_K is a Normal distribution, of mean μ_K and variance σ_K^2, left-censored at the value 1. The censoring of this distribution models the non-null mass probability at the left-limit value of the distribution, at values 0 and 1. Recall that in one experiment the limit value was at 1, not 0. For the sake of parsimony we consider a single left-censored distribution at 1 (which, obviously, automatically collects the censoring at 0).

- the π_i's are *mixing proportions*, with $\pi_i > 0$ and $\sum_0^K \pi_i = 1$. This mixing proportions are the weights of the different components of the mixture.

We define the parameter vector $\psi = (\pi, \theta)'$, where $\pi = (\pi_0, \pi_1, \ldots, \pi_K)$ is the vector of mixing proportions and $\theta = (\mu_1, \ldots, \mu_K, \sigma_1^2, \ldots, \sigma_K^2)$ is the vector of parameters of the normal distributions underlying the mixture model. The model we adopt for estimation sets π to be group-specific, but imposes the equality of θ across groups. It is reasonable to assume that there is a common pattern of reasoning accross groups of individuals playing the BC-game, therefore we let means and variances to be equal across groups. However, the proportion of players at each level of reasoning may be different accross experiments. This strategy allows also to obtain sensible estimates for complex mixture distributions even in the groups with small sample size.

3.2 ML estimation and the EM algorithm

From the finite mixture model described above, the log-likelihood function of θ is

$$l(\theta) = \sum_i \log \left(\sum_{k=0}^K \pi_k f_k(y_i\,;\theta) \right),$$

where i varies across all sample units. Since this log-likelihood function involves the log of a sum of terms that are (highly non-linear) functions of parameters and data, its maximization using standard optimization routines is not feasible in general; for this maximization, we will resort on the EM algorithm (Dempster, Laird and Rubin 1977). [6] We consider the data augmented with variables $d_i = (d_{i1}, \ldots, d_{iK})'$, where d_{ik} are dummy variables identifying the component membership (i.e., for each i, $d_{ik} = 0$, except for one particular k, when $d_{ik} = 1$). Obviously, the d_i's are non-observable. Assuming that d_i has a multinomial distribution with parameters $(\pi_0, \ldots, \pi_K)'$, the log-likelihood of the complete data is:

$$l_C(\theta) = \sum_i^n \sum_{i=0}^K d_{ik} \left(\log \pi_k + \log f_k(y_i; \theta_k) \right)$$

The EM approach computes ML estimates using the following algorithm.

1. For given values of $\hat{\pi}_{ik}$ and $\hat{\pi}_k$, maximize with respect to θ the function

   $\sum_i^n \sum_{i=0}^K \hat{\pi}_{ik} \left(\log \hat{\pi}_k + \log f_k(y_i; \theta_k) \right)$

2. For given θ, update the $\hat{\pi}_{ik}$ (estimated conditional probabilities of case i belonging to k) and the $\hat{\pi}_{ik}$ (marginal probabilities) using the formula

$$\hat{\pi}_{ik} = \frac{\pi_k f_k(y_i; \theta_k)}{\sum_{i=0}^K \pi_k f_k(y_i; \theta_k)} \quad \text{and} \quad \hat{\pi}_k = \frac{1}{n} \sum_i^n \hat{\pi}_{ik} \tag{1}$$

Starting from initial estimates $\hat{\pi}_{ik}$'s and $\hat{\pi}_k$, the EM algorithm consists in iterating 1) and 2) till convergence.

---

[6]Recently Arcidiacono and Jones (2003) have proposed an extension of the EM algorithm where the parameters of the finite mixture distribution can be estimated sequentially during each maximization step. In our case, however, we did not find necessary to resort to this alternative. See also McLachalan and Peel (2000) for different extensions and applications of the EM algorithm.

The optimization in 1) implies the maximization of a $(K+1)$ group model with weighted data. That is, we maximize

$$\sum_i^n \sum_{i=0}^K \hat{\pi}_{ik} \log f_k(y_i; \theta_k) = \sum_{i=0}^K \left(\sum_i^n \hat{\pi}_{ik} \log f_k(y_i; \theta_k) \right).$$

Note, however, that our model imposes equality across groups (the six groups of experiments) of the parameters that define the normal distributions of the mixture, while it allows for group specific mixing proportions, π_{ikg}, $g = 1, \ldots, 6$. This implies the substitution of (1) by

$$\hat{\pi}_{ikg} = \frac{\pi_{kg} f_k(y_i; \theta_k)}{\sum_1^K \pi_k f_k(y_i; \theta_k)} \quad \text{and} \quad \hat{\pi}_{kg} = \frac{1}{n_g} \sum_i^{n_g} \hat{\pi}_{ikg}, \quad g = 1, \ldots, 6. \quad (2)$$

In terms of Bayes theorem, $\hat{\pi}_{ikg}$ is the posterior probability of case i of group g to be in component k, $k = 0, 1, \ldots, K$. The posterior probabilities can be used to assign each observation to a component, by applying the simple rule that element i is assigned to component k if $\hat{\pi}_{ik} > \hat{\pi}_{ik'}$ for any $k' \neq k$. Note that in our approach, the posterior probabilities of belonging to component k change with the group g.

Information statistics can be computed using the general expression

$$C = -2\log L + qM,$$

where L is the likelihood of the data, M is some constant and q is the number of parameters to be estimated. The preferred model is the one with the smallest information criterium C, so the term qM is a penalty for over-parametrization of the model. In the present paper we set $M = 2$, which implies the use of the Akaike's information criterium (AIC) as the guide for choosing of our preferred mixtures model (see, e.g., Bozdogan (1970)).

11

4   Results of the Analysis

Using AIC to assess the fit of the model, we find that the preferred model includes five (truncated) normal distributions, in addition to the uniform and the normal censored components. The actual values of the AIC for the mixture models with four, five and six (truncated) normal distributions (plus the uniform and one left-censored distributions) are equal, respectively, to 6.7619, 6.7602 and 6.9022 (multiplied by 10^4), supporting the choice of five (truncated) normal distributions. When in this model we suppress the uniform component, then AIC jumps from 6.7602 to 6.7902 (both values multiplied by 10^4), which represents a substantial deterioration in the fit and indicates the need for the uniform component.

Using initial parameter estimates based on sample statistics (sample quantiles and variances), the EM algorithm achieves convergence after 775 iterations. The evolution of (minus) the likelihood function during the iterations process is shown in Figure 4.

Table 2 shows the estimates of the means and variances of the composing distributions, as well as the estimates of the mixing proportions across groups. Of the five components that correspond to the truncated normal distributions, three are uncannily centered at the values predicted by the IBRd hypothesis (estimated: 33.35; 22.89; 14.98; theoretical: 33.33; 22.22; 14.81). Note also that deviations around these means are moderate.

A fourth normal component is a very flat distribution, centered at 35.9 with a large SD of 9.37. This we interpret as indicating that the uniform distribution fails to capture all the Level-0 players. While the uniform dis-
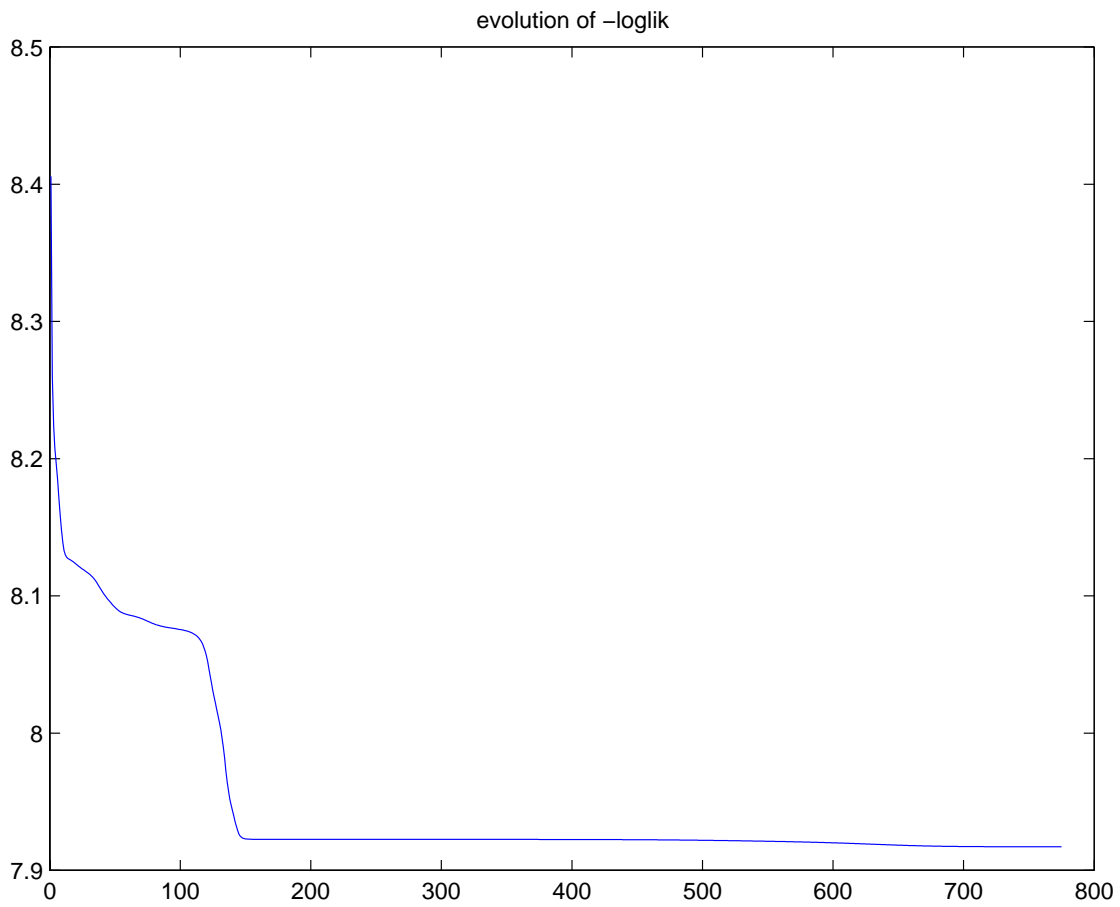
Figure 4: Evolution of the (minus) log-likelihood during iterations of the EM algorithm

13

tribution appears to take care of some random or haphazard choices between 0 and 100, the need for this normal component suggests that many of these choices are biased towards the lower half of the interval.[7] We conclude that Level-0 decisions are better described by both the uniform and this flat normal distribution. This interpretation would suggest that the number of Level-0 players is larger than previously thought.[8]

The fifth normal is centered at 7.35, below the theoretical prediction for Level-4 players. The interpretation for this normal distribution is not as straightforward as for other distributions. It could be the distribution of Level-4 choices, with a mean smaller than the theoretic value of 9.88. However, analyzing about 1000 comments submitted by participants in different BC experiments (see Bosch-Domènech et al. (2002)), we found that less than 1% reasoned at Level-4. Instead, participants reasoned either at most until Level-3, or jumped all the way to Level-∞. Among the choices belonging to subjects reaching Level-∞, only about 20% corresponded to the equilibrium and 60% were in the interval between the equilibrium and 10. This leads us to interpret this fifth distribution as capturing the choices of Level-∞ players rebounding from the equilibrium.

Finally, the estimated mean and standard deviation of the censored distribution are respectively 0.59 and 1.91. This distribution also accounts for choices of Level-∞ players. The proportion of censored observations in the different groups, both for the fitted and empirical distributions, are shown

---

[7]Actually, in game-theoretical parlance, choices above 66.66 are dominated.

[8]Using BC data on a sample of undergraduate students Nagel (1995) and Ho et al. (1998) calculate, respectively, a 13.1% and a 28.3% proportion of level-0 players. Using our sample of undergraduates we obtain that the relative size of level-0 players is 57.05%.

Table 2: Parameter estimates of the multiple-sample mixture model

	components						
	f_0	f_1	f_2	f_3	f_4	f_5	f_6
μ_k	*	35.91	33.35	22.89	14.98	7.35	0.59
σ_k	*	9.37	0.34	2.75	3.19	3.07	1.91
Reasoning levels	L-0	L-0	L-1	L-2	L-3	L-inf	L-inf
	proportions π_{kg} (in %)						
Lab	25.88	31.17	6.93	21.70	7.30	5.75	1.27
Classroom	17.56	18.11	14.79	18.57	12.47	9.83	8.67
Take-home	15.52	18.11	7.88	20.39	23.45	8.13	6.53
Theorist	12.93	11.66	3.20	9.49	10.89	19.51	32.31
Internet	13.74	16.36	9.25	15.01	13.77	7.60	24.26
Newspaper	15.31	15.96	8.32	15.35	14.71	14.57	15.78
Column mean	16.82	18.56	8.39	16.75	13.76	10.90	14.80

* uniform distribution

Table 3: The % of censoring in each group for the infinity level component

Groups	Lab	Classroom	Take-home	Theorist	Internet	Newspaper
Fitted %	0.75	5.07	3.82	18.90	14.19	9.23
Observed %	1.16	6.52	6.72	25.34	22.00	9.28

in Table 3. We observe that the proportion of censoring (i.e. the proportion of choices at the limit of the interval of choices) varies across groups, with the proportions being largest and smallest for the Theorist and Lab groups, respectively.

The components of the mixture distribution are depicted in Figure 5, where we show the probability density function of the various composing distributions, with the estimated mean values of the normal distributions displayed in the x-axis of the graph.

Table 2 also shows the estimates of the mixing proportions for each group. According to our interpretation, the first two columns of results in Table 2,
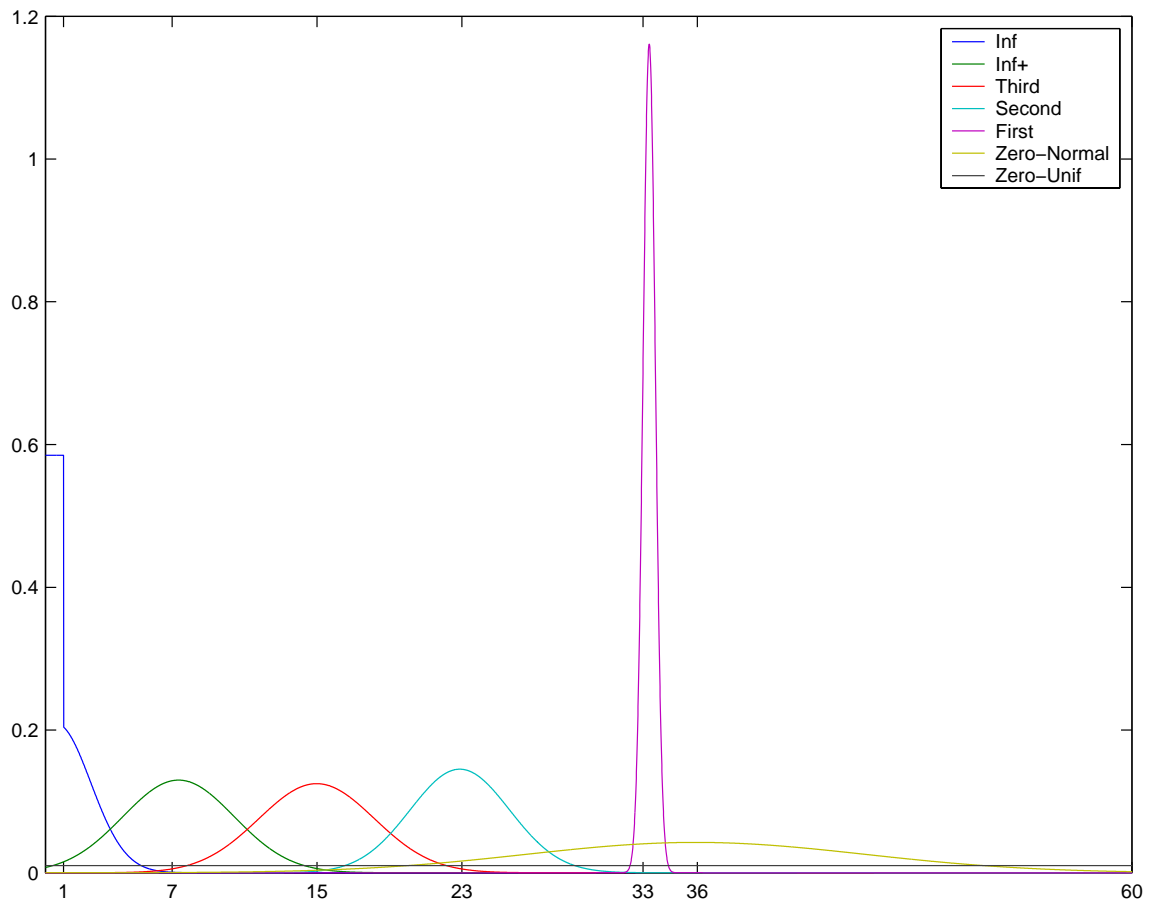
Figure 5: Components of the mixture distribution

16

taken together, would indicate the frequency of random, haphazard and un-explained choices. This proportion of Level-0 players range from about 25% among theorists to as much as close to 60% among undergraduate students. The number of Level-1 subjects tends to stay just below 10% in all groups, while Level-2 and Level-3 vary from 15% to 20% in most groups. Finally, Level-∞ participants appear in larger proportions among theorists, to as much as 51%, they consist in a fairly important chunk of newspaper readers, up to 30%, and in a small proportion of students in the lab, about 7%.

Combining the mixing proportions for each group, as they appear in Table 2, with the components of the mixture common to all the groups, as depicted in Figure 5, we obtain the fitted mixture distributions that are specific to each group, as shown in Figure 6. These fitted distributions correspond to the group-specific empirical distributions of Figure 2 and help to perceive the variation across groups of the proportions of individuals at the different levels of reasoning. It is remarkable that a unique set of components of the mixture allows us to fit the data from different groups by simply changing the mixing proportions across these groups.

An interesting feature is the increasing variance from Level-1 to Level-∞. People who reach Level-1 choose very tightly around 33. Those reaching Level-2 choose around 22, but not so tightly. The variance of the choice at Level-3 is even larger and it is largest in the choices of Level-∞ individuals, when we take the compound variance of the two distributions f_5 and f_6 of Table 2. [9]

---

[9]This is in contrast with Ho et al. (1998) and Stahl (1996), where variances were constrained to follow a decreasing pattern.
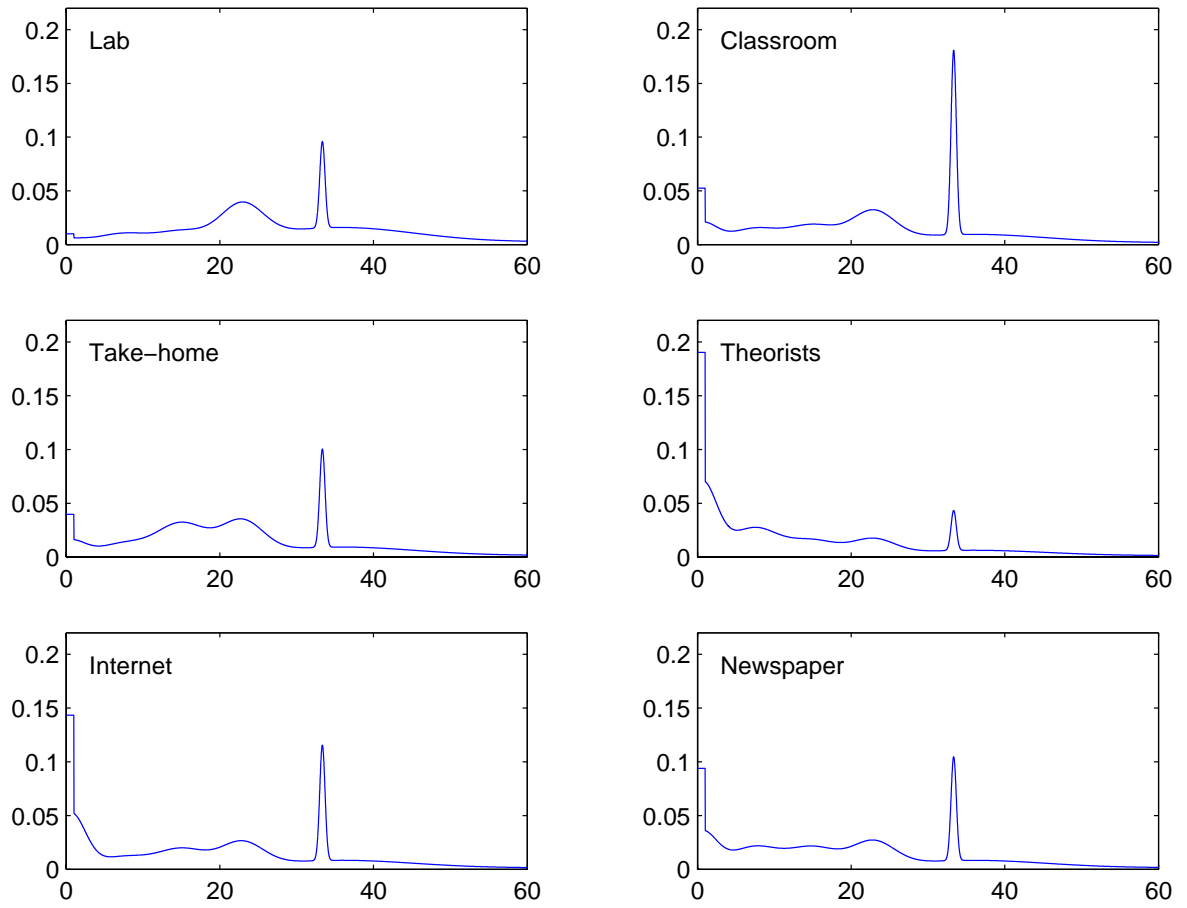
Figure 6: Fitted mixture distribution for each group

A plausible interpretation of this result is that as subjects take further steps of reasoning they become more and more aware of the complexity of the game, and assume that the rest of participants may make more and more dispersed choices. In any case, subjects at Level-k must believe that the dispersion of others' choices is centered around the choice of Level-$(k-1)$ players. Otherwise we would not see the sharp peaks we observe in the empirical data. Curiously, the increasing dispersion indicates that subjects at Level-k mistakenly believe that the dispersion of choices around Level-$(k-1)$ choice is larger than what in fact is.

To conclude, it appears that the estimated location of the composing distributions of the mixture gives empirical support to the IBRd hypothesis. The analysis also shows that the proportions of subjects with different levels of reasoning vary across groups.

5   Comparison with the literature

The literature on the estimation of data generated by BC experiments is quite diverse in its use of alternative statistical procedures. In her seminal paper on the BC, Nagel (1995) separates agents in bins centered around the theoretical values of the iterated best replies, $50p^k$, where k represents the iteration level and p the predetemined number that, when multiplied by the mean of all choosen numbers, yields the winning number. Stahl (1996) uses a boundedly rational learning rule assuming that, in the first period, the choice in each level k is distributed according to a truncated normal distribution with means specified (not estimated) at $50p^k$, and all variances following a

decreasing rule. Ho, Weigelt and Camerer (1998) specify a model in which the mean and variance of Level-k choices are functions of the mean and variance of choices at the previous level, so that the only parameters of the model are the mean and variance of Level-0 choices. This highly restricted model is then estimated by maximum likelihood.

These papers share many common features. The empirical models have as fundamental elements the decision rules used by subjects, the calculation errors or noise, and the beliefs about other players' strategies or types. Although some models take explicit account of errors in the individuals' choices (see El-Gamar and Grether (1995), or Haruvy, Stahl and Wilson (2000)), with BC data, the hypothesis of best response to type Level-$(k-1)$ players on the part of Level-k subjects provides a hierarchical model that becomes the basic tool to describe the set of decision rules.

Recently Camerer, Ho and Chong (2003) proposed a non-degenerated distribution of beliefs about other players choices. They assume that subjects believe that no other player uses as many levels of reasoning as themselves and assume also that players guess the relative proportion of other players at the different (lower) levels of reasoning. Since the number of levels of reasoning is an integer, Camerer, et al. (2003) argue that the Poisson distribution is a reasonable parametric distribution of other players reasoning levels. While this model fits well samples of data from different games, it cannot account for the multi-peaked distribution of choices typical of BC games.

In our empirical model we also assume that individuals share a common pattern of reasoning independently of the particular set-up of the BC experiment. Our choice of distribution functions is guided by the nature of

the data: truncated distributions between 0 and 100, since the choice set is constrained by these numbers, and a censored distribution to deal with the fact that there is non-null mass probability at values 0 or 1. The uniform distribution seems appropriate to take care of random choices.

All parameters of these distributions are estimated, and the number of distributions is not determined in advance. This approach is in contrast with the previous analysis just mentioned, where means and variances of a predetermined number of distributions are constrained to follow a particular sequence.

6   Conclusions

This paper provides a mixture distribution analysis of data obtained from experiments on the BC game, with diverse samples of subjects. The analysis is based on a model of censored and truncated normal distributions plus a uniform distribution, but does not impose any further structure on the model specification. The means and variances of the composing distributions of the mixture are let free, to be estimated, and so are the proportions of subjects at different levels of reasoning. Even the number of distributions involved is not predetermined. This is in contrast with previous statistical analysis of BC data.

A feature of our analysis is the assumption that individuals playing the BC game share a common pattern of reasoning, independently of the specific set-up of the experiment. However, we allow for variations across groups of experiments in the proportion of players using different depths of reasoning. In statistical terms this implies a unique specific composition of mixtures

21

across groups of experiments, with the mixing proportions of the components varying across groups. It is remarkable how much variation can be accounted for by a change in the mixing proportions. This set-up also permits the fitting of a complex mixture model to groups with relatively small sample sizes.

We apply this mixture distribution model to data gathered from experiments with newspapers readers, involving thousands of subjects in different countries, as well as from experiments run in labs with subject pools of undergraduate students, graduate students and economists. We estimate the mean and variance of each composing distribution, as well as the mixing proportions for each group of experiments. In view of the estimated locations of the composing distributions, our results support the hypothesis that individuals reason according to Iterated Best Reply (IBRd). Our results also show substantial variation across groups of the proportion of subjects using different levels of reasoning.

References

Arcidiacono, A. and Jones, J. (2003), 'Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm', *Econometrica*, 71, 3, 933-946.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., and Satorra, A. (2002), 'One, Two, Three, Infinity, ...: Newspaper and Lab Beauty-Contest Experiments ', *American Economic Review*, December, Vol 92 No.5, pp 1687-1701.

Bozdogan, H. (1987). ' Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions', *Psychometrika, 52, 345-370*

Camerer, C., Ho, T., Chong, J. (2003), 'A Cognitive Hierarchy Theory of One-Shot Games and Experimental Analysis'. *Quarterly Journal of Economics*, Forthcoming.

Dempster, A. P., N.M. Laird and D.B. Rubin (1977), 'Maximum Likelihood from Incomplete Data via de EM algorithm (With Discussion) ', *Journal of the Royal Statistical Society B, 39, 1-38*

Ho, T., Camerer, C., and Weigelt, K. (1998) 'Iterated Dominance and Iterated Best-Response in Experimental 'P-Beauty-contests', *American Economic Review, 88, 4,* pp. 947-969.

McLachlan, G. and Peel, D. (2000), *Finite Mixture Models* , John Wiley & Sons, New York.

Nagel, R. (1995) 'Unraveling in Guessing Games: An Experimental Study.' *American Economic Review*, 85 (5), 1313-1326.

Stahl, D.O. (1996) 'Rule Learning in a Guessing Game.' *Games and Economic Behavior*, 16(2), pp. 303-330.

Titterington, D., Smith, A. and Makov, U. (1992), *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.