Exp Econ
DOI 10.1007/s10683-010-9251-7

A finite mixture analysis of beauty-contest data using generalized beta distributions

Antoni Bosch-Domènech · José G. Montalvo · Rosemarie Nagel · Albert Satorra

Received: 14 July 2008 / Accepted: 7 July 2010

Abstract This paper introduces a mixture model based on the beta distribution, without pre-established means and variances, to analyze a large set of Beauty-Contest data obtained from diverse groups of experiments (Bosch-Domènech et al. 2002). This model gives a better fit of the experimental data, and more precision to the hypothesis that a large proportion of individuals follow a common pattern of reasoning, described as Iterated Best Reply (degenerate), than mixture models based on the normal distribution. The analysis shows that the means of the distributions across the groups of experiments are pretty stable, while the proportions of choices at different levels of reasoning vary across groups.

Keywords Beauty-contest experiments · Decision theory · Reasoning hierarchy · Finite mixture distribution · Beta distribution · EM algorithm

JEL Classification C14 · C16 · C72 · C91 · D03

1 Introduction

The paper applies a specific finite mixture model to identify patterns of belief formation and choice. The heterogeneity of subjects' decisions and beliefs favors clustering procedures to separate the observed outcomes, and mixture models are a convenient

We thank the editor and two referees for insightful comments that helped improve the manuscript. Thanks are also due to the Spanish Ministerio de Educación y Ciencia for financial help under research projects SEJ2005-03891/ECON (AB-D), SEJ2006-135 and SEJ2007-64340, and by the Barcelona GSE research network and the Fellowship Icrea Academia, Generalitat de Catalunya.

A. Bosch-Domènech · J.G. Montalvo · R. Nagel · A. Satorra (✉)
Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain
e-mail: albert.satorra@upf.edu

Published online: 28 July 2010

statistical tool for this type of investigation. Specifically, we analyze a data set collected from Beauty-Contest (BC) experiments by Bosch-Domènech et al. (2002) using a finite mixture of generalized beta distributions. The data in Bosch-Domènech et al. (2002) were obtained from seventeen different experiments. Often, experiments in economics draw their subjects from a student population and are typically run in laboratories. But the seventeen experiments whose data we are analyzing were performed in very diverse environments, involving different subject pools, sample sizes, payoffs, and settings: some data were collected in a laboratory and thus with high control, others in classrooms with undergraduate students, or in conferences, by e-mail and, most numerously, through newsgroups or among newspaper readers, a new source of data collection, enriching the growing area of field experiments.[1] We are, therefore, dealing with a rich and heterogeneous data set.

In a basic BC game, each player simultaneously chooses a number in an interval. The winner is the person whose number is closest to p times the mean of all chosen numbers, where p is a predetermined and known number.[2] The BC game facilitates an evaluation of the 'agents reasoning' and their beliefs on the other players' strategies. A nice feature of the game is that players are not guided by social norms, like fairness or altruism, which are often invoked when payoff-maximizing solutions cannot describe the observed behavior. The game is thus a game of pure competition. Another advantage of the game is that the high number of players that can participate in it makes collusion or cooperation difficult, therefore simplifying the interpretation of the decisions. Rational-expectations equilibrium "solves" the game by assuming common knowledge. That solution implies that all agents choose the lowest possible number, after applying the process of 'iterated elimination of weakly dominated strategies (IEDS)' starting from 100. But as can be seen from Fig. 1, this is not what necessarily happens in the seventeen experiments reported here, where $p = 2/3$ and the interval, in sixteen out of the seventeen experiments, is $[0, 100]$; in one experiment the choice set is $[1, 100]$.[3]

Alternatively to IEDS, a reasoning process that appears to describe better the observed behavior is the so-called Iterated Best Reply with degenerate beliefs (i.e., the belief that the choices of all others are at one precise value), denoted by IBRd, which classifies choices according to the depth, or number of levels, of reasoning.[4] Specifically, IBRd postulates that a Level-0 player chooses randomly in the given interval $[0, 100]$, with the mean being 50. At other levels, it is assumed that every player believes that she is exactly one level of reasoning deeper than the rest of players. A Level-1 player responds to her belief that everybody else is a Level-0 player choosing $50p$. A Level-2 player chooses $50p^2$, a Level-k player chooses $50p^k$, and so on.[5] Finally, a player who takes infinite steps of reasoning, and believes that all

[1] See Glen Harrison and John List (2004).

[2] For surveys of BC experiments see Camerer et al. (2004) and Nagel (1998, 2004).

[3] In the experiments with interval $[0, 100]$, $p < 1$ and the number of players greater than two, the only solution surviving the process of iterated elimination of weakly dominated strategies is to pick 0. Obviously, when the interval is $[1, 100]$, the rational solution is at 1.

[4] See, e.g., Nagel (1995), Stahl (1996) or Bosch-Domènech et al. (2002).

[5] As is customary in level-k models, we focus on the best response that minimizes the distance to the expected target value. Breitmoser (2010) challenges this conventional wisdom claiming that level-k models,

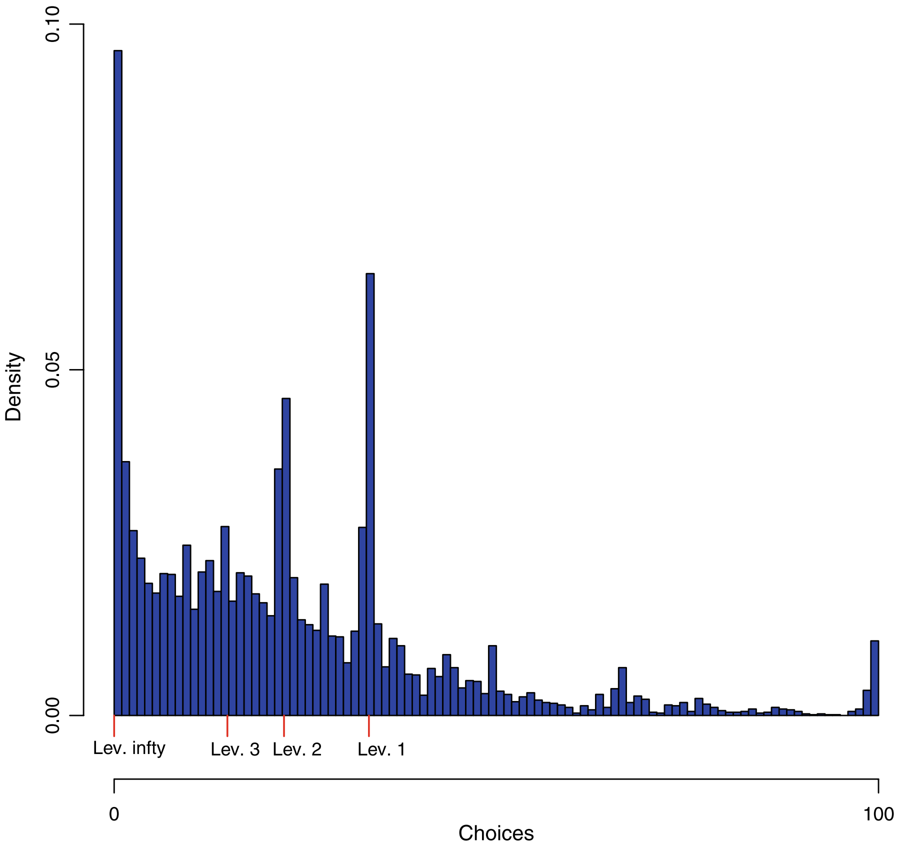

Fig. 1 Histogram for the aggregate of all the experiments. The points Level-1, Level-2, Level-3 and Level-∞ correspond to the choices of subjects with first, second, third and infinite levels of reasoning

players take infinite steps, chooses zero, the rational-expectations equilibrium. The hypothesis that players follow the reasoning process described as IBRd, 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 $33.33, 22.22, 14.81, 9.88, \ldots$ and, in the limit, 0.[6]

Starting with Nagel (1995), the analysis of the data generated by BC experiments uses diverse statistical procedures. In their seminal papers, Stahl and Wilson (1994,

as described in the literature, are not exactly, nor uniquely, rationalizable by belief systems based on iterated best response. While this is an interesting claim whose discussion lies beyond the scope of this paper, we base our choice of model on the actual observations from the experiments described in our paper plus the comments submitted by subjects in a number of experiments (see Bosch-Domènech et al. 2002 and Burchardi and Penczynski 2010), which show that a plurality of participants reason as hypothesized by the level-k model with degenerate beliefs that we propose.

[6]Because the process of iterated reasoning is similar to the one mentioned by J. M. Keynes in the context of a beauty-contest thought experiment (Keynes, 1936, p. 155), the class of p-mean-games and its variations have been called 'Beauty Contest games'.

1995) had the insight to connect the theoretical model of iterated levels of reasoning with the statistical method of mixtures, and based their analysis of a data set from 3×3 normal form games on this insight.[7] Later on, Stahl (1996) used the mixture model on data from Nagel's (1995) repeated BC games. In his paper he was interested in finding a learning rule and, to this end, he fitted a mixture model with normally distributed stochastic errors with means and variances highly restricted, so that the dynamics of the individual "learning" processes could be assessed. Specifically, he assumed that, at period 1, the individual choices in each Level-k were distributed with means specified at $50p^k$, and all variances followed a decreasing rule. Stahl's (1996) model included a uniform distribution to take care of the haphazard choices. Since his data came from experiments with students who seldom chose the equilibrium, he did not have to account for an important accumulation of choices at or near zero, and could postulate a model with a small number of iterations that did not include a final iteration to the equilibrium. In fact, the statistical procedure allowed him to choose between models with three or four levels of iteration, but was not designed to account for equilibrium (Level-∞) choices.

Our data, instead, come from a number of *one-period* experiments including a diversity of people in different settings, totaling more than 8,000 observations, which we organize in six groups, according to the type of experiment that provided them, namely Lab, Classroom, Take-home, Theorists, Internet, and Newspaper experiments (see Table 1).[8] As a result, we had data all over the range of possible choices and, in particular, at and around zero. With the same data set, in Nagel et al. (1999) (see also footnotes 26 and 28 in Bosch-Domènech et al. 2002) we fitted a mixture model similar to Stahl (1996) to the one-period data, with the *difference* that we did not constrain the means and variances of the composing distributions and, in addition, we included the Level-∞ (equilibrium) as a component of the model. This Level-∞ was a normal distribution truncated at zero. Building on this previous analysis of ours, we propose a model that, in addition to contemplating the Level-∞ and letting the means and variances free, benefits from the higher degree of flexibility provided by assuming that the stochastic errors follow a beta distribution. Contrary to the normal model, the beta distribution does not impose symmetry and adds the elegance bonus of using a single family of distributions in the mixture, as the uniform distribution is a specific beta distribution. Equally important, the beta distribution allows a more flexible representation of the Level-∞ component than the truncated normal distribution.[9]

In the meantime, other papers have followed in the approach of Stahl and Wilson (1994, 1995), fitting data from a variety of experimental games, notably Costa-Gomes

[7]Mixture models have a long history in statistics, and a comprehensive description of the statistical theory of finite mixture modeling can be found in Titterington et al. (1992) and McLachlan and Peel (2000).

[8]For details of the six groups see Bosch-Domènech et al. (2002).

[9]In economics the use of the beta distribution is not new. McKelvey and Palfrey (1992), for instance, use it in a different context, assuming that the heterogeneous beliefs of the participants in a centipede game are independently drawn from a beta distribution. In econometric analysis of labor and income distribution data, the beta distribution has a long tradition, see e.g. Heckman and Willis (1977) or Mcdonald and Xu (1995) and, more recently, Bognanno (2001) or Dastrup et al. (2007).

Table 1 The six 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	119
3 Take-home	2	Undergraduate students in take-home tasks, UPF	138
4 Theorists	4	Game Theory students and experts in game theory in conferences and e-mail	146
5 Internet	1	Newsgroup in internet	150
6 Newspapers	3	Readers of FT, E and S	7889
		Financial Times	*1468*
		Expansión	*3696*
		Spektrum der Wissenschaft	*2725*

et al. (2001), Camerer et al. (2004), Costa-Gomes and Crawford (2006), and Crawford and Iriberri (2007). Also, Ho et al. (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. In addition, Stahl (1998) estimated a levels-of-reasoning model assuming a non-degenerate distribution of beliefs about other players choices. In a similar vein, Camerer et al. (2004) assumed that subjects believe that no other player uses as many levels of reasoning as themselves and that players guess the relative proportion of other players at the lower levels of reasoning, arguing that the Poisson distribution is a reasonable parametric distribution of other players reasoning levels. While this model fits well some samples of data from different games, it does not account for stochastic responses, as in all other models described above. Recently, Harrison and Rutström (2009) use a mixture model to propose a statistical reconciliation between two dominant theories of choice under risk, EUT and Prospect Theory. Their paper offers an insightful alternative to the debate over the competing models by assuming that the data analyzed are generated by a mixture of decision rules associated to both theories.

From the computational perspective, all of the above applications use general-purpose optimization algorithms, such as the simplex method of Nelder and Mead

(1965).[10] But when handling finite mixture models, such algorithms require to carefully monitor the starting values to ensure proper convergence. Instead, in the present paper we use the Expectation Maximization (EM) algorithm (Dempster et al. 1977) that takes full advantage of the form of the likelihood function. Moreover, as a by-product of the EM algorithm, a Bayes rule classification method for assigning cases to the different levels of the reasoning process is made available.

2 Data description

The histogram for the whole distribution, when all the groups are pooled together (see Fig. 1), shows that the peaks closely correspond to the numbers chosen by individuals who would have reasoned according to the IBRd hypothesis, at levels one, two, perhaps three, and infinity. In addition, inspecting the choices for each group, as depicted in Fig. 3 of Bosch-Domènech et al. (2002), we see substantial variation across groups. The first group, Lab-experiments with undergraduates, is clearly distinguished from the rest, because the Nash equilibrium is rarely selected. Yet, when subjects have some training in game theory, the proportion of subjects choosing the equilibrium increases. 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, as should be expected from the heterogeneous level of training of their readers.

3 Finite mixture model and estimation procedure

The basic problem in fitting a statistical model to the BC data is the existence of unobserved heterogeneity (the different levels of reasoning), in addition to the multiple-group structure. This section develops the interpretation of the BC data as a mixture of beta distributions and provides the statistical strategy to estimate such a model.

To begin, the components of the mixture are anchored at the different levels of reasoning: 33.33 for Level-1, 22.22 for Level-2, 14.81 for Level-3,..., 0 for Level-∞. In addition, choices at Level-0 are captured by one specific beta distribution: the uniform distribution at (0, 100).

[10]We thank a referee for pointing out an older literature in statistics showing that the beta distribution can be well approximated by the logistic transformation of the normal, a useful fact since in many settings there exists code that works for the normal distribution, thus with existing software a simple non-linear transformation could be used to estimate the beta. See Lesaffre et al. (2007) for an updated review of the logistic transformation of the normal, and Andersen et al. (2009) for novel methods for analyzing non-linear mixed logit models.

3.1 A finite mixture model based on beta distributions

A variable X is said to belong to the family of beta distributions if its probability density function (pdf) is

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)},$$

$\alpha > 0$ and $\beta > 0$; here, $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$ is the *beta function*, and $I_{(0,1)}$ is an indicator function. We say that X is a standard beta distribution, $X \sim \text{Beta}(\alpha, \beta)$. In this paper we use the generalized beta distribution $Y \sim \text{Beta}(\alpha, \beta, a, h)$ arising from the linear transformation $Y = a + hX$, $X \sim \text{Beta}(\alpha, \beta)$ and $h > 0$, where the support of variables X and Y are (0, 1) and $(a, a+h)$ respectively. The shape of the distribution is dictated by the values of α and β. When both α and β are > 1, the distribution is unimodal; when $\alpha < 1$ (or $\beta < 1$) then $f_Y(y) \to \infty$ when $y \to a$ (or when $y \to a + h$). When both α and β are < 1, the density function is U shaped and tends to ∞ both when $y \to a$ and when $y \to a + h$. Since the distribution has bounded support, all its moments exist; the mean and variance of $\text{B}(\alpha, \beta, a, h)$ are $\mu = a + \frac{h\alpha}{\alpha+\beta}$ and $\sigma^2 = \frac{h^2\alpha(\alpha+1)}{(\alpha+\beta)^2(\alpha+\beta+1)}$, respectively.

We denote the multiple-group 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.

Associated to a model with L levels of reasoning, we consider the following L-component mixture probability density function for y:

$$f(y, \vartheta) = \pi_0 f_0(y) + \pi_1 f_1(y, \theta_1) + \cdots + \pi_K f_K(y, \theta_K) + \pi_\infty f_\infty(y, \theta_\infty),$$

where $K = L - 2$.

- f_0 is the uniform distribution in 0 to 100; that is, $f_0 \sim \text{Beta}(1, 1, 0, 100)$, associated to the Level-0 of reasoning.
- $f_k \sim B(\alpha_k, \beta_k, a_k, h)$, $k = 1, \ldots, K$, where $a_k = ((\frac{2}{3})^k H) - h/2$ are fixed values determined by the model specification, and α_k and β_k are free parameters, to be estimated, that determine the shape of the distribution. Note that f_k has support $(a_k, a_k + h)$, so $a_k + h/2$ is the midpoint of the support, called the *anchor point* of component f_k. As explained below, the *width* h of the support of f_k is taken to be 20. H is the number at which the reasoning process of an individual—who is not Level-0—starts. The value of H will be 50 or 100 depending on whether we adopt the IBRd or the IEDS theories to model this reasoning process.
- $f_\infty \sim B(\alpha_\infty, \beta_\infty, 0, h)$, with α_∞ and β_∞ being parameters to be estimated, associated to the Level-∞ of reasoning.
- The π_k's are the mixing proportions. They are positive numbers and add up to 1, i.e. $\sum_{k=0}^K \pi_k + \pi_\infty = 1$. The mixing proportions are the weights of the different components of the mixture distribution.

Moreover, $\vartheta = (\pi, \theta)'$, where $\pi = (\pi_0, \pi_1, \ldots, \pi_K, \pi_\infty)$, $\theta = (\theta_1, \ldots, \theta_K, \theta_\infty)$ and $\theta_k = (\alpha_k, \beta_k)$. The different components of the mixture accounts for the (unobserved) levels of reasoning, from the uniform distribution in (0, 100), accounting for Level-0, to the f_∞ accounting for Level-∞. The analysis estimates both the 'weights' π_k of the different components of the mixture and the shape parameters $\theta_k = (\alpha_k, \beta_k)$.

Concerning the analysis of the different groups, we consider two models. One sets the vector of mixing proportions π as well as the vector of shapes θ to be specific for each group. The second model sets the vector θ (the shapes) as well as the vector of mixing proportions π equal across groups.

The participants in the Newspaper experiments were asked to provide comments about their decision process. We use this information to fix the width h of each of the beta distributions as well as the number $L = K + 2$ of levels of reasoning. The comments of the participants in the experiment indicate that subjects who declared to be at level of reasoning $k > 0$ gives numbers deviating no more than 10 units from the corresponding anchor point $a_k + h/2$, the center of the support of the kth component.[11] This suggests a range of 20 for h.[12] Were the true range smaller, this would be picked up by the fit, since α and β are free parameters to be estimated. Accordingly, except for Level-0, that has range 100, we assign $h = 20$ to all the other components. We decided to start at $L = 4$ because the received comments showed a very small number of Level-3 participants (below 5%). We now proceed to describe the estimation of the model.

3.2 Model fitting

We estimate the mixture model by maximum likelihood (ML). From the previous subsection, it follows that the log-likelihood function is

$$\ell(\vartheta) = \sum_{i=0}^{n} \log\left(\sum_{k=0}^{K+1} \pi_k f_k(y_i; \theta_k)\right), \tag{1}$$

where i varies across sample units, k varies over all components of the mixture, and $f_{K+1} = f_\infty$ and $\pi_{K+1} = \pi_\infty$. This log-likelihood function is highly non-linear, so its maximization using standard optimization routines may be difficult. Instead, we resort to the EM algorithm, which exploits the fact that when the π's are given, the likelihood function is trivially optimized.

Consider the data augmented with an unobserved multinomial class-membership or vector of indicators $d_i = (d_{i0}, d_{i1}, \ldots, d_{iK}, d_{i\infty})'$, where $d_{ik} = 1$ or 0 and $\sum_{k=0}^{K+1} d_{ik} = 1$. The vector d_i is assumed to have a multinomial distribution with parameters $\pi = (\pi_0, \ldots, \pi_K, \pi_\infty)$. The (complete-data) log-likelihood is thus

$$\ell_C(\vartheta) = \sum_{i=1}^{n} \sum_{k=0}^{K+1} d_{ik}(\log \pi_k + \log f_k(y_i; \theta_k)).$$

The EM approach computes ML estimates using the following algorithm:

[11] This is confirmed by Burchardi and Penczynski (2010) who run a team guessing game in which a team exchanged written messages about which number to submit as a team.

[12] Only a small number of players at Level-∞ (less than 3% of the whole sample) deviate more than the range of 20.

E-step: From current estimates $\pi_k^{(t)}$ and $\theta_k^{(t)}$ of the parameters (at tth iteration), compute the expected value $Q = Q(\theta)$ of the (complete data) log-likelihood $\ell_C(\theta)$. This expectation turns out to be

$$Q = \sum_{i=0}^{n} \sum_{k=0}^{K+1} \hat{\pi}_{ik} (\log \pi_k + \log f_k(y_i; \theta_k)), \tag{2}$$

where

$$\hat{\pi}_{ik} = \frac{\pi_k^{(t)} f_k(y_i; \theta_k^{(t)})}{\sum_{k=0}^{K+1} \pi_k^{(t)} f_k(y_i; \theta_k^{(t)})}. \tag{3}$$

Note that Q can be written as:

$$Q = \sum_{i=1}^{n} \sum_{k=0}^{K+1} \hat{\pi}_{ik} (\log \pi_k) + \sum_{k=0}^{K+1} \left(\sum_{i=1}^{n} \hat{\pi}_{ik} \log f_k(y_i; \theta_k) \right). \tag{4}$$

M-step: We update $\pi_k^{(t)}$ and $\theta_k^{(t)}$ with the maximizers of Q of (2). This is accomplished by

- updating $\theta_k^{(t)}$ with

$$\theta_k^{(t+1)} = \arg\max \sum_{i=1}^{n} \hat{\pi}_{ik} \log f_k(y_i; \theta_k), \tag{5}$$

- updating $\pi_k^{(t)}$ with

$$\pi_k^{(t+1)} = \sum_{i=0}^{n} \hat{\pi}_{ik}/n \tag{6}$$

($\hat{\pi}_{ik}$ as defined in (3)).

EM iterate steps 1 and 2 till convergence is achieved.

Note that the M-step implies the simple maximization with respect to θ_k of a one-component likelihood function with a weighted sum of individual log-likelihoods (see (5)); in contrast to the usual log-likelihood of the mixture model, this is easily accomplished with basic optimization routines.

In addition to providing a convenient algorithm, the EM procedure gives, for every subject, the posterior probabilities of belonging to each of the components (or levels of reasoning). From a Bayesian perspective, the $\hat{\pi}_{ik}$ of (3), at the end of the iteration, can be interpreted as the estimated conditional probabilities of case i belonging to component k, while the $\hat{\pi}_k$ of (6) are the marginal probabilities of each component. By averaging the conditional probabilities $\hat{\pi}_{ik}$ within each group, we obtain the marginal probabilities $\hat{\pi}_{kg}$ of each component in each group gth, $g = 1, \ldots, G$. Both, the marginal probabilities for all the groups and conditional for each group are reported below.

Since the EM algorithm produces maximum likelihood estimates, we can compute the maximum log-likelihood ($\log L$) of the data, that is the value of (1) evaluated at the ML parameter estimates. To compare the fit of competing models, we suggest the Akaike Information Criteria (AIC) $AIC = -2\log L + 2q$, where q is the number of parameters of the estimated model.[13]

4 Results of the analysis

We fit the model, separately, for the whole sample and for each of the six groups. The results are presented in Table 2 which shows the estimates of the means and variances of the composing distributions, as well as the estimates of the mixing proportions across groups. Standard errors (se) of all the sample estimates have been computed using bootstrap (see, e.g., Efron and Tibshirani 1986).[14] About 30% of the choices in the overall sample correspond to a Level-0 of reasoning. The remaining choices are distributed among levels 1, 2 or ∞ in different proportions. Not surprisingly, a large number (62%) of "Theorists" choose Level-∞, while only 4% of "Lab" subjects reach this level of iteration.

The standard errors for the estimates of the various components' weights give a (95% confidence interval) variation of a tight ± 0.05 for the Newspaper experiment, and a much more variable ± 0.10 for the smaller sample size experiments. A similar pattern arises when computing the 95% confidence intervals for means and variances.

Figure 2 shows the density functions of the four components of the mixture model for the overall sample as well as the fitted mixture. To test whether each group should be analyzed separately, we compute the log-likelihood ratio test for the null hypothesis of a single-group model versus a model with parameters specific for each group. From the values in the last column of Table 2 we obtain $\chi^2 = 234.76$ with 10 degrees of freedom, p-value <0.0001. For model comparison, a different option that is free from sample size dependence and applies also to non-nested models is an information criteria such as the AIC. The AIC difference for the "all sample" model and the model with parameters specific for each group equals 134.7, favoring again the multiple-group model. The maximized log-likelihood values and the estimates of πs and θs for each group are reported in Table 2.

An interesting feature is the increasing variance from Level-1 to Level-∞. With few exceptions, people who reach Level-1 choose tightly around 33. Those reaching Level-2 choose around 22, but, in general, not so tightly. Level-∞ presents, in general, a higher variance.[15] This observation is supported even when taking into account the variation of the standard errors of the variances, shown in Table 2. A plausible interpretation of this result is that as subjects take further steps of reasoning

[13] See, e.g., Bozdogan (1987) for a review on AIC and other alternatives.

[14] We used 200 as the number of bootstrap samples.

[15] This is in contrast with Ho et al. (1998) and Stahl (1996), where variances were postulated to follow a decreasing pattern, but in accordance with the observations of Kübler and Weizecker (2004) and Goeree and Holt (2004).

Table 2 Fit of the four-component ($K = 2$) mixture model

	Level (anchor)								Log-lik
	L∞ (0)		L2 (22.22)		L1 (33.33)		L0 (50)		
	Est	(*se*)	Est	(*se*)	Est	(*se*)	Est	(*se*)	
All sample (8528)									
Mean	6.65	(*0.39*)	22.78	(*0.23*)	33.35	(*0.27*)	50.00	†	–34 023.33
sd	6.51	(*0.18*)	4.22	(*0.92*)	0.37	(*1.04*)	28.87	†	
π_k	0.39	(*0.02*)	0.22	(*0.03*)	0.09	(*0.03*)	0.31	(*0.02*)	
Newspaper (7889)									
Mean	6.71	(*0.33*)	22.80	(*0.23*)	33.36	(*0.22*)	50.00		–31 517.79
sd	6.48	(*0.17*)	4.22	(*0.73*)	0.37	(*0.80*)	28.87		
π_k	0.39	(*0.02*)	0.22	(*0.02*)	0.09	(*0.03*)	0.30	(*0.02*)	
Lab (86)									
Mean	9.04	(*0.99*)	23.22	(*1.34*)	33.72	(*0.31*)	50.00		–363.72
sd	2.05	(*1.00*)	2.57	(*1.05*)	2.07	(*0.90*)	28.87		
π_k	0.04	(*0.03*)	0.25	(*0.05*)	0.21	(*0.09*)	0.50	(*0.06*)	
Classroom (119)									
Mean	9.31	(*2.41*)	20.14	(*1.13*)	32.70	(*0.98*)	50.00		–480.36
sd	7.07	(*2.32*)	4.21	(*1.63*)	3.42	(*1.45*)	28.87		
π_k	0.24	(*0.09*)	0.35	(*0.11*)	0.18	(*0.06*)	0.23	(*0.07*)	
Take home (138)									
Mean	7.40	(*1.41*)	22.81	(*0.67*)	33.17	(*0.52*)	50.00		–541.03
sd	6.85	(*0.76*)	3.36	(*1.39*)	0.19	(*1.03*)	28.87		
π_k	0.23	(*0.05*)	0.24	(*0.06*)	0.15	(*0.08*)	0.38	(*0.08*)	
Theorist (146)									
Mean	4.67	(*0.62*)	22.00	(*1.26*)	34.15	(*1.79*)	50.00		–487.27
sd	6.01	(*0.65*)	1.65	(*1.521*)	3.81	(*2.01*)	28.87		
π_k	0.62	(*0.06*)	0.09	(*0.03*)	0.08	(*0.03*)	0.21	(*0.06*)	
Internet (150)									
Mean	4.14	(*1.46*)	22.38	(*1.06*)	32.85	(*1.08*)	50.00		–515.78
sd	5.76	(*0.64*)	5.05	(*1.70*)	0.44	(*0.07*)	28.87		
π_k	0.37	(*0.06*)	0.26	(*0.04*)	0.12	(*0.03*)	0.26	(*0.04*)	

† Equal to zero because it refers to a value set in the analysis

they become more and more aware of the complexity of the game, and assume that the rest of participants may make more dispersed choices.

We also want to assess whether the IBRd model (where $H = 50$) fits better the data than the (not nested) IEDS model (where $H = 100$). For the same number of levels of reasoning, IBRd clearly outperforms IEDS for all samples. Table 3 shows,

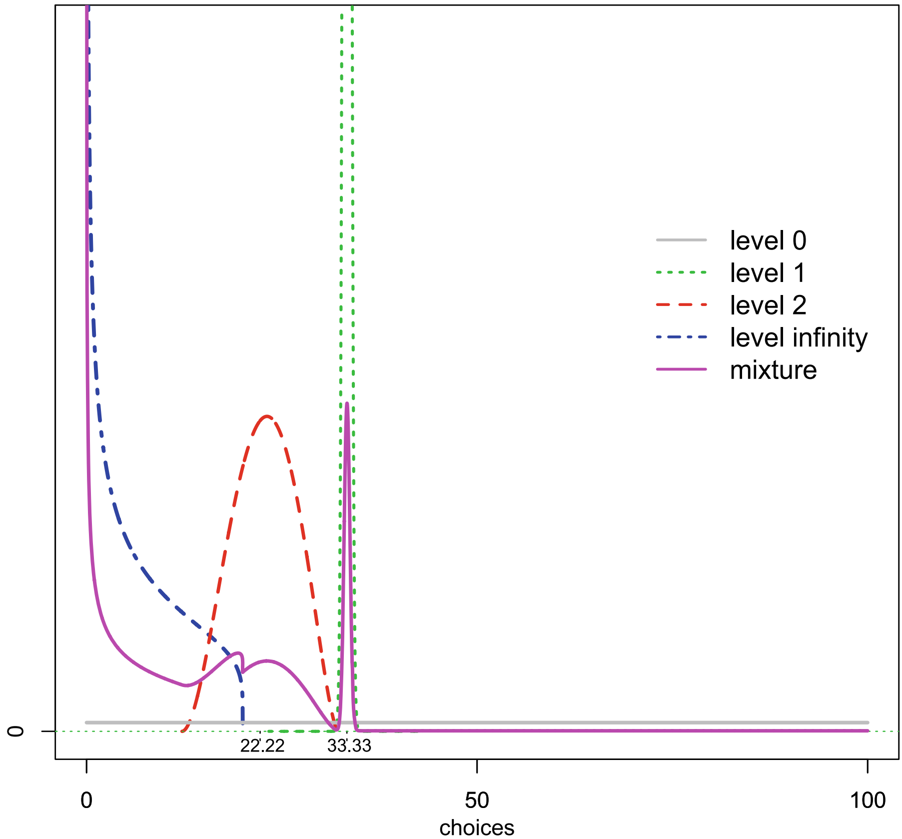

Fig. 2 The graph show the fitted probability density functions of the components for levels of reasoning 0, 1, 2 and ∞ (All sample). Level-0 is the straight line parallel to the x-axis, Level-1 is centered at 33.33 and Level-2 at 22.22, Level-∞ is the density more in the left. The probability density for Level-1 is shown truncated for clarity of presentation of the whole graph. The graph shows also the probability density function of the fitted mixture that results from applying the (fitted) weights .31, .09, .22 and .39 on components 0, 1, 2 and ∞

in particular, the AIC differences when comparing IBRd with four levels of reasoning ($K = 2$) to IEDS with five levels ($K = 3$). Since the difference is positive, the preferred model is IBRd.[16]

Finally, for the IBRd model, the AIC taking into account the number of components shows no gain when increasing from $K = 2$ ($L = 4$ components) to $K = 3$ ($L = 5$ components).

These results compare favorably with our previous specification in Nagel et al. (1999) which uses (truncated) normal distributions instead of beta distributions, where we obtained for the Newspaper data a maximized log-likelihood of $-32,045.7$ (see Table 4 in that paper) for the best fitting model, a model with five components,

[16]The same conclusion is obtained if the model comparison is based on the likelihood principle, or if the Schwarz information criterion is used instead of the AIC.

Table 3 Comparison of the models IBRd ($K=2$) and IESD ($K=3$) by the AIC criteria. Positive difference favors IBRd

Groups	Sample size	AIC_{IRBd}	AIC_{IESD}	Difference
All sample	8 525	68 066.66	68 264.71	198.05
six groups	8 525	67 931.90	68 214.96	283.06
Newspaper	7 889	63 055.59	63 286.86	231.27
Lab	86	747.44	751.33	3.88
Classroom	119	980.71	1004.07	23.35
Take-home	138	1102.05	1104.28	2.23
Theorist	146	994.55	999.23	4.68
Internet	150	1051.56	1069.21	17.65

a value that is 528.7 less (worse) than the one in the present model (see second row of Table 2), with only four components based on beta distributions, a more parsimonious model.[17] For the same data and the same number of components, the approach of Stahl (1996) would yield an even smaller likelihood than the reported likelihood of Nagel et al. (1999), since Stahl's model imposes additional restrictions on the means and variances of the composing normals and does not add the Level-∞ component.[18]

5 Conclusions

This paper provides an analysis of a finite mixture of beta distributions, without pre-established means and variances, for data obtained from experiments on the BC game, with diverse samples of subjects amounting to more than 8,000 observations. In contrast with the mixture of normal distributions, that imposes symmetry or truncation on the composing distributions, the mixture of beta distributions allows varying shapes for the composing distributions of the mixture. Therefore, for our specific experimental data, the beta family of distributions confers more flexibility to the statistical analysis, thus providing a better fit to this set of BC data than previous attempts based on a mixture of normals.

Our model assumes that all individuals playing the BC game share a common pattern of reasoning, described as Iterated Best Reply (degenerate), and the composing distributions are anchored according to the specifications of the theoretical model.

[17] In relation to the estimated proportions, we do not obtain noticeable differences in both approaches for Level-0, Level-1 and Level-2 (27, 11, 23 in the normal versus 30, 9, 22 for the beta); thus Level-3 plus Level-∞ for the normal case gives the same proportion as the Level-∞ of the beta model (39%). Also the beta model gives a variance more than six times smaller than the normal approach for Level-1, while the variances are similar for Level-2, but for Level-∞ the variance is almost twice as large for the beta approach (as Level-3 disappears in the beta model).

[18] Since Camerer et al. (2004) model does not contain a stochastic random term, their approach cannot be compared with ours in terms of log-likelihood. Costa-Gomes and Crawford (2006) and Costa-Gomes et al. (2001) analyze the data from 3×3 normal form games, thus providing insights about a different sort of strategic behavior than the one confronting BC players.

However, their means and variances are let free, to be estimated, and so are, obviously, the proportions of choices at the different levels of reasoning. Similarly, the number of distributions involved is not predetermined. The estimation shows that about 30% of our 8,000 subjects use Level-0, 10% Level-1, 20% Level-2 and 40% Level-∞. Thus, it appears that once a player has moved up from Level-0 she, most often, tends to further proceed beyond Level-1. The analysis indicates that the hypothesis that individuals reason according to the Iterated Best Reply (degenerate) model fares well when compared to the Iterated Elimination of Dominated Strategies model. It also shows that, in the Beauty Contest game, the number of Level-3 choices is not large, suggesting that most players who manage to reach Level-3 decide to jump from there to the equilibrium solution.

As a general conclusion for wider application, it appears that for some experimental data, the beta distribution can be a useful alternative to the normal and, as it avoids truncation and restrictions of symmetry, it yields a simpler analysis with better fit, at no added computational cost.

References

Andersen, S., Harrison, G. W., Hole, A. R., & Rutström, E. E. (2009). *Non-linear mixed logit and the characterization of individual heterogeneity* (Working Paper 09-02). Department of Economics, College of Business Administration, University of Central Florida.

Bognanno, M. L. (2001). Corporate tournaments. *Journal of Labor Economics*, *19*(2), 290–315.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two, (three), infinity, ...: Newspaper and lab beauty-contest experiments. *American Economic Review*, *92*(5), 1687–1701.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370.

Breitmoser, Y. (2010). *Hierarchical reasoning versus iterated reasoning in p-beauty contest guessing games* (MPRA Paper No. 19893). Munich Personal RePEc Archive.

Burchardi, K. B., & Penczynski, S. P. (2010). *Out of your mind: Eliciting individual reasoning in one shot games*. Mimeo, London School of Economics.

Camerer, C., Ho, T., & Chong, J. (2004). A cognitive hierarchy theory of one-shot games and experimental analysis. *Quarterly Journal of Economics*, *119*, 861.

Costa-Gomes, M., & Crawford, V. (2006). Cognition and behavior in guessing games: An experimental study. *American Economic Review*, *96*, 1737–1768.

Costa-Gomes, M., Crawford, V., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, *69*, 1193–1235.

Crawford, V., & Iriberri, N. (2007). Level-*k* auctions: Can boundedly rational strategic thinking explain the winner's curse and overbidding in privatevalue auctions? *Econometrica*, *75*, 1721–1770.

Dastrup, S. R., Hartshorn, R. Y., & Mcdonald, J. B. (2007). The impact of taxes and transfer payments on the distribution of income: A parametric comparison. *Journal of Economic Inequality, 5*, *3*, 1569–1721.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via de EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, *39*, 1–38.

Efron, B., & Tibshirani, R. (1986). Bootstrap measures for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*, 54–77.

Goeree, J. K., & Holt, C. A. (2004). A model of noisy introspection. *Games and Economic Behavior*, *46*(2), 365–382.

Harrison, G., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *XLII*, 1013–1059.

Harrison, G. W., & Rutström, E. E. (2009). Expected utility and prospect theory: One wedding and decent funeral. *Experimental Economics*, *12*(2), 133–158.

Heckman, J. J., & Willis, R. J. (1977). A betalogistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy*, *85*(1), 27–58.

Ho, T., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best-response in experimental 'p-beauty-contests'. *American Economic Review*, *88*(4), 947–969.
Keynes, J. M. (1936). *The general theory of employment, interest and money*. London: Macmillan.
Kübler, D., & Weizacker, G. (2004). Limited depth of reasoning and failure of cascade formation in the laboratory. *Review of Economic Studies*, *71*, 425–441.
Lesaffre, E., Rizopoulos, D., & Tsonaka, R. (2007). The logistic transform for bounded outcome scores. *Biostatistics*, *8*(1), 72–85.
Mcdonald, J. B., & Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, *66*, 133–152.
McKelvey, R. D., & Palfrey, T. R. (1992). An experimental study of the centipede game. *Econometrica*, *60*(4), 803–836.
McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, *85*(5), 1313–1326.
Nagel, R. (1998). A survey on experimental beauty-contest games: Bounded rationality and learning. In D. Budescu, I. Erev, & R. Zwick (Eds.), *Games and human behavior, Essays in honor of Amnon Rapoport* (pp. 105–142). Hillsdale: Lawrence Erlbaum Associates, Inc.
Nagel, R. (2004). How to improve reasoning in experimental beauty-contest games: A survey. In D. Friedman, & A. Cassar (Eds.), *Economics Lab: An intensive course on experimental economics. Lessons from the Trento summer school in experimental economics*. London: Routledge.
Nagel, R., Bosch-Domènech, A., Satorra, A., & Montalvo, J. G. (1999). *One, two, (three), infinity…. Newspaper and lab beauty-contest experiments* (Economics and Business WPS, Paper # 438). Barcelona: Universitat Pompeu Fabra.
Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.
Stahl, D. O. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior*, *16*, 303–330.
Stahl, D. O. (1998). Is step-j thinking an arbitrary modelling restriction or a fact of human nature? *Journal of Economic Behavior and Organization*, *37*, 33–51.
Stahl, D. O., & Wilson, P. W. (1994). 'Experimental evidence on players' models of other players. *Economic Behavior & Organization*, *25*, 309–327.
Stahl, D. O., & Wilson, P. W. (1995). 'On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, *10*, 213–254.
Titterington, D., Smith, A., & Makov, U. (1992). *Statistical analysis of finite mixture distributions*. New York: Wiley.