Computers & Operations Research 36 (2009) 1726 – 1730

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/cor

Polynomial calculation of the Shapley value based on sampling

Javier Castro[a,*], Daniel Gómez[a], Juan Tejada[b]

[a]Dpto. de Estadística e I.O. III, Escuela Universitaria de Estadística, Universidad Complutense de Madrid, Av. Puerta de Hierro s/n, 28040 Madrid, Spain
[b]Dpto. de Estadística e I.O. I, Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, Plaza de Ciencias 3, 28040 Madrid, Spain

ARTICLE INFO

Available online 15 April 2008

Keywords:
Game theory
Shapley value
Sampling algorithm

ABSTRACT

In this paper we develop a polynomial method based on sampling theory that can be used to estimate the Shapley value (or any semivalue) for cooperative games. Besides analyzing the complexity problem, we examine some desirable statistical properties of the proposed approach and provide some computational results.

1. Introduction

One of the most important solution concepts in cooperative games was defined by Shapley [1]. This solution concept is now known as the Shapley value. The Shapley value is useful when there exists a need to allocate the worth that a set of players can achieve if they agree to cooperate.

Although the Shapley value has been widely studied from a theoretical point of view, the problem of its calculation still exists. In fact, it can be proved that the problem of computing the Shapley value is an NP-complete problem (see Deng and Papadimitriou [2], Fernández et al. [3], or Faigle and Kern [4], for more details).

Several authors have been trying to find algorithms to calculate the Shapley value precisely for particular classes of games. In Bilbao et al. [5] for example, where a special class of voting game is examined, theoretical antimatriod concepts are used to polynomially compute the Shapley value. In Granot et al. [6] a polynomial algorithm is developed for a special case of an operation research game. In Castro et al. [7], it is proved that the Shapley value for an airport game can be computed in polynomial time by taking into account that this value is obtained using the serial cost sharing rule. In contrast, however, few efforts have focused on the approximation of the Shapley value. Considering the wide application of game theory to real world problems, where exact solutions are often not possible, a need exists to develop algorithms that facilitate this approximation.

Only a few references can be found in which approximations are developed to estimate the Shapley value. Although the multilinear extension defined by Owen [8] is an exact method for simple games (we will say that a game v is a *simple game* when the possible gain for any coalition is assumed to be either 0 or 1), the calculation of the corresponding integral is not a trivial task. So, when this integral is approximated (using the central limit theorem) this methodology could be considered as an approximation method. In Fatima et al. [9], a randomized polynomial method for determining the approximate Shapley value is presented for voting games. However, the aim of this work is to develop an efficient algorithm that can estimate the Shapley value for a large class of games.

Sampling (see Cochran [10] or Lohr [11] for example) is a process or method of drawing a representative group of individuals or cases from a particular population. Sampling and statistical inference are used in circumstances in which it is impractical to obtain information from every member of the population. Taking this into account, we use sampling in this paper to estimate the Shapley value and any semivalues. These estimations are efficient if the worth of any coalition can be calculated in polynomial time.

The remainder of the paper is organized as follows. In Section 2, we introduce the preliminaries and some notation. In Section 3, we develop an algorithm for the estimation of the Shapley value with some desirable properties. In Section 4, we present some computational results for some well-known cooperative games; in particular, a comparison with the multilinear extension is done. We finish the paper with some conclusions and final remarks.

2. Preliminaries

In this section we introduce some sampling and theoretical game terminology in order to provide a better understanding of the rest of the paper.

An n-person game in characteristic function form is defined as a 2-tuple (N, v), N being the set of players and v a real valued function defined on the subsets of N satisfying $v(\emptyset) = 0$. Given an n-person game (N, v), $v(S)$ represents the worth of the set of players S. When this worth represents the cost that players must be charged, the

* Corresponding author. Tel.: +34 913943962; fax: +34 913944064.
E-mail addresses: jcastroc@estad.ucm.es (J. Castro), dagomez@estad.ucm.es (D. Gómez), jtejada@mat.ucm.es (J. Tejada).

0305-0548/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2008.04.004

game is known as a cost game and we will denote these kind of games as (N, c).

Given an n-person game (N, v), a semivalue of a player i (Dubey et al. [12]) is a weighted sum of its marginal contributions $v(S \cup \{i\}) - v(S)$, $S \subset N \backslash \{i\}$. Formally, a semivalue of an n-person game (N, v) is defined as

$$\gamma_i(v) = \sum_{S \subset N, i \notin S} P_S \ (v(S \cup \{i\}) - v(S)), \quad i = 1, \ldots, n,$$

where

$$\sum_{S \subset N, i \notin S} P_S = \sum_{s=0}^{n-1} \binom{n-1}{s} P_S = 1, \ P_S \geqslant 0, \quad \forall s = 0, \ldots, n-1,$$

P_S representing the relative importance in the solution concept of the coalitions with cardinal s.

One of the most important semivalues is the Shapley [1] value that is defined as follows:

$$Sh_i(v) = \sum_{S \subset N \ i \notin S} \frac{(n-s-1)! \ s!}{n!} (v(S \cup \{i\}) - v(S)), \quad i = 1, \ldots, n.$$

An alternative definition of the Shapley value can be expressed in terms of all possible orders of the players N. Let $O : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation that assigns to each position k the player $O(k)$. Let us denote by $\pi(N)$ the set of all possible permutations with player set N. Obviously, the cardinality of $\pi(N)$ is $n!$. Given a permutation O, let us denote by $Pre^i(O)$ the set of predecessors of the player i in the order O (i.e. $Pre^i(O) = \{O(1), \ldots, O(k-1)\}$, if $i = O(k)$).

Thus, the Shapley value can be expressed in the following way:

$$Sh_i(v) = \sum_{O \in \pi(N)} \frac{1}{n!} (v(Pre^i(O) \cup i) - v(Pre^i(O))), \quad i = 1, \ldots, n.$$

A generalization of the Shapley value that takes the *a priori* importance of each of the players into account is the weighted Shapley [1] value. To define this concept, let $w = (w_1, \ldots, w_n)$ be a weight vector, where $w_i > 0$ represents the weight of the player i for $i = 1, \ldots, n$. Given this weight vector w and a permutation O, a probability function that represents the probability of the different orders in $\pi(N)$ can be defined as follows:

$$P_W(O) = \prod_{k=1}^{n} \left(w_{O(k)} \Bigg/ \sum_{l=1}^{k} w_{O(l)} \right).$$

Finally the weighted Shapley value is defined as

$$Sh_i(v) = \sum_{O \in \pi(N)} P_W(O)(v(Pre^i(O) \cup i) - v(Pre^i(O))), \quad i = 1, \ldots, n.$$

Let us observe that in the classical definition of the Shapley value it is supposed that all different orders have equal probability.

In the sampling process, we will denote a finite population by $P = \{a_1, a_2, \ldots, a_p\}$ where a_i is the sampling unit. The characteristics observed in each sampling unit will be denoted by $x(a_i) = (x(a_i)_1, \ldots, x(a_i)_r)$. $\theta = (\theta_1, \ldots, \theta_r)$ will be the vector of parameters that we are trying to estimate and $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_r)$ its estimate.

In this paper we will consider a sample M as an element of $\underbrace{P \times P \times \cdots \times P}_{m \text{ times}}$, i.e., a sample that has been obtained with replacement.

Given a parameter θ we will often abuse the notation (when there is no ambiguity) by writing $\hat{\theta}$ for the estimate and the estimator.

3. The estimation of the Shapley value

In this section, we will obtain, in polynomial time, an estimation of the Shapley value with some desirable properties. To estimate the Shapley value, we will use a unique sampling process for all players $i \in \{1, \ldots, n\}$. The sampling process (*ApproShapley*) is defined as follows:

(1) The population of the sampling process P will be the set of all possible orders of N players, i.e. $P = \pi(N)$.
(2) The vector parameter under study is $Sh = (Sh_1, \ldots, Sh_n)$.
(3) The characteristics observed in each sampling unit, $O \in \pi(N)$, are the marginal contributions of the players in the order O, i.e.

$$x(O) = (x(O)_1, \ldots, x(O)_n) \quad \text{where } x(O)_i = v(Pre^i(O) \cup \{i\}) - v(Pre^i(O)).$$

(4) The estimate of the parameter Sh, $\hat{Sh}$, will be the mean of the marginal contributions over the sample M, i.e.

$$\hat{Sh} = (\hat{Sh}_1, \ldots, \hat{Sh}_1) \quad \text{where } \hat{Sh}_i = \frac{1}{m} \sum_{O \in M} x(O)_i.$$

(5) Finally, the selection process used to determine the sample M will take any order $O \in \pi(N)$ with probability $\frac{1}{n!}$.

Algorithm ApproShapley
Begin
  Determine m
  $Cont := 0$ and $\hat{Sh}_i := 0 \ \forall i \in N$.
  While $Cont < m$
  Begin
    Take $O \in \pi(N)$ with probability $1/n!$
    For all $i \in N$
    Begin
      Calculate $Pre^i(O)$
      Calculate $x(O)_i := v(Pre^i(O) \cup \{i\}) - v(Pre^i(O))$
      $\hat{Sh}_i := \hat{Sh}_i + x(O)_i$
    End
    $Cont := Cont + 1$
  End
$\hat{Sh}_i := \frac{\hat{Sh}_i}{m} \ \forall i \in N$.
End

In order to assure that $\hat{Sh}$ can be calculated in polynomial time, we have to assume that the worth of any coalition S, $v(S)$, can be calculated in polynomial time.

Having described the sampling method, we will dedicate the remainder of the section to analyzing the main properties of this approach.

Proposition 3.1. *The estimator $\hat{Sh}_i$ is unbiased i.e. $E[\hat{Sh}_i] = Sh_i$ and its variance is given by*

$$Var[\hat{Sh}_i] = \frac{\sigma^2}{m} \quad \text{where } \sigma^2 = \sum_{O \in \pi(N)} (x(O)_i - Sh_i)^2 \frac{1}{n!}.$$

Proof. The proof is straightforward taking into account that, the estimator $\hat{Sh}_i$ is a sample mean and the parameter Sh_i is a population mean. □

Corollary 3.1. *The estimator $\hat{Sh}_i$ is consistent in probability, i.e.*

$$\lim_{n \to \infty} P(|\hat{Sh}_i - Sh_i| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Corollary 3.2. *If a player i is a dummy player for a game (N,v) (i.e. $v(T\cup\{i\})=v(T)+v(\{i\}), \forall T\subset N\backslash\{i\}$), then $\hat{Sh}_i=Sh_i$.*

Proof. This proof is straightforward taking into account that the value of $x(O)_i$ is constant for any O in $\pi(N)$. □

Proposition 3.2. *The estimate $\hat{Sh}$ is efficient in the allocation, i.e.*:

$$\sum_{i=1}^{|N|}\hat{Sh}_i=v(N).$$

Proof. This proof is straightforward taking into account that the sum of the marginal contributions in any order coincides with $v(N)$. □

The rest of this section is devoted to determine the sample size m. To do that, we want to guarantee that the error in the estimation process is lower than e with a probability greater than $1-\alpha$.

Following the central limit theorem it holds that the estimator $\hat{Sh}_i \sim N(Sh_i,\sigma^2/m)$ and thus, if $m \geqslant Z^2_{\alpha/2}\sigma^2/e^2$, then $P(|\hat{Sh}_i - Sh_i|\leqslant e)\geqslant 1-\alpha$, with $Z_{\alpha/2}$ being the value such that $P(Z\geqslant Z_{\alpha/2})=\alpha/2$, and $Z\sim N(0,1)$.

Taking into account that σ^2 is unknown, it is necessary to provide an upper bound of this value in order to determine the sample size. To bound σ^2 let us define the maximum and minimum values that can be reached by the variable $\hat{Sh}_i$, i.e. $x^i_{\max}=\max_{S\subset N\backslash\{i\}} x(S)$ and $x^i_{\min}=\min_{S\subset N\backslash\{i\}} x(S)$. Now, let us observe that for any random variable bounded between two values ($x^i_{\min}$ and $x^i_{\max}$ in this case), the maximum variance is reached when this variable takes the two extreme values with the same probability $\frac{1}{2}$, and thus the following inequality holds:

$$\sigma^2 \leqslant \frac{1}{2}\left(x^i_{\max}-\frac{x^i_{\max}+x^i_{\min}}{2}\right)^2+\frac{1}{2}\left(x^i_{\min}-\frac{x^i_{\max}+x^i_{\min}}{2}\right)^2 = \frac{(x^i_{\max}-x^i_{\min})^2}{4}.$$

Remark 3.1. Let us observe that it is possible to calculate the weighted Shapley value with vector of weights w in a similar way, modifying the probabilities of the different orders as follows:

$$P_W(O)=\prod_{k=1}^{n}\left(w_{O(k)}\Big/\sum_{l=1}^{k}w_{O(l)}\right)\quad \text{for each } O\in\pi(N).$$

4. Computational results

In this section, we will approach the Shapley value for some well-known games found in the literature in which the worth can be calculated in polynomial time for any coalition. To measure the error of this sampling approach, we will use some examples for which we know the Shapley value.

We will take $\alpha=0.01$ to calculate the theoretical error bound (e_{th}). We should remember that this error represents the value that satisfies: $P(|Sh_i-\hat{Sh}_i|\leqslant e_{\text{th}})\geqslant 1-\alpha=0.99$.

4.1. A symmetric voting game

Let (N,v) be a voting game where $N=\{1,\ldots,1000\}$ and $\forall\, S\subset N$,

$$v(S)=\begin{cases}1 & \text{if } |S|>500,\\ 0 & \text{otherwise.}\end{cases}$$

Taking into account that the players of this game are symmetrical, it is very easy to see that the Shapley value is 0.001 for each one of them. Table 1 shows the maximum error, the average error and the theoretical error bound for this estimate. To calculate the theoretical error, we have considered that $\sigma^2_{\hat{Sh}_i}\leqslant(x^i_{\max}-x^i_{\min})^2/4=\frac{1}{4}$.

Table 1
Errors in the symmetric voting game.

m	10^3	10^4	10^5	10^6	10^7	10^8
$e_{\max}$	0.004	0.001	0.00034	0.000106	0.000036	0.000011
e_{ave}	0.000684	0.000252	0.000083	0.000026	0.000008	0.000003
e_{th}	0.0475	0.015	0.00475	0.0015	0.000475	0.00015

4.2. A non-symmetric voting game

Let (N,v) be the non-symmetric game defined in Owen [13] for a voting process in the United States. The player set is $N=\{1,\ldots,51\}$ and the characteristic function of this game is given by

$$v(S)=\begin{cases}1 & \text{if } \sum_{i\in S}w_i>\sum_{j\in N}w_j/2,\\ 0 & \text{otherwise,}\end{cases}$$

where $w=\{w_1,\ldots,w_{51}\}=\{45,41,27,26,26,25,21,17,17,14,13,13,12,12,12,11,\underbrace{10,\ldots,10}_{4\text{ times}},\underbrace{9,\ldots,9}_{4\text{ times}},8,8,\underbrace{7,\ldots,7}_{4\text{ times}},\underbrace{6,\ldots,6}_{4\text{ times}},5,\underbrace{4,\ldots,4}_{9\text{ times}},\underbrace{3,\ldots,3}_{7\text{ times}}\}$.

To estimate the Shapley value, Owen uses the multilinear extension. Let us observe that taking into account the large number of players it is necessary to approximate the calculations in the multilinear extension as it is done in [13, p. 297]. Table 2 show the true Shapley value and the approximations given by the multilinear extension and sampling process ($m=10^8$).

Table 3 shows the maximum error, the average error and the theoretical error bound for this estimate. To calculate the theoretical error, we have taken into account that $\sigma^2_{\hat{Sh}_i}\leqslant(x^i_{\max}-x^i_{\min})^2/4=\frac{1}{4}$.

It is important to note that, for a sampling size greater than or equal to 10^7 the estimate obtained by sampling is better than the one given in Owen [13] since the maximum error and the average error in the multilinear extension are 0.00021 and 0.000057, respectively.

4.3. An airport game

Let (N,c) be an airport game where $N=\{1,\ldots,100\}$. The characteristic function of this game, v, is defined as follows, $\forall\, S\subset N$, $c(S)=\max_{i\in S} c_i$, where $c=\{c_1,\ldots,c_{100}\}=\{\underbrace{1,\ldots,1}_{8\text{ times}},\underbrace{2,\ldots,2}_{12\text{ times}},\underbrace{3,\ldots,3}_{6\text{ times}},\underbrace{4,\ldots,4}_{14\text{ times}},\underbrace{5,\ldots,5}_{8\text{ times}},\underbrace{6,\ldots,6}_{9\text{ times}},\underbrace{7,\ldots,7}_{13\text{ times}},\underbrace{8,\ldots,8}_{10\text{ times}},\underbrace{9,\ldots,9}_{10\text{ times}},\underbrace{10,\ldots,10}_{10\text{ times}}\}$.

In Castro et al. [7] it has been proved that the Shapley value of this game coincides with the Serial Cost Sharing Rule for this problem, so we can calculate the Shapley value. Table 4 show the Shapley value and the estimation given by the sampling process with sample size $m=10^8$.

Table 5 shows the maximum error, the average error and the theoretical error bound for this estimate. To calculate the theoretical error, we have taken into account that $\sigma^2_{\hat{Sh}_i}\leqslant(x^i_{\max}-x^i_{\min})^2/4=\frac{10^2}{4}=25$.

4.4. A shoes game

Let (N,v) be a *shoes game* where $N=\{1,\ldots,100\}$. The characteristic function of this game is defined as follows, $\forall\, S\subset N$, $v(S)=$

Table 2
True Shapley value versus multilinear and sampling approaches in the non-symmetric voting game (I).

Player	True Shapley	Multilinear	Sampling
1	0.08831	0.08852	0.088257
2	0.07973	0.07976	0.079723
3	0.05096	0.05113	0.050952
4	0.04898	0.04915	0.048959
5	0.04898	0.04915	0.048976
6	0.04700	0.04716	0.047010
7	0.03917	0.03928	0.039199
8	0.03147	0.03157	0.031455
9	0.03147	0.03157	0.031420
10	0.02577	0.02586	0.025759
11	0.02388	0.02396	0.023847
12	0.02388	0.02396	0.023859
13	0.02200	0.02207	0.022037
14	0.02200	0.02207	0.021991
15	0.02200	0.02207	0.022018
16	0.02013	0.02019	0.020128
17	0.01827	0.01833	0.018281
18	0.01827	0.01833	0.018278
19	0.01827	0.01833	0.018277
20	0.01827	0.01833	0.018262
21	0.01641	0.01647	0.016452
22	0.01641	0.01647	0.016417
23	0.01641	0.01647	0.016424
24	0.01641	0.01647	0.016401
25	0.01456	0.01461	0.014572
26	0.01456	0.01461	0.014577
27	0.01272	0.01276	0.012716
28	0.01272	0.01276	0.012716
29	0.01272	0.01276	0.012703
30	0.01272	0.01276	0.012736
31	0.01088	0.01092	0.010905
32	0.01088	0.01092	0.010901
33	0.01088	0.01092	0.010871
34	0.01088	0.01092	0.010888
35	0.009053	0.009078	0.0090567
36	0.007230	0.007243	0.0072156
37	0.007230	0.007243	0.0072248
38	0.007230	0.007243	0.0072389
39	0.007230	0.007243	0.0072208
40	0.007230	0.007243	0.0072332
41	0.007230	0.007243	0.0072324
42	0.007230	0.007243	0.0072160
43	0.007230	0.007243	0.0072478
44	0.007230	0.007243	0.0072206
45	0.005412	0.005431	0.0054173
46	0.005412	0.005431	0.0054107
47	0.005412	0.005431	0.0054221
48	0.005412	0.005431	0.0054045
49	0.005412	0.005431	0.0054231
50	0.005412	0.005431	0.0054140
51	0.005412	0.005431	0.0054084

Table 3
Errors in the non-symmetric voting game.

m	10^3	10^4	10^5	10^6	10^7	10^8
$e_{\max}$	0.00883	0.004495	0.000969	0.000283	0.000115	0.000053
e_{ave}	0.003104	0.001073	0.000321	0.000082	0.00003	0.000014
e_{th}	0.0475	0.015	0.00475	0.0015	0.000475	0.00015

$\min\{|S_{\mathrm{left}}|,|S_{\mathrm{right}}|\}$, where $|S_{\mathrm{left}}|$ is the number of left shoes and $|S_{\mathrm{right}}|$ the number of right shoes. In this example, $|S_{\mathrm{left}}|=|S_{\mathrm{right}}|=50$. For the sampling process, we have supposed, without loss of generality, that the first 50 players correspond to the left shoes and the other 50 to the right shoes. It is very easy to see that the Shapley value of this game is $Sh_i=\frac{1}{2},\ \forall i\in N$.

Table 6 provides the maximum error, the average error and the theoretical error bound for this estimate. To calculate the theoretical error, we have taken into account that $\sigma^2_{Sh_i}\leqslant(x^i_{\max}-x^i_{\min})^2/4=\frac{1}{4}$.

Table 4
Shapley value versus sampling in the airport game (I).

Player	Shapley	Sampling	Player	Shapley	Sampling
1	0.01	0.01000240	26	0.033369565	0.03336215
2	0.01	0.01002122	27	0.046883079	0.04681264
3	0.01	0.00999404	28	0.046883079	0.04693904
4	0.01	0.01001061	29	0.046883079	0.04686664
5	0.01	0.00998937	30	0.046883079	0.04683868
6	0.01	0.01000476	31	0.046883079	0.04688131
7	0.01	0.01000974	32	0.046883079	0.04694949
8	0.01	0.00998920	33	0.046883079	0.04686780
9	0.020869565	0.02086825	34	0.046883079	0.04687685
10	0.020869565	0.02086269	35	0.046883079	0.04681076
11	0.020869565	0.02085620	36	0.046883079	0.04693215
12	0.020869565	0.02087864	37	0.046883079	0.04683531
13	0.020869565	0.02089870	38	0.046883079	0.04687293
14	0.020869565	0.02086929	39	0.046883079	0.04688354
15	0.020869565	0.02089350	40	0.046883079	0.04686851
16	0.020869565	0.02087001	41	0.063549745	0.06363301
17	0.020869565	0.02086078	42	0.063549745	0.06354695
18	0.020869565	0.02085292	43	0.063549745	0.06358871
19	0.020869565	0.02085290	44	0.063549745	0.06351779
20	0.020869565	0.02087947	45	0.063549745	0.06346493
21	0.033369565	0.03335218	46	0.063549745	0.06350608
22	0.033369565	0.03335009	47	0.063549745	0.06354124
23	0.033369565	0.03339154	48	0.063549745	0.06356064
24	0.033369565	0.03329275	49	0.082780515	0.08280608
25	0.033369565	0.03338934	50	0.082780515	0.08276462
51	0.082780515	0.08284313	76	0.139369662	0.13940729
52	0.082780515	0.08268617	77	0.139369662	0.13934942
53	0.082780515	0.08276607	78	0.139369662	0.13951629
54	0.082780515	0.08285130	79	0.139369662	0.13936278
55	0.082780515	0.08290920	80	0.139369662	0.13935631
56	0.082780515	0.08279981	81	0.189369662	0.18945158
57	0.082780515	0.08277814	82	0.189369662	0.18946746
58	0.106036329	0.10594986	83	0.189369662	0.18939991
59	0.106036329	0.10600076	84	0.189369662	0.18936684
60	0.106036329	0.10592028	85	0.189369662	0.18945528
61	0.106036329	0.10605632	86	0.189369662	0.18939632
62	0.106036329	0.10598094	87	0.189369662	0.18929814
63	0.106036329	0.10627204	88	0.189369662	0.18936075
64	0.106036329	0.10611809	89	0.189369662	0.18925926
65	0.106036329	0.10614854	90	0.189369662	0.18938667
66	0.106036329	0.10605967	91	0.289369662	0.28941387
67	0.106036329	0.10604222	92	0.289369662	0.28922169
68	0.106036329	0.10618083	93	0.289369662	0.28929850
69	0.106036329	0.10598174	94	0.289369662	0.28949639
70	0.106036329	0.10612261	95	0.289369662	0.28946205
71	0.139369662	0.13937247	96	0.289369662	0.28920360
72	0.139369662	0.13923064	97	0.289369662	0.28943995
73	0.139369662	0.13934868	98	0.289369662	0.28930091
74	0.139369662	0.13927767	99	0.289369662	0.28926759
75	0.139369662	0.13934098	100	0.289369662	0.28925455

Table 5
Errors for the airport game.

m	10^3	10^4	10^5	10^6	10^7	10^8
$e_{\max}$	0.06663	0.03726	0.00621	0.00191	0.00064	0.00024
e_{ave}	0.01441	0.00535	0.00121	0.00048	0.00015	0.00005
e_{th}	0.31	0.098	0.031	0.0098	0.0031	0.00098

4.5. A minimum spanning tree game

Let $G=(N\cup\{0\},E)$ be a valued graph where $N=\{1,\dots,100\}$, and the cost associated to an edge (i,j) is

$$C_{ij}:=\begin{cases}1 & \text{if } i=j+1,\ i=j-1,\ i=1\wedge j=100,\\ & \quad i=100\wedge j=1,\\ 101 & \text{if } i=0 \text{ or } j=0,\\ \infty & \text{otherwise.}\end{cases}$$

A minimum spanning tree game (N,c) is a cost game, where for a given $S\subset N$, $c(S)$ is the sum of the edge cost of the minimum spanning tree, i.e. $c(S)=$ Minimum Spanning Tree of the graph $G|_{S\cup\{0\}}$, which is the partial graph restricted to the players S and the source node 0. It is easy to see that the Shapley value for this game is $Sh_i=2,\ \forall i\in N$.

Table 6
Errors for the shoes game.

m	10^3	10^4	10^5	10^6	10^7	10^8
e_{max}	0.046	0.01285	0.00285	0.001265	0.000355	0.0001255
e_{ave}	0.01095	0.00385	0.00095	0.00031	0.000115	0.000035
e_{th}	0.031	0.0098	0.0031	0.00098	0.00031	0.000098

Table 7
Errors for the minimal spanning tree game.

m	10^3	10^4	10^5	10^6	10^7	10^8
e_{max}	7	1.7	0.625	0.2024	0.079	0.0214
e_{ave}	2.454	0.5982	0.1864	0.0624	0.024	0.0066
e_{th}	6.23	1.97	0.623	0.197	0.0623	0.0197

Table 7 shows the maximum error, the average error and the theoretical error bound for this estimate. To calculate the theoretical error, we have taken into account that $\sigma^2_{Sh_i} \leqslant (x^i_{\max} - x^i_{\min})^2/4 = (101 - (-100))^2/4 = 10100.25$.

5. Conclusions and final remarks

As it has been previously noted, one of the most important solution concepts in cooperative games is the Shapley value. Calculation of the Shapley value is an open topic in game theory, where despite the fact that for some particular cases the Shapley value has been well solved, in general, the complexity of the calculation is exponential. Throughout this paper we have emphasized the necessity of presenting an approximation to this value whose calculation may be possible in polynomial time. The sampling process presented here is an important tool in order to calculate the Shapley value polynomially. The method proposed in this paper gives an estimation of the Shapley value with desirable properties such as being unbiased, being consistent, has no error for dummy players and it is possible to calculate the theoretical error in a probabilistic way.

We would like to stress again that the only constraint of the algorithm is the fact that the worth of any coalition must be computed polynomially. Let us observe that this condition is less restrictive than the approximations that appear in the literature on the Shapley value.

In this paper we have focused on the Shapley value solution concept. It is important to observe that the methodology that has been used to estimate the Shapley value can be extended to any semivalue. To estimate a semivalue $\gamma = (\gamma_1, \dots, \gamma_n)$ of a coalitional game (N, v) we will use n independent sampling processes, one for each player $i \in N$.

For a player $i \in N$, the sampling procedure (*ApproSemivalue*) to calculate a semivalue γ_i can be defined as:

(1) The population of this sampling process P will be the set of subsets of $N \backslash \{i\}$.
(2) The parameter under study is $\theta_i = \gamma_i$.
(3) The characteristic observed in each sampling unit $S \subset N \backslash \{i\}$ is the marginal contribution of the player i in the coalitional game (N, v), i.e. $x_i(S) = v(S \cup \{i\}) - v(S)$.
(4) The estimate $\hat{\theta}_i$ will be the mean of the marginal contributions over the sample M, i.e.

$$\hat{\theta}_i = \hat{\gamma}_i = \frac{1}{m} \sum_{S \in M} x_i(S) \quad \text{where } m = |M| \text{ is the sample size.}$$

(5) Finally, we will determine the sample M in m steps. In each step, an element S of P will be included in M with probability P_s, where P_s is the coefficient of the semivalue. In order to do this, first, we determine the cardinality of the sampling unit s with probability $\binom{n-1}{s} P_s$. After this, we select with equal probability one of the $\binom{n-1}{s}$ possible sampling units with cardinality s.

Algorithm ApproSemivalue
Begin
  For all $i \in N$
  Begin
    Determine m
    Cont $:= 0$ and $\hat{\gamma}_i := 0$
    While *Cont* $< m$
    Begin
      Take $S \subset N \backslash \{i\}$ with probability P_s
      Calculate $x_i(S) := v(S \cup \{i\}) - v(S)$
      $\hat{\gamma}_i := \hat{\gamma}_i + x_i(S)$
      Cont $:=$ *Cont* $+ 1$
    End
    $\hat{\gamma}_i := \frac{\hat{\gamma}_i}{m}$
  End
End

Remark 5.1. The estimator $\hat{\gamma}_i$ obtained in this algorithm also satisfies, in the same sense, the following previously established properties: the estimator can be calculated in polynomial time (if $v(S)$ can be calculated in polynomial time), is unbiased, is consistent, has no error for dummy players and the error has the same bound.

The two processes defined in this paper are not equivalent for the case in which $\gamma = Sh$. If our aim is to calculate the Shapley value, this new algorithm is worse. On one hand *ApproShapley* guarantees that the calculation is efficient in the allocation (*ApproSemivalue* does not). On the other hand, the first algorithm does not consider n independent sampling processes.

Acknowledgments

This research has been partially supported by the *Plan Nacional de I+D+i* of the Spanish Government, under the project MTM2005-09184-C02-01.

References

[1] Shapley LS. A value for n-person games. In: Kuhn HW, Tucker AW, editors. Contributions to the theory of games II, Annals of mathematics studies, vol. 28. Princeton, NJ: Princeton University Press; 1957. p. 307–17.
[2] Deng X, Papadimitriou CH. On the complexity of cooperative solution concepts. Mathematics of Operations Research 1994;19(2):257–66.
[3] Fernández JR, Algaba E, Bilbao JM, Jiménez A, Jiménez N, López JJ. Generating functions for computing the Myerson value. Annals of Operations Research 2002;109:143–58.
[4] Faigle U, Kern W. The Shapley value for cooperative games under precedence constraints. International Journal of Game Theory 1992;21(3):249–66.
[5] Bilbao JM, Fernández JR, Jiménez Losada A, López JJ. Generating functions for computing power indices efficiently. TOP 2000;8(2):191–213.
[6] Granot D, Kuipers J, Chopra S. Cost allocation for a tree network with heterogeneous customers. Mathematics of Operations Research 2002;27(4):647–61.
[7] Castro J, Gómez D, Tejada J. A polynomial rule for the problem of sharing delay costs in PERT networks. Computers and Operation Research 2008;35(7):2376–87.
[8] Owen G. Multilinear extensions of games. Management Science Series B—Application 1972;18(5):64–79.
[9] Fatima SS, Wooldridge M, Jenniggs NR. An analysis of the Shapley value and its uncertainty for the voting game. Lectures notes in artificial intelligence, vol. 3937, 2006. p. 85–98.
[10] Cochran WG. Sampling techniques. New York: Wiley; 1977.
[11] Lohr H. Sampling: design and analysis. Duxbury; 1999.
[12] Dubey P, Neyman A, Weber RJ. Value theory without efficiency. Mathematics of Operations Research 1981;6(1):122–8.
[13] Owen G. Game theory. New York: Academic Press; 1995.