R tools to link Game Theory and Statistics by sampling

Alejandro Saavedra-Nieves

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

VII XORNADA DE USUARIOS DE R EN GALICIA

Motivation: the International Monetary Found

The IMF (January, 2002)

- Analysis of the power of the 179 members when $q = 50\%$.
- A group passes a law when its aggregate weight is larger than q.

Country	Voting rights	Banzhaf-Owen value
United States	17.11	0.6471
Japan	6.14	0.1752
Germany	6.00	0.1712
France	4.95	0.1411
United Kingdom	4.95	0.1411
Austria	0.87	0.0248
Belarus	0.19	0.0054
Belgium	2.13	0.0608
Czech Republic	0.39	0.0111
Hungary	0.49	0.0140
Kazakhstan	0.18	0.0051
Luxembourg	0.14	0.0040
Slovak Republic	0.18	0.0051
Slovenia	0.12	0.0034
Turkey	0.46	0.0131
Armenia	0.05	0.0014
Bosnia and Herzegovina	0.09	0.0026
Bulgaria	0.31	0.0088
...	...	...

Alonso-Meijide, J. M., Bowles, C. (2005). Generating functions for coalitional power indices: An application to the IMF. Annals of Operations Research, 137(1), 21-44.

TU-games

Game theory: mathematical theory of interactive decision problems

A TU-game is a pair (N, v):

- N is the set of players and,
- $v : 2^N \longrightarrow \mathbb{R}$ is the characteristic function with $v(\emptyset) = 0$.

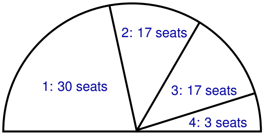
TU-games

Game theory: mathematical theory of interactive decision problems

A TU-game is a pair (N, v):

- N is the set of players and,
- $v : 2^N \longrightarrow \mathbb{R}$ is the characteristic function with $v(\emptyset) = 0$.

A parliament



Majority: 34

- $v(S) = 1$ (winning), if $S \supset T$, $T \in \{\{1,2\},\{1,3\},\{2,3\}\}$.
- Otherwise, $v(S) = 0$ (not winning).

S is a swing if

$$v(S \cup i) - v(S) = 1, \text{ for all } S \subseteq N \setminus \{i\}.$$

That is, i converts S into a winning coalition.

TU-games

Game theory: mathematical theory of interactive decision problems

A TU-game is a pair (N, v):

- N is the set of players and,
- $v : 2^N \longrightarrow \mathbb{R}$ is the characteristic function with $v(\emptyset) = 0$.

A main goal: definition and analysis of rules to allocate $v(N)$

TU-games

Game theory: mathematical theory of interactive decision problems

A TU-game is a pair (N, v):

- N is the set of players and,
- $v : 2^N \longrightarrow \mathbb{R}$ is the characteristic function with $v(\emptyset) = 0$.

A main goal: definition and analysis of rules to allocate $v(N)$

Shapley value	Owen value
Banzhaf value	Banzhaf-Owen value

Marginal contribution of i to S: $v(S \cup i) - v(S)$, for all $S \subseteq N \setminus \{i\}$.

The Banzhaf value and the Banzhaf-Owen value

The Banzhaf value for (N, v) is $Bz_i(N, v) = \frac{1}{2^{n-1}} \sum_{S \subseteq N \setminus \{i\}} (v(S \cup \{i\}) - v(S))$, for all $i \in N$.

Let $P = \{P_1, \ldots, P_m\}$ be a partition of N.

The Banzhaf-Owen value (Owen, 1982) for (N, v, P), for every $i \in N$, is

$$BzO_i(N, v, P) = \sum_{R \subseteq P \setminus P_{(i)}} \frac{1}{2^{m-1}} \sum_{S \subseteq P_{(i)} \setminus \{i\}} \frac{1}{2^{p_i - 1}} \left(v\Big(\bigcup_{P_l \in R} P_l \cup S \cup \{i\}\Big) - v\Big(\bigcup_{P_l \in R} P_l \cup S\Big) \right),$$

with $i \in P_{(i)} \in P$ and $p_i = |P_{(i)}|$.

- For $i \in N$, $T \subseteq N \setminus \{i\}$ is *compatible with* P:

$$T = \bigcup_{P_l \in R} P_l \cup S \text{ for } R \subseteq P \setminus P_{(i)} \text{ and } S \subseteq P_{(i)} \setminus \{i\}.$$

Banzhaf, J.F. (1964). Weighted voting doesn't work: a mathematical analysis. Rutgers Law Review 19, 317-343.

Owen, G. (1982). Modification of the Banzhaf-Coleman index for games with a priori unions. In: Power, Voting and Voting Power (M.J. Holler, ed.), Physica-Verlag, 232-238.

Motivation: the International Monetary Found

The IMF (January, 2002)

Country	Voting rights	Banzhaf-Owen value
United States	17.11	0.6471
Japan	6.14	0.1752
Germany	6.00	0.1712
France	4.95	0.1411
United Kingdom	4.95	0.1411
Austria	0.87	0.0248
Belarus	0.19	0.0054
Belgium	2.13	0.0608
Czech Republic	0.39	0.0111
Hungary	0.49	0.0140
Kazakhstan	0.18	0.0051
Luxembourg	0.14	0.0040
Slovak Republic	0.18	0.0051
Slovenia	0.12	0.0034
Turkey	0.46	0.0131
Armenia	0.05	0.0014
Bosnia and Herzegovina	0.09	0.0026
Bulgaria	0.31	0.0088
...	...	...

- Analysis of the power of the 179 members when $q = 50\%$.
- A group passes a law when its aggregate weight is larger than q.

Alonso-Meijide, J. M., Bowles, C. (2005). Generating functions for coalitional power indices: An application to the IMF. Annals of Operations Research, 137(1), 21-44.

Motivation: a weighted majority game

- $N = \{1, \ldots, n\}$: the countries.
- $h_1, \ldots, h_n$: the weights of N.
- $q > 0$ imposes the majority.

A weighted majority game is given, for each $S \subseteq N$, by

$$v(S) = 1, \text{ when } \sum_{i \in S} h_i \geq q. \text{ Otherwise, } v(S) = 0.$$

- Members are grouped by constituences. Each constituence is an union.

For Austria ($m = 24$ and $p_i = 10$)

$2^{24-1}2^{10-1}$ available compatible coalitions

Proposal: sampling techniques?

Estimating the Banzhaf-Owen value

Problem: estimating a mean on a finite population

Let (N, v, P) be a game with $P = \{P_1, \dots, P_m\}$.

- $BzO_i(N, v, P) = Bz_i(N^*, v)$, with $N^* = \left\{k \ : P_k \in P \setminus P_{(i)}\right\} \cup \{j \ : j \in P_{(i)}\}$.

A procedure based on sampling techniques

1. We generate at random a sample $\mathcal{T} = \{T_1, \dots, T_\ell\}$ of ℓ coalitions in $N^* \setminus \{i\}$.
2. For each $T_j \in \mathcal{T}$,
$$x(R_j, S_j)_i = v(\underset{P_l \in R_j}{\cup} P_l \cup S_j \cup \{i\}) - v(\underset{P_l \in R_j}{\cup} P_l \cup S_j),$$
being $R_j \subseteq P \setminus P_{(i)}$ and $S_j \subseteq P_{(i)} \setminus \{i\}$ such that $T_j = \{k \ : \ P_k \in R_j\} \cup S_j$.
3. The estimation of BzO_i is $\overline{BzO}_i = \frac{1}{\ell}\sum_{j=1}^{\ell} x(R_j, S_j)_i$, for all $i \in N$.

Which is the best option of sampling?

A pseudocode for obtaining coalitions

Take N and an integer k such that $1 \leq k \leq 2^n$.

A pseudocode to determine coalitions

1. Initialize $x = (x_1, x_2, \ldots, x_n)$.
2. Take $a = k - 1$ and $cont = 1$.
3. Do
$$x_{cont} = a - 2\lfloor \frac{a}{2} \rfloor$$
and update $a = \lfloor \frac{a}{2} \rfloor$ and $cont = cont + 1$, being $\lfloor x \rfloor$ the floor function. Repeat Step 3 until $cont > n$.
4. Finally, the coalition S associated with k is given by
$$S = \{i \in N : \; x_i = 1\}.$$

New function: `determine_coalition(k)`

sampling coalitions = sampling numbers

Sampling with replacement | Sampling without replacement | Systematic sampling

An estimator for the mean under simple random sampling with replacement: $\overline{BzO}_i$

- It is the most usual way of estimating a mean population.
- Theoretical results are well-known for bounding the estimation error.
- We are applying this methodology in a finite population scenario.

Bachrach, Y., Markakis, E., Resnick, E., Procaccia, A. D., Rosenschein, J. S., and Saberi, A. (2010). Approximating power indices: theoretical and empirical analysis. *Autonomous Agents and Multi-Agent Systems*, 20(2), 105-122.

Castro, J., Gómez, D., and Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. Computers & Operations Research, 36(5), 1726-1730.

Sampling with replacement

Sampling without replacement

Systematic sampling

Simple random sampling with replacement

Proposal 1: `sample(1:ntotal,replace=TRUE)`

Proposal 2: `library(sampling)`, with `srswr(ell,ntotal)`.

An estimator for the mean under simple random sampling without replacement: $\overline{BzO_i}$

- In the exact calculation, each coalition is evaluated only once.
- The hypothesis of non-replacement ensures a lower variance for the estimator (in comparison to the case with replacement).
- It requires the storing of information about the coalitions previously sampled.

Saavedra-Nieves, A., and Fiestras-Janeiro, M. G. (2020). Sampling methods to estimate the Banzhaf–Owen value. *Annals of Operations Research (to appear).*

Sampling with replacement → **Sampling without replacement** → Systematic sampling

Simple random sampling without replacement

- Proposal 1: `sample(1:ntotal,replace=FALSE)`
- Proposal 2: `library(sampling)`, with `srswor(ell,ntotal)`.

An estimator for the mean under subsampling: $\overline{\overline{BzO}}_i$

1. We take without replacement $\mathcal{R} = \{R_1, \ldots, R_{\ell_r}\}$ of ℓ_r coalitions $R_j \subseteq P \setminus P_{(i)}$.
2. For $R_j \in \mathcal{R}$, we take without replacement $\mathcal{S}_{R_j} = \{S_{j1}, \ldots, S_{j\ell_s}\}$ of ℓ_s coalitions $S_{jk} \subseteq P_{(i)} \setminus \{i\}$.
3. For (R_j, S_{jk}), we obtain $x(R_j, S_{jk})_i$.
4. The estimation of BzO_i is $\overline{\overline{BzO}}_i = \frac{1}{\ell_r}\sum_{j=1}^{\ell_r}\left(\frac{1}{\ell_s}\sum_{k=1}^{\ell_s} x(R_j, S_{jk})_i\right),$ for all $i \in N$.

Subsampling techniques

- Stage 1: `sample(1:ntotal_R,replace=FALSE)`
- Stage 2: `sample(1:ntotal_S,replace=FALSE)`, for each sampling unit obtained at Stage 1.
- Properties and results on determining sampling sizes are provided in Cochran (2007).

An estimator for the mean under systematic sampling: $\overline{BzO}_i^{sys}$

- The non-replacement hypothesis increases the computational complexity in the approximation.
- We have to store information about those compatible coalitions previously sampled.
- The difficulties in obtaining the variance complicate the task of establishing bounds of the error.

Systematic sampling

Saavedra-Nieves, A. (2020). A systematic sampling procedure for estimating the Banzhaf-Owen value. *Operations Research Letters*, 48 (6), 725-731.

Empirical analysis

- We approximate the Banzhaf-Owen value in examples where it is exactly obtained by specific formulas.
- We compare the Banzhaf-Owen value and the estimations under these approaches.

Empirical analysis

- We approximate the Banzhaf-Owen value in examples where it is exactly obtained by specific formulas.

We compare the Banzhaf-Owen value and the estimations under these approaches.

Simulation results

- Similar results in the approximation of the Banzhaf-Owen value with all methodologies.

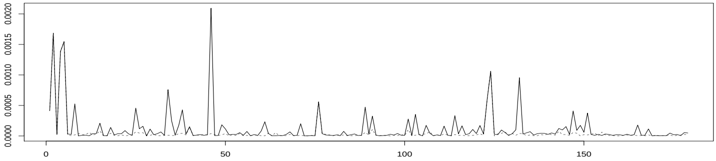


Figure: Absolute error for the 179 players in the IMF under subsampling (black) and under *SRSwor* (dashed line).

Empirical analysis

- Theoretical bounds of the absolute error result conservative in practice.
- Subsampling techniques require less computation time to obtain these estimations.

1000 estimations of the Banzhaf-Owen value: Austria in the IMF

- $\ell = 10^7$

Absolute error		Minimum	Average	Maximum
$q = 50\%$	*SRSwr*	$3.70 \cdot 10^{-9}$	$3.90 \cdot 10^{-5}$	$1.47 \cdot 10^{-4}$
	SRSwor	$3.70 \cdot 10^{-9}$	$3.91 \cdot 10^{-5}$	$1.45 \cdot 10^{-4}$
	Subs.	$3.70 \cdot 10^{-9}$	$1.90 \cdot 10^{-4}$	$7.78 \cdot 10^{-4}$
	Sys	$1.04 \cdot 10^{-7}$	$4.27 \cdot 10^{-5}$	$1.05 \cdot 10^{-4}$
$q = 70\%$	*SRSwr*	$2.80 \cdot 10^{-8}$	$2.30 \cdot 10^{-5}$	$1.23 \cdot 10^{-4}$
	SRSwor	$2.80 \cdot 10^{-8}$	$2.30 \cdot 10^{-5}$	$1.23 \cdot 10^{-4}$
	Subs.	$1.72 \cdot 10^{-7}$	$1.18 \cdot 10^{-4}$	$4.64 \cdot 10^{-4}$
	Sys	$1.28 \cdot 10^{-7}$	$1.79 \cdot 10^{-5}$	$5.50 \cdot 10^{-5}$
$q = 85\%$	*SRSwr*	$7.00 \cdot 10^{-9}$	$3.09 \cdot 10^{-6}$	$1.76 \cdot 10^{-5}$
	SRSwor	$7.00 \cdot 10^{-9}$	$3.08 \cdot 10^{-6}$	$1.45 \cdot 10^{-5}$
	Subs.	$9.30 \cdot 10^{-8}$	$1.59 \cdot 10^{-5}$	$7.96 \cdot 10^{-5}$
	Sys	$7.00 \cdot 10^{-9}$	$2.02 \cdot 10^{-6}$	$6.91 \cdot 10^{-6}$

Time (sec)		Minimum	Average	Maximum
$q = 50\%$	*SRSwr*	218.90	227.30	245.10
	SRSwor	206.47	229.19	249.27
	Subs.	15.67	16.12	17.61
	Sys	197.60	213.00	238.70
$q = 70\%$	*SRSwr*	219.50	228.20	244.50
	SRSwor	210.50	225.60	241.90
	Subs.	7.21	8.77	10.11
	Sys	203.80	229.60	256.50
$q = 85\%$	*SRSwr*	218.90	226.90	244.40
	SRSwor	223.70	234.20	281.80
	Subs.	2.34	2.43	2.69
	Sys	197.70	213.10	237.70

The Galician milk conflict

The dairy sector in Galicia*

- The dairy sector generates 1.5% of GDP
- Dairy farms manage 35% of the farmland.
- Galicia is the leading dairy power in Spain ($>$ 50% of farms and 40% of production)

A conflict in Galicia

- It is characterized by high production and low competitiveness.
- The average price is the lowest in Spain.
- European Union regulates it by a system of quotas.

*Data source: Consellería de Medio Rural da Xunta de Galicia.

Saavedra-Nieves, A., and Saavedra-Nieves, P. (2020). On systems of quotas from bankruptcy perspective: the sampling estimation of the random arrival rule. *European Journal of Operational Research*, 285(2), 655 - 669.

The Galician milk conflict

- After the European milk quotas in March 2015...

Year	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
2013	32.65	32.66	32.76	32.84	33.06	33.03	34.44	34.86	35.59	38.57	38.93	39.17
2014	39.24	38.90	38.65	36.05	35.61	35.37	33.72	33.73	33.64	32.24	32.10	31.95
2015	30.52	30.60	30.30	28.80	28.40	27.90	27.60	27.70	28.30	28.70	28.70	28.80

Table: Averaged prices of the milk in Galicia (in euros per 100 litres) in the period 2013-2015.

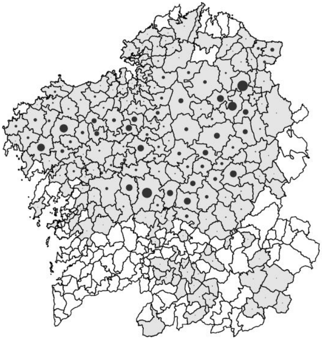
How to increase the price of milk?

A low production scenario

- Reducing the milk production in Galicia?
- If we know a maximum production per municipality, how this reduction affect each of the 190 involved?
- We build a new system of quotas for councils.

*Data source: Consellería de Medio Rural da Xunta de Galicia.

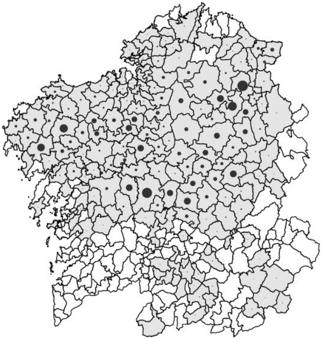
A low milk production scenario



Some assumptions

- The total milk production for Galicia decreases with respect to March 2015.
- We assume that the individual capabilities for councils are not reduced.
- Each council intends to maintain the maximum possible quota.

A low milk production scenario



A new bankruptcy problem

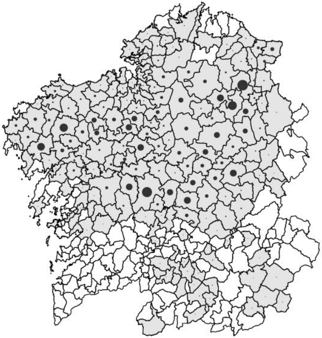
- Set of agents: the 190 most representative councils.
- Estate: the tons of milk in 2014-2015 for Galicia reduces by $\rho\%$, $\rho \in (0, 100]$
- Claims: the capabilities of milk production (individual production of the councils, March 2015).

*Data source: Consellería de Medio Rural da Xunta de Galicia.

Some assumptions

- The total milk production for Galicia decreases with respect to March 2015.
- We assume that the individual capabilities for councils are not reduced.
- Each council intends to maintain the maximum possible quota.

A low milk production scenario



A new bankruptcy problem

- Set of agents: the 190 most representative councils.
- Estate: the tons of milk in 2014-2015 for Galicia reduces by $\rho\%$, $\rho \in (0, 100]$
- Claims: the capabilities of milk production (individual production of the councils, March 2015).

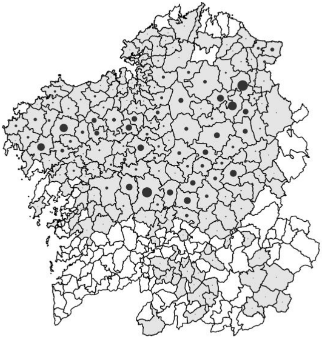
*Data source: Consellería de Medio Rural da Xunta de Galicia.

Computing the Shapley value is a difficult task!

2^{190} **coalitions to be evaluated**

Proposal: a sampling procedure to estimate the Shapley value

A low milk production scenario: a case study



What about variability?

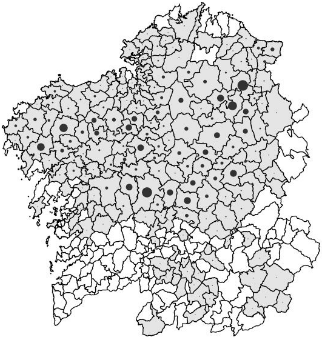
We approximate the Shapley value for the most representative councils in Galicia.

Top 10 of councils with the largest milk production for Galicia

Council	A Pastoriza	Lalín	Castro de Rei	Santa Comba	Mazaricos	Chantada	Cospeito	Sarria	Silleda	Arzúa
Maximum	52499.22	49633.35	43188.07	40002.56	36723.09	32731.26	32668.26	32068.42	31319.75	30539.21
Average	52472.49	49593.97	43161.96	39973.33	36696.91	32708.51	32645.75	32050.23	31300.44	30519.81
Minimum	52425.39	49561.72	43135.33	39942.24	36668.44	32681.45	32626.93	32030.53	31281.85	30502.95

Table: Summary of 100 estimations of the milk quotas for the councils with $\rho = 40\%$.

A low milk production scenario: a case study



What about variability?

- We approximate the Shapley value for the most representative councils in Galicia.
- We obtain 100 estimations and we do some basic statistics.
  - Differences in the results seem bearable.

Top 10 of councils with the largest milk production for Galicia

Council	A Pastoriza	Lalín	Castro de Rei	Santa Comba	Mazaricos	Chantada	Cospeito	Sarria	Silleda	Arzúa
Maximum	52499.22	49633.35	43188.07	40002.56	36723.09	32731.26	32668.26	32068.42	31319.75	30539.21
Average	52472.49	49593.97	43161.96	39973.33	36696.91	32708.51	32645.75	32050.23	31300.44	30519.81
Minimum	52425.39	49561.72	43135.33	39942.24	36668.44	32681.45	32626.93	32030.53	31281.85	30502.95

Table: Summary of 100 estimations of the milk quotas for the councils with $\rho = 40\%$.

Conclusions and further research

- Sampling methods to estimate coalitional values with a priori unions.

Saavedra-Nieves, A., García-Jurado, I., and Fiestras-Janeiro, M. G. (2018). Estimation of the Owen value based on sampling. In E. Gil, E. Gil, J. Gil, and M. Á. Gil (Eds.), The Mathematics of the Uncertain: A Tribute to Pedro Gil (pp. 347–356). Springer.

Saavedra-Nieves, A. and Fiestras-Janeiro, M. G. (2020). Sampling methods to estimate the Banzhaf-Owen value *Annals of Operations Research (to appear)*.

Saavedra-Nieves, A. (2020). A systematic sampling procedure for estimating the Banzhaf-Owen value. *Operations Research Letters*, 48 (6), 725-731.

Saavedra-Nieves, A., and Saavedra-Nieves, P. (2020). On systems of quotas from bankruptcy perspective: the sampling estimation of the random arrival rule. *European Journal of Operational Research*, 285(2), 655 - 669.

- Theoretical analysis from a statistical point of view.

Empirical analysis on examples where coalitional values are known.

Future work

Extension to other coalitional values or other allocation procedures.

Creation of a package for estimating coalitional values.