Density estimation using game theory[1]

Ignacio García-Jurado
Department of Statistics and OR, Faculty of Mathematics, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain.

Luciano Méndez-Naya
Department of Econometrics, Faculty of Economics, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain.

César Sánchez-Sellero
Department of Statistics and OR, Faculty of Mathematics, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain.

Abstract: In this note we show that the mathematical tools of cooperative game theory allow a successful approach to the statistical problem of estimating a density function. Specifically, any random sample of an absolutely continuous random variable determines a transferable utility game, the Shapley value of which proves to be an estimator of the density function of binned kernel and WARPing types, with good computational and statistical properties.

Key Words: Cooperative Games, Density Estimation, Shapley Value.

Mailing Author: Ignacio García-Jurado (igjurado@usc.es).

Running Title: Density Estimation.

---

[1]Authors acknowledge the financial support of Spanish Ministry for Science and Technology and FEDER through projects BFM2002-03213 and BEC2002-04102-C02-02 and of Xunta de Galicia through projects PGIDT00PXI20104PR and PGIDT03PXIC20701PN. They also thank the comments of two anonymous referees.

1 Introduction

The relationship between game theory and statistics is generally seen as very one-sided: whereas the influence of probability theory and bayesian statistics on game theory is evident, the only well-known contribution of game theory to statistical thought is the minimax principle. Noncooperative games had a certain influence on statistical theory (Blackwell and Girshick (1953); Schwarz (1994)), but cooperative game theory has been applied to statistics very rarely (Land and Gefeller's (1997) and (2000) treatment of an epidemiological problem as a cost allocation game is one of those applications). This is to some extent surprising if one bears in mind that fairness, one of the central themes of cooperative game theory, is clearly desirable in statistics in the sense that a good statistical estimator should in some sense provide an estimate that is *fair*, given the available data.

The aim of this note is to provide a new illustration of the connection between cooperative game theory and statistics. We show that the mathematical tools of cooperative game theory allow a successful approach to the statistical problem of estimating a density function. Specifically, we show that any random sample of an absolutely continuous random variable can be used to construct a TU (transferable utility) game, the Shapley value of which (Shapley (1953)) is an estimator of the density function of binned kernel type (Hall and Wand (1996)) and WARPing type (Härdle (1991)). Density estimation is an important problem in statistics that has generated a large literature in the last decades. Its purpose is to provide an accurate estimation of the density function of a random variable on the basis of a random sample. Nowadays, density estimation is a basic tool for exploratory data analysis and for other important fields within statistics. For a complete introduction to density estimation, Silverman (1986) can be consulted.

In section 2 below we describe the density estimation problem, model it as a cooperative game whose Shapley value is the required density estimator, and obtain a more convenient expression for this estimator. In section 3 we provide an axiomatic characterization of the Shapley density estimator. Finally, in section 4 we make some comments on the statistical and computational properties of the Shapley density estimator.

2 Density estimation as a game-theoretical problem

In a density estimation problem we have a random sample $X_1, ..., X_m$ of m independent and identically distributed observations of an absolutely continuous random variable with density function f. f is unknown and the problem is to estimate f using the random sample. A variety of nonparametric density estimators have been proposed and studied since the publication of the pioneer works by Parzen (1962) and Rosenblatt (1956) on the so-called kernel methods. For a survey on density estimation see Silverman (1986). A kernel estimator is any real function $\hat{f}$ of the form

$$\hat{f}(x) = \frac{1}{mh} \sum_{j=1}^{m} K\left(\frac{x - X_j}{h}\right)$$

where K is a density function symmetric around zero and h is the so-called bandwidth or smoothing parameter. Note that, defined in this way, $\hat{f}$ is a density function (a non-negative real function which integrates to one). The necessity of the smoothing parameter comes from the fact that every observation X_i in the sample not only shows that there is a positive density on the real number X_i but also that there is a positive density on a neighborhood of the real number X_i (note that the original variable is continuous and we have a *finite* sample to estimate its density).

The selection of h is an important, though difficult, issue. There is a large number of papers on the selection of the bandwidth parameter for given classes of estimators. For a survey on this topic see Cao et al (1994). Very often, the statistical literature has treated this problem as a different one. One issue is to identify families of density estimators with good statistical and computational properties. Another different issue, which is usually analyzed in a second stage, is to find a good selector of the smoothing parameter for a given family of density estimators. In this paper we focus on the first of these issues and obtain a density estimator based on the Shapley value. Our main target is to illustrate a connection between cooperative game theory and statistics by showing that this new estimator is a competitive one from a statistical point of view. Moreover, we provide an axiomatic characterization for it. Axiomatic characterizations might be useful for statisticians, who have sometimes several procedures to solve a problem and no clear reasons to choose one among those procedures.

To model the density estimation problem in game-theoretical terms, we proceed essentially as follows: we consider the real line $\mathbb{R}$ as an infinite set of players of a cooperative game with characteristic function v such that for any coalition A (i.e. any subset A of $\mathbb{R}$), $v(A)$ is determined by $X_1, .., X_m$ and A itself; we then define our estimator $\hat{f}$ of f as the payoff vector allocated by an appropriate game-theoretical solution concept. The precise nature of $\hat{f}$ evidently depends on both the solution concept used and the way in which $v(A)$ is determined by $X_1, .., X_m$ and A. In this paper we adopt a simple approach to the latter question, taking $v(A)$ proportional to the cardinality of $\{X_1, .., X_m\} \cap A'$, where A' is a set containing A. To avoid game-theoretical complications, and in the interests of computational efficiency (see section 4), we actually group the points of $\mathbb{R}$ in a countable number of "indivisible coalitions" J_i that act as the effective players.

As noted above, we start by dividing $\mathbb{R}$ in intervals of the same length δ, i.e. $\mathbb{R} = \cup_{i \in \mathbb{Z}} [\delta i, \delta(i+1))$. We denote $[\delta i, \delta(i+1))$ by J_i (for all $i \in \mathbb{Z}$). Now, for every $S \subset \mathbb{Z}$, define

$$\bar{v}(S) = \frac{1}{m\delta} \sum_{j=1}^{m} I_{\cup_{i \in S} \bar{J}_i}(X_j)$$

where, for all $i \in \mathbb{Z}$, $\bar{J}_i$ is the set $\cup \{J_r \mid |r - i| \leq k\}$, k being a non-negative integer, and I_A denotes the indicator function of A, for any set $A \subset \mathbb{R}$ ($I_A(x) = 1$ if $x \in A$, $I_A(x) = 0$ if $x \in \mathbb{R} \setminus A$). About $\bar{v}$ note that:

1. $\bar{v}(\mathbb{Z}) = \frac{1}{\delta}$. In fact, we want to allocate $\frac{1}{\delta}$ to the intervals of $\{J_i \mid i \in \mathbb{Z}\}$ because in this way we can define $\hat{f}(x)$ as the share of $\frac{1}{\delta}$ that J_i obtains ($x \in J_i$) and, then, $\int_{-\infty}^{+\infty} \hat{f}(x)dx = 1$.

2. For every $S \subset \mathbb{Z}$, $\bar{v}(S)$ is $\frac{1}{m\delta}$ times the number of observations belonging to an interval k-close to an interval in $\{J_i \mid i \in S\}$. This can be interpreted as the maximum density that can be allocated to S. Observe that k plays here the role of the smoothing parameter.

Now we can make a TU-game (N, v) out of the density estimation problem characterized by the sample $X_1, ..., X_m$:

- $N = \{i \in \mathbb{Z} \mid \bar{v}(i) > 0\}$ (note that N, which is a finite set, can also be written as $N = \{i \in \mathbb{Z} \mid \bar{v}(S \cup \{i\}) - \bar{v}(S) > 0 \text{ for some } S \subset \mathbb{Z}\}$).
- v is the restriction of $\bar{v}$ to $\{S | S \subset N\}$.

About this game we can make the following comments:

1. If $k = 0$ the game is additive. If $k \geq 1$ the game is subadditive and, moreover, $v(N) < \sum_{i \in N} v(i)$. As k increases, the J_i intervals share their observations with more neighboring intervals, producing a smoothing effect in the game.
2. It is easy to see that v is a concave game.
3. v can be seen as a cost game. Since $v(S)$ can be interpreted as the maximum density that should be allocated to S, a fair allocation x must belong to the core of v, i.e.
$$\sum_{i \in S} x_i \leq v(S)$$
for all $S \subset N$.

Since v is a concave game, its Shapley value $\phi(v)$ lies in its core (see Shapley (1971)). Thus, a promising estimator of the density function is that based on $\phi(v)$. Formally, the Shapley estimator we propose is given by:

$$\hat{f}_S(x) = \begin{cases} \phi_i(v) & \text{if } x \in J_i,\ i \in N \\ 0 & \text{if } x \in J_i,\ i \notin N. \end{cases}$$

Since the number of players in the game may be large, calculation of $\hat{f}$ directly from its definition and that of the Shapley value can be very onerous. The following theorem provides a more convenient expression.

Theorem 1 *Let (N, v) be the TU-game associated with the density estimation problem characterized by the sample $X_1, ..., X_m$. Then, for any $i \in N$,*

$$\phi_i(v) = \frac{1}{m\delta} \sum_{r \in N^k(i)} \frac{n(r)}{2k+1}$$

where $n(r)$ denotes the number of observations belonging to the interval J_r and $N^k(i)$ is the set $\{r \in N \mid |r - i| \leq k\}$.

Proof. From the definition of the Shapley value,

$$\phi_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi(N)} \frac{1}{m\delta} \sum_{r \in P^k(i,\pi)} n(r)$$

where $\Pi(N)$ is the set of permutations of N, n is the cardinality of the set N and

$$P^k(i,\pi) = \{r \in N^k(i) \mid \pi(i) \leq \pi(s) \text{ for all } s \in N^k(r)\}.$$

Now, taking into account that, for each $r \in N^k(i)$, the cardinality of the set

$$\{\pi \in \Pi(N) \mid \pi(i) \leq \pi(s) \text{ for all } s \in N^k(r)\}$$

is $n!/(2k+1)$, then

$$\sum_{\pi \in \Pi(N)} \sum_{r \in P^k(i,\pi)} n(r) = \sum_{r \in N^k(i)} \frac{n!}{2k+1} n(r)$$

and the theorem follows.□

3 An axiomatic characterization of the Shapley estimator

Taking into account the properties of the Shapley value, it is possible to provide axiomatic characterizations of the Shapley estimator. We present one in this section which follows the ideas in Myerson (1977).

Assume that $\delta > 0$ is fixed. A δ-estimator is a map which assigns to every random sample $\mathbf{X} = X_1, \ldots, X_m$ a density function[2] $\hat{f}_{\mathbf{X}}$ which satisfies $\hat{f}_{\mathbf{X}}(x) = \hat{f}_{\mathbf{X}}(y)$ for all $x, y \in J_i$ and all $i \in \mathbb{Z}$. For simplicity, we denote by $\hat{f}_{\mathbf{X}}(i)$ the evaluation of $\hat{f}_{\mathbf{X}}$ in any $x \in J_i$. Observe that, since $\hat{f}_{\mathbf{X}}$ is a density function, it holds that $\hat{f}_{\mathbf{X}}(x) \geq 0$ for all $x \in \mathbb{R}$ and that $\sum_{i \in \mathbb{Z}} \hat{f}_{\mathbf{X}}(i) = 1/\delta$. Note also that the Shapley estimator is, in fact, a family of δ-estimators (one for each $k \in \mathbb{N}$; remember that k denotes the smoothing parameter).

[2]For notational convenience, in this section we denote by $\hat{f}_{\mathbf{X}}$ the density estimation for a given sample $\mathbf{X}$.

Let us see some interesting properties for the δ-estimators. Let $\mathbf{X} = X_1, \ldots, X_m$ be a random sample and suppose that the smoothing parameter k is fixed. We say that $i, j \in \mathbb{Z}$ are directly connected if there exists an element of the sample X_l such that $X_l \in \bar{J}_i \cap \bar{J}_j$. We say that $C \subset \mathbb{Z}$ is a connected set if, for every $i, j \in C$ there exists a finite sequence $\{i_1, \ldots, i_r\} \subset C$ such that $i_1 = i$, $i_r = j$ and i_s, i_{s+1} are directly connected for every $s \in \{1, \ldots, r-1\}$. We say that C is a connected component of $\mathbb{Z}$ if C is a maximal connected subset of $\mathbb{Z}$. The first property that we consider is *component efficiency*, which states that the density that the δ-estimator assigns to the intervals in a connected component is the one corresponding to the observations belonging to those intervals.

CE (Component Efficiency). A δ-estimator is said to satisfy component efficiency for k if, for every random sample $\mathbf{X}$ and every connected component C,

$$\sum_{i \in C} \hat{f}_{\mathbf{X}}(i) = \frac{m_C}{m} \frac{1}{\delta}$$

where m_C is the number of observations lying in $\cup_{i \in C} J_i$.

Note that if a δ-estimator satisfies CE for k, then it allocates density equal to zero to the intervals in whose neighborhoods there are no observations of the sample, more precisely to those J_i such that $\sum_{r \in N^k(i)} n(r) = 0$.

The second property that we introduce is the *fairness property*. Informally, this property states that, when estimating a density function, the increment of the density due to a particular observation is the same for all intervals in the neighborhood of this observation.

F (Fairness). A δ-estimator is said to satisfy fairness for k if, for every random sample $\mathbf{X}$, every $i, j \in \mathbb{Z}$ directly connected and every $X_l \in \bar{J}_i \cap \bar{J}_j$, it holds that

$$\hat{f}_{\mathbf{X}}(i) - \hat{f}_{\mathbf{X} \setminus X_l}(i) = \hat{f}_{\mathbf{X}}(j) - \hat{f}_{\mathbf{X} \setminus X_l}(j),$$

where $\mathbf{X} \setminus X_l$ is the sample identical to $\mathbf{X}$ except for the fact that X_l has been shifted far from the observations of $\mathbf{X}$ (more precisely, it has been shifted to an interval J_r such that $\bar{J}_r \cap \bar{J}_s = \emptyset$ for all J_s containing some observation of $\mathbf{X}$).

A statistical argument to further explain these properties is the following. As it was remarked, every sample observation X_i shows that there is a positive density not only on the real number X_i but also on a neighborhood of it (given by the smoothing parameter). Having this in mind, CE can be interpreted as that the positive density given by the sample is allocated to the right intervals. F means that the density corresponding to an observation X_i is allocated equally to all those intervals to which it must be allocated. These two properties are, in our opinion, quite natural and appealing. Moreover, they characterize the family of the Shapley estimators, as the following theorem shows.

Theorem 2 *Take $\delta > 0$. For every value of the smoothing parameter k, the corresponding Shapley estimator is the unique δ-estimator satisfying CE and F.*

Proof. Clearly, the Shapley estimator satisfies CE and F. In order to prove uniqueness, assume that there exist two different δ-estimators satisfying CE and F (let us call them estimators 1 and 2). Since they are different, there must exist a sample $\mathbf{X}$ of size m such that $\hat{f}^1_{\mathbf{X}} \neq \hat{f}^2_{\mathbf{X}}$ and such that it has a maximal number of connected components among those samples of size m for which the two estimators provide different estimations. For this $\mathbf{X}$, take a connected component C and $i, j \in C$ such that i and j are directly connected. Since both estimators satisfy F,

$$\hat{f}^r_{\mathbf{X}}(i) - \hat{f}^r_{\mathbf{X}}(j) = \hat{f}^r_{\mathbf{X}\setminus X_l}(i) - \hat{f}^r_{\mathbf{X}\setminus X_l}(j) \tag{1}$$

for all $r \in \{1, 2\}$ and all $X_l \in \bar{J}_i \cap \bar{J}_j$. Note now that

$$\hat{f}^1_{\mathbf{X}\setminus X_l}(i) - \hat{f}^1_{\mathbf{X}\setminus X_l}(j) = \hat{f}^2_{\mathbf{X}\setminus X_l}(i) - \hat{f}^2_{\mathbf{X}\setminus X_l}(j) \tag{2}$$

because, if X_l is the unique observation in $\cup_{k\in C} J_k$, then all the elements in equation (2) are zeros and, otherwise, $\mathbf{X} \setminus X_l$ has, at least, one connected component more than $\mathbf{X}$ and then (2) follows from maximality of $\mathbf{X}$. Hence, in view of (1) and (2), it holds that

$$\hat{f}^1_{\mathbf{X}}(i) - \hat{f}^2_{\mathbf{X}}(i) = \hat{f}^1_{\mathbf{X}}(j) - \hat{f}^2_{\mathbf{X}}(j).$$

It is clear that repeating this argument a finite number of times, one concludes that $\hat{f}^1_{\mathbf{X}}(i) - \hat{f}^2_{\mathbf{X}}(i)$ is constant for all $i \in C$. Since both estimators satisfy CE that constant must be zero. This clearly implies that $\hat{f}^1_{\mathbf{X}} = \hat{f}^2_{\mathbf{X}}$, which is a contradiction. □

4 Properties of the Shapley estimator

In the sequel, the statistical and computational properties of this estimator will be studied in comparison with a kernel estimator.

The Shapley estimator results to be within well-known families of density estimators: it is a binned kernel density estimator and it is a WARPing-type density estimator.

Binned kernel density estimators use some set of binning functions $w_i^\delta(x)$ (where $\sum_{i\in\mathbb{Z}} w_i^\delta(x) = 1$ for all $x \in \mathbb{R}$ and $\delta > 0$) to distribute the weight of each sample point X_j among the intervals J_i (the bins); concentrate the accumulated weight of each bin at its center, g_i; and then apply a kernel estimator to the resulting weighted "sample" $\{(g_i, N_i)\}$, where $N_i = \sum_{j=1}^m w_i^\delta(X_j)$:

$$\hat{f}_B(x) = \frac{1}{mh}\sum_{i\in\mathbb{Z}} N_i K(\frac{x-g_i}{h}).$$

The Shapley estimator is a binned kernel estimator with $w_i^\delta(x) = I_{J_i}(x)$, $N_i = n(i)$, $K = \frac{1}{2}I_{[-1,1)}$ and $h = \delta(k+\frac{1}{2})$.

WARPing (Weighted Averaging of Rounded Points) estimators use a discrete function $w(r,i)$ ($r, i \in \mathbb{Z}$; $\sum_{i\in\mathbb{Z}} w(r,i) = 1$ $\forall r \in \mathbb{Z}$) to distribute the total weight of the sample points in each bin among the neighboring bins:

$$\hat{f}_W(x) = \frac{1}{m\delta}\sum_{r\in\mathbb{Z}} w(r,i)n(r) \quad \forall\, x \in J_i.$$

If δ is small enough, smoothing is mainly effected by the function $w(r,i)$, which is often defined in terms of a smoothing parameter. The Shapley estimator is a WARPing estimator with $w(r,i) = \frac{1}{(2k+1)}$ if $r \in N^k(i)$, $w(r,i) = 0$ otherwise. Scott (1985) showed that one particular type of WARPing estimator, the averaged shifted histogram, is similar to the histogram in computational efficiency and to kernel estimators in statistical efficiency. By computational efficiency we simply mean little computational costs, whereas statistical efficiency refers to the accuracy of the estimation. Results on the efficiency of binned kernel estimators have been obtained by Hall and Wand (1996).

The objective of binning is to reduce computational costs. Computationally, the most efficient estimator is the histogram, which uses bins without any distribution of sample weight or other smoothing, but its statistical efficiency is poor by any reasonable criterion. On the other side, the family of kernel estimators should be mentioned. Roughly speaking, these estimators have good properties from the point of view of statistical efficiency, but they are relatively poor from the point of view of computational efficiency. Take, for instance, the kernel estimator with uniform kernel (which is the limiting form of the Shapley estimator as δ tends to zero),

$$\hat{f}_{KU}(x) = \frac{1}{2mh} \sum_{j=1}^{m} I_{[x-h,x+h)}(X_j).$$

Its computational efficiency is relatively poor because, although the simplicity of its kernel reduces the cost of smoothing in comparison with other kernels, it makes no use of bins, and thus has to be calculated for each point x for which its value is required.

Meanwhile, the Shapley estimator only requires to compute the frequency on each interval J_i and to compute the summations. Note also that the Shapley estimator is constant on each interval J_i and that we can make use of the recursive relation

$$\hat{f}_S(i+1) = \hat{f}_S(i) + \frac{1}{2mh}(n(i+k+1) - n(i-k))$$

where $\hat{f}_S(i)$ denotes the Shapley estimator evaluated at any point in the interval J_i.

Hence, the Shapley estimator allows for substantial computational savings. Moreover, taking reasonable values of δ, the statistical efficiency of the Shapley estimator is competitive with respect to the kernel estimator with uniform kernel. A deeper discussion of this assertion would require to introduce technical details that we decided not to include here[3].

The MISE (mean integrated squared error) is the most commonly used measure to assess statistical efficiency in this context. It is defined by

$$MISE = E(\int (\hat{f}(x) - f(x))^2 \, dx).$$

[3]Those readers interested in these details may ask the authors for an extended version of this paper, which includes an involved analysis of the statistical properties of the Shapley estimator.

In the extended version of this paper, the following result concerning the MISE of the Shapley estimator is proved, and some of its consequences are widely commented.

Theorem 3 *Assume that f is five times differentiable and that f and its first five derivatives are bounded and integrable. If k and δ are defined as functions of the sample size m in such a way that the smoothing parameter $h = \delta(k + \frac{1}{2})$ tends to zero as m tends to infinity, then*

$$\begin{aligned} MISE\left(\hat{f}_S\right) &= \frac{\delta^2}{12}\int f'(x)^2\,dx + \left(\frac{h^4}{36} - \frac{h^2\delta^2}{72} - \frac{\delta^4}{960}\right)\int f''(x)^2\,dx \\ &+ \frac{1}{2mh} - \frac{1}{m}\int f(x)^2\,dx + O\left(h^5\right) + O\left(\frac{h}{m}\right) \end{aligned}$$

where, as usual, $O(a_m)$ denotes a sequence such that the sequence $\frac{O(a_m)}{a_m}$ is bounded.

5 Final discussion

In summary, cooperative game theory provides a successful approach to the statistical problem of estimating a density function. The Shapley estimator has good computational and statistical properties which show it to be a competitive compromise between the histogram and the kernel estimator with uniform kernel. The major difference between traditional approaches to density estimation and the game-theoretic approach described in this article is that, whereas the sample weight distribution schemes of traditional approaches ensure that weight is conserved (for example, by using a density function as the kernel of a kernel estimator), the worth function of the game (N, v) assigns each set $S' = \cup_{i\in S} J_i$ the sample weight that "properly" corresponds to a k-neighborhood of S', without S' losing any of its own weight; it is left to the Shapley value to restore conservation of weight by adjudicating a fair share of the worth of the grand coalition to each player in the game.

We believe that this connection between game theory and statistics is an inspiring one, and that it may be indicative of a broader relationship that can be fruitful in the future.

References

Blackwell D, Girshick MA (1954) Theory of Games and Statistical Decisions. John Wiley, New York

Cao R, Cuevas A, González-Manteiga W (1994) A comparative study of several smoothing methods in density Estimation. Computional Statistics and Data Analysis 17:153-176

Hall P, Wand MP (1996) On the accuracy of binned kernel density estimators. Journal of Multivariate Analysis 56:165-184

Härdle W (1991) Smoothing Techniques with Implementation in S. Springer-Verlag

Land M, Gefeller O (1997) A game-theoretic approach to partitioning attributable risks in epidemiology. Biometrical Journal 39:772-792

Land M, Gefeller O (2000) A Multiplicative Variant of the Shapley Value for Factorizing the Risk of Desease. In: Patrone F, García-Jurado I, Tijs S (eds.) Game Practice: Contributions from Applied Game Theory. Kluwer Academic Publishers, pp. 143-158

Myerson R (1977) Graphs and cooperation in games. Mathematics of Operations Research 2:225-229

Parzen E (1962) On estimation of a probability density function and mode. Annals of Mathematical Statistics 33:1065-1076

Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics 27:832-837

Schwarz G (1994) Game Theory and Statistics. In: Aumann R, Hart S (eds.) Handbook of Game Theory (Vol. 2). Elsevier, Amsterdam, pp. 769-779

Scott DW (1985) Averaged shifted histograms: Effective nonparametric density estimators in several dimensions. Annals of Statistics 13:1024-1040

Shapley LS (1953) A value for n-person games. In: Tucker A, Kuhn H (eds.) Contributions to the theory of games II. Princeton University Press, pp. 307-317

Shapley LS (1971) Cores of convex games. International Journal of Game Theory 1:11-26

Silverman BW (1986) Density Estimation for Statistics and Data Analysis. Chapman and Hall