

AN ALGORITHMIC DETERMINATION OF OPTIMAL MEASURE FROM DATA AND SOME APPLICATIONS

N. K. AGBEKO AND A. HÁZY

ABSTRACT. In [8], by determining fuzzy measure from the data, the authors did not mention under what circumstances this can be done. Following the idea of [8] we propose by means of optimal measure theory an alternative way which is to determine optimal measure from the data. We also show that this determination can always be done.

1. INTRODUCTION

All along the pair $(\Omega, \mathcal{F})$ will stand for an arbitrarily fixed measurable space and $[0, 1]^{\mathcal{F}}$ the collection of all set functions $\mu: \mathcal{F} \rightarrow [0, 1]$.

Readily we note that among the functions in $[0, 1]^{\mathcal{F}}$ probability measure constitutes the most studied field, due to Kolmogorov's impacting pace. Yet, there are still at least as many as continuum other types of functions in $[0, 1]^{\mathcal{F}}$ to be dealt with.

In fact, by various statistical tools it is customary to find from a representing data a probability measure followed by a given sample, which is a finite sequence of random variables mapping Ω into $\mathbb{R}$. At times if the normality is assumed confidence intervals for the theoretical mean (or variance) can be computed.

Nevertheless, the question arises whether or not from a given data only probability measure can and should be looked for.

In the literature some approaches have been initiated to the solution of the raised problem. Fuzzy measure can be mentioned here as an instance. We notice that fuzzy measure which also belongs to set $[0, 1]^{\mathcal{F}}$ saw the light in the 60s. However, we must point out that the axioms of fuzzy measure are

2000 *Mathematics Subject Classification*. Primary 58C05, 26E60; Secondary 26E35, 34E18.

Key words and phrases. optimal measure, optimal average.

The second author has been supported by the Hungarian Scientific Research Fund (OTKA) Grants K-62316, NK-68040 and NK-81402.

not uniquely given as in the case of probability measure. In [8], by determining fuzzy measure from the data the authors did not mention under what circumstances this can be done.

To the image of probability measure and mathematical expectation (or Lebesgue integral) we introduced (using an axiom of one of the types of fuzzy measures) the so-called optimal measure in the set $[0, 1]^{\mathcal{F}}$ and a non-linear functional referred to as optimal average or integral. Among others we obtained the corresponding Fubini as well as Radon-Nikodym theorems. For more about these notions and the derived results see [1, 2, 3, 4].

Alongside with the idea of [8], the object of our present communication, which is a part of the first author's PhD thesis [5], is to propose by means of optimal measure theory an alternative way which consists of the determination of optimal measure from the data. We also show that in our framework this determination can always be done. A MAPLE code program can be found at the end of this paper.

In the sequel the symbols $\vee$ and $\bigvee$ will stand for the maximum operators, and $\bigwedge$ for the minimum operator.

Here below we shall list some basic definitions and results.

Definition 1.1 (cf. [1]). A set function $p: \mathcal{F} \rightarrow [0, 1]$ will be called optimal measure if it satisfies the following three axioms:

Axiom 1. *The identities $p(\Omega) = 1$ and $p(\emptyset) = 0$ hold.*

Axiom 2. *For all measurable sets B and E , we have $p(B \cup E) = p(B) \vee p(E)$.*

Axiom 3. *Function p is continuous from above, i.e. whenever $(E_n) \subset \mathcal{F}$ is a decreasing sequence, then*

$$p\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n \rightarrow \infty} p(E_n) = \bigwedge_{n=1}^{\infty} p(E_n).$$

A general form of Definition 1.1 is formulated here below.

Definition 1.2 (cf. [2]). By a quasi-optimal measure we mean a set function $q: \mathcal{F} \rightarrow [0, \infty]$ which satisfies Axioms 1–3, with the hypothesis $q(\Omega) = 1$ in Axiom 1 replaced by the hypothesis $0 < q(\Omega) < \infty$.

A typical example of quasi-optimal measure is provided in Proposition 2.1 of [2].

Definition 1.3 (cf. [2]). By a p -atom we mean a measurable set H , $p(H) > 0$ such that whenever $B \in \mathcal{F}$ and $B \subset H$, then $p(B) = p(H)$ or $p(B) = 0$.

Definition 1.4 (cf. [2]). A p -atom H is decomposable if there exists a subatom $B \subset H$ such that $p(B) = p(H) = p(H \setminus B)$. If no such subatom exists, we shall say that H is indecomposable.

Theorem 1.1 (The Structure Theorem). *Let p be an optimal measure defined on the measurable space $(\Omega, \mathcal{F})$. Then there exists a collection $\mathcal{H}(p) = \{H_n : n \in J\}$ of disjoint indecomposable p -atoms, where J is some countable (i.e. finite or countably infinite) index set, such that for every measurable set $B \in \mathcal{F}$ with $p(B) > 0$ we have*

$$(1) \quad p(B) = \bigvee_{n \in J} p(B \cap H_n).$$

Moreover, if the index set J is countably infinite, then the only limit point of the set $\{p(H_n) : n \in J\}$ is 0.

Next, we shall recall the definition of the optimal average or integral as follows.

Let

$$s = \sum_{i=1}^n b_i \chi(B_i)$$

be an arbitrary nonnegative measurable simple function, where

$$\{B_i : i = 1, \dots, n\} \subset \mathcal{F}$$

is a partition of Ω .

Definition 1.5 (cf. [1]). The quantity

$$\bigwedge_{\Omega} s dp := \bigvee_{i=1}^n b_i p(B_i)$$

will be called optimal average of s .

We proved ([1], Theorem 1.0) that the optimal average of any nonnegative measurable simple function does not depend on its decomposition. Due to Proposition 2.0 in [1] we were able to extend Definition 1.5 to arbitrary random variables.

Definition 1.6. The optimal average of a measurable function f is defined by $A|f| := \bigwedge_{\Omega} |f| dp = \sup \bigwedge_{\Omega} s dp$, where the supremum is taken over all measurable simple functions $s \geq 0$ for which $s \leq |f|$.

The following remark is readily obtained.

Remark 1.1 (cf. [2]). If a function $f : \Omega \rightarrow \mathbb{R}$ is measurable, then it is constant almost everywhere on every indecomposable atom.

Proposition 1.1 (cf. [2], Proposition 2.6). *Let $p \in \mathcal{P}$ and f be any measurable function. Then*

$$\bigwedge_{\Omega} |f| dp = \sup \left\{ \bigwedge_{H_n} |f| dp : n \in J \right\},$$

where $\mathcal{H}(p) = \{H_n : n \in J\}$ is a p -generating countable system. Moreover, if $A|f| < \infty$, then

$$\int_{\Omega} |f| dp = \sup \{c_n \cdot p(H_n) : n \in J\},$$

where $c_n = f(\omega)$ for almost all $\omega \in H_n$, $n \in J$.

2. ALGORITHMIC DETERMINATION OF OPTIMAL MEASURE FROM DATA

2.1. Some preliminary. In fuzzy sets theory the crux was how to determine the values of the fuzzy measure in a given real problem. To achieve that goal the Sugeno integral was used together with the so-called genetic algorithm to solve it (see [8]), say. The Sugeno integral with respect to a given fuzzy measure μ is regarded as a multi-input single-output system. The input is the integrand, i.e. the vector $(f(\omega_1), \dots, f(\omega_n))$, while the output is the value of its Sugeno integral

$$E := (S) \int f d\mu = \sup \{\alpha \wedge \mu(F_\alpha) : \alpha \in [0, 1]\},$$

where f (with range contained in $[0, 1]$) is a measurable function defined on a finite measurable space $(\Omega, \mathcal{F})$ and

$$F_\alpha := \{\omega \in \Omega : f(\omega) \geq \alpha\}.$$

By repeatedly observing the system

$$(f(\omega_1), \dots, f(\omega_n))$$

results the following

$f_{11}(\omega_1)$	$f_{12}(\omega_2)$	$\dots$	$f_{1k}(\omega_n)$	E_1
$f_{21}(\omega_1)$	$f_{22}(\omega_2)$	$\dots$	$f_{2k}(\omega_n)$	E_2
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$f_{k1}(\omega_1)$	$f_{k2}(\omega_2)$	$\dots$	$f_{kk}(\omega_n)$	E_k

and the authors of [8] obtained an approximate fuzzy measure μ with

$$E_i = (S) \int f_i d\mu, \quad (i = 1, \dots, k),$$

such that the expression

$$e := \sqrt{\frac{1}{k} \sum_{i=1}^k \left(E_i - (S) \int f_i d\mu \right)^2}$$

is minimized. For more about the genetic algorithm see [6, 7], for example.

2.2. An analogous approach. An analogous crucial question also arises to know how to determine the range of optimal measure from the data. We shall first formulate some useful problems.

Problem 1. Let $(\Omega, \mathcal{F})$ be the measurable space with $\Omega = \{1, \dots, n\}$ and $\mathcal{F} = 2^\Omega$, i.e., $\mathcal{F}$ is the power set of Ω . Write $B_1 := \{1\}, \dots, B_n := \{n\}$ and let f be a random variable assuming the theoretical values in $[0, \infty)$. Observe k times this measurable function with results $f_1, \dots, f_k$, i.e.,

B_1	B_2	$\dots$	B_n		
$f_1(1)$	$f_1(2)$	$\dots$	$f_1(n)$	Q_1	$Q_1^{(r*)}$
$f_2(1)$	$f_2(2)$	$\dots$	$f_2(n)$	Q_2	$Q_2^{(r*)}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
$f_k(1)$	$f_k(2)$	$\dots$	$f_k(n)$	Q_k	$Q_k^{(r*)}$

where

$$Q_i^{(r)} := \frac{1}{n} \sum_{j=1}^n f_{ij}^r$$

respectively

$$Q_i^{(r*)} := \frac{1}{n} \sum_{j=1}^n |f_{ij} - Q_i^{(1)}|^r$$

with $f_{ij} := f_i(j)$ being the observed value of f in the i^{th} trial on event B_j , ($j \in \{1, \dots, n\}$, $i \in \{1, \dots, k\}$ and $r \geq 1$). The question is to know which one of these sample r^{th} moments $Q_i^{(r)}$ (resp. r^{th} central moments $Q_i^{(r*)}$) can "best" approximate the theoretical r^{th} moment (resp. r^{th} central moment) of the random variable f .

To solve Problem 1 we propose to look for an approximation of the theoretical optimal measure p for which

$$\int_{\Omega} f_i^r dp \approx Q_i^{(r)},$$

respectively

$$\int_{\Omega} |f_i - Q_i^{(1)}|^r dp \approx Q_i^{(r*)} \quad (i = 1, \dots, k),$$

such that the expression

$$err_1 := \sqrt{\sum_{i=1}^k \varepsilon_i^2} = \sqrt{\sum_{i=1}^k \left(Q_i^{(r)} - \int_{\Omega} f_i^r dp \right)^2}$$

respectively,

$$err_2 := \sqrt{\sum_{i=1}^k \varepsilon_i^2} = \sqrt{\sum_{i=1}^k \left(Q_i^{(r*)} - \int_{\Omega} |f_i - Q_i^{(1)}|^r dp \right)^2}$$

is minimized. Write p_0 for the optimal measure p for which the least square is minimal. Now, it is not difficult to see that

$$\bigvee_{i=1}^k \left| Q_i^{(r)} - \int_{\Omega} f_i^r dp_0 \right| < err_1.$$

Let i_0 be the index where the maximum is attained, i.e.,

$$\left| Q_{i_0}^{(r)} - \int_{\Omega} f_{i_0}^r dp_0 \right| = \bigvee_{i=1}^k \left| Q_i^{(r)} - \int_{\Omega} f_i^r dp_0 \right|.$$

Then we can conclude that with respect to the optimal measure p_0 the i_0 th sample provides us with the best possible sample r th moments. Similar conclusion can be drawn for the theoretical r th central moment.

As we know statistical spaces are not restricted in general to the real line nor to the real vector spaces. For this reason we need to formulate the following problem. We shall then indicate how to reduce Problem 2 to Problem 1.

Problem 2. Let $(X, \mathcal{S})$ be a measurable space with $\mathcal{S}$ being an arbitrary σ -algebra. Fix a measurable partition $D_1, \dots, D_n$ of X and consider a random variable $h: X \rightarrow [0, \infty)$, assuming theoretical values. Observe k times this measurable function with the following results:

D_1	D_2	$\dots$	D_n		
$h_1(1)$	$h_1(2)$	$\dots$	$h_1(n)$	$Q_1^{(r)}$	$Q_1^{(r*)}$
$h_2(1)$	$h_2(2)$	$\dots$	$h_2(n)$	$Q_2^{(r)}$	$Q_2^{(r*)}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
$h_k(1)$	$h_k(2)$	$\dots$	$h_k(n)$	$Q_k^{(r)}$	$Q_k^{(r*)}$

where

$$Q_i^{(r)} := \frac{1}{n} \sum_{j=1}^n h_{ij}^r$$

respectively

$$Q_i^{(r*)} := \frac{1}{n} \sum_{j=1}^n \left| h_{ij} - Q_i^{(1)} \right|^r$$

with $h_{ij} := h_i(j)$ being the observed value of h in the i^{th} trial on event D_j where $j \in \{1, \dots, n\}$, $i \in \{1, \dots, k\}$ and $r \geq 1$. The question is to know which one of these sample r^{th} moments $Q_i^{(r)}$ (resp. r^{th} central moments $Q_i^{(r*)}$)

can "best" approximate the theoretical r^{th} moment (resp. r^{th} central moment) of the random variable h .

To solve Problem 2, first write

$$\mathcal{S}_0 := \sigma(D_1, \dots, D_n).$$

We note that $\mathcal{S}_0$ is a finite σ -algebra and the random variable h is also $\mathcal{S}_0$ -measurable. Clearly, $\mathcal{S}_0$ and 2^Ω are equinumerous, where $\Omega = \{1, \dots, n\}$. Then Problem 2 can be reduced to Problem 1 if we define $f_{ij} := h_{ij}$ where $i = 1, \dots, k$ and $j = 1, \dots, n$.

2.3. Why do the above solutions always exist? In the similar papers where different types of set functions $\mu \in [0, 1]^\mathcal{F}$ were estimated from data no guarantees have been shown that the least square minimization can always be done. Hereafter we shall show that in the case of optimal measure theory the least square minimization can always be used to provide estimated optimal measures from data.

Theorem 2.1. *Let $0 < s = b_1\chi(B_1) + b_2\chi(B_2) + \dots + b_n\chi(B_n)$ be any measurable simple function defined on $(\Omega, \mathcal{F})$, where $B_1, \dots, B_n$ is a partition of Ω . Then for all numbers $\varepsilon \in (0, 1)$ and $c \in (0, \infty)$ there exists a quasi-optimal measure q such that*

$$\left| c - \int_{\Omega} s dq \right| < \varepsilon.$$

Proof. Suppose in the contrary that there are some numbers $\varepsilon_0 \in (0, 1)$, $c_0 \in (0, \infty)$ and for every quasi-optimal measure q ,

$$\left| c_0 - \int_{\Omega} s dq \right| > \varepsilon_0.$$

Let $\{n_1, \dots, n_n\}$ be a permutation of $\{1, \dots, n\}$ which arranges decreasingly the numbers $b_1, \dots, b_n$, that is $b_{n_n} > \dots > b_{n_2} > b_{n_1}$. Clearly, to the value b_{n_j} corresponds the measurable set B_{n_j} , $j \in \{1, \dots, n\}$. On the other hand it is also trivial that

$$s = b_1\chi(B_1) + \dots + b_n\chi(B_n) = b_{n_1}\chi(B_{n_1}) + \dots + b_{n_n}\chi(B_{n_n}).$$

Now, fix $\omega_j \in B_{n_j}$, $j \in \{1, \dots, n\}$ and define the function $q_0 : \mathcal{F} \rightarrow [0, \infty)$, by

$$q_0(E) = \max \left\{ \frac{c_0}{b_{n_j}} : \omega_j \in E \right\}.$$

Obviously, q_0 is a quasi-optimal measure. Let

$$\mathcal{H}(q_0) = \{H_{n_j} : j \in \{1, \dots, n\}\}$$

be a generating system of q_0 . We then trivially have that

$$q_0(H_{n_j}) = q_0(B_{n_j}) = \frac{c_0}{b_{n_j}}, \quad j \in \{1, \dots, n\}.$$

Hence

$$\varepsilon_0 < \left| c_0 - \bigwedge_{\Omega} sdq \right| = \left| c_0 - \bigvee_{j=1}^n b_{n_j} q_0(H_{n_j}) \right| = 0,$$

which, however, contradicts the choice of ε_0 . This concludes the proof. $\square$

3. AN ALGORITHM TO SOLVE THE FIRST PROBLEM

The following algorithm provides a solution for the first problem in the case $r = 1$. The general form can be obtained by making little changes in both the algorithm and the Maple code program.

The Inputs: n positive integer
 $\Omega = \{1, \dots, n\}$
 $k \times n$ matrix $F = [f(i, j)]_{i,j=1}^{n,k}$
 n -dimensional vector Q
error bound ε
 $B_j = \{j\}, j = 1 \dots n$
 X = the power set of Ω .
Generate the set σ of all permutations of $\{1, \dots, n\}$.

Step 1.

Generate a decreasing sequence
 $\alpha(j) \in (0, 1]$, with $\alpha(1) = 1$.

Step 2.

For any permutation $\{n_1, \dots, n_n\} \in \sigma$
Put $p(B_j) = \alpha(n_j)$, for $j = 1, \dots, n$
Compute the optimal average:
 $A(i) = \max\{f(i, j) * p(B_j) : j = 1, \dots, n\}$ for $i = 1, \dots, k$
Compute the corresponding error: $err = \sqrt{\left(\sum_{j=1}^n (Q(i) - A(i))\right)^2}$

Step 3.

If $err < \varepsilon$ for some permutation then
Find the index i_0 such that $|Q(i_0) - A(i_0)| = \max\{|Q(i) - A(i)| : i = 1, \dots, k\}$
Determine $p(B) = \max\{\alpha(n_j) : j \in B\}$ for each $B \in X$
Else GOTO Step 1

The outputs:

- 1.) **Best sample:** $f(i_0, 1), \dots, f(i_0, n)$
- 2.) **The approximated optimal**

measure:

2^Ω	$p(B)$
$\{\}$	0
B_1	$p(B_1)$
$\vdots$	$\vdots$
B_i	$p(B_i)$
$\vdots$	$\vdots$

4. A MAPLE CODES SOLUTION OF PROBLEM 1

```

> with(combinat) :
> with(stats) :
> with(VectorCalculus) :
> st := time() :
> bestsample := proc(k, n, epsz)
  global Omega, A, B, X, S, F, alpha, Q, err, p, measure, i0, setmeasure;
  local i, j, vel, per, step, err1;
  Omega := {};
  setmeasure := array(1..2^n, 1..2);
  for i from 1 to n do
    Omega := Omega union {i};
  od;
  S := subsets(Omega) :
  i := 1;
  while not S[finished] do
    X[i] := S[nextvalue]();
    i := i + 1
  od;
  F := array(1..k, 1..n);
  for i from 1 to k do
    for j from 1 to n do
      F[i, j] := abs(stats[random, normald](1));
    od;
  od;
  print('Matrix F : '); print(F);
  for j from 1 to n do
    B[j] := j
  od;

```

```

Q := array[1..k];
for i from 1 to k do
    Q[i] := 0;
    for j from 1 to n do
        Q[i] := Q[i] + F[i, j]/n;
    od;
od;
print('The mean values of the rows of matrix F, i.e. vector Q : '); print(Q);
err1 := 10000000;
while epsz < err1 do
    alpha := [1, op(sort(RandomTools[Generate](list(float(range = 0..1), n - 1))
        , '>'))];
    step := 1;
    err1 := 10000000;
    print('The elements of vector alpha'); print(alpha);
    while (epsz < err1 and step <= n!) do
        per := permute(n)[step];
        for j from 1 to n do
            p[j] := alpha[per[j]]
        od;
        for i from 1 to k do
            for j from 1 to n do
                if j = 1 then A[i] := F[i, j] * p[j];
                else if A[i] < F[i, j] * p[j] then A[i] := F[i, j] * p[j]; fi;
            od;
        od;
        err := 0;
        for i from 1 to k do
            err := (Q[i] - A[i])^2;
        od;
        err := sqrt(err);
        step := step + 1;
        if err < err1 then err1 := err fi;
    od;
    print('The error obtained from this alpha : '); print(err1);
od;

```

```

print('The permutation which provides the optimal measure : '); print(per);
i0 := 1;
for i from 2 to k do
    if abs(Q[i] - A[i]) > abs(Q[i0] - A[i0]) then i0 := i; fi;
od;
print('The value of i0 : '); print(i0);
for i from 1 to 2n do
    measure := 0;
    for j from 1 to n do
        if in(j, X[i]) then measure := max(measure, p[j]) fi;
    od;
    setmeasure[i, 1] := X[i];
    setmeasure[i, 2] := measure;
od;
print('The measure of the sets : '); print(setmeasure);
print('The run time : '); time() - st;
end proc;

```

5. EXAMPLE

‘Matrix F‘

$$\begin{bmatrix} 0.4194899055 & 0.6079846246 & 0.9765680503 & 0.09453300345 \\ 1.231491638 & 0.3284744201 & 1.255955432 & 1.705108694 \\ 1.128982417 & 1.052401359 & 0.008623922093 & 0.1198713918 \end{bmatrix}$$

‘The mean values of the rows of matrix F, i.e. vector Q:‘

table ([1 = 0.5246438961, 2 = 1.130257546, 3 = 0.5774697724])

‘The elements of vector alpha:‘

[1, 0.5248995821, 0.0000000009879313609, 0.0000000002610718605]

‘The error obtained from this alpha:‘

0.01513262650

‘The permutation which provides the optimal measure:‘

[2, 3, 1, 4]

‘The value of i_0 :‘

1

‘The measure of the sets:‘

$\{\}$	0
$\{1\}$	0.5248995821
$\{2\}$	0.0000000009879313609
$\{3\}$	1
$\{4\}$	0.0000000002610718605
$\{1, 2\}$	0.5248995821
$\{1, 3\}$	1
$\{1, 4\}$	0.5248995821
$\{2, 3\}$	1
$\{2, 4\}$	0.0000000009879313609
$\{3, 4\}$	1
$\{1, 2, 3\}$	1
$\{1, 2, 4\}$	0.5248995821
$\{1, 3, 4\}$	1
$\{2, 3, 4\}$	1
$\{1, 2, 3, 4\}$	1

‘The run time:‘

2.687

REFERENCES

- [1] N. K. Agbeko. On optimal averages. *Acta Math. Hungar.*, 63(2):133–147, 1994.
- [2] N. K. Agbeko. On the structure of optimal measures and some of its applications. *Publ. Math. Debrecen*, 46(1-2):79–87, 1995.
- [3] N. K. Agbeko. How to characterize some properties of measurable functions. *Math. Notes (Miskolc)*, 1(2):87–98, 2000.
- [4] N. K. Agbeko. Mapping bijectively σ -algebras onto power sets. *Math. Notes (Miskolc)*, 2(2):85–92, 2001.
- [5] N. K. Agbeko. *The role of the maximum operator in the theory of measurability and some applications*. PhD thesis, Miskolc, 2009.

- [6] R. Baragona, F. Battaglia, and C. Calzini. Genetic algorithms for the identification of additive and innovation outliers in time series. *Comput. Statist. Data Anal.*, 37(1):1–12, 2001.
- [7] C.-C. Lin and A.-P. Chen. Fuzzy discriminant analysis with outlier detection by genetic algorithm. *Comput. Oper. Res.*, 31(6):877–888, 2004.
- [8] Z. Wang, K.-S. Leung, and J. Wang. Determining nonnegative monotone set functions based on Sugeno’s integral: an application of genetic algorithms. *Fuzzy Sets and Systems*, 112(1):155–164, 2000.

Received May 12, 2009.

UNIVERSITY OF MISKOLC,
DEPARTMENT OF APPLIED MATHEMATICS
MISKOLC-EGYTEMVÁROS
E-mail address: `matagbek@uni-miskolc.hu`

UNIVERSITY OF MISKOLC,
DEPARTMENT OF APPLIED MATHEMATICS
MISKOLC-EGYTEMVÁROS
E-mail address: `matha@uni-miskolc.hu`