Selecting the Most Representative Sample is NP-Hard: Need for Expert (Fuzzy) Knowledge

J. Esteban Gamez, François Modave, and Olga Kosheleva

***Abstract*—One of the main applications of fuzzy techniques is to formalize the notions of "typical", "representative", etc.**

The main idea behind fuzzy techniques is that they formalize expert knowledge expressed by words from natural language.

In this paper, we show that if we do not use this knowledge, i.e., if we only use the data, then selecting the most representative sample becomes a computationally difficult (NP-hard) problem. Thus, the need to find such samples in reasonable time justifies the use of fuzzy techniques.

I. Introduction to the problem

In many practical situations, it is desirable to find the statistical analysis of a certain population, but this population is so large that it is not practically possible to analyze every individual element from this population. In this case, we select a *sample* (subset) of the population, perform a statistical analysis on this sample, and use these results as an approximation to the desired statistical characteristics of the population as a whole.

For example, this is how polls work: instead of asking the opinion of all the people, pollsters ask a representative sample, and use the opinion of this sample as an approximation to the opinion of the whole population.

The more "representative" the sample, the larger our confidence that the statistical results obtained by using this sample are indeed a good approximation to the desired population statistics. Typically, we gauge the representativeness of a sample by how well its statistical characteristics reflect the statistical characteristics of the entire population. For example, in the sample of human voters, it is reasonable to require that in the selected sample, the average age, the average income, and other characteristics are the same as in the population in a whole. Of course, the representativeness of averages is not enough: e.g., the voting patterns of people whose salary is exactly the national average are not necessarily a good representation of how people will work on average. For that, we need the sample to include both poorer and reacher people – i.e., in general, to be representative not only in terms of averages but also in terms of, e.g., standard deviations (i.e., equivalently, in terms of variances).

In practice, many techniques are used to design a representative sample; see, e.g., [2]. In this paper, we consider this problem as the exact optimization problem, and we show that this problem is computationally difficult (NP-hard).

Comment. Similar results are known: for example, we know that a similar problem of maximizing diversity is NP-hard; see, e.g., [1].

J. Esteban Gamez and François Modave are with the Department of Computer Science, University of Texas at El Paso, emails estebangamez@gmail.com, and Olga Kosheleva is with the Department of Teacher Education, University of Texas at El Paso, email olgak@utep.edu.

This work was supported in part by This work was supported in part by an internal grant from New Mexico State University, by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, and by Texas Department of Transportation grant No. 0-5453.

II. Towards formulation of the problem in exact terms

Let us assume that we have a population consisting of N objects. For each of N objects, we know the values of k characteristics x_1, x_2, ..., x_k. The value of the first characteristic x_1 for i-th object will be denoted by $x_{1,i}$, the value of the second characteristic x_2 for the i-th object will be denoted by $x_{2,i}$, ..., and finally, the value of the characteristic x_k for the i-th object will be denoted by $x_{k,i}$. As a result, we arrive at the following formal definition:

Definition 1. *By a* population*, we mean a tuple*

$$p \stackrel{\text{def}}{=} \langle N, k, \{x_{j,i}\}\rangle,$$

where:

- *N is an integer; this integer will be called the* population size*;*
- *k is an integer; this integer is called the* number of characteristics*;*
- *$x_{j,i}$ ($1 \le j \le k, 1 \le i \le N$) are real numbers; the real number $x_{j,i}$ will be called the* value *of the j-th characteristic for the i-th object.*

Based on these known values, we can compute the population means

$$E_1 = \frac{1}{N}\cdot\sum_{i=1}^{N} x_{1,i}, \quad E_2 = \frac{1}{N}\cdot\sum_{i=1}^{N} x_{2,i}, \quad \ldots,$$

and the population variances

$$V_1 = \frac{1}{N}\cdot\sum_{i=1}^{N}(x_{1,i}-E_1)^2,$$

$$V_2 = \frac{1}{N}\cdot\sum_{i=1}^{N}(x_{2,i}-E_2)^2, \quad \ldots$$

We can also compute higher order central moments.

Definition 2. *Let $p = \langle N, k, \{x_{j,i}\}\rangle$ be a population, and let j be an integer from 1 to k.*

- *By the* population mean E_j *of the j-th characteristic, we mean the value*

$$E_j = \frac{1}{N} \cdot \sum_{i=1}^{N} x_{j,i}.$$

- *By the* population variance V_j *of the j-th characteristic, we mean the value*

$$V_j = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{j,i} - E_j)^2.$$

- *For every integer $d \geq 1$, by the even order* population central moment $M_j^{(2d)}$ of order $2d$ *of the j-th characteristic, we mean the value*

$$M_j^{(2d)} = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{j,i} - E_j)^{2d}.$$

Comment. In particular, the population central moment $M_2^{(2)}$ of order 2 (corresponding to $d = 1$) is simply the population variance.

In addition to the values $x_{1,i}$, $x_{2,i}$, ..., we are given a size $n < N$ of the desirable sample. For each sample $I = \{i_1, \ldots, i_n\} \subseteq \{1, 2, \ldots, N\}$ of size n, we can compute the sample means

$$E_1(I) = \frac{1}{n} \sum_{i \in I} x_{1,i}, \quad E_2(I) = \frac{1}{n} \sum_{i \in I} x_{2,i}, \quad \ldots$$

and the sample variances

$$V_1(I) = \frac{1}{n} \sum_{i \in I} (x_{1,i} - E_1(I))^2,$$

$$V_2(I) = \frac{1}{n} \sum_{i \in I} (x_{2,i} - E_2(I))^2, \ldots$$

Definition 3. *Let N be a population size.*

- *By a* sample*, we mean a non-empty subset*

$$I \subseteq \{1, 2, \ldots, N\}.$$

- *For every sample I, by its* size*, we mean the number of elements in I.*

Definition 4. *Let $p = \langle N, k, \{x_{j,i}\}\rangle$ be a population, let I be a sample of size n, and let j be an integer from 1 to k.*

- *By the* sample mean $E_j(I)$ *of the j-th characteristic, we mean the value*

$$E_j(I) = \frac{1}{n} \cdot \sum_{i \in I} x_{j,i}.$$

- *By the* sample variance $V_j(I)$ *of the j-th characteristic, we mean the value*

$$V_j(I) = \frac{1}{n} \cdot \sum_{i \in I} (x_{j,i} - E_j(I))^2.$$

- *For every $d \geq 1$, by the* sample central moment $M_j^{(2d)}(I)$ of order $2d$ *of the j-th characteristic, we mean the value*

$$M_j^{(2d)}(I) = \frac{1}{n} \cdot \sum_{i \in I} (x_{j,i} - E_j(I))^{2d}.$$

Comment. Similarly to the population case, the sample central moment $M_2^{(2)}$ of order 2 (corresponding to $d = 1$) is simply the sample variance.

We want to select the *most representative* sample, i.e., the sample for which the sample statistics $E_1(I), E_2(I), \ldots, V_1(I), V_2(I), \ldots$ are the closest to the population statistics $E_1, E_2, \ldots, V_1, V_2, \ldots$

Definition 5. *Let $p = \langle N, k, \{x_{j,i}\}\rangle$ be a population.*

- *By an E-*statistics tuple *corresponding to p, we mean a tuple*

$$t^{(1)} \stackrel{\text{def}}{=} (E_1, \ldots, E_k).$$

- *By an (E, V)-*statistics tuple *corresponding to p, we mean a tuple*

$$t^{(2)} \stackrel{\text{def}}{=} (E_1, \ldots, E_k, V_1, \ldots, V_k).$$

- *For every integer $d \geq 1$, by a* statistics tuple of order $2d$ *corresponding to p, we mean a tuple*

$$t^{(2d)} \stackrel{\text{def}}{=} (E_1, \ldots, E_k, M_1^{(2)}, \ldots, M_k^{(2)},$$
$$M_1^{(4)}, \ldots, M_k^{(4)}, \ldots, M_1^{(2d)}, \ldots, M_k^{(2d)}).$$

Comment. In particular, the statistics tuple of order 2 is simply the (E, V)-statistics tuple.

Definition 6. *Let $p = \langle N, k, \{x_{j,i}\}\rangle$ be a population, and let I be a sample.*

- *By an E-*statistics tuple *corresponding to I, we mean a tuple*

$$t^{(1)}(I) \stackrel{\text{def}}{=} (E_1(I), \ldots, E_k(I)).$$

- *By an (E, V)-*statistics tuple *corresponding to I, we mean a tuple*

$$t^{(2)}(I) \stackrel{\text{def}}{=} (E_1(I), \ldots, E_k(I), V_1(I), \ldots, V_k(I)).$$

- *For every integer $d \geq 2$, by a* statistics tuple of order $2d$ *corresponding to I, we mean a tuple*

$$t^{(2d)}(I) \stackrel{\text{def}}{=} (E_1(I), \ldots, E_k(I),$$
$$M_1^{(2)}(I), \ldots, M_k^{(2)}(I),$$
$$M_1^{(4)}(I), \ldots, M_k^{(4)}(I) \ldots, M_1^{(2d)}(I), \ldots, M_k^{(2d)}(I)).$$

Comment. In particular, the statistics tuple of order 2 corresponding to a sample I is simply the (E, V)-statistics tuple corresponding to this same tuple.

We will show that no matter how we define closeness, this problem is NP-hard (computationally difficult).

Let us describe the problem in precise terms. To describe which tuple

$$t(I) \stackrel{\text{def}}{=} (E_1(I), E_2(I), \ldots, V_1(I), V_2(I), \ldots)$$

is the closest to the original statistics tuple

$$t \stackrel{\text{def}}{=} (E_1, E_2, \ldots, V_1, V_2, \ldots),$$

we need to fix a *distance function* $\rho(t(I), t)$ describing how distant are the two given tuples. Similarly to the usual distance, we would like this distance function to be equal to 0 when the tuples coincide and to be positive if when the tuples are different. So, we arrive at the following definitions.

Definition 7. *By a* distance function*, we mean a mapping ρ that maps every two real-valued tuples t and t' of the same size into a real value $\rho(t, t')$ in such a way that $\rho(t, t) = 0$ for all tuples t and $\rho(t, t') > 0$ for all $t \neq t'$.*

As an example, we can take Euclidean metric between the tuples $t = (t_1, t_2, \ldots)$ and $t' = (t'_1, t'_2, \ldots)$:

$$\rho(t, t') = \sqrt{\sum_j (t_j - t'_j)^2}.$$

Now, we are ready to formulate the problem.

Definition 8. *Let ρ be a distance function. By a* E-sample selection problem *corresponding to ρ, we mean the following problem. We are given:*

- *a population $p = \langle N, k, \{x_{j,i}\}\rangle$, and*
- *an integer $n < N$.*

Among all samples $I \subseteq \{1, \ldots, N\}$ of size n, we must find the sample I for which the distance $\rho(t^{(1)}(I), t^{(1)})$ between the corresponding E-statistical tuples is the smallest possible.

Definition 9. *Let ρ be a distance function. By a* (E, V)-sample selection problem *corresponding to ρ, we mean the following problem. We are given:*

- *a population $p = \langle N, k, \{x_{j,i}\}\rangle$, and*
- *an integer $n < N$.*

Among all samples $I \subseteq \{1, \ldots, N\}$ of size n, we must find the sample I for which the distance $\rho(t^{(2)}(I), t^{(2)})$ between the corresponding (E, V)-statistical tuples is the smallest possible.

Definition 10. *Let ρ be a distance function, and let $d \geq 1$ be an integer. By a* $2d$-th order sample selection problem *corresponding to ρ, we mean the following problem. We are given:*

- *a population $p = \langle N, k, \{x_{j,i}\}\rangle$, and*
- *an integer $n < N$.*

Among all samples $I \subseteq \{1, \ldots, N\}$ of size n, we must find the sample I for which the distance $\rho(t^{(2d)}(I), t^{(2d)})$ between the corresponding $(2d)$-th order statistical tuples is the smallest possible.

III. Main results

Proposition 1. *For every distance function ρ, the corresponding E-sample selection problem is NP-hard.*

Proposition 2. *For every distance function ρ, the corresponding (E, V)-sample selection problem is NP-hard.*

Proposition 3. *For every distance function ρ and for every integer $d \geq 1$, the corresponding $(2d)$-th order sample selection problem is NP-hard.*

IV. Proof of Propositions 1–3

A. Main idea: reduction to subset sum, a known NP-hard problem

We prove NP-hardness of our problem by reducing a known NP-hard problem to it: namely, a *subset sum* problem, in which we are given m positive integers $s_1, \ldots, s_m$, and we must find the signs $\varepsilon_i \in \{-1, 1\}$ for which

$$\sum_{i=1}^{m} \varepsilon_i \cdot s_i = 0;$$

see, e.g., [3].

A reduction means that to every instance $s_1, \ldots, s_m$ of the subset sum problem, we must assign (in a feasible, i.e., polynomial-time way) an instance of our problem in such a way that the solution to the new instance will lead to the solution of the original instance.

B. Reduction: explicit description

Let us describe this reduction: we take $N = 2n$, $k = 2$, $n = m$, and we select the values $x_{j,i}$ as follows:

- $x_{1,i} = s_i$ and $x_{1,m+i} = -s_i$ for all $i = 1, \ldots, m$;
- $x_{2,i} = x_{2,m+i} = 2^i$ for all $i = 1, \ldots, m$.

We will show that for this new problem, the most representative sample I has $\rho(t(I), t) = 0$ if and only if the original instance of the subset sum problem has a solution.

C. General analysis

Indeed, by definition of a distance function, the equality $\rho(t(I), t) = 0$ is equivalent to $t(I) = t$, i.e., to the requirement that for the sample I, means (and variances) within the sample are exactly the same as for the entire population.

D. Consequences for the second component

Let us start by analyzing the consequences of this requirement for the mean of the second component. For the entire population of size $N = 2m$, for each i from 1 to m, we have two elements, i-th and $(m + i)$-th, with the value $x_{2,i} = x_{2,m+i} = 2^i$. Thus, for the population as a whole, this mean is equal to

$$E_2 = \frac{2 + 2^2 + \ldots + 2^m}{m}.$$

For the selected subset I of size m, this mean should be exactly the same: $E_2(I) = E_2$. Thus, we must have

$$E_2(I) = \frac{2 + 2^2 + \ldots + 2^m}{m}.$$

By definition,

$$E_2(I) = \frac{1}{m} \cdot \sum_{i \in I} x_{2,i}.$$

Thus, we conclude that

$$S_2(I) \stackrel{\text{def}}{=} \sum_{i \in I} x_{2,i} = 2 + 2^2 + \ldots + 2^m.$$

What can we now conclude about the set I?

First of all, we can notice that in the sum $2+2^2+\ldots+2^m$, all the terms are divisible by 4 except for the first term 2. Thus, the sum itself is not divisible by 4.

In our population, we have exactly two elements, element 1 and element $m+1$, for which $x_{2,1} = x_{2,m+1} = 2$. For every other element, we have $x_{2,i} = x_{2,m+i} = 2^i$ for $i \geq 2$ and therefore, the corresponding value is divisible by 4.

In regards to a selection I, there are exactly three possibilities:

- the set I contains none of the two elements 1 and $m+1$;
- the set I contains both elements 1 and $m+1$; and
- the set I contains exactly one of the two elements 1 and $m+1$.

In the first two cases, the contribution of these two elements to the sum $S_2(I)$ is divisible by 4 (it is 0 or 4). Since all other elements in the sum $S_2(I)$ are divisible by 4, we would thus conclude that the sum itself is divisible by 4 – which contradicts to our conclusion that this sum is equal to $2 + 2^2 + \ldots + 2^m$ and is, therefore, not divisible by 4.

This contradiction shows that the set I must contain exactly one of the two elements 1 and $m+1$. Let us denote this element by k_1. For this element, $x_{2,k_1} = 2$. Subtracting x_{2,k_1} and 2 from the two sides of the equality

$$S_2(I) = \sum_{i \in I} x_{2,i} = 2 + 2^2 + \ldots + 2^m,$$

we conclude that

$$S_2(I - \{k_1\}) = \sum_{i \in I - \{k_1\}} x_{2,i} = 2^2 + 2^3 + \ldots + 2^m.$$

In the new sum $2^2 + 2^3 + \ldots + 2^m$, all the terms are divisible by $2^3 = 8$ except for the first term 2^2. Thus, the sum itself is not divisible by 8.

In our remaining population $\{2, \ldots, m, m+2, \ldots, 2m\}$, we have exactly two elements, element 2 and element $m+2$, for which $x_{2,2} = x_{2,m+2} = 2^2$. For every other element, we have $x_{2,i} = x_{2,m+i} = 2^i$ for $i \geq 3$ and therefore, the corresponding value is divisible by 3.

In regards to a selection I, there are exactly three possibilities:

- the set I contains none of the two elements 2 and $m+2$;
- the set I contains both elements 2 and $m+2$; and
- the set I contains exactly one of the two elements 2 and $m+2$.

In the first two cases, the contribution of these two elements to the sum $S_2(I - \{k_1\})$ is divisible by 8 (it is 0 or 8). Since all other elements in the sum $S_2(I - \{k_1\})$ are divisible by 8, we would thus conclude that the sum itself is divisible by 8 – which contradicts to our conclusion that this sum is equal to $2^2 + 2^3 + \ldots + 2^m$ and is, therefore, not divisible by 8.

This contradiction shows that the set I must contain exactly one of the two elements 2 and $m+2$. Let us denote this element by k_2. For this element, $x_{2,k_2} = 2^2$. Subtracting x_{2,k_2} and 2^2 from the two sides of the equality

$$S_2(I - \{k_1\}) = \sum_{i \in I - \{k_1\}} x_{2,i} = 2^2 + 2^3 + \ldots + 2^m,$$

we conclude that

$$S_2(I - \{k_1, k_2\}) = \sum_{i \in I - \{k_1, k_2\}} x_{2,i} = 2^3 + 2^4 + \ldots + 2^m.$$

Now, we can similarly conclude that the set I contains exactly one element from the pair $\{3, m+3\}$, and in general, for every i from 1 to m, we can conclude that the selection set I contains exactly one element k_i from the pair $\{i, m+i\}$.

E. Consequences for the first component

Let us now analyze the consequences of this requirement for the mean of the first component. For the entire population of size $N = 2m$, for each i from 1 to m, we have two elements, i-th and $(m+i)$-th, with the opposite values $x_{1,i} = s_i$ and $x_{2,m+i} = -s_i$. Thus, for the population as a whole, this mean is equal to $E_1 = 0$.

For each i from 1 to m, the selection set contains exactly one element of these two: $k_i = i$ and $k_i = m + i$. Thus, $E_1(I) = 0$ means that the corresponding sum is equal to 0: $\sum\limits_{i=1}^{m} x_{1,k_i} = 0$. Here, $x_{1,k_i} = \varepsilon_i \cdot s_i$, where:

- $\varepsilon_i = 1$ if $k_i = i$, and
- $\varepsilon_i = -1$ if $k_i = m + i$.

Thus, we conclude that $\sum\limits_{i=1}^{m} \varepsilon_i \cdot s_i = 0$ for some $\varepsilon_i \in \{-1, 1\}$, i.e., that the original instance of the subset problem has a solution.

F. Equivalence

Vice versa, if the original instance of the subset problem has a solution, i.e., if $\sum\limits_{i=1}^{m} \varepsilon_i \cdot s_i = 0$ for some $\varepsilon_i \in \{-1, 1\}$, then we can select $I = \{k_1 \ldots, k_m\}$, where:

- $k_i = i$ when $\varepsilon_i = 1$, and
- $k_i = m + i$ when $\varepsilon_i = -1$.

One can easily check that in this case, we have $E_1(I) = E_1$, $E_2(I) = E_2$, $V_1(I) = V_1$, $V_2(I) = V_2$, and, in general, $M_1^{(2d)}(I) = M_1^{(2d)}$ and $M_2^{(2d)}(I) = M_2^{(2d)}$.

G. Conclusion

The reduction is proven, so the problem of finding the most representative sample is indeed NP-hard.

V. Auxiliary results

A. Motivations

In our proofs, we considered the case when the desired sample contains half of the original population. In practice, however, samples form a much smaller portion of the population. A natural question is: what if we fix a large even number $2P \gg 2$, and look for samples which constitute the $(2P)$-th part of the original population? It turns out that the resulting problem of selecting the most representative sample is still NP-hard.

B. Definitions

Definition 11. *Let ρ be a distance function, and let $2P$ be a positive even integer. By a* problem of selecting an E-sample of relative size $\frac{1}{2P}$*, we mean the following problem:*

- *We are given a population $p = \langle N, k, \{x_{j,i}\}\rangle$.*
- *Among all samples $I \subseteq \{1, \ldots, N\}$ of size $n = \frac{n}{2P}$, we must find the sample I for which the distance $\rho(t^{(1)}(I), t^{(1)})$ between the corresponding E-statistical tuples is the smallest possible.*

Definition 12. *Let ρ be a distance function, and let $2P$ be a positive even integer. By a* problem of selecting an (E, V)-sample of relative size $\frac{1}{2P}$*, we mean the following problem:*

- *We are given a population $p = \langle N, k, \{x_{j,i}\}\rangle$.*
- *Among all samples $I \subseteq \{1, \ldots, N\}$ of size $n = \frac{n}{2P}$, we must find the sample I for which the distance $\rho(t^{(2)}(I), t^{(2)})$ between the corresponding (E, V)-statistical tuples is the smallest possible.*

Definition 13. *Let ρ be a distance function, let $d \geq 1$ be an integer, and let $2P$ be a positive even integer. By a* problem of selecting an $(2d)$-th order sample of relative size $\frac{1}{2P}$*, we mean the following problem:*

- *We are given a population $p = \langle N, k, \{x_{j,i}\}\rangle$.*
- *Among all samples $I \subseteq \{1, \ldots, N\}$ of size $n = \frac{n}{2P}$, we must find the sample I for which the distance $\rho(t^{(2d)}(I), t^{(2d)})$ between the corresponding statistical tuples of order $2d$ is the smallest possible.*

C. Results

Proposition 4. *For every distance function ρ and for every even integer $2P$, the corresponding problem of selecting an E-sample of relative size $\frac{1}{2P}$ is NP-hard.*

Proposition 5. *For every distance function ρ and for every even integer $2P$, the corresponding problem of selecting an (E, V)-sample of relative size $\frac{1}{2P}$ is NP-hard.*

Proposition 6. *For every distance function ρ, for every integer $d \geq 1$, and for every even integer $2P$, the corresponding problem of selecting a $(2d)$-th order sample of relative size $\frac{1}{2P}$ is NP-hard.*

D. Proof of Propositions 4–6

The proof is similar to the proofs of Propositions 1–3.

The main difference is that for each i from 1 to m, we now have not two but $2P$ different objects

$$i, m+i, 2m+i, \ldots, k \cdot m + i, \ldots, (2P-1) \cdot m + i$$

with the same value

$$x_{2,i} = x_{2,m+i} = \ldots =$$

$$x_{2,k\cdot m+i} = \ldots = x_{2,(2P-1)\cdot m+i} = (2P)^i.$$

(And this common value is also different.)

Among these $2P$ objects with the same value of the second characteristic $x_{2,.}$, for the first half, we have $x_{1,.} = s_i$ and for the second half, we have $x_{1,.} = -s_i$, i.e.:

$$x_{1,i} = x_{1,m+i} = \ldots = m_{1,(P-1)\cdot m+i} = s_i;$$

$$x_{1,P\cdot m+i} = x_{1,(P+1)\cdot m+i} = \ldots = m_{1,(2P-1)\cdot m+i} = -s_i.$$

By using divisibility by $(2P)^2$ (instead of divisibility by 2^2), we conclude that the best fitting sample is the one which has exactly one element of each group. Thus, from $E_1(I) = E_1$, we similarly conclude that the original instance of the subset problem has a solution – and hence that the new problems are indeed NP-hard.

VI. Conclusions

One of the applications of fuzzy techniques is to formalize the notions of "typical", "representative", etc. The main idea behind fuzzy techniques is that they formalize expert knowledge expressed by words from natural language.

In this paper, we have shown that if we do not use this knowledge, i.e., if we only use the data, then selecting the most representative sample becomes a computationally difficult (NP-hard) problem. Thus, the need to find such samples in reasonable time justifies the use of fuzzy techniques.

References

[1] C. C. Kuo, F. Glover, and K. S. Dhir, "Analyzing and modeling the maximum diversity problem by zero-one programming", *Decision Sciences*, vol. 24, no. 6, pp. 1171–1185, 1993.

[2] H. Lohr, *Sampling: Design and Analysis*, Duxbury Press, 1999.

[3] C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, San Diego, 1994.