Journal of Scientific Research and Studies Vol. 5(2), pp. 31-33, February, 2018
ISSN 2375-8791

http://www.modernrespub.org/jsrs/index.htm

Short Communication

A counterexample of the statement P = NP

Peter Kopanov

Faculty of Mathematics, Plovdiv University, 4000 Plovdiv, Bulgaria. E-mail: pkopanov@yahoo.com

Accepted 01 February, 2018

In this paper, it is shown that the classes of *P* and *NP* do not coincide by constructing a relatively simple example of unsolvability in polynomial time in a particular well-known *NP*-complete problem. For a particular NP-complete problem, a random process is constructed that generates a solution with random numbers. Such a solution cannot be found with a polynomial algorithm, but it can be verified with one if it is known. In this way, the impossibility of equality P = NP is shown.

Key words: NP-complete problem, polynomial algorithm, independent and identically distributed random variables, random choice.

INTRODUCTION

The problem "Is P = NP?" was officially formulated by Cook (1971a). The precise statement of the P versus NP problem was introduced in 1971 by Stephen Cook in his seminal paper (Cook, 1971b). We will not give details of the current results. We mention only that until now there is neither a rigorous proof of $P \neq NP$, nor an algorithm in polynomial time that solves at least one of the over 3000 known *NP*-complete problems. Due to the latter, most computer scientists indeed believe that $P \neq NP$.

Attempts to prove or dispute the Statement of P = NP have been conducted frequently. We will note some of them, which are described in the works (Deolalikar, 2010; Anand, 2008; Kupchik, 2004; Meek, 2008; Tärnlund, 2009; Feinstein, 2004; Bringsjord and Taylor, 2004; Feinstein, 2007) since they are more recognized.

Here it will be shown that the assumption that a concrete *NP*-complete problem belongs to the *P* class will contradict the random and independent choice of a concrete solution to the problem. Indeed, if the assumption was correct, it would imply that we can guess a sequence *A* of zeroes and ones through another sequence of numbers *B* that has been constructed independently of *A*.

Objective

The aim of this study is to show that Problem 20 from the list (Karp, 1972) of Karp with 21 *NP*-complete problems cannot be solved generally in polynomial time.

PRELIMINARIES

The problem is named as PARTITION and its formulation is as follows:

INPUT: $(c_1, c_2, \ldots, c_s) \in \mathbb{Z}^s$.

PROPERTY: There is a set $I \subset \{1, 2, 3, \ldots, s\}$ such that $\sum_{h\in I} c_h = \sum_{h\notin I} c_h$.

(that is, we want to check whether a given set of s integers can be partitioned into two subsets, the elements of which have equal sum).

Finding a solution I to the problem (in terms of the input) is equivalent to finding a binary vector $(a_1, a_2, \ldots, a_s)$ of length *s*, such that

$$a_i=1 \Leftrightarrow i \in I, \quad a_i=0 \Leftrightarrow i \notin I. \qquad (1)$$

The total number of the binary vectors of length *s* is 2^s. For finding *I*, we have to check whether these exists a binary vector $(a_1, a_2, \ldots, a_s)$ satisfying the equality

$$\sum_{k=1}^{s} a_k . c_k = \frac{1}{2} . \sum_{k=1}^{s} c_k \qquad (2)$$

In this case, the statement *P=NP* would imply that there exists an algorithm in polynomial time, which in terms of an input $(c_1, c_2, \ldots, c_s)$ finds an output vector $(a_1, a_2, \ldots, a_s)$ (if such exists), corresponding to *I* as in Equation (1).

Of course, a direct (but exponential) algorithm is to completely verify all 2^s binary vectors of length *s* in the equality of Equation (2).

If *P=NP*, then not all binary *s*-length vectors would need to be verified - only some of them would be sufficient; the number of verifications would be a polynomial on the variable *s*. Therefore, there would exist an algorithmic connection between the input vector $(c_1, c_2, ..., c_s)$ and the eventual output vector $(a_1, a_2, ..., a_s)$.

We will show that the latter generally does not hold, through a concrete construction.

This section contains only an intuitive explanation of the idea, the rigorous proof with details begins from the next section.

CONSTRUCTION OF THE SOLUTION

Set $m=2^{100}-1$ (this particular suitable value is chosen only for explicitness).

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent and identically distributed discrete random variables, each with a probability mass function:

$$\mathbb{P}\{X_n = j\} = \frac{1}{2m},\ j \in \mathbb{Z},\ 0 < |j| \le m$$

The sequence $\{X_n\}_{n=1}^{\infty}$ can also be defined through two other sequences of independent identically distributed discrete random variables $\{Y_n\}_{n=1}^{\infty}$ and $\{Z_n\}_{n=1}^{\infty}$, given by the following mass functions:

$$\mathbb{P}\{Y_n = j\} = \frac{1}{2},\ j \in \{-1,1\}$$

$$\mathbb{P}\{Z_n = j\} = \frac{1}{m},\ j \in \mathbb{Z},\ 1 \le j \le m$$

The above definitions imply the following equalities for the three sequences:

$X_n = Y_n Z_n$, $Y_n = \mathrm{sgn}(X_n)$, $Z_n = |X_n|$

A crucial step is:

Proposition 1. *There exists a process of generating* $\{Y_n\}_{n=1}^{\infty}$ *and* $\{Z_n\}_{n=1}^{\infty}$, *such that it shows that the variables* Y_n *and* Z_n *are independent.*

Proof. Let $\{T_n\}_{n=1}^{\infty}$ be a sequence of independent and identically distributed Bernoulli random variables, each with a probability $\frac{1}{2}$ as an outcome, i.e.

$$\mathbb{P}\{T_n = j\} = \frac{1}{2},\ j \in \{-1,1\}$$

Consider *l*=101 - in particular, *l* satisfies $m=2^{l-1}-1$.

For $n \in \mathbb{N}$, set

$Y_n = 2T_{l(n-1)+1}-1$

$Z_n = \sum_{i=2}^{l} 2^{l-i} T_{l.(n-1)+i}$

(if $T_{l.(n_0-1)+i}=0$ for all $2 \le i \le l$, then ignore $X_{n_0}=0$ - it will not influence the distinct elements of the sums $S_n = \sum_{j=1}^{n} X_j$ which we shall consider later. In all other cases, note that the inequality $0<|X_n|\le m$ always holds).

Now we verify that these definitions of Y_n and Z_n agree with the respective mass functions:

1. We have $\mathbb{P}\{T_{l(n-1)+1} = j\} = \frac{1}{2},\ j \in \{0,1\}$.

Therefore $\mathbb{P}\{Y_n = -1\} = \mathbb{P}\{Y_n = 1\} = \frac{1}{2}$ indeed.

2. We have $\mathbb{P}\{T_{l(n-1)+i} = j\} = \frac{1}{2},\ j \in \{0,1\}$ and the values which Z_n can attain are exactly the integers from 1 to *m*, each with equal probability. Therefore $\mathbb{P}\{Z_n = j\} = \frac{1}{m},\ j \in \mathbb{Z},\ 1 \le j \le m$, as required.

Now note that Z_n is a linear combination of the independent and identically distributed random variables $T_{l(n-1)+i}$, $2 \le i \le l$ and that Y_n is a linear function on the variable $T_{l(n-1)+1}$, which is independent of the previous ones. Therefore Y_n and Z_n are independent, as desired.

Now consider the sums

$S_n = X_1 + X_2 + ... + X_n$

We have the following:

Lemma 2. *For* $\{X_n\}_{n=1}^{\infty}$ *of independent identically distributed random variables, each having a symmetric distribution, and* $S_n = X_1+X_2+...+X_n$ *we have* $\mathbb{P}\{S_n = 0$ *for infinitely many n} = 1.*

Proof. See (Shepp, 1962; Feller, 1968, 1971)

Remark. In fact, the above result is strong for our purposes; we will only use that there necessarily exists an index *s* with $S_s=0$.

Now we are ready to conclude:

Theorem 3. *The problem PARTITION does not belong to the P class.*

Proof. Since our variables X_k have a symmetric distribution, by Lemma 2; there exists an index *s* with $S_s = 0$. Then from the input $(c_1, c_2, ..., c_s)$ with $c_k = Z_k$ we get

$$\sum_{k=1}^{s} X_k = 0 \Leftrightarrow \sum_{k=1}^{s} Y_k Z_k = 0 \Leftrightarrow \sum_{k=1}^{s} \left(\frac{1+Y_k}{2}\right) Z_k = \frac{1}{2}\sum_{k=1}^{s} Z_k$$

and hence, referring to $\sum_{k=1}^{s} a_k . c_k = \frac{1}{2} . \sum_{k=1}^{s} c_k$, we get

- $I=\{k \in \mathbb{N} : Y_k=1\}$
- $a_k = \frac{1+Y_k}{2}$, $1 \le k \le s$

Now note that we may assume that the solution I is unique for the given input; else, we start again from the beginning of the process of generating Y_n, Z_n with the aim to reach an input that corresponds to a unique solution .

Now, if PARTITION belonged to the P-class, then the uniquely determined vector $(a_1, a_2, ..., a_s)$ and the vector $(c_1, c_2, ..., c_s)$ would have been polynomially dependent; however, this and the definitions of a_k, c_k contradict the independence of Y_n and Z_n as random variables, established in Proposition 1.

Conclusion

This article shows that solutions of a particular NP-complete problem can be generated using independent random numbers. Equality P=NP contradicts to the existence of independent random numbers.

Acknowledgements

I would like to express my special gratitude to Miroslav Marinov. Without his insistence, the construction would not have been described.

REFERENCES

Anand BS (2008). A trivial solution to the P versus NP problem. To appear in the proceedings of the 2008 International Conference on Foundations of Computer Science, July 14-17, 2008, Las Vegas, USA.

Bringsjord S, Taylor J (2004). P=NP. arXiv:cs/0406056v1.

Cook S (1971a). The P versus NP Problem. The Clay Math Institute Official Problem Description.

Cook S (1971b). The complexity of theorem proving procedures. Proceedings of the Third Annual ACM Symposium on Theory of Computing:151–158.

Deolalikar V (2010). P ≠ NP. HP Research Labs, Palo Alto .

Feinstein CA (2004). EVIDENCE THAT P ≠ NP. arXiv:cs/0310060v7.

Feinstein CA (2007). A new and elegant argument that P ≠ NP. arXiv:cs/0607093v2.

Feller W (1968). An introduction to probability theory and its applications, Volume I, Wiley; 3rd Edition.

Feller W (1971). An introduction to probability theory and its applications, Volume II, Wiley; 3rd Edition.

Karp RM (1972). Reducibility among combinatorial problems. In R.E. Miller and J.W. Thatcher (editors). Complexity of Computer Computations. New York: Plenum:85–103.

Kupchik M (2004). P versus NP problem solution. Preprint submitted to Elsevier Science.

Meek J (2008). P is a proper subset of NP. arXiv:0804.1079v12.

Shepp LA (1962). Symmetric random walk. Trans. Am. Math. Soc. 104(1):144-153.

Tärnlund S (2009). P is not equal to N P. arXiv:0810.5056v2.