1-1-2022

Search Under Uncertainty Should be Randomized: A Lesson From the 2021 Nobel Prize in Medicine

Martine Ceberio
The University of Texas at El Paso, mceberio@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:

Technical Report: UTEP-CS-22-03

Search Under Uncertainty Should be Randomized: A Lesson From the 2021 Nobel Prize in Medicine *

Martine Ceberio and Vladik Kreinovich[0000−0002−1244−1650]

Department of Computer Science
University of Texas at El Paso, El Paso, TX 79968, USA
{mceberio,vladik}@utep.edu
https://martineceberio.fr/
https://www.cs.utep.edu/vladik/longvita.html

Abstract. In many real-life situations, we know that one of several objects has the desired property, but we do not know which one. To find the desired object, we need to test these objects one by one. In situations when we have no additional information, there is no reason to prefer any testing order and thus, a usual recommendation is to test them in any order. This is usually interpreted as ordering the objects in the increasing value of some seemingly unrelated quantity. A possible drawback of this approach is that it may turn out that the selected quantity is correlated with the desired property, in which case we will need to test all the given objects before we find the desired one. This is not just an abstract possibility: this is exactly what happened for the research efforts that led to the 2021 Nobel Prize in Medicine. To avoid such situations, we propose to use randomized search. Such a search would have cut in half the multi-year time spent on this Nobel-Prize-winning research efforts.

Keywords: Search under uncertainty · Randomized search · 2021 Nobel Prize in Medicine

1 Formulation of the Problem

1.1 Search under uncertainty: a general problem

In many practical situations, we strongly suspect that one of several objects satisfies the desired property, but we do not know which one – and we have

* This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology.
It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

no reason to assume that some of these objects are more probable to have the desired property. To find the desired object, we must therefore try all the objects one by one until we find the object that has the desired property. Once we have found the desired object, we can stop the search.

In some situations, checking the desired property is very time- and/or resource-consuming. It is therefore desirable to minimize the number of such checks. So, a natural question is: in what order should we test the given objects?

1.2 What is known

In this situation, we know the following (see, e.g., [1]):

- The best situation is when the very first object that we check has the desired property. In this case, we only need a single check.
- The worst situation is when the very last object that we check has the desired property – or even none of the tested objects has the desired property. In this situation, we need to perform n checks, where n is the number of objects.

It is also possible to compute the average number of checks. Indeed, since we have no reason to believe that some objects are more probable to satisfy the desired property, it is reasonable to assume that each object has the exact same probability that this object has the desired property. This ideas – going back to Laplace – is known as the *Laplace Indeterminacy Principle*; see, e.g., [4].

Under this assumption:

- with probability $1/n$, the first object has the desired property, so we will need only 1 check;
- with probability $1/n$, the second object has the desired property, so we will need 2 checks;
- $\ldots$, and
- with probability $1/n$, the last object has the desired property, so we will need n checks.

The average number of checks is therefore equal to

$$\frac{1}{n} \cdot (1 + 2 + \ldots + n) = \frac{1}{n} \cdot \frac{n \cdot (n+1)}{2} = \frac{n+1}{2}.$$

1.3 The usual (seemingly reasonable) recommendation

We have no information about the objects, we do not know which objects are more probable to be desired and which are less probable.

From this viewpoint, it seems like we cannot make any recommendation about the order of testing, so any order should be OK.

1.4 What we do in this paper

In this paper, we argue that this usual recommendation is somewhat misleading, and that in situations of uncertainty, it is better to use randomized order.

This will not be just a theoretical conclusion – we show that following the new recommendation would have sped up the research that led to the 2021 Nobel Prize in Medicine.

2 Analysis of the Problem and the New Recommendation

2.1 How people usually interpret the "any order" recommendation

When people read "any order", they usually follow some seemingly unrelated order – e.g., alphabetic order of the objects' names or order by using the values of some seemingly unrelated quantity of different objects.

2.2 Sometimes, this works, but sometimes, it doesn't

If the quantity used in ordering is really unrelated to the property that we want to test, then this interpretation works well. However, since we know nothing about the desired property, it could be that the quantity used in the ordering is actually correlated with the desired quantity.

And if this correlation is positive, and we sort the objects in the increasing order of the selected quantity, then the desired object will be the last one – i.e., we arrive at the worse-case situation.

Of course, in this case, if we sort the objects in the decreasing order of the selected quantity, then we arrive at the best-case situation: we will pick the desired object right away or at least almost right away. So maybe it makes sense to try first the first and the last objects in the selected order, then the second and the last but one, etc.? This would cover both the cases of positive and negative correlation, but what if the dependence of the desired quality on the selected quantity is quadratic, with the maximum for the midpoint value? In this case, we again arrive at the worst-case situation.

How can we avoid this?

2.3 Natural idea

Since ordering by a fixed quantity may lead to the worst-case situation, a natural idea is *not* to use any deterministic approach, not to use any quantity to sort the objects, but to use *random* order. In other words:

- as a first object to test, we select any of the n given objects, with equal probability $1/n$;
- if the first selected object does not satisfy the desired property, then, for the second checking, we select any of the remaining $n-1$ objects with equal probability $1/(n-1)$;
- if the second selected objects does not satisfy the desired property either, then, for the third checking, we select any of the remaining $n-2$ objects with equal probability $1/(n-2)$, etc.

2.4 How many checks will we need if we follow this recommendation

Under such random order, the average number of checks is equal to $(n+1)/2$ – this follows from the same argument that we had before.

It is possible that we will get into the worst-case situation, but the probability of this happening is the probability that the desired object will be the last in the random order – i.e., $1/n$. For large n, this probability is very small.

2.5 But is all this really important?

Theoretically, the new recommendation is reasonable, but a natural question arises: how important is the difference between this new recommendation and the usual one?

In the next section, we show that this difference is not purely theoretical: namely, we show that following the new recommendation would have sped up the research that led to the 2021 Nobel Prize in Medicine.

3 Case Study

3.1 Research that led to the Nobel Prize

The 2021 Nobel Prize in Physiology or Medicine was awarded to Dr. David Julius and Dr. Ardem Patapoutian who discovered receptors for temperature and touch; see, e.g., [2, 3, 5, 6, 8]. In this research, first, they identified 72 genes as potential sensors of mechanical force. To find out which of these genes is a receptor for touch, they silenced each of these genes one by one and checked whether the cells would still react to poking.

Interestingly, when they silences each of the first 71 genes in their ordering, the cell was still reacting to touch. It was only when they silenced ("switched of") the last, 72nd gene, that they saw that the reaction to poking disappeared – which showed that this gene was a receptor for touch.

It is not easy to switch off a gene, so this research took several years.

3.2 How this research could be sped up

Clearly, in this study, the researchers hit what we described as the worst-case situation, when the quantity selected for ordering was correlated with the desired quality.

If instead of selecting a deterministic order, they would have used a random order, they would, on average, have cut the number of tests – and thus, the research duration – in half, with the probability of encountering the worst-case situation as small as $1/72 \approx 1.5\%$.

References

1. Cormen, T. H., Leiserson, C. E., Rivest, R. L, Stein, C.: Introduction to Algorihms. MIR Press, Cambridge, Massachusetts (2009)
2. Coste, B., Mathur, J., Schmidt, M., Earley, T. J., Ranade, S., Petrus, M. J., Dubin, A. E., Patapoutian, A.: Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. Science **330**, 55–60 (2010)
3. Howard Hughes Medical Institute (HHMI): David Julius and Ardem Patapoutian Awarded the 2021 Nobel Prize in Physiology or Medicine. https://www.hhmi.org/news/david-julius-and-ardem-patapoutian-awarded-2021-nobel-prize-physiology-or-medicine, posted on October 4, 2021
4. Jaynes, E. T., Bretthorst, G. L.: Probability Theory: The Logic of Science. Cambridge University Press, Cambridge, UK (2003)
5. The Nobel Prize Committee, Press release: The Nobel Prize in Physiology or Medicine 2021. https://www.nobelprize.org/prizes/medicine/2021/press-release/, Posted on October 4, 2021.
6. Ranade, S. S., Woo, S. H., Dubin, A. E., Moshourab, R. A., Wetzel, C., Petrus, M., Mathur, J., Bégay, V., Coste, B., Mainquist, J., Wilson, A. J., Francisco, A. G., Reddy, K., Qiu, Z., Wood, J. N., Lewin, G. R., Patapoutian, A.: Piezo2 is the major transducer of mechanical forces for touch sensation in mice. Nature **516**, 121–125 (2014)
7. Sheskin, D. J.: Handbook of Parametric and Non-Parametric Statistical Procedures. Chapman & Hall/CRC, London, UK (2011)
8. Woom S. H., Lukacs, V., de Nooij, J. C., Zaytseva, D., Criddle, C. R., Francisco, A., Jessell, T. M., Wilkinson, K. A., Patapoutian, A.: Piezo2 is the principal mechonotransduction channel for proprioception. Nature Neuroscience **18**, 1756–1762 (2015)