Solving certain problems of a hierarchical money-distribution system with pseudo-random integer vectors and multidimensional scaling

Szabolcs, Takács

This thesis considers several, well-separated fields of study. However, these fields are treated as a single unit when supporting the solution of a large-scale problem which originated from a real life application.

During the project a pseudo-random integer vector generator has been tested and improved, using statistical tests which are even predefined by us or adopted to the random vector case. As an evidence of validation the results have been compared with a well-known and widely used algorithm. (MATLAB's© algorithm to create random vectors or numbers of integer members).

The analysis of the data of the core problem used the Multidimensional Scaling (MDS). The results indicated new questions which have successfully widened the already wide spectrum of this field. By reformulating this fundamentally statistical problem as an optimization problem, it became possible to analyze the original question in a significantly more general context than it is possible with the generalizations of statistical methods.

The third section focuses on the original problem's mathematical behavior and characteristics. It turned out that if we represent the problem as a classical flow problem, then it has unique characteristics affecting the later decisions of the user of the model. Due to these special structural properties, it was necessary to examine the task with sensitivity analysis. A stochastic method to study and resolve these kind of problems has been suggested. At the end of this section, a method for calculating a robust solution is provided.

These various aspects have been combined into the method providing the basis for the created application. During this process, novel, previously non-published methods of analysis and approaches have been presented, including new test scenarios. The resulting method is currently in use by the Employment Center which defined the original problem.

The main new results of the method are as follows:

1. New statistical methods for testing pseudo random vectors
2. Developing an algorithm for generating a list of pseudo-random integers
3. Generalization of the multidimensional scaling method (MDS)
4. The application of the MDS in a new field
5. Structural examination and generalization of the outcome in a hierarchically organized project

6. The application of the developed generator in a robust optimization problem
7. Utilization of the MDS in the interpretation of the results

The following subsections provide a more detailed description for each of these items. We concentrate on the summary of the results; for a more detailed description, see the corresponding chapters of the thesis.

1 Results in the field of pseudo-random integer vector generators

Pseudo-random number or vector generating algorithms are used in a large number of applications. The testing of the quality of an algorithm like this is an important aspect of each application. The first chapter of this thesis describes the process of quality assurance of a specified algorithm. The results can be summarized as follows:

- **(1.1. Thesis)** An algorithm (Knuth's "monotonicity harmony test" [7] for testing algorithms generating pseudo-random integer numbers) is adapted for testing algorithms generating pseudo-random sequences of integer vectors.
- **(1.2. Thesis)** In the predefined set of tests, the algorithm developed in the thesis did not perform worse than a well-known and widely used algorithm provided by MATLAB.
- **(1.3. Thesis)** Adapting algorithms for randomly generated number series and examining some new test methods, an algorithm for generating sequences of random integer vectors with an arbitrary pre-defined cycle length has been created, achieving a pseudo-random vector generator the self repeating cycle of which can be set to any arbitrary rate [5].

These results are elaborated in the thesis as follows:

1.1 Adaptation of the tests

The original tests can be found e.g. in Knuth's book [7] or Marsaglia's studies [8]. These tests are used to test the level of randomness of pseudo random series, therefore cannot be used without modifications in our case. Some of the tests are designed to perform the analysis of the distribution of the series, while the others look at the interdependency of the members of the series.

1.2 Test results

The test results are summarized and published [5]. The main results are:

1. No differences in the behavior of the algorithm were found in lower or in higher dimensions.

2. The parameter of shearing (i.e. the process of making the data sparser − which was applied to decrease and examine the interdependency) showed that it should not be selected higher than 8. At a lower shearing parameter (e.g. listing every second member) interdependency can still be found. Taking every 8th vector the internal interdependency of them is negligible. Working with higher spacing the quality of the results do not increase but the speed of the method becomes slower.

3. According to our experience the Kolmogorov-Smirnov test proved more effective than the χ^2 statistics for testing uniformity in this algorithm. The series-test outlined in Knuth's book [7] was found to be the strongest for testing interdependency.

   In both tests, both the algorithm in MATLAB© and the test developed by us achieved a result around 5% - so both algorithms have failed with the same ratio.

   The tests were predefined, therefore the modifications and examinations applied in our algorithm were driven and inspired by the results obtained in these tests. The process and reasons for the modifications are published in [5].

1.3 Publications

The original algorithm can be found in Vizvári's research report [12] or in Ramirez's book [11]. Our paper [5] mentioned above contains our results and the modifications.

2 Results in generalizations of the MDS method

The MDS is applied in many areas of science and it has been generalized for many applications. We have not found any similar generalizations published before.

The results described in our thesis are as follows:

- **(2.1. Thesis)** We adapted the multidimensional scaling for distances different from euclidian metrics with unorthodox methods. We attempted to find a completely different approach from Kruskal's robust method [10].

- **(2.2. Thesis)** We examined our method in different structures and with different distance definitions [1], [2]. The original MDS was a tool to provide critique for solutions for previous examples and was also used as an interpretation technique throughout the entire process of solving the original problem.

2.1 The results of the unorthodox method and the application

The generalization of the method is elaborated and examined for l_1 and l_∞ metrics on several predefined configurations and published in our paper [2]. As the coordinates of these points are known, the ability of the methods to restore the original points are easily analyzed.

This analysis is required by the study in which the distances between various objects were analyzed, filtering out the anomalies that occurred and trying to identify the cause of these anomalies.

According to our experience, it often occurs that when an application is made for a method, the data assigned to the application − for demonstrational or testing purposes − never appears in a real life situation.

Several problems which are known from the literature, have been found such that although it was claimed that the data described a real geometric configuration, the MDS was unable to reconstruct it.

The method has been applied on new areas with new, i.e. non-Euclidian, metrics. In these cases the original MDS method provides a less precise geometry of the objects. This made it necessary to modify the method to allow for the change of the distances which makes it necessary to cope with a more complex system with longer solution times.

2.2 Results in interpretation

The properties of the MDS for interpretation are known from several sources, and are applied at several places in the thesis, primarily when looking for objects with pre-defined distances.

Another utilization of the MDS is presented − published in another paper [3] − when a method for visualization and classification is presented, which is inspired by the application of the MDS in a psychological research.

The third area is the application developed in this thesis. The MDS is used for the visualization of the relation of the branches of the Employment Center; this way,the MDS helps to make the interpretation of the application's outputs easier for the users.

2.3 Publications

The classical interpretations and generalizations of the MDS are widely known and extensively published, and thus it is not feasible to provide a full comprehensive overview of the literature; instead, the thesis aims to collect those publications that are relevant to the solutions of the problems concerned by the thesis.

As to our best knowledge, there are no published results of the MDS in some of the areas visited by the thesis. These results have been published in a paper [1] and a research report [2]. As it was only known that the distances for which we utilized the method are non-Euclidian, so the generalizations are adequate.

The idea behind utilizing the MDS based on the distances of the objects inspired several other applications. For example, it provided a boost for a new classification application in psychology [3].

3 Application development

The results detailed in the first two chapters of the thesis serve as the basis for modelling a regional money distribution problem and to provide an optimal money distribution system for its users.

The thesis presents the hierarchical system which describes a typical government's money distribution system's characteristics. In this system several important structural parts are identified where a mathematical model provides an economically acceptable description, fixing the mathematical properties of the model. The injustice caused by the anomalies of the model could be moderated by involving additional resources.

The thesis provides reasonable strategies to moderate the injustice caused by the structural characteristics. The users of the applications are satisfied; the method is accepted and these strategies are utilized.

The following results are obtained:

- **(3.1. Thesis)** The structural background of the original problem has been revealed, and a mathematical model is described. The solution satisfies both the economic and political expectations.
- **(3.2. Thesis)** The thesis provides a stochastic model and a robust method. This robust method is used as a reference; it is compared against a simulation method and with the merge of these two a more precise robust method is achieved.

In what follows, the results of these methods are summarized with the exception of the numerical demonstration.

3.1 Mathematical analysis

The first description of the problem was an optimization model. This description did not contain several structural characteristics of the problem, therefore it had to be reformulated.

At this stage, it was already possible to prove that in this case − and in most similar hierarchical structured systems with the same (not too tight) attributes − the solution with the cheapest operational costs has peculiar characteristics.

The thesis shows that the structure of the problem implies that the state of the participants of the money distribution system will be either minimal or full load − and only a few exceptional members will be able to influence the decision making process. This means that only one branch of this hierarchical system will be in a not full load or minimal load state.

This characteristic of the system only came into focus in the deterministic case. However, there are more than one point in the system where there could be ground for uncertainty. The method balances this uncertainty by using stochastic methods.

This is the reason why the generation of random integer valued vectors becomes important from the point of view of the application. The implemented system is independent from the available computer resources of the user.

The involvement of stochastic methods caused some new side effects as new structural questions were revealed. These were successfully resolved in this specific case. Based on Prékopa's robustness-definition [9], new solutions were generated to handle most of the uncertainties, i.e. a solution that would be feasible for any realization of the random data. Using this approach, although the cheapest solution was identified among the robust solutions, it was unacceptable

for the political leadership, since it gave too much preference for some branches and tools of the system implying that other tools and branches became forbidden.

The result of the robust method has been used as a reference solution for the simulation, since it is the cheapest solution which is feasible without making the system unstable. It has also been used to create mixed solutions that provided significantly different alternative solutions from the original simulation's output.

The simulation provides a method which will generate a solution that is expected to be acceptable: the project has its restrictions based on the rate of return (to stay within a predefined range) that is politically acceptable, and it does not give a significantly more expensive solution than the robust one.

It is important to note, that the solutions given both by the robust and the simulation methods allocate less amount of money than available in the budget.

The solution provided by the robust method is the cheapest one among all known feasible solutions. Thus the convex combinations of any such alternative solutions also satisfy these attributes.

It is obvious that the solution obtained by averaging the solutions provided by the simulations might not be a feasible solution because some of the requirements might not be satisfied (in the current problem, the violated constraints usually have been the condition of a given unit's rate of return).

It is important to note that the whole project's rate of return is unaffected which means that the overall project will be a success. In a real world situation it would mean that one of the participant members performs below the required limit but it is balanced by another one which performs over the requirement limit.

3.2 Introducing the application, publications

The provided application would not be complete without the development of additional tools of the method to make it possible for the users to tune and repeatedly apply the method on other similar tasks. The thesis presents the application's programming environment and demonstrates the results of each processing step. The thesis also provides the specification of the implemented program, with the purpose to make the calculations reproducible based on the description and the operations. However, the source code remains unshared.

The developed application utilizes the pseudo-random vector generator procedure presented in the earlier chapters. Also, the MDS takes a major part in the interpretation and the visualization. Its representation is a 2 dimensional map where the members represented as points and their distances help in the comparison and evaluation. The main results are published in a paper [6].

4 Summary

Overall, the thesis presents a modified, novel pseudo-random integer valued vector generator, along with new testing methods and the adaptation of well-known, earlier test methods to the vector case.

The thesis identifies new areas of application of the MDS method.

Both procedures took part in the solution of a real world problem which was solved by implementing an application. Both the method and the application have been applied with success. Based on the user feedbacks the provided solutions satisfy the requirements and are applicable even with limited computing resources.

The evaluation criteria for the developed methods were determined in advance either by defining the goals or by the users. All of the modifications and improvements to the processes were approached with the goal of satisfying these criteria. Based on the empirical tests, the developed application is reliable, the results provided are acceptable and are suitable for a real world use.

References

5 Publications relevant to the thesis

[1] Niroomand, S.; **Takács Sz.**; Vizvári B.: *To lay out or not to lay out?* Annals of Operations Research, Vol. 191., 183−192, 2011.

[2] Niroomand, S.; **Takács Sz.**; Vizvári B.: *On the Generalization of the MDS Method*, Rutcor Research Report, 22., 2012.

[3] Surányi Zs.; Babocsay Á.; **Takács Sz.**; Vargha A.: *Új klasszifikációs módszerek a személyiségpszichológiában*, Pszichológia, Vol. 31(4)., 317−340, 2011.

[4] **Takács Sz.**: *Egy nem hagyományos statisztikai eljárás bemutatása az OECD PISA adatbázison - esettanulmány*, Alkalmazott Matematikai Lapok, Vol. 27., 157−174, 2010.

[5] **Takács Sz.**: *Eljárás jó minőségű véletlen egész vektorok generálására*, Alkalmazott Matematikai Lapok, Vol. 28., 41−58, 2011.

[6] **Takács Sz.**: *A hátrányos helyzetűek foglalkoztatását segítő projekt tervezésének támogatása szimulációs modellel*, Munkaügyi Szemle, Stúdium, 2012. szeptember.

6 Other references

[7] Knuth, D.: *A programozás művészete*, Vol. 2., Műszaki Könyvkiadó, Budapest, 51−87, 1987.

[8] Marsaglia, G.: *Diehard Battery of Tests of Randomness*, http://stat.fsu.edu/pub/diehard/, 1995.(letöltés: 2012. 08. 30.).

[9] Prékopa A.: *Stochastic Programming*, Kluwer, Dordrecht, 1995.

[10] Móri T.; Székely G.: Többváltozós statisztikai analízis, Műszaki Tankönyvkiadó, Budapest, 1986.

[11] Ramirez-Alfonsin, J. L.: *The Diophatine Frobenius Problem*, Oxford Lecture Series in Mathematics and its Applications, Vol. 30., Oxford University Press, 2005.

[12] Vizvári B.: *Generation of Uniformly Distributed Random Vectors of Good Quality*, Rutcor Research Report, 1994.