Some suggestions for graduate students and scholars undertaking quantitative interdisciplinary research: remarks from the practice

Dode Prenga
Department of Physics, Faculty of natural Sciences, University of Tirana
dode.prenga@fshn.edu.al

Safet Sula
Department of Physics, Faculty of natural Sciences, University of Tirana
safetsula@yahoo.com.

Abstract

Quantitative researches encompass large a field of works that embrace mathematics, engineering, physics, other natural science , economy and finance, quantitative sociology and so on. Concerning differences and benefiting from similarities is a state of art for researchers and it is more highlighted for young scholars working in interdisciplinary applications. Elements of classical and advanced statistics as seen from computing perspective, simulations, special and general techniques and models are the frontward of the start for a successful analysis. In this aspect there are many challenges for young scientist that must be addressed carefully. This became more imperative in the framework of applied informatics.

1. Introduction

Researches in natural science affront students and scholars with a permanent challenge, how to shorten the path from data to the appropriate results. In recent years many methods and techniques from natural science have been successfully used on other discipline as econometrics, sociology etc., giving rise to the interdisciplinary branches as econo-physics, socio physics and so on. Computing and quantitative analysis have been recommended as initial step of research in such fields by many guides or books as described in [Rob16]. Young researchers need to run-through the simulation techniques that might affront them with more complicated situations. Some of them mimic physical system as annealing simulation (cooling Monte Carlo) or use biological behavior to speed up procedures of numerical convergences. In this case, one needs more solicited knowledge on such natural science too. However, many calculation problems have been addressed through advanced and specified techniques that are available as discussed for example in [Suz13], [Jan12], [Ott17] etc. So, newly debuting researcher in interdisciplinary field necessitating quantitative analysis, computing techniques, algorithms or simulation procedures would probably find fine solution by carefully reviewing computational literature. However, the proper analysis and personalized view on concrete problem remain always in the heart of research work. In this case a good strategy could be based on avoiding inappropriate approaches too. It worth to capitulate some aspects of this process and below we will discuss it by commenting concrete situation encountered. Notice that in interdisciplinary studies, the quantitative approach usually starts by assuming a model which poses an additional question about it validity. In practice many of such aspects would be addressed and managed by the research team leader and would surely subject of detailed expertise, although it happen that if following a courageously a more independent path of the research, young scientist might run in inappropriate analysis, and problematic interpretation of the results. In this regard, in the case of interdisciplinary researches there are always space for better practice and strategies. We want to illustrate some such cases in the following.

2. Standardized models and software

Usually in the preparatory work for a concrete research, scholars try to apply a known model. It is a common advice from mentors and team leaders to use models that are proven to work on the analysis of the systems under study. Occasionally the level and routine of research drop down to the application and therefore the work would be further directed in some verification of the results; secondary estimation etc. This is common in spectral or time series analysis, investigating the reaction coefficients in a model, measurement or data elaboration and so on. Being those activities so common in data analysis steps, they are considered in dedicated software that attends standard models and various thematic issues. Likely, in physics or chemistry measurement, the instruments are accompanied with an interface and the software that perform directly the data analysis. Next, in econometrics, this is realized using more general tools, the standard statistical software as SPSS, EVIEW, SAS, LISREL, and ONYX. They offer adequate modeling and calculation capacities, as described in the appropriate web pages. Other software wants the user to be more active as is the case of R, PYTHON etc., and some others software introduce many mathematical and computational tools as MATLAB (or its LINUX counterpart, OCTAVE), MATHEMATICA etc. Notice that basic languages as C, C++, FORTRAN, BASIC, PASCAL etc., are plenipotentiary for every computation requirements but they need professional skills in programming and computation. Detailed remarks on how to use them are largely elaborated and easy reached in open sources. For this reason, in the first tentative, students are advised to apply the preprogramed ones models or to use functions form the software libraries. No need to spend time for something that others have perfectly done, but one has to know that they exist however. Nevertheless, it is crucial that when using dedicated softs and models, each assumption and every condition of the models should be completely fulfilled. Sometimes this is not rigorously possible. So, a good strategy for valuable quantitative analysis could be the building of algorithm from the researcher in a more interactive environment, aside of dedicated software application. In following we want to lists some precautions on such cases.

2.1 First stop on random numbers

Random numbers are a key element in calculation algorithms. The process of obtaining unaffected simulated systems outcomes is realized by the help of random numbers that drive the algorithm to the new unconditioned value. So we pick up a random value for the variable and calculate the modeled value. In other application the probability to select between alternatives is fixed by comparison a given number to a random one. It is clear that the quality or randomness for random numbers could be crucial for the unbiased (as desired) outcome. Also, it is of great interests in cryptographic security, where it is necessary to examine the real randomness of various "random" number generators. Next, the randomness of casted number is decisive for Monte Carlo simulations numerical integration. Some young researchers believe that the machine random number generators are quite accurate, but in reality this is not the case: it is difficult to get a computer to do something by chance. Remark that a computer follows instructions blindly and is therefore predictable. In practical simulations, the random numbers are taken from pseudo-random-number generators (PRNG) but in more sensitive application randomness is based on a very un-predictable physical event as radioactivity or atmospheric noise. Up here we admit that testing randomness is a crucial advice for a good start. In literature there many arguments in testing randomness as Dieharder test or other recommended alternatives [Wan03]

a. *Frequency Test: Monobit*
b. *Frequency Test: Block*
c. *Runs Test*
d. *Test for the Longest Runs of Ones in a Block*
e. *Binary Matrix Rank Test*
f. *Discrete Fourier Transform (Spectral Test)*
g. *Non-Overlapping Template Matching Test*
h. *Overlapping Template Matching Test*
i. *Maurer's Universal Statistical Test*
j. *Linear Complexity Test*
k. *Serial Test*
l. *Approximate Entropy Test*
m. *Cumulative Sums Test*
n. *Random Excursions Test*

Each type of tests a-n above and others not included herein can be implemented in specific subroutines, but comparison between generated random arrays which have been confirmed by tests, seems to be not an easy tasks. Moreover it needs detailed knowledge on each. To improve the above mentioned calculation, we should realize ourselves a better PRN generator. In this case we must apply step by step testing to fix the better generation. To visualize the un-randomness for PRN generated in computers let's start from the evident fact that in generating random normally distributed numbers we would expect that the outcome should be normally distributed. We can test directly for the Gaussianty as suggested in many textbooks of statistics using the kurtosis

$$K(x) = \frac{E(x-\mu)^4}{\sigma^4} - 3 \quad (1)$$

Relation (1) is easy to apply but has many complementary assumptions that are difficult to be tested. Therefore we have applied another idea by direct measuring the distance of the distribution under analysis from normal distribution using q-functions introduced in [Uma10]

$$p(x) = \frac{1}{Z}[(1 + \frac{(1-q)}{5-3q}\left(\frac{x-\mu}{\sigma}\right)^2]^{\frac{1}{1-q}} \quad (2)$$

Equation (2) reproduce the Gaussian for q=1 so the difference q-1 estimate directly the distance from normal distribution. In Fig 1 we show the fit of the distribution of machine PRN. As routinely practices by physicist, we use log-log presentation which highlights the differences in the extremities of the graphs. The deviance of the generated number's distribution from normal ones in the large values limit is easy noticeable by naked eye. By using (2) for an array of 10^6 generated normally distributed random numbers we obtained q~1.020 in the generation using *for ...randn() end* loop in MATLAB; q~1.017 in the array generation using *normrnd()* command whereas by a simple BoxMuller algorithm using *rand()* as starting points, we had q~1.015.

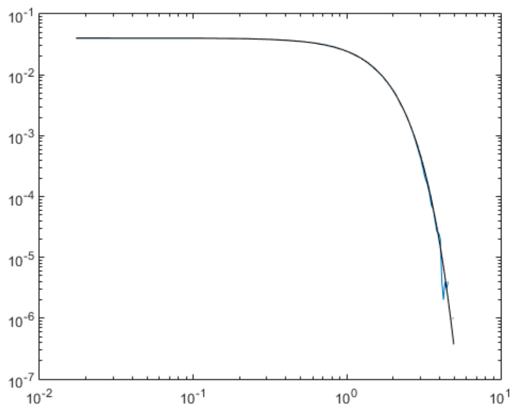


Figure 1: Log-log plot of normal random number

Based on the arguments of [Tsa09] or [Uma10], the distribution for numbers generated by the last algorithm is more Gaussian. By nature we cannot measure the randomness directly, but judging from the resulting distribution, the random numbers produced in the second is expected to be better. To this end, we suggest to the young researcher to construct themselves random number generators and hence they would always have a profit from *the machine ability to produce in it own the PRN and method perfection to generate PRN sequences.* Next they'd better do

- *test the randomness before application*
- *pre-calculate the overall effect of non-randomness*

2.1 Avoiding distribution's assumption misuse

Statistical analysis is so common in interdisciplinary modeling and fitting procedures. So, it happens that the assumed theoretical distribution is accepted without proof as describing the system or process under study. Or the assumption of the normally distributed deviances in the fitting process was not put in doubt too. However, under some specific circumstances, there are sufficient arguments that the final error induced by the violation of normality assumption is not determinant [Gen72]. Other view as in [Hu13] suggest to the researchers to go deeper in error analysis. In practice other common assumption are the homogeneity paradigm; time-invariant processes and so on. Here one needs a careful evidence for distributions and other herein mentioned assumption which in turn result in a quite an easy task, but the benefit could be remarkable.

2.1.1 Some worked example for real systems

Intriguingly the intuitive assumption that distribution for values of variables arising from a long time natural process would be a lognormal, has remained in the basis of many regulation and predictions. So, the famous Black-Shoe derivation for the distribution of the return of prices namely $r = \dfrac{p_t - p_{t-1}}{p_{t-1}}$ has been found un-applicable even being very attractive in its first appearance. It is suggested that in this case the distribution could be q-Gaussian of the form (2) [Bor04]. Following this idea, a lognormal analogue of q-Gaussian (2) has been verified with good statistical significance even for exchange rate of ALL as we represented in [Pre14]. In another such analysis presented in [Sul16] we showed that the probability that an extreme flood in Drini cascade calculated using lognormal distribution is as 8 times smaller than the one calculated from the empiric fit distribution obtained using 20 year daily side floods as registered. Practically the discharges from the lakes as response of near to extreme raining had occurred so frequently last years that coincides to the calculation of expected occurrence of one time over more than 100 years. In many other real systems we observed that the best fitted functions are in the parametric form like (2) or its lognormal q-counterpart

$$p_q(x) \sim \frac{1}{x^q}\left(1-\beta(1-q)\left(\frac{x^{1-q}-1}{1-q}\right)-\mu\right)^{2}\Bigg)^{\frac{1}{1-q}} \quad (3)$$

usually fits better than expected functions say Gaussian, lognormal, Weibull etc.

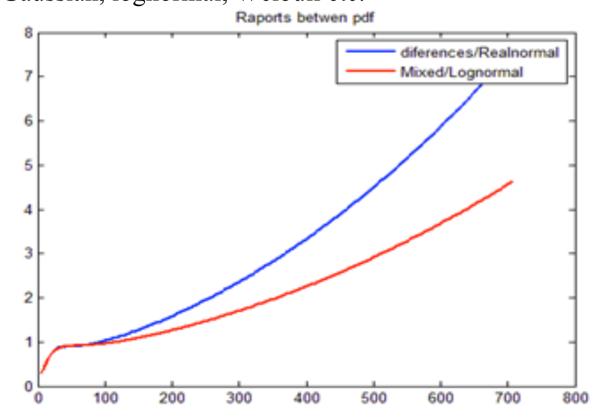


Figure 2: Illustration of differences between standard and alternative distributions approach

So, un-proofed assumption that the distribution on the data would be Gaussian or lognormal or Weibull etc. should be avoided in applications until a test would confirmed it. Otherwise it could happen than oversimplification of the systems or tendencies to confirm generalized expectation would leads to the following conclusions seen in a paper recently. It is not surprise if a real erroneous use of normal distribution paradigm would produce the result of Fig.3

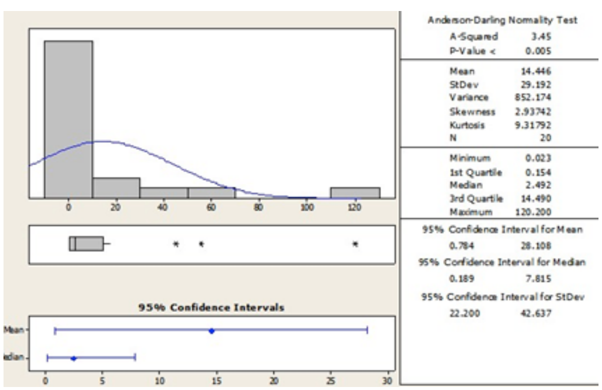


Figure 3: Misuse of standard distribution.

To this end, we highlight the logic step of distribution analysis to test them starting form un-stationary ones which are most likely to be found in real systems.

2.1.2 Measurement and data analysis assumptions

Another inadequacy in the data elaboration stage could be the assumption that the distribution is stable. This is worse in the case of real systems with limited number of points and characteristic heterogeneity. We specifically mention here

- Data gathered from measurement process in engineering, natural sciences researches etc.
- Data gathered via inquires in social end economical sciences

We observed that in some more detailed analysis the un-verified distributions assumption leads to speculative conclusions or even in wrong measurement practice. Scholars report the level of contamination in an area without offering supporting arguments for stationary of the state where measurements have been performed. It seems that mathematically is taken

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \leftrightarrow E(x) \equiv \int_{x-\sup port} x\rho(x)dx \quad (4)$$

and thus the mean is the best representative of variable x in its population. Notice that the right side of (4) exists only if dhe probability density function (the distribution) of variable x is finite that is the case of stationary distribution. If not, value E does not exist at all, so we cannot perform any statistical report on the measurement. In (4) the variable x could be the direct value measured or an output parameter as error in regression procedures. Hence, in those cases the verification stationary for the distribution ρ(x) is compulsory. Otherwise, the mean could be refereed as the *best value of the sample measured, but not representative for the population*. Mathematically the stability for distribution would be measured by parameter α-Levy but in calculation procedures it requires the fit for a complicated t-Student to the empirical data. Instead on can suggested an easy way-out from this situation making use of relations (2) above and testing parameter q. It is related to the α-Levy and there is e simple relationship with degree of freedom in the T-student by the rule $\nu = \frac{3-q}{q-1}$.

Fortunately, from the computing point of view, the form (2) can be fitted easy with standard nonlinear fitting algorithm, whereas T-student is more complicated. Next one can perform the evaluation of the stability for the distribution under analysis by simply using the condition of variance finiteness making use of the formula $\sigma = \frac{1}{(5-3q)\beta}$ calculated in [Uma10]. Stability requirement is $1 \le q \le \frac{5}{3}$ but a broader rule say $1 \le q \le 2$ has been suggested therein. Moreover, if q>3 there is no distribution at in statistical sense. In this case the relation (4) became meaningless hence the arithmetical average value has to be declared as the mean of the data from the measurement and never should be confounded with population's mean which does not exist.

2.1.3 Bin optimization procedures

Finding the appropriate distribution should not be considered as a trivial task. Usually the regressions are too easy in the first sight. But here is another point to step in. The trick entails the way we approach the underlying distribution for given data frequencies. So in practice, the set of data series is ordered in J categories or classes that (as a rule) are of equal size

$$J = \frac{x_{\max} - x_{\min}}{h} \qquad (5)$$
$$d(j) = n(x_i) \in \left[x_{\min} + (j-1)h, x_{\min} + jh,\right]$$

The process (5) is called histogram or discretization of the data distribution. But a (hidden) question remains mostly unanswered and unreported as well: how is chosen the parameter h in (5)? Mathematically speaking, the underlying (natural) distribution d should not be affected from the binning procedure (5) and in analytic view one request that moments of variable x have not to be affected. So far, this has been considered straightforwardly and optimization rules have been included in software or programs, but again, there exist cases that those steps have not been performed. A detailed analysis on methods and techniques for histogram optimization is provided in [Shi10]. Correct binning step should use Stokes rules or Friedman-Diaconincs formula

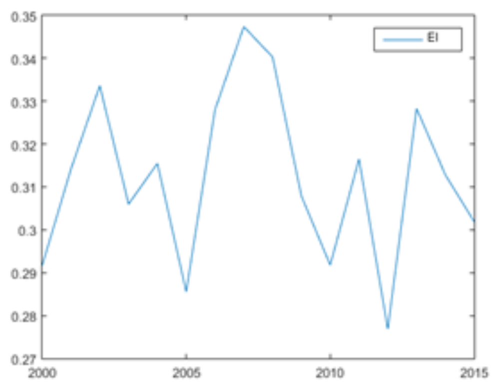
$$h \sim (3.49 \div 3.73)\sigma N^{-\frac{1}{3}} \qquad (6)$$

where σ is standard deviation and N is the total number of values in the data set. However in (6) it is assumed that deviations from the real data were normally distributed which should be analyzed as we discussed in the preceding paragraph. We have noticed that in practice, neglecting (6) unfortunately is not an isolated error and in some cases young researchers have no idea about it importance. To complicate things, related to relations (6) some programs offered themselves a bin number (usually 20) or clearly ask to the user to input the bin number. Statistical softs applies directly (6) or similar formula without signaling us, and so avoiding the subjective bin-size. But again, (6) is valid if deviances are normally distribution that might not be true. Thereof, a very good suggestion for correctness in data analysis is the optimization of the bin size.

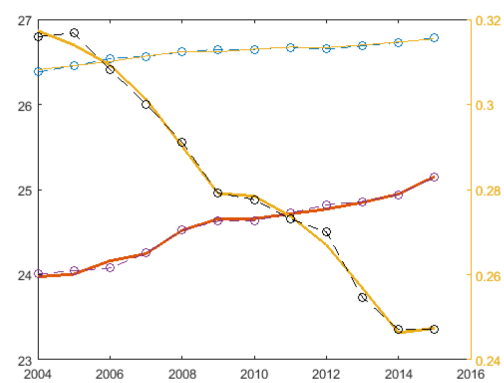
2.2 More Flexible when working with models and conditions of validity

Using well known models is a good practice but in this case pre-programmed ones could mislead to wrong interpretation. Many dedicated softs as SPSS, SAS, EVIEW in statistical analysis or LISREL, ONYX etc., in structural equation studies, offers various solutions for econometric, socio-dynamic problems and related subjects. Their routine includes many preparatory steps and assumptions (again some of them need to be tested separately by the user). In this case a good advice is to build algorithms ourselves. Here is an example what can happen. In the calculation of the informal economy as a hidden variable, we had in disposal a small portion of data, only 18 series (years 1998-2016). The model known as MIMIC (multi cause, multi indicators) adopted by EVIEW or LISREL have been used by other researchers consequently and widely recommended in such calculation. But specifically those programs request a sufficient number of data series for statistical analysis (at least above 50 points in our knowledge). Second they apply directly the unit roots removing procedures. Next the result obtained as output needs further elaboration. If one tries to program the routine by ourselves, a detailed description is provided in [Jor75]. In our example, we observed that the result obtained using deferent methods does not match. This was the result of not fulfillment of presumed assumption by our data set. In particular the use of differences to remove unit roots as recommended, from the other side has reduced significantly the data series from 14 to 12, and for some variables included in model the stationary has not been verified! However the very small number of points led to high uncertainty on statistical test. To overcome the problem we preferred the calculation using our routine that performed those additional steps.

a. *analysis of tiny data (monthly records) by which the dynamics of the quantities has been identified (in an high level); especially there have been tow regimes in the interval considered so we used data that belongs to the same regime for the fit*
b. *accounting for those two effects fitting has been accepted for lower confidence level*
c. *number of factor variables , responses and latent ones have been calculated using factor analysis*



a. Informal Economy by MIMIC 8-1-3 model：



b. Reproduction of the indicators: yellow line, unemployment rate, blue line, ln(GDP); red line logarithm of narrow money,

Figure 4: Example of easy step by step analysis (case study: Informal Economy estimated by MIMIC model)

So we write the algorithm in MATLAB as direct application of the model elaborated in [Jor72], [Gol64] including preparatory steps (a-c). The results we obtained using deferent approach (currency approach and MIMIC model in the concrete work) matched much better. Moreover the reproduced variables fit very well with original ones confirming the goodness of the calculation in this case as seen in the Fig 4. In another calculation related to the consumer behavior we observed inadequate outcomes when using variables directly as from the measured. Calculation was performed basing on standard logistic model used in econometrics and generally in the models involving categorical variables [Kus18]. Again it is preferred to construct the program considering specifics of the system and its variables using the same idea as above.

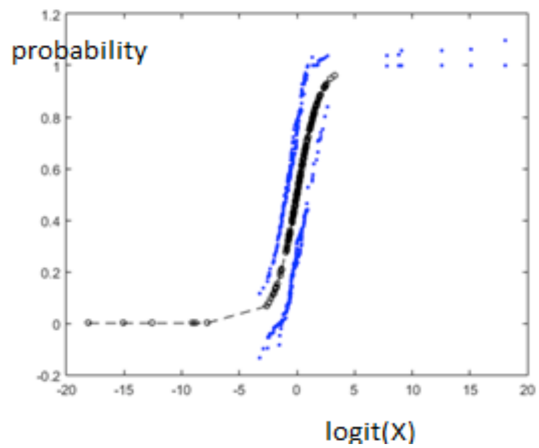


Figure 4: Another example of normalizing models: Fitting logistic model in consumer behavior

So, the same result has been reported using the logistics and probit approach, that signify an improvement of the calculation. In general we can underline and highlight the importance of carefulness with models and especially

a. *Detailed verification of all dedicated softs assumption. Avoiding any non-logical operation on the data series*
b. *Constructing single-purpose algorithm instead of using multi-purpose pre-programed ones*
c. *Going deep in the mathematic of the problem before applying retunes*
d. *Analyzing the overall state of the system*

2.3 Trying Calculation challenges

Some non-linear function or equations cause headache to the practitioners. Let consider for example the problem of fitting parametrical functions like the ones including non-homogenous unit if variable as

$$y \sim y_0 + (x - x_c)^m \left[A + B\cos(\log\omega(x - x_c) + \varphi)... \right] \quad (7)$$

The form (7) is verified as underlying bobble dynamics in financial asset or indexes, failures, explosions etc. [Sor01]. Regressions including nonlinear ones do not work in this case. Taboo search is not reported as effective too [Sor01]. Moreover, the deviation is an Uhlenberg process that cannot be tested as we do for chi–deviances; hence the statistics for a fit is not available by standard procedures. To deal with numerical analysis of near to characteristic behavior we used recently [Pre16] a more complicated form of (7) by extending relations (5)

$$P(t) = a + b(t - t_c)^m + c(t - t_c)^m \cos(\omega \left[\frac{(t - t_c)^{1-q} - 1}{1 - q} \right]^{1-q} + \phi) \quad (8)$$
$$+ d(t - t_c)^m \cos(2\omega \left[\frac{(t - t_c)^{1-q} - 1}{1 - q} \right]^{1-q} + \phi)...$$

To solve those problems it is suggested a genetic algorithm model which is detailed in [Sor01]. It is based on two step calculation or 'slaving parameters". We write an ad hoc such a routine and the fit has been found very accurately in the case of the dynamics of exchange rates [Pre16], [Pre14] or anxious-like behavior in the water level during intensive floods in Komani Lake [Sul16]. Genetic algorithm is found successful for many such fitting difficulties. For interest of the readers we motioned that genetic algorithm mimics the Darwinian evolution. So in the core of the program, one impose by a given probability a mutation in the solution vectors $v = [m, x_c, \omega, \varphi]$, and if the result is not good, one changes the distribution of the random numbers used to impose the mutation. We realized that by using beta distribution to produce random numbers, the convergence of our ad-hoc algorithm has been realized even for more complicated forms of (7) resulting in (8) which we called near to characteristic behavior in [Pre16]. Similarly, the taboo searches can work for other situation especially where the possibility of returning in the old solution is permanent. In such cases it is very important for the researcher to explore many specific techniques and trying again to challenge the problem by madding up routines.

3. Non-neglecting Calculation and Simulation Performance

Advanced studies include simulation and hard calculation even in the graduate level. Students can try directly in open sources as Wolfram Alpha to calculate difficult integrals or they can use MATHEMATICA, MATLAB services etc. In numerical calculus including integration many method exists and with little effort nearly all problems for not advanced studies could be answered using each of them. But choosing the appropriate method or algorithm might result in consuming time and energy for students. Clearly there exist no general receipt in these cases and it is just the duty of the research leading, but again some advices could help. For many purposes the two above mentioned software (and surely many others) are really mines with opportunities. Just needs to explore them. But again statistical and mathematical tools are indispensable. Here are some considerations from a recent work.

3.1.1 More effort in analytic relations

Analytic solutions are always the most desired outcome in the study of systems. Let mention here a simple physical system containing two vectors (magnets). Later on it is proposed to model opinion formation in a pair of individuals. Statistical mechanics calculation start with partition function that in this case reads

$$Z = \int_\Gamma \exp\left(-\frac{H}{kT}\right) d\Gamma \qquad (9)$$

Where

$$H = -\sum_{i,j} J m_i m_j + \sum_j \mu B m_i \quad (10)$$

the Hamiltonian, m is magnet vectors and B is the magnetic inductions. Here $m^2=1$. Physical quantities in principle will be calculated using appropriate formula of physics, once the partition function Z is evaluated in analytic form. Calculation of (9) having H given by (10) a genuine trick proposed in [Cif99] just to replace $m_1 m_2 = (m_1 + m_2)^2 - 2$ for 2-continus spin magnet system rends (10) the form

$$H = -\frac{J}{2}\left(M^2 - 2\right) - BM\cos(B, M) \quad (11)$$

that turns calculation (9) to be in analytic form! In statistical physics analytic forms of Z are the most “wanted” cases! Here $M=m_1+m_2$ is the sum of tow vectors. All calculation has been performed in [Cif16]. So, in a case-application in socio-dynamics, and practically for calculating of the opinion using an ad-hoc model, we used a more complicated inter-coupled Hamiltonian in the form proposed in [Pre18]

$$U = J - \frac{O^2}{2} - FO\cos\left(F, O\right)\left[1 - \alpha\frac{J}{2}\left(O^2 - 2\right)\right] \quad (12)$$

where $O=O_1+O_2$ is the resulting vector of opinion and U is the utility function using terms proposed in [Sta09]. Here making use of properties of Bessel function, some adding integrals realized in [Cif16], one realized to find analytic form of the Z integral and following statistical mechanics formula finally we obtained the average opinion per individuals as following

$$\left\langle O_x \right\rangle = \frac{1}{2}\frac{\displaystyle\int_0^2 \exp\left(\frac{\beta J O^2}{2}\right)\frac{I_1\left(\beta F A(O)\right)}{\sqrt{4 - O^2}}A(O)dO}{\displaystyle\int_0^2 \exp\left(\frac{\beta J O^2}{2}\right)\frac{I_0\left(\beta F A(O)\right)}{\sqrt{4 - O^2}}dO} \quad (13)$$

where $A(O) = O\left(1 - \frac{\alpha J}{2}\left[O^2 - 2\right]\right)$.

Next we proceeded with numeric integration of (13) concerning in the zeros and infinite values. For the interest of the reader, we mention that the MATLAB offer adding facilities when dealing with integrands so (13) have been calculated numerically and the result is represented in the Fig.7. This problem solved by using Matlab and some adding knowledge about functions involved in there, is a good argument for suggesting crossing of the methods and techniques. We observe that without mathematical the trick offered in [Cif99] analytic forms weren’t impossible and so the following calculation in (14). However, exploring about solutions of specific problems even very difficult could result successful because there is always someone that can solve easy our problem.

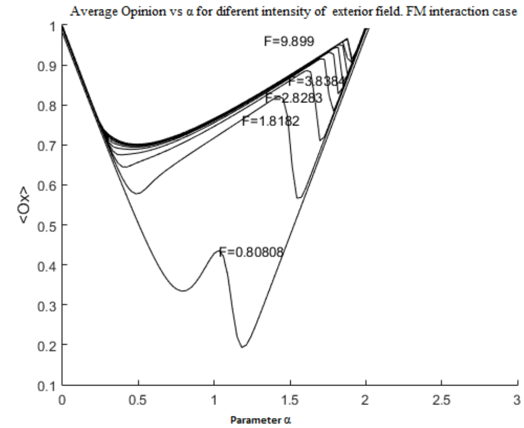


Figure 5: Illustration of the calculation of opinion (13) performed by directly using Matlab routines.

3.1.2 Actual softs offer near everything: symbolic operations might help significantly

A full analysis of the system having utility (12)

Similar calculation could strain the researcher because the calculation of Hessian needs differencing (20) and analyzing the behavior of parametric equation, studying the logic solution, imposing constraints etc. Fortunately this is not a case: by using symbolic equation and differencing in MATLAB (MATHEMATICA etc.) we easily identified fixed points, null clines and everything from nonlinear dynamics analysis of the system. So if we try to obtain the solution of

$$\left(O_c, \phi_c\right)_i = Arg\begin{cases} \dfrac{\partial}{\partial\varphi}\left[-\dfrac{J}{2}\left(O^2 - 2\right) - FO\cos\phi\left[1 - \alpha\dfrac{J}{2}\left(O^2 - 2\right)\right]\right] = 0; \\ \dfrac{\partial}{\partial O}\left[-\dfrac{J}{2}\left(O^2 - 2\right) - FO\cos\phi\left[1 - \alpha\dfrac{J}{2}\left(O^2 - 2\right)\right]\right] = 0; \\ 0 \le \varphi_c \le \pi \\ 0 \le O_c \le 2 \end{cases}$$

which give null clines and in the analysis of second order derivatives involved in the Hessian, we observe that traditional effort are very likely to fail. Moreover, symbolic operation in this case facilitate remarkably the analysis by giving the opportunity of solving complicated systems including inequalities, plotting complicated graphs. In the Fig.8 is shows such a step on searching for stationary state for the system (12) at zero degree temperature.

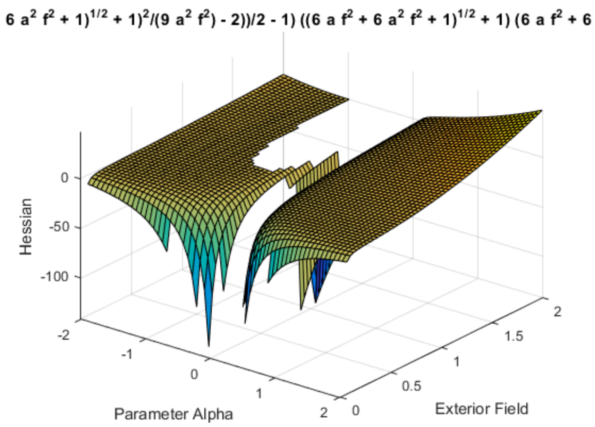


Figure 6: Qualitative Analysis of stationary state

This example suggest that a better knowledge about particular programs would be a very helpful when dealing with complicated algebra in calculation.

3.1.3 Exploration on simulation platforms

In many applications, the first idea coming in mind could be speeding up the study, so practically one start with general algorithm and easiest ones. Not surprisingly this can lead the research on some valley of the solution, making every effort to amend properly the algorithm, useless. As routinely used in numerical simulation, Monte Carlo technique is the broadest method used. In those similar cases it very important to explore as many as possible algorithms and methods. Typically algorithm might slow down or might never converge due to the number of states around particular point in the solution space. We will explain in short this idea by just evoking the calculation of the average opinion of system (12). According to the literature suggestions, we used the WOLF algorithm. The core algorithm has the following steps:

1. *one start from a random configuration of magnets assimilated in the angles between a vector and exterior field (φ)*
2. *pick a magnet (i) and calculate the energy of the cell involving all surrounding magnets*
3. *randomly select a direction θ, and turn all spins upward to this direction*
4. *Calculate the energy in new configuration, if it is smaller, the move is accepted, else it is accepted with Metropolis probability.*
5. *Stop if no more improvement could be done*

Basically this algorithm is fruitful for complexes calculation and it worked for some simplified case of equation (12). Other alternatives are available too. But if one use the simplified Metropolis-Hasting method we observe a non-sufficient convergence for simples XY2D model. Notice that new researchers want to follow the simplified MH procedure (11) instead of taking care of full detailed balance assumption. In this case a good advice is to measure directly the acceptance ratio. In many Monte Carlo algorithm would have acceptance ratio around 0.5 or lower, but it is not a receipt however. The suggestion on those cases is to explore patiently on possible specific algorithms rather using general algorithm. In our example in first tentative we had an acceptation ratio as high as 0.8, and by using right formula on probability detailed balance this ratio was decreased to 0.5. Later on a modified version called MALA (Metropolis-Adjusted Langeven Algorithm) as detailed in [Jan12], [Suz13] etc. However the solution of the problem was not finally concluded until we used the Wolf algorithm. Surely this could be a common circumstance for many students or new researcher therefore we insist in the suggestion of being real careful in implementation of every specific of quantitative methods. This is very useful for such researcher dealing in the interdisciplinary studies and especially for them that have in their basic background do not have a solid mathematical programming formation.

4. Conclusions

Successful quantitative studies needs for important and individual efforts in informatics, programing, applied mathematics and computation techniques. In the data analysis researches, graduate students and new scientist must pay much effort for a prior deep knowledge of the system, its characteristics and the nature of the state where the measurements have been made. In general, using preprogramed algorithm or programs would not be the first choice and the benefit of building algorithm themselves could be apparent for new researchers. Investing in a deeper mathematical model analysis would be a very good start in the case of young researchers with solid natural science background. New scientist dealing with interdisciplinary studies would have better results if exploring patiently on the possibilities of modern engineering programs, including forums as well

References

[Rob16] M.A. Robinson. Quantitative research principles and methods for human-focused research in engineering design. '*Research methods' publications*. May 2016. DOI: 10.1007/978-3-319-33781-4_3

[Wan03] Ying Wang. Nonparametric Tests for Randomness. *Ece 461 Project Report*, MAY 2003

[Tsa09] Constantino Tsallis .Computational applications of non-extensive statistical mechanics. *Journal of Computational and Applied Mathematics 227* (2009) p 51-58.

[Uma10] Sabir Umarov, Constantino Tsallis, Murray Gell-Mann, Stanly Steinberg. Generalization of symmetric -stable Lévy distributions for q>1. *Journal of mathematical physics* 51, 033502 2010

[Gen72] Gene V Glass Percy D. Peckham, James R. Sanders Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance . *Review of education research vol. 42, no. 3.* 1972

[Hu13] Shu-Ping Hu PRT-152 Fit, Rather Than Assume, a CER Error Distribution. *ICEAA Annual Conference and Training Workshop New Orleans, LA*.2013

[Bor04] Liza Borland. A Theory of Non-Gaussian option Pricing: capturing the smile and the skew (2004) www. *Spiedl. org/data/ conference s/SPIEP*

[Pre14] Dode Prenga. Dinamika dhe vetorganizimi në disa sisteme komplekse. *Monografi*. Pegi, 2014

[Sul16] Safet Sula, Dode Prenga. Improving the Analysis of Hydrologic Time Data Series in our basins and economic impact of hydro-industries. *ISTI Faculty of Economy, University of Tirana*. 2016.

[Pre16] D. Prenga, M. Ifti. Complexity Methods Used in the Study of Some Real Systems with Weak Characteristic Properties. *AIP Conf. Proc. 1722,* 080006 (2016)

[Sor01] D. Sornette, A. Johansen."Significance of Log-periodic Precursors to Financial Crashes," *Quantitative Finance, 1: 452*. 2001

[Shi10] Hideaki Shimazaki · Shigeru Shinomoto . Kernel bandwidth optimization in spike rate estimation. *J Comput Neurosci* (2010) 29:171–182. DOI 10.1007/s10827-009-0180-4

[Cif16] Orion Ciftja, Dode Prenga. Magnetic properties of a classical XY spin dimer in a "planar" magnetic field. *Journal of Magnetism and Magnetic Materials 416* (2016) 220-225.

[Cif99] Orion Ciftja and Marshall Luban. Equation of state and spin-correlation functions of ultra-small classical Heisenberg magnets. P*hysical Review B Volume 60, Number 14 1* October 1999-II.

[Sta12] Dietrich Stauffer. A Biased Review of Sociophysics. *Journal of Statistical Physics. 151*. 10.1007/s10955-012-0604-9.

[Gol64] Goldberger, Arthur S. (1964). *Econometric Theory*. New York: John Wiley & Sons. Pp. 238–243

[Jor75] Kalr Jorskog, Arthur Goldbwerg. Estimation of a model with multiple indicator and multiple causes of ingle latent variable. *Journal of the American statistical association. Volume 70, issue 351*(Sep. 1975)

[Kus18] Elmira Kushta, Dode Prenga, Fatmir Memaj. Analysis of consumer behavior in a small size market unit: case study for Vlora District, Albania. *IJSRM*,2018

[Ott17] Michela Ottobre, Natesh S. Pillai, Konstantinos Spiliopoulosar. Optimal Scaling of the MALA algorithm with Irreversible Proposals for Gaussian targets. 2017 aXiv:1702.01777 [stat.ME]

[Jan12] Wolfhard Janke Monte Carlo Simulations of Spin *Systems Institut f¨ur Physik, Johannes Gutenberg-Universit¨at, D-55099 Mainz, Germany*. https://pdfs.semanticscholar.org

[Suz13] Hideyuki Suzuki. Monte Carlo simulation of classical spin models with chaotic billiards. *Physical Review E 88, 052144* (2013)

[Pre18] Unpublished. Possible use of XY model in calculations of early stage of opinion formation. *Submitted in the Journal of Modern Physics B*. 2018