Applications of Multidimensional Time Model for Probability Cumulative Function for parameter and risk reduction.

Joint Statistical Meeting, Chicago, IL.

Michael Fundator, Board on BCSS of National Academy of Sciences.

Abstract.

Multidimensional Time Model for Probability Cumulative Function can be reduced to finite-dimensional time model, which can be characterized by Boolean algebra for operations over events and their probabilities and index set for reduction of infinite dimensional time model to finite number of dimensions of time model considering the fractal-dimensional time that is arising from alike supersymmetrical properties of probability, This can lead to various applications for parameter evaluation and risk reduction in such big complex data structures as complex dependence structures, images, networks, and graphs, missing and sparse data, such as various DNA analyses.

Keywords: finite-dimensional time model, normally distributed, Chebyshev-Hermite polynomials

"Let Tp denote a nonparametric multivariate test in p dimensions. Let $\pi_{p\to q}$ denote a method of reducing the dimensionality to q ,such as by using the first q principal components, or by multidimensional scaling, or by "lining" or "planning". The q-dimensional form of the multivariate test might stil be applicable to the resulting q-dimensional data… test Tq($\pi_{p\to q}$)…conjecture that, when p-dimensional data is sparse, Tq($\pi_{p\to q}$) will often be in some sense a better test than Tp… ***this simple suggestion should not be confused with the method of "projecting" down to one dimension (lining), and then using a univariate test… it is best to think of q as at least 2. The suggestion is so simple that it has probably been suggested several times, but I'd like to know of a reference.("***The Good Book" v.2 C356. Multivariate Tests after dimensionality reduction p. 277)

1.Introduction.

It seems that the brief introduction of the historical sketch of the development of the Theory of Brownian Motion and some historical problems that very often overlooked of would be very helpful to make understanding of this work more clear:

1 One such problem as to decide if a number is a special value of a well known function was solved by Gauss, when he made an observation, that 1.85407 457 is a rational value of an elliptic integral, and this certainly was a great contribution to the development of analysis.

2 Probability theory is the quantitative framework for scientific inference. It codifies how observations (data) combine with modeling assumptions (prior distributions and likelihood functions) to give evidence for or against a hypothesis and values of unknown quantities. There is continuing debate about how prior assumptions can be chosen and validated, but the role of probability as the language of uncertainty is rarely questioned. That is, as long as the subject of inference is a physical variable. What if the quantity in question is a mathematical
statement, the solution to a computational task? Does it make sense to assign a probability measure p(x) over the solution of a linear system of equations Ax = b if A and b are known? If so, what is the meaning of p(x), and can it be identified with a notion of 'uncertainty'?
If one sees the use of probability in statistics as a way to remove "noise" from "signal", it seems misguided to apply it to a deterministic mathematical problem. But noise and stochasticity are themselves difficult to define precisely. Probability theory does not rest on the notion of randomness (aleatory uncertainty), but extends to quantifying epistemic uncertainty, arising solely from missing information.

3 Connections between deterministic computations and probabilities have a long history. Erdös and Kac showed that the number of distinct prime factors in an integer follows a normal distribution. Their statement is precise, and useful for the analysis of factorization algorithms, even though it is difficult to "sample" from the integers. It is meaningful without appealing to the concept of epistemic uncertainty.

4 However, besides their approach we can also emphasize an example by A.Y. Khinchin, that the direct consequence of the Theorem of Lagrange, which says, that every natural number can be expressed as the sum of at most four squares, is that, if four sequences
$0, 1^2, 2^2, 3^2, \ldots, k^2, \ldots..$
are added together, the resulting sequence contain all natural numbers, it was his main introductory point for Waring's Conjecture.Probabilistic and deterministic methods for inference on physical quantities have shared dualities from very early on: Legendre introduced the method of least squares in 1805 as a deterministic best fit for data without a probabilistic interpretation. Gauss' 1809 probabilistic formulation of the exact same method added a generative stochastic model for how the data might be assumed to have arisen. Legendre's least-squares is a useful method without the generative interpretation, but the Gaussian formulation adds the important notion of uncertainty (also interpretable as model capacity) that would later become crucial in areas like the study of dynamical systems.
"It has been customary certainly to regard as an axiom the hypothesis that if any quantity has been determined by several direct observations, made under the same circumstances and with equal care, the arithmetical mean of the observed values affords the most probable value, if not rigorously, yet very nearly at least, so that it is always most safe to adhere to it." — Gauss (1809, section 177) *Theoria motvs corporvm coelestivm in sectionibvs conicis Solem ambientivm [Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections]*
5 Another open problem is a question of normality, if the decimal expansions of π, e, $\sqrt{2}$, $\sqrt{10}$, and many other mathematical constants such as γ, or log2 all have the property that the limiting frequency of any digit is one tenth, and the limiting frequency of any *n*-long string of decimal digits is 10^{-n} (and similarly for binary expansions).

There are following examples for decimal expansion of π:

The sequence 0123456789 occurs beginning at digits 17 387 594 880, 26 852 899 245, 30 243 957 439, 34 549 153 953, 41 952 536 161, and 43 289 964 000

The sequence 9876543210 occurs beginning at digits 21 981 157 633, 29 832 636 867, 39 232 573 648, 42 140 457 481, and 43 065 796 214

The sequence 27182818284 (the first few digits of e) occurs beginning at digit 45 111 908 393.

J. Havil in his book ``The Irrationals'' gives this impressive example, that sequence 0123456789 appears for the first time starting at the 17,387,594,880th digit; whereas 0691143420 continues to prove elusive. ``His book has ISBN 978-069114342-2 and was published 2 years ago by Princeton University Press. The website gives the record for this year as 10 trillion, with current of 13.3 trillion.

This problem is very much related to such open problems in the Theory of Brownian Motion as if all Brownian paths are possible, and the like problems of transience and recurrence of Random walk in 2 or more dimensions.

It clearly seen from the above examples, that the decimal expansion of π does not follow strictly Gaussian distribution, but instead is supposed to follow some kind of slightly non-Gaussian distribution.

2. Insights into near-Gaussian distributions.

Before considering any discussion about the above phenomenon or any possible approach to analyze or investigate it, it seems appropriate to quote Karl Pearson*, who wrote 110 years ago on p. 189 "*My custom of terming the curve the Gauss–Laplacian or normal curve saves us from proportioning the merit of discovery between the two great astronomer mathematicians." One of the definitions of Peirce of "normal" as of what would, in the long run, occur under certain circumstances, clearly implies Principle of prediction and LLN.

"It is undeniable that, in a large number of important applications, we meet distributions which are at least approximately normal. Such is the case, e.g., with the distributions of errors of physical and astronomical measurements, a great number of demographical and biological distributions, etc." Cramer.

The first investigation of slightly non-Gaussian distributions was undertaken by Chebyshev around a century and a half ago, who studied in detail a family of orthogonal polynomials which form a natural basis for the expansions of these distributions. A few years later the same polynomials were also investigated by Hermite and they are called Chebyshev-Hermite or simply Hermite polynomials, their definition was first given by Laplace.

These methods use Edgeworth's form that is equivalent to the Gram-Charlier Type A series with use cumulant analysis for the representation of the distribution function in terms of different types of sums of functions of Gaussian processes.

A standard method of exploring high-dimensional datasets is to examine various low-dimensional projections thereof. In fact, many statistical procedures are based explicitly or implicitly on a projection pursuit. Under weak regularity conditions on a distribution $P = P(n)$ on R^n, most d-dimensional orthonormal projections of P are similar (in the weak topology) to a mixture of centered, spherically symmetric Gaussian distributions on R^d if n tends to infinity while d is fixed.

PROPOSITION1. Theorem1The use of supremum of a function over an interval, or maximum, or minimum instead of sum of the numbers completely loses the notion of multidimensionality and reduces it possibly to 1 dimension instead of multidimensional.

3. Analysis of Axioms of Probability of Von Mises.

Theory of von Mises was based on 2 axioms:

1. $\forall \varepsilon_i$ (element appering in the sequence) $\varepsilon_1, \varepsilon_2 \ldots \varepsilon_{i\ldots}$, ε_i has limiting frequency depending on ε_i
2. For $\forall(\tau_1, \tau_2, \ldots \tau_i \ldots)$ (possibly infinite subseq of).$\in$ $(\varepsilon_1, \varepsilon_2 \ldots \varepsilon_{i\ldots})$, with other selection method than prior knowledge of the values of elements selected, the limiting frequencies should be the same.

Property 1. is known as the LLN which in measure-theoretic probability theory is a theorem, holding for almost all sequences x.

Property 2. stands for the rules for selection method that called "selection rules", and selection rules that are different from "prior knowledge" are called "proper selection rules" in contrast to "improper".

After 18 years of debates, the logician Alonzo de Church proposed additional relation to the von Mises system of axioms that only “effectively calculable “ selections should be admitted and thus the set of admissible place selections should consist of the “computable” or “partial recursive functions”. With this addition of central notion of “recurrence” to the system of axioms of von Mises was introduced the Theory of Algorithms (or Recursive Function Theory or Computability Theory). “Kolmogorov-Loveland stochasticity” selection rule that was introduced some 25 years later was very valuable addition.

Richard von Mises mentions “the four fundamental operations” on pages 38-58 of his book “Probability, Statistics and Truth”:

1. Selection;

2. Mixing;

3. Partition; and

4. Combination,

the last 3 defined so as to correspond to usual addition, division, and multiplication rules, and the 1st one is defined as the attributes unchanged and the sequence of elements reduced by place selection, which correspond to the unchanging of distribution, and can be viewed as identity element and related to the previous operations.

Proposition2. Analysis, of how the integration over the events would work in the system of axioms of probability of von Mises, is leading to the decision, that it would definitely be an algebra with operations summation and multiplication. So we can assign ordinal number 4 to the general model.

Proposition3. Theorem2Let k and l be arbitrary natural numbers. Then there exists a natural number n(k,l) such that, if an arbitrary segment, of length n(k,l) , of the sequence of natural numbers is divided in any manner into k classes (some of which may be empty), then an arithmetic progression of length l appears in at least one of these classes.(Khinchin ”Three pearls of Number theory”)

Consider Stone representation of Boolean algebra, which is represented by an algebra with known axioms for Boolean algebra and can be characterized by quadruplets **B** = <X, 0, *, ~>, where 0 is an element from a set X, and * is a binary operation and ~ is an unary operation, which would be a Boolean algebra with 1 as a unit on the operations $\wedge$, $\vee$, and ~. Besides that it has four unary operations, two of which are constant operations, another is the identity, and negation and besides the number of n-ary operations, the number of the dimensions that infinite-dimensional model can be reduced to through application of Boolean prime ideal theorem and Stone duality, can be indexed by an index set.

Proposition4. Multidimensional Time Model for Probability Cumulative Function can be reduced to finite-dimensional time model, which can be characterized by Boolean algebra for operations over events and their probabilities and index set for reduction of infinite dimensional time model to finite number of dimensions of time model considering the fractal-dimensional time that is arising from alike supersymmetrical properties of probability,

It is interesting to consider such point that multidimensional pattern could be related to time dimensions through the introduction in the theory of Brownian motion by E i n s t e i n the consideration of a range of time intervals τ, the possibility of a term proportional to τ in the expression for Moments of Brownian motion B_τ^n(n> 1) is related to the fact that the values of X at moments $t_1, t_2 \dots t_n$ which lie sufficiently close together are no longer independent; and Moments of Brownian motion B_τ^n(n> 1) in fact are represented by a volume integral

$$\int \dots \int X(t_1)X(t_2)\dots X(t_n)\, dt_1 dt_2 \dots dt_n$$

over an n-dimensional cube; the contribution to this integral due to a narrow cylinder extending along the diagonal $t_1 = t_2 = \dots = t_n$ may give a term proportional to τ, as it was mentioned in Kramers' work on chemical reactions rate about particle escape through potential barrier.
This is closely related to the following questions and phenomena in the theory of multidimensional Brownian motion.

4. Various properties of multidimensional Brownian motion.

1. A fixed two-dimensional projection of a three-dimensional Brownian motion is almost surely neighborhood recurrent.
2. Three-dimensional Brownian motion can hit infinite cylinder with probability 1, but it does not hit all cylinders. There almost surely exists avoided infinite cylinder.
3. The problem of finding lower bounds for the Hausdorff dimension of the intersection sets is best approached using the technique of stochastic co-dimension, which is to take a suitable random test set, and check whether it is zero or positive. This approach is based on using the family of percolation limit sets as test sets.
4. Brownian motion according to R. von Mises in "Probability, Statistics and Truth" page 186:
"About a hundred years ago, the English botanist Brown observed under the microscope that certain organic liquids contain small particles moving to and fro in an incessantly agitated manner. It was discovered later that this so-called 'Brownian motion' is common to all sufficiently small particles suspended in a gas or in a liquid, and that it represents a mass phenomenon following the laws of probability calculus. Since we are only interested in the fundamental logical structure of this problem, we can simplify our conception by considering a two-dimensional scheme. We assume that the particles move in a zigzag course in the horizontal plane, excluding any up or downward motion, or else, we may say that we consider only the projection of the three-dimensional motion onto a horizontal plane."

5. **Proposition5 Theorem 3(Levy 1940). Almost surely, $L_2(B[0; 1]) = 0$.**
The range of planar Brownian motion has zero area
Suppose $\{B(t) : t \geq 0\}$ is planar Brownian motion.
Denote the Lebesgue measure on R^d by L_d.
Proposition5 Theorem 1(Levy 1940). Almost surely, $L_2(B[0; 1]) = 0$.

Theorem 4.
(a) For $d \geq 4$, almost surely, two independent Brownian paths in Rd have an empty intersection, except for a possible common starting point.
(b) For $d \leq 3$, almost surely, the intersection of two independent Brownian paths in Rd is nontrivial, i.e. contains points other than a possible common starting point.

Theorem 5
(a) For $d > 3$, almost surely, three independent Brownian paths in Rd have an empty intersection, except for a possible common starting point.
(b) For $d = 2$, almost surely, the intersection of any finite number p of independent Brownian paths in R^d is nontrivial, i.e. contains points other than a possible common starting point.

5. Computer vision, biology, medicine, and DNA analyses.

In computer vision, the target probability p(x) is often defined on a graph representation G=<V, E>. , thus can be divided in two types of graph structures, and thus the Markov chains are designed accordingly.
1. Descriptive models on a plat graph where all vertices are semantically at the same level, e.g. various Markov random fields *image segmentation, graph partition/coloring, shaping*
2. Generative models on a hierarchic And-Or graph with multiple levels of vertices
where a high level vertex is divided into various components at the low level, e.g. Markov trees, sparse coding, *object recognition*, *image parsing, etc*
In advanced models, these two structures are integrated because the vertices at each level of a generative model are connected by contextual horizontal links which represent various relations among the vertices. These computer vision simulations are considered very important for DNA structures representation, as could be clearly seen from the following examples. The biomedical applications of fractal concepts have led to a wealth of new insights in biology and physiology, including a new formulation of the concept of health. The complexity inherent in physiological structures and processes has been described by random fractals. Fractal scaling in various physiological contexts contributed to the analysis of the DNA sequencing, the dendritic branching of neurons and blood vessels, the mammalian lung, the beating of the heart, the dynamics of proteins, ion channel gating and radioactive clearance curves from the body in order to reveal an underlying unity to physiological processes.
DNA is made up of two polymeric strands composed of monomers that include a nitrogenous base (A-adenine, C-cytosine, G-guanine, and T-thymine), deoxyribose sugar, and a phosphate group. The sugar and phosphate groups, which form the backbone of each strand, are located on the surface of DNA while the bases are on the inside of the structure.

References:

1. Gauss (1809, section 177) *Theoria motvs corporvm coelestivm in sectionibvs conicis Solem ambientivm [Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections]*
2. Peirce, Charles S. (c. 1909 MS), Collected Papers v. 6, paragraph 327
3. Richard von Mises "Probability, Statistics, and Truth" Dover/81 ISBN: 978-0-486-24214-9
4. Karl Pearson*,"'Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson'. A rejoinder". Biometrika* **4** *(1): 169–212*
5. Khinchin "Three pearls of Number theory"
6. Alonso de Church "On the concept of a random sequence"
7. W. Feller "Introduction to Probability Theory" v-1, 2, Wiley/99 ISBN: 978-81-265-1805-
8. P. Erdos "On the strong Law of Large Numbers"
9. Adelman, Burdzy and Pemantle (1998): Sets avoided by Brownian motion.Ann. Probab. 26, 429-464.
10. Wolfram MathWorld Pi didits
11. H. A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions. Physica, 7,4, 284-304 (1940)
12. P. Morters Lecture notes.
13. Lutz Dumbgen, P. D. Conte-Zerial On low-dimensional projections of high-dimensional distributions.
14. Hilbert D. "Uber die DarstellungdefiniterFormenalsSumme von Formenquadraten" Math Ann. 32, 342-350 (88).
15. I. Schur, Bemerkungungenzur Theorie der beschranktenBilinearformenmitendlichvielenVer□andlichen, J. reineangew. Math., 140 (11)
16. D. V. Widder, An inequality related to one of Hilbert's, J. London Math. Soc. **4** ,194-198.

17. Frazer C. Note on Hilber's inequality, J. London Math. Soc. 21/46 pp..7-9
18. G. H. Hardy, J. E. Littlewood and G. P´olya, Inequalities, 2nd ed., CUP, Cambridge, 1988
19. Mitrinovic, Dragoslav S. Analytic Inequalities.
20. H.S. Wilf, Finite sections of some classical inequalities, Ergebnisse der Mathematik, Band 52, Springer
21. Minkowski, Hermann *(1896). Geometrie der Zahlen. Leipzig: Teubner.*
22. Milman, Vitali D. (1986). "Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espacesnormés. [An inverse form of the Brunn-Minkowski inequality, with applications to the local theory of normed spaces]". C. R. Acad. Sci. Paris Sér. I Math. **302** (1): 25–28
23. *Landau, E. (1913). "UngleichungenfürzweimaldifferenzierbareFunktionen". Proc. London Math. Soc.* ***13****: 43–49.* doi:10.1112/plms/s2-13.1.43
24. N. G. de BruijnCarleman's Inequality for finite series. Koninkl. Nederl. Akademie. Proceedings Series A,66,4
25. Philipp Hennig, Michael A Osborne, and Mark Girolami Probabilistic Numerics and Uncertainty in Computations rspa.royalsocietypublishing.org
26. S.Blinnikov , R.Moessner Expansions for nearly Gaussian distributions Astrophysics
27. Halmos, Paul Lectures on Boolean Algebras. van Nostrand.

Hilbert some 120 years ago, though did not use Cauchy's inequality, but used rather some complicated methods to prove that $\sum\sum\frac{a_m b_n}{m+n} \leq C(\sum a_m^2)^{1/2}(\sum b_m^2)^{1/2}$, with C= 2π , where the summations run from 1 to +∞, which was later reduced by I. Schur to C= π .There were some improvements to constant C in Hilbert's inequality such, as by H. Frazer, who showed that it can be reduced to (n+1) sin $\frac{\pi}{(n+1)}$,

and N. G. de Bruijn and H. S. Wilf showed that the best possible constant for the discrite case of summations running from 1 to N in $\sum\sum\frac{a_m a_n}{m+n} \leq C(\sum a_n^2)^{1/2}$ can be replaced by $C_N= \pi - \frac{1}{2}\pi^5 (\log N)^{-2}+O(\log\log(\log N)^{-3})$, $N\to+\infty$.

The later remarkable result is very important in light of the Bruijn-Wilf type best constant for n[th] finite section of Carleman's inequality $\sum_{k=1}^{n}(a_1 a_2 \ldots a_k)^{1/k} < C_n \sum_{k=1}^{n} a_k$. C_n=e-2 π^2 e1/ln(n)^2+O(1/ln(n)^3)C_2=1/2(1+sqrt(2)) C_3=4/3 And for the Hardy's inequality C_n=4-16 π^2 / ln(n)^2+ O(lnlnn/ln(n)^3) ,The extensions of Widder with the remark of Hardy are the following inequalities:

$$\sum\sum\frac{\log(m/n)}{m-n}a_m b_n \leq \pi(\sum\sum\frac{a_m a_n}{m+n})^{1/2}(\sum\sum\frac{b_m b_n}{m+n})^{1/2}$$

$$\sum\sum\frac{|\log(m/n)|}{\max(m,n)}a_m b_n \leq 2(\sum\sum\frac{a_m a_n}{\max(m,n)})^{1/2}(\sum\sum\frac{b_m b_n}{\max(m,n)})^{1/2}$$

where the summations run from 0 to +∞ , and the coefficient on the left-hand side is interpreted as 1/n when *m* = n;

In this form it is very much resembling the famuos Landau-Kolmogorov inequality, which was inroduced by E. Landau 2 years after I. Schur's improvement.that for $\|f\|$ to be the supremum of $|f(x)|$, a real-valued function f defined on $(0,\infty)$. and for a usually defined norm of real-valued function f defined on $(0,\infty)$ as $\|f\| = \sqrt{\int_0^\infty [f(x)]^2\,dx}$. there are different constants $C(n,k)$ in inequality

$$\|f^{(k)}\| \leq C(n,k)\,\|f\|^{1-k/n}\,\|f^{(n)}\|^{k/n}.$$

Similar notions can arise in the discrete inequalities of Ky Fan, Taussky, and Todd that date 50 to 60 years ago: $\sum_{i=0}^{n-1}(x_i - 2x_{i+1} + x_{i+2})^2 \geq 16\sin^4\frac{\pi}{2(n+1)}\sum_{i=0}^{n} x_i^2$;for the sequence of real numbers $x_0=0, x_1,\ldots x_n, x_0=x_{n+1}=0$, with equality holding if and only if $x_i=c\sin\frac{i\pi}{n+1}$, i=1,...,n, where c is a real constant and for the sequence of real numbers $x_1=0, x_1,\ldots x_n$: $\sum_{i=1}^{n-1}(x_i - x_{i+1})^2 \geq 4\sin^2\frac{(i-1)\pi}{2n-1}\sum_{i=2}^{n} x_i^2$.

99 years passed before V. Milman established reverse form of the Brunn–Minkowski inequality that was proposed by Hermann Brunn 1 year before Hilbert's inequality.

PROPOSITION. Therefore, the use of supremum of a function over an interval, or maximum, or minimum instead of sum of the numbers completely loses the notion of multidimensionality and reduces it possibly to 1 dimension instead of *double-number* dimensional.

It is very certain, that this effect is given not only in mathematical formulae, but it has some psychological and methodological outcomes, that affect ability to recognize multidimensionality in the problem or mathematical model, as a way to solve it.