Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

Article

Testing Statistical Hypotheses in Light of Mathematical Aspects in the Analysis of Probability

Michael Fundator

Division of Behavioral and Social Sciences of National Academy of Sciences, Washington, DC, USA; michaelfundator@gmail.com

Abstract: The philosophy of testing statistical hypothesis is a natural consequence and functional extension of mathematical analysis of Probability. Along with the concept of recurrence when applied to random sequences and functions, it leads to the analysis of a priori and posterior which implies testing statistical hypothesis. Testing statistical hypothesis also involves algebraic, functional and dimensional considerations, which are found in the works of Laplace. Aspects of mathematical analysis such as universality of solutions, Laws of Large Numbers, Entropy, Information, and various functional dependencies are the main factors explained in the five properties that lead to implication of testing statistical hypothesis. Various interesting examples with modern scientific significance from genetics, astrophysics, and other areas give methodological access to answers of different problems and phenomena which are involved in the logic of testing statistical hypothesis.

Keywords: Principal of prediction; random sequences; recurrence; Law of Large Numbers; exponential; normal; bivariate; distribution; Entropy; Information

---

1. Introduction

Researchers in modern molecular biology use hundreds times more hypothesis tests in one project than previously, and as a consequence the question whether to discuss a fallacy of single hypothesis testing, or to correct for multiple testing became oblivious in this area, leaving place for other methodological discussions and questions, such as how to correct them [40-43]. Hypothesis tests are not free of error, however, and for every hypothesis test there is a risk of falsely rejecting a hypothesis that is true, i.e. a type I error, and of failing to reject a hypothesis that is false, i.e. a type II error.The False Discovery Rate (FDR) is a new statistical procedure to control the number of mistakes made when performing multiple hypothesis tests. Large-scale multiple testing has been applied in fields such as genomics[, astrophysics [, brain imaging, and spatial epidemiology [38].

As a decision theoretical problem testing statistical hypothesis can be part of Bayesian or frequentist approach, it can vary from parametrical to semi- or non- parametrical models, which should be mentioned, although this paper is mainly concerned with the first ones.

There are five main aspects of mathematical analysis of probability in the testing of statistical hypothesis (TSH).Testing statistical hypothesis is a method that helps in making statistical decisions based on experimental data. The method however works basically as an assumption about the population parameter. Although the procedure seems to be very simple, one may have the perception that it is difficult hence try to apply it in special circumstances, e.g. facing one of 5 types of bivariate exponential distribution, [23-27,32], or one of 5 types of bivariate gamma distribution[28-31,18] in the theory of ratios of random variables.

The 5 properties involve the analysis of a priori and posterior along with the concept of recurrence, random sequences and functions [1-7, 10-14]. TSH with familiar concepts of decision theory loss and risk functions are a continuation of basic algebraic operations over random variables. The basic algebraic operations also use the random variables along with differential and integral calculus over them, with further extension to a multi-dimensional model[1-3]. This requires application of special mathematical interest in search of generalized and uniform functions and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

solutions [16, 17]. TSH also can be applied to check for the universality of solutions of different problems.

Just as functional operators are related to different types of LLN, TSH is related to concepts of Entropy and Information in different goodness of fit tests [21, 22]. TSH is often presented in conjunction to mixed models repeated measures, or regression models, or cluster randomized trails

There are five steps in hypothesis testing that look like step by step instructions manual [37]:

1. Making assumptions about the data and fitting model.
2. Stating the null and alternative hypotheses. Selecting significance level (alpha).
3. Selecting the sampling distribution and specifying the test statistic.
4. Computing the test statistic and probability value.
5. Making a decision and interpreting the results.

In comparison, there is summary of a step by step procedure aimed at finding the equation of a tangent line to a curve at an indicated point:

1. Find the first derivative of f(x).
2. Substitute value of x_1 point into f '(x) to find the slope at x_1.
3. Substitute value of x_1 into f(x) to find the y_1 coordinate of the tangent point.
4. Use the point-slope formula to find the equation for the tangent line.

$$y - y_1 = m\,(x - x_1).$$

5. Check the results, possibly by drawing a graph.

Though seemingly both procedures are not related directly, we can find similarities in series expansions, e.g. for derivation of test statistic for Hamiltonian system of one, two, or 3-dimensional lattice, such as classical perturbation theory [38].

1.1. History of the Laws of Large Numbers [8,9,44]

Statistical prediction in modern sciences traces its roots in ancient mathematics. To introduce this, it would be reasonable to start from the theorems of Probability and Statistics. Such theorems support the calculation of probability, estimation of and testing statistical hypotheses (TSH) for the next occurrence of the Heads and Tails, or +s and –s. For instance, the theorem of Bernoulli for the probability of coin tossing outcomes of Heads and Tails depending on their frequencies. Consider Erdos-Renyi law of large numbers for general sequences of independent identically distributed (i.i.d.) random variables. As an extension of the Erdos' result that was obtained the same year when Erdos and Selberg found an elementary probabilistical proof of Prime Number Theorem. The theorem extended Kac, Salem, and Zygmund result for functions with conditions

$$f(x+1) = f(x),\ \int_0^1 f(x) = 0,\ \int_0^1 f^2(x) = 1\ ,$$

$$\int_0^1 (f(x) - \varphi_n(f))^2 = O\left(\frac{1}{(\log n)^{\epsilon}}\right) \text{ for some } \epsilon > 0, \qquad (1)$$

where $\varphi_n(f)$ is the nth partial sum of the Fourier series of f(x), with $n_1 < n_2 < \ldots < n_k < \cdots$ a sequence of numbers satisfying $n_{k+1}/n_k > c > 1$, then for almost all x

$$\lim_{N\to\infty} \frac{1}{N}\left(\sum_{k=1}^{N} f(n_k x) = 0\right) \qquad (2)$$

to the Theorem of existence of f(x) and sequence n_k such that

$$limsup_{N\to\infty} \frac{1}{N}\left(\sum_{k=1}^{N} f(n_k x) = \infty, \qquad (3)\right.$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

This result surprisingly points to a philosophical view of prediction at infinity. The question, if the limit in (1) is true for all f(x) was already proved by Raikov for sequences $n_k = 2^k$.

And then he further used the gap between the conditions:

$$\int_0^1 (f(x) - \varphi_n(f))^2 = O(\frac{1}{(loglogn)^{2+\epsilon}}), \qquad (4)$$

And

$$\int_0^1 (f(x) - \varphi_n(f))^2 < \frac{1}{(logloglogn)^{\epsilon}} \qquad (5)$$

to prove even stronger version

$$limsup_{N\to\infty} \frac{1}{N(loglogN)^{\frac{1}{2}-\epsilon}} (\sum_{k=1}^{N} f(n_k x)) = \infty, \qquad (6)$$

And

$$limsup_{N\to\infty} \frac{1}{N(logN)^{\frac{1}{2}+\epsilon}} (\sum_{k=1}^{N} f(n_k x)) = 0 \qquad (7)$$

This result established the weak sense of prediction for the coin tossing outcomes, or the length of the longest heads run that could be used further for the development of method of TSH from the axiomatical approach to mathematical theory of probability.

The history of application of the law of large numbers (LLN) to statistical analysis started possibly some 450 years ago, when Gerolamo Cardano stated without proof that the accuracies of empirical statistics tend to improve with the number of trials. John Arbuthnot was the first to publish statistical test on fraction of boys and girls born year after year [44], three years before the LLN for coin tossing random variable was first proved by Jacob Bernoulli around. And only after more than 120 years Poisson used the name "la loi des grands nombres" ("The law of large numbers"). Five other versions of LLN were derived by Chebyshev, Markov, Borel, Kolmogorov, and Khinchin. They differ from one another by convergence in probability (weak LLN) , almost sure convergence, or with probability 1 (strong LLN, e.g. Kolmogorov LLN), if the i.i.d. random variables in the sequence should have a variance (e.g. Kolmogorov LLN), if the variables can be correlated, and the like variations of the LLN. Many of them are proved with the help of Chebyshev inequality.

$$P(|\dot{X} - \mu| \geq k\sigma) \leq 1/k^2 \qquad (8)$$

Borel's LLN is the extension of Bernoulli's Theorem that the limiting frequency of the repeating event tends to probability of this event with probability 1. This is direct consequence of Kolmogorov LLN. Khintchin's weak LLN is the convergence in probability of sample average to the expected value.

$$\lim_{n\to\infty} P(|\dot{X} - \mu| > \varepsilon) = 0 \qquad (9)$$

It was established so far that statistical analysis leads to the week sense of prediction, which we call Principle. The Principle of prediction leads to different statistical analyses [40]. There is no surprise why in the further analysis of Laplace's several works on probability, we find that Chapter IV of Laplace's "Philosophical Essay on Probabilities" is called "Concerning hope". The Treatise is one of the important works on the subject of probability where possibly, a half of the Treatise is concerned with statistical methods and applications. In the beginning of Chapter XVII "Concerning the various means of approaching certainty", he describes the principles that are components of the Principle of prediction. These principles include induction, analogy and hypotheses, supported by numerous comparisons, which he called "principal means for arriving at truth". In his view analysis and natural philosophy of sciences, Newton's binomial theorem and the principle of universal gravity are the consequences of induction. To show that induction should be followed as far as anticipated with logic which is strengthened by analogy and hypotheses testing, he brings as an example Fermat Conjecture that states that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

$2^{2^n}+1$ is a prime for all n, which was caused by induction and was recognized by Euler that for n=32 gives 4,294,967,297, a number divisible by 641 (pp. 177,178).

In the account of Laplace on the history of the subject, Pascal and Fermat were the first to state the principles and the methods of probability (pp. 167,185).

In light of the above account about induction and Fermat Conjecture, a reasonable continuation is the Descartes' theorem. Descartes was a very close friend of Fermat and one of the founders of French Academy. The theorem states, that if there are 3 mutual tangent circles to each other, then the 4th circle, which can possibly be circumscribed or inscribed could be constructed. The 4th circle would be tangent to all 3 circles and the 4th curvature could be calculated from the first 3 from equation:

$(k_1+ k_2 + k_3 + k_4)^2 = 2(k_1^2 + k_2^2 + k_3^2 + k_4^2)$, $k_i= \pm 1/r_i$, with r_i being radius of the i-th circle (10)

Euler showed, that the special case, when the k_i are perfect squares of a_i^2, is equivalent to the 3 simultaneous equations of Pythagorean triples.

$$(a_1^2 + a_2^2 + a_3^2+a_4^2)^2=2(a_1^4 + a_2^4 + a_3^4+a_4^4) \quad (11)$$

$$2(a_1a_2)^2 + 2(a_3a_4)^2 = (a_1^2 + a_2^2 - a_3^2-a_4^2)^2 \quad (12)$$

$$2\,(a_1a_3)^2 + 2(a_2a_4)^2 = (a_4^2 - a_2^2 + a_3^2-a_4^2)^2 \quad (13),$$

$$2(a_1a_4)^2 + 2(a_2a_3)^2 = (a_1^2 - a_2^2 - a_3^2+a_4^2)^2 \quad (14)$$

It is also interesting to notice that Fermat, who was in one Mersenne circle with Descartes, and later in exchanged correspondence with Pascal laid foundations of the theory of probability. Fermat gave the smallest Pythagorean triple with both the hypotenuse c and the sum of the sides a + b as perfect squares such triple has sides

a =4,565,486,027,761; b =1,061,652,293,520; and c =4,687,298,610,289,

with a +b = $2{,}372{,}159^2$ and c = $2{,}165{,}017^2$, although there are infinitely many such Pythagorean triples.

It is also interesting to notice that there is a general formula that gives all solutions of Fermat cubic

$$a_1^3 + a_2^3 + a_3^3 = a_4^3 \ , \quad (15)$$

$$(3x^2+5xy-5y^2)^3+(4x^2-4xy+6y^2)^3 + (5x^2 - 5xy - 3y^2)^3 = (6x^2 - 4xy + 4y^2)^3 \quad (16)$$

Farey sequence can be viewed as a natural continuation of Descartes' circles theorem, after more than 120 years, when L.R. Ford presented a very important property of Farey's series. The property indicates that circles of radius $1/b^2$ drawn above each reduced fraction a/b, and touching the number line at this point never overlap despite expectations, although they touch very often. It was first studied by C. Haros a year after Gauss' successful calculation of the orbit of Ceres, and thus was introduced 200 years ago with Farey's paper

After preliminary discussion about the correspondence of the Principle of prediction with different statistical analyses and fundamentals of statistics, including testing statistical hypotheses, we have to notice that Laplace in several works on probability, devoted a lot of attention to many techniques and results of statistics. He was very much concerned with the calculus of probabilities and Central Limit Theorem (CLT), including testing statistical hypotheses.

1.2. Axiomatization of probability, random sequences, and recurrence.[5,6,10-14].

In order to answer efficiently this question we have to turn attention to the foundational and axiomatical principles of probability that were the subject of the intense discussion, which started 100 years ago, and with inclusion of new names it drew a lot of attention for the next 20 years. It continued to resurface occasionally in different theories for the next 40 years. The main subject of the debate was application of mathematical rigor to random sequences.

Emile Borel was one of the first mathematicians to formally address randomness with his work on normal numbers 10 years before Richard von Mises gave the first definition of algorithmic

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

randomness, which was inspired by the LLN and introduced a notion of a random sequence almost 100 years ago.

It was based on 2 axioms:

1. $\forall\varepsilon_i$ (element appering in the sequence) $\varepsilon_1, \varepsilon_2 \dots \varepsilon_{i\dots}$, ε_i has limiting frequency depending on ε_i
2. For $\forall(\tau_1, \tau_2, \dots \tau_i \dots)$ (possibly infinite subseq of)$\subset$ $(\varepsilon_1, \varepsilon_2 \dots \varepsilon_{i\dots})$, 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 is the requirement that frequency stability be preserved under the operation of extracting infinite subsequences. These rules for selection method are called "selection rules", and selection rules that are different from "prior knowledge" are called "proper selection rules" in contrast to "improper".

1.3. Development of rigorous approach to probability theory [5,6,10-14].

After 18 years of development the logician Alonzo de Church proposed that only "effectively calculable" selections should be admitted. He also claimed that the set of admissible place selections or sequences should be based on the "computable" or "primitive recursive functions" in accordance with Wald's Theorem. With this addition of central notion of "recurrence" to the system of axioms of von Mises, the Theory of Algorithms (or Recursive Function Theory or Computability Theory) was introduced. "Kolmogorov-Loveland stochasticity" selection rule that was introduced some 25 years later was a very valuable addition. J.L. Doob, in his obituary on William Feller discussed some very new concept in von Mises' development of the rigorous approach to the probability theory. The concept is related to Frechet's introduction of measure on an abstract space. At that time probability was a subject with no clear knowledge, and no understanding between the mathematical and what was real.

It implies that in the understanding of von Mises, it was not always possible to describe the mathematical problem in terms of measurable functions. Rigorous approach of this phenomenon is applied in Subchapter 3.1 on scientific applications to multiple hypotheses testing [34, 40-43]. Instead one had to turn to the a priori knowledge or an a priori probability which is defined as a probability that is derived purely by deductive reasoning. That was another problem in the system of axioms of von Mises, which was fixed by introduction of recurrence.

This is where the concept of testing hypothesis comes next, following Laplace's applications of his tests for binomial probabilities with the possible extension to other models.

To compensate for not considering a priori knowledge or probability, Laplace considered recurrent series of Moivre in his "Philosophical Essay on Probabilities", which served as an introduction and partial explanation to early written "Analytic Theory of Probability"?

2 Five Properties of mathematical analysis of probability that are leading to testing statistical hypothesis.

2.1.Property I [5,6,10-14].

Proposition1. Probabilistical analysis leads to the analysis of a priori and posterior, which in their turn along with the concept of recurrence, applied to random sequences and functions, imply testing statistical hypothesis, as mathematical continuation and application of the analysis of the foundational and axiomatical principles of probability to abstract and applied statistical problems related to Principle of prediction. We should call it Property I of mathematical analysis of probability that is leading to testing statistical hypothesis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

2.2.Property II [1-15,19,20,40, 45,47]

Proposition 2. Natural continuation of algebraic considerations, concerning basical operations of summation and multiplication over random variables as functions over events and their probabilities along with differential and integral calculus over them with further extension to multi-dimensional calculus that is called calculus of probabilities is leading to an additional operation of testing statistical hypothesis, such as different comparisons tests of mean, variance, goodness of fit, etc. As alike functional operation TSH is often presented in conjunction to mixed models repeated measures, or regression models, or cluster randomized trails.

There were certainly other aspects of mathematical concepts considered in Laplace's' works that imply very familiar concepts of decision theory loss and risk functions. He certainly was not interested only in derivation of distribution functions, but also considered random variables, as functions, as it clearly seen from the end of Chapter IV and Chapter V "Concerning the analytical methods of the calculus of probabilities". And though most of his examples were based on binomial random variables, he was interested in basic algebraic operations over them along with the differential and integral calculus on them with further extension to 2-dimensional calculus, then to 3- dimensional, and finally to as many dimensions as is chosen for the particular model, his main objective in the "Analytic Theory of Probability" was Central Limit Theorem (CLT) [2],[3].

In as much as he describes operations of summation and multiplication, it can be extended in the very natural way to normal random variables with mean μ and variance σ^2 with probability density function denoted

$$X \sim N(\mu, \sigma^2) \qquad \text{with} \quad f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (17)$$

One of the definitions of Peirce of "normal" as of what would, in the long run, occur under certain circumstances [4], clearly implies Principle of prediction and LLN that were already discussed above. With standard normal i.i.d. random variables $X_1, X_2 \ldots X_i \sim N(0, 1)$ that can be used for comparisons of sample means and many other tests. Their sum of squares

$\sum_{i=1}^{n} X_i^2{}_i = S_n \sim \chi_n^2$ is distributed as chi-squared distribution with n degrees of freedom that can be used for testing statistical hypothesis for sample variance of a normal random variables and for goodness of fit test, and in this context it was rediscovered by Karl Pearson. Chi-squared distribution also can be used for confidence intervals estimation of Poisson distribution (Garwood chi-square) along with normal distribution function. However, Chi-squared distribution, which is special case of Gamma distribution function, has special case for 2 degrees of freedom that becomes exponential distribution function, which is, following W. Feller v.2 p.9, essential in derivation of Poisson process $f_n(t) = e^{-\alpha t}(\alpha t)/n!$ (18)

and was first discovered by John Michell some 70 years before Poisson with consideration of distribution for the random scattering of points in the region without introduction of Poisson distribution. Poisson process with $\lambda = \alpha t$ would become Poisson distribution, and it was previously obtained by Abraham de Moivre as a limiting case of binomial with $\lambda = np$.

Introduction of Stein's estimators for exponential family of distributions is another connection between Poisson process in dimension ≥ 3 and normal or Gaussian distributions[44,46].
If X is from exponential family of distributions with density

$f_\theta(x) = \exp\{\theta x - \varphi(\theta)\}k(x)$, $x \in \mathbb{R}$, let $t(X) = -\frac{k'(X)}{k(X)}$ for any absolutely continuous function g on $\mathbb{R}$, such that $E|g'(X)| < \infty$, then $E\{(t(X) - \theta)g(X)\} = E\{g'(X)\}$

For regression model, $\mathbf{Y} = \mathbf{XC} + \mathbf{e}$, where $\mathbf{Y}$ is a N x 1 vector of N observations on the variable to be explained, X is a N x K full-column-rank matrix of N observations on K fixed explanatory variables, $\mathbf{C}$ is a column vector of regression coefficients, and $\mathbf{e}$ is a N x 1 error vector with a multivariate normal distribution with mean vector 0 and variance-covariance matrix $\sigma^2\mathbf{I}_N$, with σ^2 unknown. The least-squares (LS) estimator for $\mathbf{C}$ is $\hat{C}_{LS} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$, and Stein's estimator

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

is $\mathbb{C}_{SE}=\left[1-\frac{L(Y-XC)'(Y-XC)}{C'X'XC}\right]\mathbf{C}$, where L is a scalar, n= N-K, $\mathbf{S^2}=(\boldsymbol{Y}-\boldsymbol{XC})'(\boldsymbol{Y}-\boldsymbol{XC})/n$

Though $\mathbb{C}_{SE}$ is also baised it dominates $\hat{C}_{LS}$ so as

$E[(\mathbb{C}_{SE}-\mathbf{C})'\,\mathbf{X'X}(\mathbb{C}_{SE}-\mathbf{C})] < E[(\hat{C}_{LS}-\mathbf{C})'\,\mathbf{X'X}(\hat{C}_{LS}-\mathbf{C})]$, and $0<L<\frac{2(K-2)}{(n+2)}$, $K>2$.

An unbiased estimator of $\mathbb{C}_{SE}$ is $C_{UB}=\mathbb{C}_{SE}-\hat{C}_{LS}=-L\frac{nS^2}{\hat{C}_{LS}'\mathbf{X'X}\hat{C}_{LS}}\hat{C}_{LS}$

$Z=(\mathbf{X'X})^{1/2}\hat{C}_{LS}$ and $\mu \propto N(\mu, \sigma^2\mathbf{I}_K)$ with $(nS^2/\sigma^2) \propto \chi^2_n$.

To test null hypothesis $\mathbf{H_0}$: $\mathbf{C=0}$ against alternative hypothesis $\mathbf{H_1}$: $\mathbf{C\neq 0}$

The uniformly most powerful invariant test statistic is $\mathbf{F}=(\hat{C}_{LS}'\,\mathbf{X'X}\hat{C}_{LS})/KS^2$,

$\mathbf{F} \propto \mathcal{F}_{K,n}$ an F-distribution with K and n degrees of freedom and noncentrality parameter,

$D=\mathbf{C'X'XC}/2\sigma^2$

And for F-ratio based on $\mathbb{C}_{SE}$ statistic $\mathbb{F}=\mathbb{C}_{SE}'[\mathbf{S^2(X'X)^{-1}}]^{-1}\,\mathbb{C}_{SE}=\mathbf{F(1-}nL/K\mathbf{F)^2}$.

Since $\mathbb{F}$ is a function of of **F**, though nonmonotonic, and some authors proposed to use $\mathbb{F}^+$, it is invariant to the same linear and orthogonal transformations, as **F** is.

The above analysis was developed during the last 60 years and greatly contributed to the theory of multihypotheses testing and FDR analysis, which resurfaced only recently after Chemberlin's discussion and would be discussed in the Chapter on scientific application of TSH.

To summarize the above, statistical analysis with testing statistical hypothesis is direct consequence of mathematical formulation of probability theory.

2.3.Property III

Proposition 3. Expectation of universality of solution is an important factor for the solution of mathematical problems, and it is subject of dream of many mathematicians in the past and present, and cannot be avoided so easily. As in a historical example with Gauss' search for orbit of a newly discovered celestial body, in which he used method of least squares, TSH is an important component for approximation of mathematical models, such as differential equations. In time of growth of the theory and applications of linear combinations and ratios of random variables that by themselves became very complicated, such as 5 different types of bivariate exponential distribution and at list 5 different types of bivariate Gamma distribution, to include 5 types of different methods, the opponents of TSH view the complication of completing calculations correctly as an obstacle in consideration of very important concept of uniformity of solution to complicated problems.

Though the theory of universal differential equation is not well developed, it is important factor for the solution of mathematical problems, and it is subject of dream of many mathematicians in the past and present, and cannot be avoided so easily. Uniform equation or solution can be related to Leibniz's pre-established harmony principle, or to universal differential equation [16-17,35-36]. It can be related to a system of linear equations with more equations than unknowns, and can be related to many numerical methods, stochastic differential equations and differential equations for large ensembles of particles such as gases and fluids, stochastic processes and multivariate statistical analysis, etc. are the examples.

The variety of different methods and types of bivariate exponential and Gamma distribution support the number of different hypotheses tests developed. Only for Poisson distribution there are at least 20 different procedures for finding confidence intervals.

2.3.1 Variety of methods of linear combinations and ratios of random variables.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

Five methods of the theory and applications of linear combinations and ratios of random variables:

1. Ratios of normal random variables appear as sampling distributions in single equation models, in simultaneous equations models, as posterior distributions for parameters of regression models
2. Weighted sums of uniform random variables
3. Ratio of linear combinations of chi-squared random variables multivariate linear functional relationship model Sums of independent gamma random variables
4. Linear combinations of inverted gamma random variables for the Behrens-Fisher problem and variance components in balanced mixed linear models Beta distributions their linear combinations
5. Linear combinations of the form T = $a_1 t_{f_1} + a_2 t_{f_2}$, where t_f denotes the Student t random variable based on f degrees of freedom and weighted sums of the Poisson parameters.

2.3.2 Five types of the bivariate exponential distribution [32, 24-27].

1. $\int_x^\infty (\int_y^\infty f(x,y)dy)dx$ = F(x; y) = 1 − exp(−a_1x −a_2y− a_{12}max(x, y)).x,y ≥ 0. (19) for Marshal and Olkin model.
2. F(x,y)=1-exp(-x) –exp(-y) + exp(-x-y-axy); x,y>0, 0≤ a ≤ 1 (20) for Gumbel model.
3. X = $U_1^2 + U_2^2$ Y = $U_3^2 + U_4^2$with with (U_1, U_3) independent from (U_2, U_4), but have the same joint normal with zero means, variances ½, and corr r(0≤ r ≤ 1) for Moran model.
4. For Freund model component 1 and component 2 are dependent in that a failure of either component changes the parameter of the life distribution of the other component.
5. Block and Basu considered a bivariate distribution whose marginals are mixtures of exponentials and having an absolutely continuous joint distribution.

2.3.3 5 types of bivariate gamma distributions. In univariate case the gamma distributions is generalization of Erlang distribution, which is the sum of i.i.d. exponentially distributed random variables:

1. McKay's bivariate gamma distribution given by the joint pdf

$$f(x, y) = (a^{p+q}/Ga(p)Ga(q))x^{p-1}(y - x)^{q-1}\exp(-ay), \quad (21)$$

where Ga(.) denotes Gamma function, for y >x >0, a >0, p >0 and q >0.

2. Cherian's bivariate gamma distribution given by the jointpdf

$$f(x, y) = (\exp(-x - y)/Ga(\theta1)Ga(\theta2)\ Ga(\theta3))\int_0^{min(x,y)}(x - z)^{\theta_1-1}\ (y - z)^{\theta_2-1}z^{\theta_3-1}\exp(z)dz$$

forx >0, y >0, θ_1 >0, θ_2 >0 and θ_3 >0. (22)

3. Kibble's bivariate gamma distribution given by the jointpdf

$$f(x, y) = \frac{(xy)^{(\alpha-1)/2}}{Ga(\alpha)(1-\rho)\rho^{(\alpha-1)/2}}\exp(\frac{x+y}{1-\rho})I_{\alpha-1}(\frac{2\sqrt{xy\rho}}{1-\rho}), \quad (23),$$

where I_f(.) denoted modified Bessel function of the first kind of order f

4. the Beta Stacy distribution is given by the joint pdf

$$f(x, y) = \frac{c}{a^{bc}Ga(b)B(p,q)}x^{p-1}(y-x)^{q-1}y^{bc-p-q}\exp\{-(y/a)^c\}, \quad (24)$$

where B(.,.) is Beta distribution function fory >x >0, a >0, b >0, c >0, p >0 and q >0.

5. Becker and Roux's bivariate gamma distribution

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

$$f(x, y)=\frac{\beta'\alpha^{a}}{Ga(a)Ga(b)}x^{a-1}\{\beta'(y-x)+\beta x\}^{b-1}\exp\{-\beta' y-(\alpha+\beta-\beta')x\},\text{if } y>x>0,$$

$$\text{or}\frac{\alpha'\beta^{b}}{Ga(a)Ga(b)}y^{b-1}\{\alpha'(x-y)+\alpha y\}^{a-1}\exp\{-\alpha' x-(\alpha+\beta-\alpha')y\},\text{if } y>x>0,$$

$$\text{for} x>0,\ y>0,\ a>0,\ b>0,\ \alpha>0,\ \beta>0,\ \alpha'>0 \text{ and } \beta'>0. \qquad (25)$$

Based on this, the opponents of TSH view the complication of completing calculations correctly as an obstacle in consideration of very important concept of uniformity of solution to complicated problems.

2.3.4. Different approaches to TSH related to different types of multivariate exponential and gamma distributions and their linear combinations and ratios.

1. Method of maximum entropy (MM)

Differential Entropy is defined as an expectation of information function

$h[f] = E[-\ln(f(x))] = -\int f(x)\ln(f(x))]\,dx$ over **X** and Shannon entropy = $-\sum_i p_i \log_a p_i$.

Because of special properties of logarithm function, it can be used for different problems , where a strait ahead approach is very difficult to find.

2. Stein's method of empirical Bayes or Bayesian entropy.

3. Fisher information for a covariance matrix. On the equivalence between Stein and De Bruijn identities

4.Trimmed estimators[49]

5. Estimation of Kullback-Leibler Information for problems within the exponential family, MLE, log likelihood ratio, C-alpha test.

2.4. Property IV

Proposition 4. The properties of Laws of Large Numbers, as was mentioned above are deeply related to the functional extension of calculus of probabilities in the form of testing statistical hypothesis, and as such their continuous extensions in form of stochastic calculus with week and strong solutions of stochastic differential equations also closely related to the TSH. As a example can be given for uniform LLN, which can be used for different types of estimators, e.g. extremum estimator , minimal distance estimator, MLE estimator, etc. Minimal distance estimator is related to many different tests, Cramer-von Mises, Kolmogorov-Smirnov, Anderson-Darling, D'Agostino, Shapiro-Wilk, and Michael's tests with their modifications, based on characteristic functions. As alike functional operators related to different types of LLN, TSH is often presented in conjunction to mixed models repeated measures, or regression models, or cluster randomized trails that are helpfu methods in modern research on missing data and large data [23].

2.5. Property V

Proposition 5. As alike functional operators related to different types of LLN, MLE, and tests based on characteristic functions,TSH can be often considered in conjunction to very important properties for analysis in Probability and other fields of scientific research, such as Entropy and Information that are also related to such questions as additivity of measures, general solution of equations, and therefore, to different questions related to LLN, e.g. sufficient condition for LLN over classes of functions to hold uniformly is finiteness of the Koltchinskii-Pollard entropy integral, or more direct relation to TSH, such as Testing Statistical Hypothesis of orders of $\alpha-\beta$ weighted information energy [33]. This relation was already noticed in the works relating Kullback-Leibler

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

entropy or information (KLD) to Newman-Person lemma. Other methods use KLD to choose M best hypotheses from the given N hypotheses[48,49], or application of KLD to the concept of multiple working hypotheses that was advocated over 120 years ago by Chamberlin[49-53].

3. Examples of scientific application.

3.1 Multiple testing dependence [34, 40-43].

For the model in matrix form: $\mathbf{Y} = \mathbf{CS(X)} + \mathbf{E}$ the problem of testing m hypotheses of the form: $H_{oi} : \mathbf{c}_i \in \Omega_0$ vs. $H_{1i} : \mathbf{c}_i \in \Omega_1$ has typically been defined in terms of P values or test statistics resulting from multiple tests.

Definition: Population-level multiple testing dependence exists,

if $\Pr(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_N \mid \mathbf{X}) \neq \Pr(\mathbf{y}_1 \mid \mathbf{X}) \times \Pr(\mathbf{y}_2 \mid \mathbf{X}) \times \cdots \times \Pr(\mathbf{y}_N \mid \mathbf{X})$.

Correspondingly estimation-level multiple testing dependence exists ,

If $\Pr(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_N \mid \hat{\mathbf{C}}, \mathbf{S}(\mathbf{X})) \neq \Pr(\mathbf{y}_1 \mid \hat{\mathbf{C}}, \mathbf{S}(\mathbf{X})) \times \Pr(\mathbf{y}_2 \mid \hat{\mathbf{C}}, \mathbf{S}(\mathbf{X})) \cdots \times \Pr(\mathbf{y}_N \mid \hat{\mathbf{C}}, \mathbf{S}(\mathbf{X}))$, where $\hat{\mathbf{C}}$ denotes the estimator of $\mathbf{C}$

The dependence is equivalent to dependence of the rows of $\mathbf{E}$, or the rows of residual matrix.

Proposition Suppose that for each $\mathbf{e}_i$ in the above model, there is no Borel measurable function g such that $\mathbf{e}_i = g(\mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \mathbf{e}_{i+1}, \ldots, \mathbf{e}_N)$ almost surely. Then, there exist matrices $\mathbf{D}_{N\times r}$, $\mathbf{G}_{r\times K}$ ($r \leq K$), and $\mathbf{U}_{N\times K}$ such that $\mathbf{Y} = \mathbf{CS} + \mathbf{DG} + \mathbf{U}$ where the rows of $\mathbf{U}$ are jointly independent random vectors so that

$\Pr(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N) = \Pr(\mathbf{u}_1) \times \Pr(\mathbf{u}_2) \times \cdots \times \Pr(\mathbf{u}_N)$.

Also, for all $i = 1, 2, \ldots, N$, $\mathbf{u}_i \neq \mathbf{0}$ and $\mathbf{u}_i = h_i(\mathbf{e}_i)$ for a non-random Borel measurable function h_i.

Corollary Under the assumptions of this Proposition, all population level multiple testing dependence is removed when conditioning on both $\mathbf{X}$ and a dependence kernel $\mathbf{G}$. Similar result holds for estimation-level multiple testing dependence, when conditioning on $\mathbf{X}$ and a dependence kernel $\mathbf{G}$, and estimators $\hat{\mathbf{C}}$ of C and $\check{\mathbf{D}}$ of $\mathbf{D}$.

Different methods were developed for strongly controlling the family-wise error rate (FWER) or FDR mentioned in introduction. Most of these methods, e. g. finite-sample strong control of several FDR procedures and the conservative point estimation of the FDR require the true null P values calculated from multiple hypothesis tests to be independent in order for the procedure to provide strong control.

There is a following important both historical and methodological connection between FDR and Stein estimators,

For $x_i \propto N(\mu_i,1)$, $i = 1, 2, \ldots, N$, the introduction of Stein result, concerns with testing N null hypotheses Benjamini and Hochberg's original 1995 paper (BH), which was published almost 40 years after $H_{0i} : x_i \propto N(\mu_i,1)$, with $\mu_i = 0$ $H_{1i} : \mu_i \neq 0$

$\hat{G}(x)$ denotes the right-sided CDF of the N x-values with $G(x) = 1-\Phi(x)$, the right-sided null hypothesis CDF, and k is the proportion of true null hypothesis,

the estimated FDR $\boldsymbol{FDR}(x) = kG(x)/\hat{G}(x)$, BH set k equal to its upper bound 1.

Proposition (BH) The expected proportion of "false discoveries" produced by this rule; that is, the rejected null hypotheses that are actually true, is less than FDR control rule chosen at control level q, say $q = .1$, to compute $x_0 = \min_x(\boldsymbol{FDR}(x) \leq q)$ The frequentist philosophy of FDR was followed by Bayesian modeling.

4. DNA, RNA sequencing, tests of χ_1^2, Poisson, and other assumptions

Data from next generation sequencing (NGS) technologies posed new computational and statistical challenges because of their massive size, complicated structure (large number of genes), limited information, and discreteness. NGS data are more complex than data from previous high

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

throughput technologies such as microarrays that were increasingly utilized for genetic testing of individuals with unexplained developmental disorders. From a statistical perspective, the theory of multiple hypotheses testing (Efron et al., 2001),
and the use of false discovery rates (FDR) for multiple testing problems (Benjamini and Hochberg, 1995; Storey, 2003; Efron, 2010), which were motivated by
microarray data, also apply to NGS data analysis with modifications to acknowledge the data specific features (Oshlack et al., 2010)
Let n_{gst} be the observed count of gene g in the sample s with treatment t, and $\theta\, gt$ is the expected value of n_{gst} . The approach is similar to that of microarray analysis, to obtain
test statistics or posterior distributions, for testing gene-wise differential expression,
$\mu_{g1} - \mu_{g2}$ between the two treatment groups. Such approach has its important application in clinical trials for the mentioned above cluster randomized trials or mixed models repeated measures models that are closely related to such fields as epidemiology and immunology and are discussed later. However, the sampling distribution of the test statistic for
NGS data analysis is difficult to obtain without restrictive distributional assumptions.
When analyzing microarray data it is quite common to assume the data follow Gaussian distribution after normalization and log transformation. As such, for each gene, the test statistic for differential expression follows a t-distribution with $(2S\text{-}2)$ degrees of freedom (Efron et al.,2001).
However, in the last 4-5 years many works used FDR and multiple TSH application
to determine if gene differential expression follows Poisson distribution and χ_1^2 , or Negative Binomial assumptions that are often used to model reads from RNA-seq data. Although a Poisson model may be appropriate for technical replicates when variability is lower, with higher variability in biological replicates or surrogate replicates, the Poisson model does not control Type-I error and underestimates the variability, as overdispersion can be observed. Because multiple tests are being performed across the set of genes, q-values are reported that control the false discovery rate using the Benjamini–Hochberg procedure [51].

Data were gathered from 75 RNA-seq experiments conducted in five different bacteria: E. coli, N.gonorrhoeae, S. enterica, S. pyogenes and X. nematophila. Altogether, the RNA-seq experiments yielded over two billion sequencing reads corresponding to 189 billion nt[51]. Without application of TSH it has been deemed impossible to observe that real data for gene sequencing does not follow assumptions of χ_1^2, Poisson, and other distributions.

4.1 TSH, multiple TSH, and FDR have their applications also in Cluster randomized trials for clinical trials, Theories related to astrophysics such as hypotheses of N-body mass estimations, a general relativistic null hypothesis test with event horizon telescope observations of the black-hole shadow, ecology, and Heat transport dynamics[37,46-49].

5. Conclusion

Large-scale multiple testing has been applied in fields such as genomics[34, 40-43], astrophysics [, and spatial epidemiology. Although the procedure seems to be very simple, one may have the perception that it is difficult hence try to apply it in special circumstances, e.g. facing one of 5 types of bivariate exponential distribution, [23-27,32], or one of 5 types of bivariate gamma

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

distribution[28-31,18] in the theory of ratios of random variables, or even in special cases with Gaussian distribution assumptions, such as in RNA-seq.

There are five main aspects of mathematical analysis of probability in the testing of statistical hypothesis (TSH).

The 5 properties involve the analysis of a priori and posterior along with the concept of recurrence, random sequences and functions [1-7, 10-14]. TSH with familiar concepts of decision theory loss and risk functions are a continuation of basic algebraic operations over random variables. The basic algebraic operations also use the random variables along with differential and integral calculus over them, with further extension to a multi-dimensional model[1-3]. This requires application of special mathematical interest in search of generalized and uniform functions and solutions [16, 17]. TSH also can be applied to check for the universality of solutions of different problems.

References:

1. A. De Moivre, "The Doctrine of Chances", 2nd ed. (London, England: H. Woodfall, 1738),
2. Pierre Simon, Marquis De Laplace "Analytical theory of Probability"
3. Pierre Simon, Marquis De Laplace "Philosophical essay on Probability"
4. Peirce, Charles S. (c. 1909 MS), Collected Papers v. 6, paragraph 327
5. Richard von Mises "Probability, Statistics, and Truth" Dover/81 ISBN: 978-0-486-24214-9
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. Josef Steinebach "On a necessary condition for the Erdos-Renyi law of large numbers" Proceedings of the American Mathematical Society Volume 68, Number 1, January/78
10. Donald Loveland "A New Interpretation of the von Mises' Concept of Random Sequence" Mathematical Logic Quarterly V 12, Is 1,pages 279–294, 1966
11. Se´rgio B. Volchan "What is a random sequence? Mathematical Association of America [Monthly 109 January 2002]
12. Michiel van Lambalgen "Randomness and foundations of probability: von Mises' axiomatisation of random sequences",
13. Michiel van Lambalgen Statistics, Probability and Game Theory IMS Lecture Notes - Monograph Series (1996) Volume 30.
14. Michiel Van Lambalgen "Von Mises' Definition of Random Sequences Reconsidered". Journal of Symbolic Logic 12/2002; 52(3). DOI: 10.2307/2274360
15. J. Doob "William Feller and twentieth century probability" Proceddings of the 4thBerckley Symposium on Probability and Statistics
16. Keith Briggs "Another universal differential equation"
17. Rubel, L. A. [1981], 'A universal differential equation', Bull. Am. Math. Soc. **4**, 345–349.
18. Saralees Nadarajah and Samuel Kotz Bivariate gamma distributions, sums and ratios Bull Braz Math Soc, New Series 37(2), 241-274. 2006, SociedadeBrasileira de Matemática
19. C. Stein, "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," in Proc. Third Berkeley Symp. On Math. Statist. and Prob., vol. 1, pp. 197-206,
20. C. Stein. Estimation of the mean of a multivariate normal distribution. Ann. Stat.,9(6):1135–1151, 1981.
21. Enrique Castillo "Functional Equations and Modelling in Science and Engineering" Springer
22. Caponnetto, E. De Vito, M. Pontil "Entropy conditions for L_r-convergence of empirical processes"
23. Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curvesfrom incomplete data. J. Amer. Statist. Assoc. 71 897{902.
24. Block, H.W. and Basu, A.P. (1974). A continuous bivariate exponential extension, Journal of the American Statistical Association, 69, 1031-1037
25. Freund, J.E. (1961). A bivariate extension of the exponential distribution, Journal of the American Statistical Association, 56, 971-977
26. Gumbel, E.J. (1960). Bivariate exponential distributions, Journal of the American Statistical Association, 55, 698-707

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2016 doi:10.20944/preprints201607.0069.v1

27. *P. A. P. Moran Testing for correlation between non-negative variates Biometrika(1967)54 (3-4):385-394*
28. P.J. Becker and .J.J. Roux, A bivariate extension of the gamma distribution. SA Statistical Journal, **15**, 1–12.
29. A.T. McKay, Sampling from batches. Journal of the Royal Statistical Society,Supplement, **1** (1934), 207–216.
30. K.C. Cherian, A bivariate correlated gamma type distribution function. Journal ofthe Indian Mathematical Society, **5** (1941), 133–144.
31. W.F. Kibble, A two-variate gamma type distribution. Sankhya, **5** (1941), 137–150.
32. Marshall, A.W. and Olkin, I. (1967), A Multivariate Exponential Distribution, J. Amer. Stat. Assoc., 63, pp. 30-44.
33. M. D. C. Pardo; J. A. Pardo Statistical applications of order α-β weighted information energy Applications of Mathematics, Vol. 40 (1995), No. 4, 305—317
34. Alicia Oshlack, Mark D Robinson Matthew D Young "From RNA-seq reads to differential expression results." Genome Biology November/10. 10 pages article with 100 references.
35. I. M. Gelfand, et al "Generalized functions" vv. 1-5 Academic Press, New York and London**.**
36. Gerald Paul, H Eugene Stanley "Partial test of the Universality Hypothesis: The case of different coupling strengths in different lattice directions. "20 pages article with 80 references
37. D. R. Anderson, K.P. Burnham Kullback-leibler Information as a basis for strong inference in ecological studies" Wildlife Research, 28, 111-119
38. G.Polenta, D.Marintucci, A.Balbi, P.de Bernardis, E.Hivon, S.Masi, P.Natoli, N.Vittorio "Unbiased Estimation of an Angular Power Spectrum" Astrophysics
39. P. J. Marshall, M. P. Hobson, S. F. Gull and S. L. Bridle Maximum-entropy weak lens reconstruction: improved methods and application to data Monthly Notices of the Royal Astronomical Society Volume 335, Issue 4, pages 1037–1048, October/02
40. Bradley Efron Large-Scale Inference:Empirical Bayes Methods for Estimation, Testing and Prediction
41. Tutorial in biostatistics: multiple hypothesis testing in genomics Statist. Med. 2012, 00 1-27 Jelle J. Goemanan, Aldo Solarib
42. Kenneth F. Manly, Dan Nettleton and J.T. Gene Hwang Hypotheses Genomics, Prior Probability, and Statistical Tests of Multiple Hypotheses.
43. Jeffrey T. Leek , John D. Storey A general framework for multiple testing dependence
44. T. A. B. Snijders Hypothesis Testing: Methodology and Limitations.
45. Ullah, A. & D.E.A. Giles. The positive-part Stein-rule estimator and tests of hypotheses. Economics Letters 26 (1988): 49-52.
46. H. Spohn et al Numerical test of hydrodynamic fluctuation theory in the Fermi-Pasta-Ulam chain PHYSICAL REVIEW E **90**, 012124 (2014)
47. H.M. Hudson "A Natural Identity for exponential families with applications in multiparameter estimation" The Annals of Statistics V.6, N 3
48. Martin Vetterli et al "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback–Leibler Distance" *IEEE*
49. M. Ghosh D. K. Dey Trimmed Estimates in Simultaneous Estimation of Parameters in Exponential Families Journal of multivariate analysis 15, 183-200 (1984)
50. P. Kannappan "On some functional equations from additive and nonadditrve measures"
51. Ryan McClure et al "Computational analysis of bacterial RNA-Seq data"
52. S. Eguchia, J. Copas "Interpreting Kullback–Leibler divergence with the Neyman–Pearson lemma" Journal of Multivariate Analysis 97 (2006) 2034 – 2040
53. B. Póczos, L. Xiong, and J. Schneider. "Nonparametric divergence estimation with applications to machine learning on distributions"

© 2016 by the author; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).