PROCEEDINGS
OF **SCIENCE**

Two approaches to Combining Significances

Sergey Bityukov[*][†]

Institute for high energy physics, 142281 Protvino, Russia
E-mail: Serguei.Bitioukov@cern.ch

Nikolai Krasnikov

Institute for nuclear research RAS, Moscow, Russia
E-mail: Nikolai.Krasnikov@cern.ch

Alexandre Nikitenko

Imperial College, London, United Kingdom, on leave from ITEP, Moscow
E-mail: Alexandre.Nikitenko@cern.ch

Vera Smirnova

Institute for high energy physics, 142281 Protvino, Russia
E-mail: Vera.Smirnova@ihep.ru

We discuss two approaches to the combining of signal significances: the approach, in which the signal significances are considered as corresponding random variables, and the approach with the using of confidence distributions. We consider several signal significances, which are often used in analysis of data in experimental physics as a measure of excess of the observed or expected number of signal events above the predicted number of background events. We show the applicability of Stouffer's method to combine of these significances.

---

1. Introduction

One of a basic problems in science is how best to consolidate data from several experiments that are designed to measure the same unknown quantity. These experiments may differ in time, geographical location, and laboratory apparatus, and may also differ in underlying theory. The following quotation seems quite relevant: "Suppose one experiment sees a 3-sigma effect and another experiment sees a 4-sigma effect. What is the combined significance? Since the question is ill-posed, the statistics literature contains many papers on the topic ?" [1].

Methodology for combining findings from repeated studies did in fact begin with the idea of combining independent tests back in the 1930's (Tippett, 1931 [2]; Fisher, 1932 [3]; Pearson, 1933 [4]). There are many approaches to this subject. Many of them is discussed in cited above review of R. Cousins [1].

We consider the using of one (Stouffer et al., 1949 [5]) of these methods for combining of significances. We show the applicability of this method in the case of Poisson flows of events under study. We also discuss the approach based on confidence distributions. This approach shows an applicability of Stouffer's method for combining of significances under certain conditions.

2. Combination of tests

In this Section we follow the content and ideas presented in Workshop on Meta Analysis [6] (Tampere-Turku, 2006). Meta-analysis, a term coined by Glass [7], is intended to provide the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings.

All the methods of combining tests depend on what is known as a $P-$value. A key point is that the observed $P-$values derived from continuous test statistics follow a uniform distribution under the null hypothesis H_0 regardless of the form of the test statistic, the underlying testing problem, and the nature of the parent population from which samples are drawn.

Quite generally, suppose $X_1, \ldots, X_n$ is a random sample from a certain population indexed by the parameter θ, and $T(X_1, \ldots, X_n)$ is a test statistic for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$, where θ_0 is a null value, and suppose also that H_0 is rejected for large values of observed $T(x_1, \ldots, x_n)$. Then if the null distribution of $T(X_1, \ldots, X_n)$ is denoted by $g(t)$, the (one-sided) $P-$value based on $T(X_1, \ldots, X_n)$ is defined as

$$P = \int_{T(x_1,\ldots,x_n)}^{\infty} g(t)dt = P[T(X_1,\ldots,X_n) > T(x_1,\ldots,x_n)|H_0] \tag{2.1}$$

which stands for the probability of observing as extreme a value of statistic $T(X_1, \ldots, X_n)$ as the observed one $T(x_1, \ldots, x_n)$ under the null hypothesis. Here $x_1, \ldots, x_n$ denote the observed realization of the X_i's. Since the null hypothesis H_0 is rejected for large values of $T(x_1, \ldots, x_n)$, this is equivalent to reject H_0 for small values of P.

In most meta analysis applications, the $P-$values are computed from the approximate normal distribution of the relevant test statistics. Thus, if $T(X_1, \ldots, X_n)$ is approximately normally distributed with mean $\mu(\theta)$ and variance $\sigma^2(\theta, n)$, the $P-$value is computed as

$$P = P[T(X_1,\ldots,X_n) > T(x_1,\ldots,x_n)|H_0] = P[N(0,1) > \frac{T(x_1,\ldots,x_n) - \mu(\theta_0)}{\sigma(\theta_0,n)}]. \qquad (2.2)$$

In general we can consider k different studies in which test problems H_{0i} versus H_{1i} are considered, $i = 1,\ldots,k$. A combined test procedure tests the global null hypothesis

$$H_0 : \ All \ H_{0i} \ true \quad i = 1,\ldots,k$$

versus the alternative

$$H_1 : \ Some \ of \ the \ H_{1i} \ true.$$

There are two general properties of a combined test procedure:

- admissibility: a combined test procedure is said to be admissible if it provides a (not necessarily the only) most powerful test against some alternative hypothesis for combining some collection of tests;

- monotonicity: a combined test procedure is said to be monotone if the combined test procedure rejects the null hypothesis H_0 for one set of $P-$values and it must also reject the hypothesis for any set of component-wise smaller $P-$values.

As shown in [8]: every monotone combined test procedure is admissible and therefor optimal for some testing situation.

Two broad classes of combined tests based on the $P-$values:

- uniform distribution methods, e.g. Tippett's method [2] and Wilkinson's method [9];

- probability transformation methods, e.g. Stouffer's method [5], Fisher's method [3], and logit method [10].

Here we provide a comprehensive review of these so-called omnibus or non-parametric statistical methods [1]. Each of the methods described below satisfies the monotonicity principle.

Minimum P method. Tippett's [2] minimum P test rejects the null hypothesis H_0 if any of the k $P-$values is less than α^*, where $\alpha^* = 1 - (1-\alpha)^{\frac{1}{k}}$. In other words, we reject H_0 if

$$min(P_1,\ldots,P_k) = P_{[1]} < \alpha^* = 1 - (1-\alpha)^{\frac{1}{k}}. \qquad (2.3)$$

Wilkinson's method. This method, due to Wilkinson [9], rejects H_0 if the rth smallest $P-$value, $P_{[r]}$ is small, i.e., less than some c for some fixed r. Since under H_0, $P_{[r]}$ follows a beta distribution with the parameters r and $k - r + 1$, it is easy to determine the cut-off point c for this test from the following equation:

---

[1]The problem of selecting a test for H_0 is complicated by the fact that there are many different ways in which the omnibus null hypothesis H_0 can be false.

$$\alpha = \int_0^c \frac{u^{r-1}(1-u)^{k-r}}{B(r,k-r+1)} du, \tag{2.4}$$

where $B(.,.)$ is the usual beta function.

Stuffer's method. This method is due to Stouffer and his colleagues [5], also called inverse normal method. It is based on the fact that the $z-$value based on the $P-$value, defined as

$$\tilde{z} = \Phi^{-1}(P) \tag{2.5}$$

is a standard normal variable under the null hypothesis H_0, where $\Phi(.)$ is a standard normal cumulative distribution function (cdf). Thus, when the $P-$values $P_1, \ldots, P_k$ converted to the $\tilde{z}-$values $\tilde{z}_1, \ldots, \tilde{z}_k$, we have independent identically distributed (iid) standard normal variables under H_0. The combined significance test is essentially based on the sum of these $\tilde{z}-$values, which has a normal distribution under the null hypothesis with mean 0 and variance k. The test statistic

$$\tilde{Z} = \frac{\sum_{i=1}^k \tilde{z}(P_i)}{\sqrt{k}} \tag{2.6}$$

is thus a standard normal variable under H_0, and hence can be compared with the critical values in the standard normal table.

Some authors (see next Section) suggest to compute the z scores from $P-$values by using the formula

$$z = \Phi^{-1}(1-P). \tag{2.7}$$

If this is done, the resulting value, say $Z-$value, i.e. significance of an enhancement, will be large for small values of P, implying thereby that H_0 is rejected when Z is large.

Fisher's method. This method, which is is a special case of the inverse chi-square transform, was described by Fisher [3], and is widely used in meta analysis. This method is based on the fact that the variable $-2lnP$ is distributed as a chi-square variable with 2 degrees of freedom under the null hypothesis whenever P has a uniform distribution. The sum of k of these values is therefore a chi-square variable with $2k$ degrees of freedom under H_0. The test thus reject H_0 when $-2\sum_{i=1}^k lnP_i$ exceeds the $100(1-\alpha)$% critical value of the chi-square distribution with $2k$ degrees of freedom.

Logit method. George [10] proposed this method using a statistic

$$G = -\sum_{i=1}^k ln\left(\frac{P_i}{1-P_i}\right)\left[\frac{k\pi^2(5k+2)}{3(5k+4)}\right]^{-\frac{1}{2}} \tag{2.8}$$

as another combined significance technique. The argument is that the logit (i.e., $ln\frac{P}{1-P}$ is distributed as a logistic variable under H_0, and further that the distribution of the sum of the logits, suitable normalized, is close to the t distribution. There are usually two approximations of the null distribution of G which can be used. First, we can approximate the null distribution of G with the t distribution based on $(5k+4)$ degrees of freedom. The test based on this approximation rejects H_0 if G exceeds the $100(1-\alpha)$% critical value of the t distribution with $(5k+4)$ degrees of freedom.

Another approximation is based on the observation that, under H_0, $ln(\frac{P_i}{1-P_i})$ could be viewed as approximately normal with a zero mean and variance of $\frac{\pi^2}{3}$. The test based on this approximation therefore rejects H_0 when

$$G = -[\sum_{i=1}^{k} ln(\frac{P_i}{1-P_i})][\frac{3}{k\pi^2}]^{\frac{1}{2}} \tag{2.9}$$

exceeds z_α.

There is no general recommendation for the choice of the combination method. All the combination methods are optimal for some testing situations. Note, combining $P-$values can lead to incorrect conclusions because

- acceptance or rejection can depend more on the choice of the statistic than on the data,

- the information in a highly informative experiment can be masked, and thereby largely disregarded.

A $P-$value itself is not as informative as the estimate and standard error on which it is based. If this more complete summary information about a study is available, it makes good sense to use it and avoid $P-$values altogether. However, methods that combine $P-$values have their place when such precise information is unavailable.

3. Zoo of significances of an enhancement

"Common practice is to express the significance of an enhancement by quoting the number of standard deviations" [11]. Let us define a significance Z (in high energy physics the significance often is denoted as S) in accordance with Eq.2.7 [1]

$$Z = \Phi^{-1}(1-p) = -\Phi^{-1}(p), \tag{3.1}$$

where

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{\frac{-t^2}{2}} dt = \frac{1+erf(\frac{Z}{\sqrt{2}})}{2}, \tag{3.2}$$

so that

$$Z = \sqrt{2} erf^{-1}(1-2p). \tag{3.3}$$

For example, $Z = 5$ corresponds to a $p-$value of $2.87 \cdot 10^{-7}$. One can see the relation between some probability p, which estimates the uncertainty in hypotheses testing, and with the corresponding number of standard deviations Z in the frame of standard normal distribution.

Z characterizes the significance of the deviation of one value from another value (for example, signal s + background b from background b). The choice of significance to be used depends on the study. There are three types of significances.

A. If s and b are expected values then we take into account both statistical fluctuations of signal and of background. Before observation we can calculate only an expected (or an internal) significance Z which is a parameter of experiment. Z characterizes the quality of experiment.

B. If $\widehat{s+b}$ is observed value and b is expected value then we take into account only the fluctuations of background. In this case we can calculate an observed significance $\hat{Z}$ which is an estimator of expected significance of experiment Z. $\hat{Z}$ characterizes the quality of experimental data.

C. If $\hat{s}$ and $\hat{b}$ are observed values with known errors of measurement then we can use the standard theory of errors for estimation of the significance of enhancement of signal.

Different significances are used. For example, the significances Z_{Bi} (Binomial) = Z_Γ (Gamma), Z_N (Bayes Gaussian), Z_{PL} (Profile Likelihood) were studied in details in paper [12]. As shown in ref. [13] two types (case A and case B) of significances can be considered under certain conditions as normal random variables with variance close to 1. The significances S_{c12} and Z_N (S_{cP} in HEP) satisfy this property.

- S_{c12} [14] corresponds to the case of hypotheses testing of two simple hypotheses $H_0 : \theta = b$ against $H_1 : \theta = s + b$, where $S_{c12} = 2(\sqrt{s+b} - \sqrt{b})$ is an expected significance (case A).

- Z_N as proposed in ref. [15] is the probability from Poisson distribution with mean b to observe equal or greater than $s + b$ events, converted to equivalent number of sigmas of a Gaussian distribution. It is the case of hypotheses testing with $H_0 : \theta = b$ against $H_1 : \theta > b$. Often Z_N is used as observed significance (case B).

Let us show the applicability of the Stouffer's method to combining of significances. We present here only the results for S_{c12}. Results for Z_N (S_{cP}) [13] are analogous.

4. The Monte Carlo study of combining significances

Let us consider the distributions of possible values of observed significance $\hat{S}_{c12}$ for experiment without signal due to fluctuation of background. Distributions of observed $\hat{S}_{c12}$ in the case of signal absence for $3 \cdot 10^6$ simulated experiments for several values of expected b ($b = 40$, 50, 60, 6, correspondingly) are presented in Fig. 1 (left). The tests allow to say in our case that this significance under hypothesis $H_0 : s = 0$ obeys the standard normal distribution in wide range of values of b.

We use the method which allows to connect the magnitude of the observed significance with the confidence density (this notion is introduced in next Section) of the parameter "the expected significance". We carried out the uniform scanning of expected significance S_{c12}, varying S_{c12} from 1 up to 16, using step size 0.075. By playing with the two Poisson distributions (with parameters s and b) and using $3 \cdot 10^5$ trials for each value of expected S_{c12} we construct the conditional distributions of the probability of the production of the observed value of significance $\hat{S}_{c12}$ by the experiments with expected significance S_{c12}. Integrated luminosity of each experiment is a constant $s + b$. The parameters s and b are chosen in accordance with the given expected significance S_{c12}, the realization N_{obs} (or $\widehat{s+b}$) is a sum of realizations N_s (or $\hat{s}$) and N_b (or $\hat{b}$).

The distributions of observed significance $\hat{S}_{c12}$ for several values of expected significance S_{c12} (with the given integrated luminosity $s + b = 70$) are presented in Fig. 1 (right). The observed
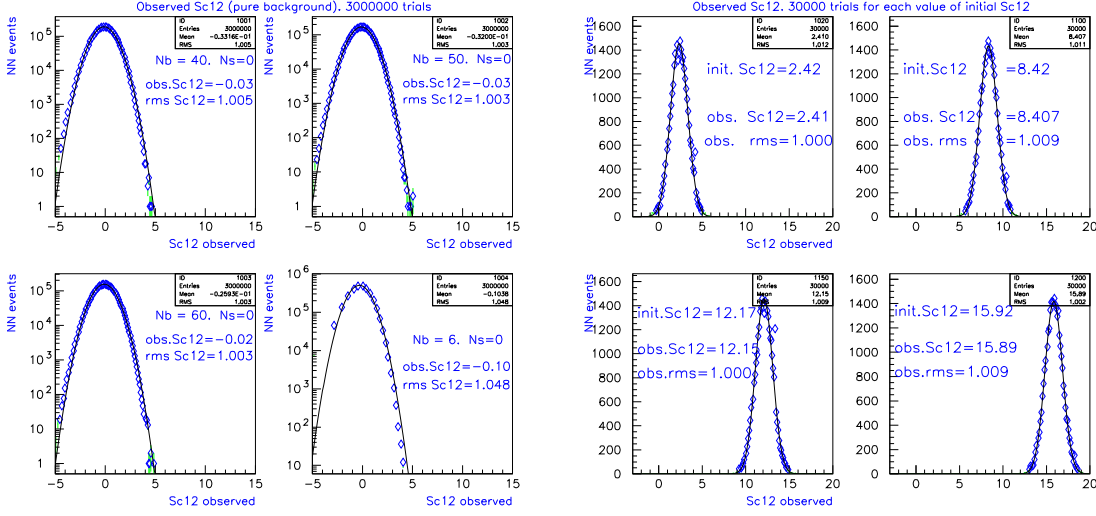
Figure 1: Distributions of observed $\hat{S}_{c12}$ in the case of signal absence for $3 \cdot 10^6$ simulated experiments for several values of expected b (left). The distributions of observed significance $\hat{S}_{c12}$ for several values of expected significance S_{c12} with the given integrated luminosity $s + b = 70$ (right).

distributions of significances are similar to the distributions of the realizations of normal distributed random variable with variance which close to 1.

The distribution of the observed significance $\hat{S}_{c12}$ versus the expected significance S_{c12} shows the result of the full scanning (Fig.2 (left)). The normal distributions with a fixed variance are statistically self-dual distributions [16]. It means that the confidence distribution of the parameter "expected significance" S_{c12} has the same distribution as the random variable which produced a realization of the observed significance $\hat{S}_{c12}$ (compare Fig. 1 (right) and Fig. 2 (right)).

The several distributions, which characterize the probability (after normalization) of the expected significances S_{c12} to produce the given observed values of $\hat{S}_{c12}$, are presented in Fig. 2 (right). These figures clearly show that the observed significance $\hat{S}_{c12}$ is an unbiased estimator of the expected significance S_{c12}.

The observed significance $\hat{S}_{c12}$ (the case of the Poisson flow of events) is a realization of the random variable which can be approximated by normal distribution with variance close to 1 (for example, it is a standard normal distribution $\mathscr{N}(0,1)$ in the case of pure background without signal). It means that with this observed significance one can work as with the realization of the random variable.

Let us define the observed summary significance $\hat{Z}_{sum}$, the observed combined significance $\hat{Z}_{comb}$ and the observed mean significance $\hat{Z}_{mean}$ for the k partial observed significances $\hat{Z}_i$ with standard deviation $\sigma(Z_i) \sim 1$:

$$\hat{Z}_{sum} = \sum_{i=1}^{k} \hat{Z}_i, \quad \sigma^2(\hat{Z}_{sum}) = \sum_{i=1}^{k} \sigma^2(Z_i), \tag{4.1}$$
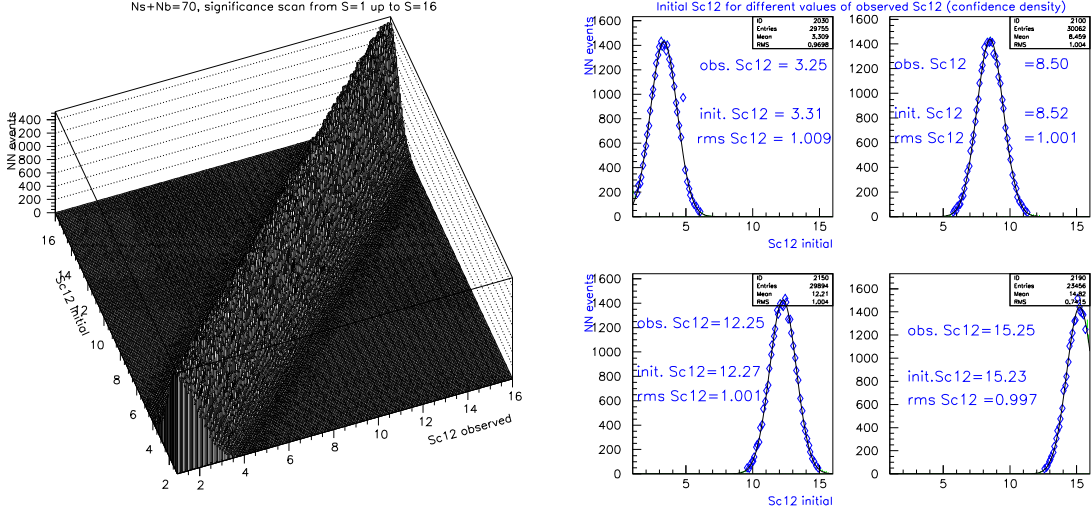
Figure 2: The distributions of observed significance $\hat{S}_{c12}$ for several values of expected significance S_{c12} with the given integrated luminosity $s+b = 70$ (left). The distributions of the expected significances, which produced the given observed significance $\hat{S}_{c12}$. These distributions after normalization are conditional probability distributions (confidence densities) of the expected significances S_{c12} to produce the given observed significance (right).

$$\hat{Z}_{comb} = \frac{\hat{Z}_{sum}}{\sqrt{\sigma^2(\hat{Z}_{sum})}}, \tag{4.2}$$

$$\hat{Z}_{mean} = \frac{\hat{Z}_{sum}}{k}. \tag{4.3}$$

The ratio of the sum of the several partial observed significances and the standard deviation of this sum is the estimator of the combining significance of several partial observed significances. The formula 4.2 is essentially the formula of Stouffer's method, because $\sigma(Z_i) \sim 1$ and, correspondingly, $\sqrt{\sigma^2(\hat{Z}_{sum})} \sim \sqrt{k}$.

It can also be shown by a Monte Carlo simulation. Let us generate the observation of the significances for four experiments with different parameters b and s simultaneously. The results of this simulation ($3 \cdot 10^5$ trials) for each experiment are presented in Fig. 3 (left). The distribution of the sums of four observed significances $\hat{S}_{c12_i}$, $i = 1,4$ of experiments (i.e. $\hat{Z}_{sum}$) in each trial and the distribution of $\hat{Z}_{comb}$ in each trial is shown too (Fig. 3 (right)).

Note, the combination of partial significances Z_1 and Z_2 combine with third partial significances Z_3 according to formula

$$\frac{Z_1 + Z_2}{\sqrt{2}} \cdot \frac{\sqrt{2}}{\sqrt{3}} + Z_3 \cdot \frac{1}{\sqrt{3}} = \frac{Z_1 + Z_2 + Z_3}{\sqrt{3}}. \tag{4.4}$$

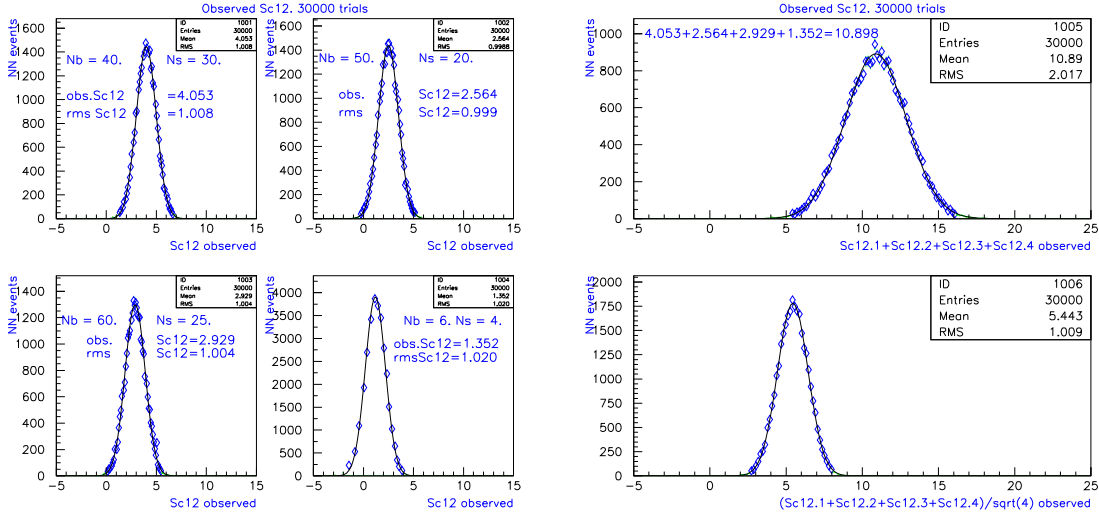The problem of weights is discussed in ref. [17] in details.

Figure 3: The results of the simulation ($3 \cdot 10^5$ trials) of the observed significance $\hat{S}_{c12}$ for four different experiments (left). The distribution of the $\hat{Z}_{sum}$ (right, up) and the distribution of the $\hat{Z}_{comb}$ (right, down) which are produced in each of trials.

5. Combining information from independent sources through confidence distributions

In this Section we follow the results presented in paper [18]. Suppose $X_1, X_2, \ldots, X_n$ are n independent random draws from a population **F** and χ the sample space corresponding to the data set $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)^T$. Let θ be a parameter of interest associated with **F** (**F** may contain other nuisance parameters), and let Θ be the parameter space.

<u>Definition</u>: *A function $H_n(\cdot) = H_n(X_n, (\cdot))$ on $\chi \times \Theta \rightarrow [0, 1]$ is called a* **confidence distribution** *(CD) for a parameter θ if*
(i) for each given $\mathbf{X}_n \in \chi$, $H_n(\cdot)$ is a continuous cumulative distribution function;
(ii) at the true parameter value $\theta = \theta_0, H_n(\theta_0) = H_n(\mathbf{X}_n, \theta_0)$, as a function of the sample $\mathbf{X}_n$, has the uniform distribution $U(0, 1)$.
We call, when it exists, $h_n(\theta) = H_n^{'}(\theta)$ a **confidence** *or CD* **density**.

Item *(i)* requires the function $H_n(\cdot)$ to be a distribution function for each given sample.

Item *(ii)* states that the function $H_n(\cdot)$ brings the information onto probability scale and thus provides confidence intervals and $p-$values.

The notion of a CD is attractive for the purpose of combining information. The main reasons are that the CDs are relatively easy to construct and interpret and there is a wealth of information on θ inside a CD.

Let $H_1(y), \ldots, H_k(y)$ be k independent CDs, with the same true parameter value θ_0. Suppose $g_c(U_1, \ldots, U_k)$ is any continuous function from $[0, 1]^k$ to R that is monotonic in each coordinate. A general way of combining, depending on $g_c(U_1, \ldots, U_k)$ can be described as follows:

Define $H_c(U_1,\ldots,U_k) = G_c(g_c(U_1,\ldots,U_k))$, where $G_c(\cdot)$ is the continuous cumulative distribution function of $g_c(U_1,\ldots,U_k)$, and $U_1,\ldots,U_k$ are independent $U(0,1)$ distributed random variables.

Denote $H_c(y) = H_c(H_1(y),\ldots,H_k(y))$. It is easy to verify that $H_c(y)$ is a CD function for the parameter θ. $H_c(y)$ is a combined CD.

Let $F_0(\cdot)$ be any continuous cumulative distribution function and $F_0^{-1}(\cdot)$ be its inverse function. A convenient special case of the function g_c is

$$g_c(U_1,\ldots,U_k) = F_0^{-1}(U_1) + \ldots + F_0^{-1}(U_k).$$

In this case, $G_c(\cdot) = F_0 * \ldots * F_0(\cdot)$, where $*$ stands for convolution. Just like the p-value combination approach, this general CD combination recipe is simple and easy to implement. Some examples of F_0 are:

- $F_0(t) = \Phi(t)$ is the cumulative distribution function of the standard normal distribution. In this case

$$H_{NM}(y) = \Phi(\frac{1}{\sqrt{k}}[\Phi^{-1}(H_1(y)) + \ldots + \Phi^{-1}(H_k(y))]).$$

  This formula leads to simple Stouffer's formula for combined $p-$values [5].

- $F_0(t) = 1 - e^{-t}$, for $t \geq 0$, is the cumulative distribution function of the standard exponential distribution (with mean 1). In this case the combined CD is

$$H_{E1}(y) = P(\chi^2_{2k} \leq -2\sum_{i=1}^{k} log(1 - H_i(y))),$$

  where χ^2_{2k} is a χ^2-distributed random variable with $2k$ degrees of freedom [2].

- $F_0(t) = \frac{1}{2}e^t \mathbf{1}_{(t \leq 0)} + (1 - \frac{1}{2}e^{-t})\mathbf{1}_{(t \geq 0)}$, denoted as $DE(t)$ from now on, is the cumulative distribution function of the standard double exponential distribution. Here $\mathbf{1}_{(.)}$ is the indicator function. In this case the combined CD is

$$H_{DE}(y) = DE_k(DE^{-1}(H_1(y)) + \ldots + DE^{-1}(H_k(y))),$$

  where $DE_k(t) = DE * \ldots * DE(t)$ is a convolution of k copies of $DE(t)$ [3].

One can see, that this approach also leads to the formula of Stouffer for the case of combining of normal distributed significances.

---

[2]It is the Fisher's omnibus method [3].

[3]Combining the double exponential maximizes Bahadur slope.

6. Conclusion

As shown, the Stouffer's method of combining significances works for significances which obey the normal distribution. The significances S_{c12}, Z_N, Z_{Bi}, and Z_{PL} satisfy to the criterion of normality in wide range of values s and b in Poisson flows.

The choice of the combination method depends on many factors. As seems, the confidence distributions are often convenient for combining information from independent sources. This approach also leads to the Stouffer's formula in our case.

Note, any of methods for combining $P-$values, which is discussed in ref. [1], can be used for combining significances.

Acknowledgments

References

[1] R.D. Cousins, *Annotated Bibliography of Some Papers on Combining Significances or p -values*, `arXiv:0705.2209 [physics.data-an]`.

[2] L. Tippett, *The Methods of Statistics*, Williams and Norgate, Ltd., London, 1st edition. Sec. 3.5, 53-6, 1931, as cited by Birnbaum and by Westberg.

[3] R.A. Fisher, *Statistical Methods for Research Workers*, Hafner, Darien, Connecticut, 14th edition, 1970. The method of combining significances to have appeared in the 4th edition, Oliver & Boyd, 1932.

[4] K.Pearson, *On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random*, *Biometrika*, **25(3/4)** (1933) 379-410.

[5] S. Stouffer, E. Suchman, L. DeVinnery, S. Star, and R.W. Jr, *The American Soldier*, volume I: Adjustment during Army Life. Princeton University Press, 1949.

[6] *Workshop on meta analysis*, Un.of Tampere, Un. of Turku, October 4-6, 2006, Instructors: B. Sinha, G. Knapp, L. Koskela, 2006.

[7] G.V. Glass, *Primary, secondary, and meta-analysis*, *Educational Researcher*, **5**, 3-8, 1976.

[8] A. Birnbaum, *Combining independent tests of significance*, *J. of the American Statistical Association*, **49**, 559-575, 1954.

[9] B. Wilkinson, *A statistical consideration in psychological research*, *Psychological Bulletin*, **48**, 156-158, 1951.

[10] E.O. George, *Combining independent one-sided and two-sided statistical tests – Some theory and applications. Doctoral dissertation*, Un. of Rochester, 1977.

[11] A.G.Frodesen, O.Skjeggestad, H.Toft, *Probability and Statistics in Particle Physics,* UNIVERSITETSFORLAGET, Bergen-Oslo-Tromso, p.97, 1979.

[12] R.D. Cousins, J.T. Linnemann, J. Tucker, *Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process, Nucl.Instr.&Meth.*, **A595**, 480-501, 2008.

[13] S. Bityukov, N. Krasnikov, A. Nikitenko, *On the Combining Significances*, `arXiv: physics/0612178`.

[14] S.I. Bityukov, N.V. Krasnikov, *New physics discovery potential in future experiments, Mod.Phys.Lett.* **A13**, 3235-3249, 1998.

[15] I. Narsky, *Estimation of upper limits using a Poisson statistic, Nucl.Instr.&Meth.*, **A450**, 444-455, 2000.

[16] S.I.Bityukov, N.V. Krasnikov, V.A. Taperechkina, V.V. Smirnova, *Statistically dual distributions in statistical inference*, in proceedings of *Statistical problems in Particle Physics, Astrophysics and Cosmology* (PhyStat'05), September 12-15, 2005, Oxford, UK, Imperial College Press, 2006, pp.102-105.

[17] T. Liptak, *On the combination of independent tests, Maguar Tud. Akad. Mat. Kutato Int. Kozl.*, **3**, 171-197, 1958.

[18] K. Singh, M. Xie, W. Strawderman, *Combining information from independent sources through confidence distributions, Annals of Statistics*, **33**, 159-183, 2005.

12