On Combining Significances. Trivial examples.

S.I. Bityukov and N.V. Krasnikov
Institute for nuclear research RAS, Moscow, Russia

"Suppose one experiment sees a 3-sigma effect and another sees a 4-sigma effect. What is combined significance?" (R. Cousins (2008). His conclusion - the question is not well-posed.)

In this talk we discuss the problem of significances combining.

Plan

- **Introduction**
- **Combining methods**
- **Examples**
- **Weighted combination**
- **The case of small statistics**
- **Systematics**
- **Conclusion**

Introduction

Significance S is related with the probability by the relation (one-sided tail probability)

$$S = \Phi^{-1}(1-p) = -\Phi^{-1}(p),$$

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

$$S = \sqrt{2} erf^{-1}(1-2p).$$

For example,
$S = 5$ **corresponds to a** $p-$**value of** $2.9 \cdot 10^{-7}$**.**
$S = 4 \Longrightarrow p = 3.2 \cdot 10^{-5}$**.**
$S = 3 \Longrightarrow p = 0.0013$**.**
$S = 2 \Longrightarrow p = 0.023$**.**
$S = 1 \Longrightarrow p = 0.16$**.**
So the problem – how to combine a set of a $p-$**values or a set of of** $S-$**significances ?**

Combining methods I

There are several methods for significances combining:

1. **Fisher's (R.A. Fisher, 1932) method based on the choice of** $P = \prod_{i=1}^{N} p_i$**. A simple way to combine** p_i **is the use of relation**

   $-2\sum_i ln\ p_i = \chi^2_{2N,p}$, **where** $\chi^2_{\nu,p}$ **denotes the upper** p **point of the probability integral of a central chi-squared of** ν **degrees of freedom.**

   For two p_1 **and** p_2 **combining (for example, F. James (2006))**

   $p(p_1, p_2) = p_1 p_2 (1 - ln(p_1 p_2))$.

   Plus a lot of generalizations: I.J. Good (1955), H. Lancaster (1961) et al.

2. **Tippett's (L. Tippett, 1931) method using the smallest** p_i

   $p = 1 - (1 - (min\ p_i))^N \approx N \cdot min\ p_i$.

   Plus generalizations: B. Wilkinson (1991), ...

Combining methods II

3. **Stouffer's (S. Stouffer et al., 1949) method adding the inverse normal of the p_i's**
   $\sum \Phi^{-1}(p_i) = \sqrt{N}\Phi^{-1}(p)$,
   equivalently
   $S = \frac{\sum S_i}{\sqrt{N}}$.
   Plus generalizations: F. Mosteller and R. Bush (1954), T. Liptak (1958), . . . , S. Bityukov et al. (2006).

Here we shall compare the Fisher and Stouffer approaches.

Our conclusion: For high energy physics with Poisson distribution in many cases the Stouffer approach is more natural.

Example I

Suppose the CMS experiment for some signature measures in july (2010) 10300 events and in august it detects 9700 events with the theoretical expectation $\lambda_{july} = \lambda_{august} = 10000$ **in Poisson distribution**

$$Pois(n, \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

For $\lambda \gg 1$**,** $n_{obs} \gg 1$ **Poisson distribution is approximated by normal distribution with mean** $\mu = \lambda$ **and variance** $\sigma^2 = \lambda$ **and we find that**

$$S_{july} = \frac{|10300 - 10000|}{\sqrt{10000}} = 3,$$

$$S_{august} = \frac{|9700 - 10000|}{\sqrt{10000}} = 3.$$

If we analyze data for july plus august we find (using the fact that the sum of two Poisson processes with λ_1 **and** λ_2 **is a Poisson process with** $\lambda = \lambda_1 + \lambda_2$**) that**

$$S_{july+august} = \frac{|n_{obs,july} + n_{obs,august} - \lambda_{july} - \lambda_{august}|}{\sqrt{\lambda_{july} + \lambda_{august}}} = 0$$

in perfect agreement with a theory.

Example I

If we combine significances using formula
$p(p_1, p_2) = p_1\ p_2(1 - ln(p_1\ p_2))$**,**
we find that
(remember that $S = 3 \Longrightarrow p_i = 0.0013, \quad i = 1, 2$**)**

$p = 0.000026$**;**

$S^{Fisher}_{july+august} = 4.05$

Example II

For other example with $n_{july} = 10300$**,** $n_{august} = 10400$
$\lambda_{july} = \lambda_{august} = 10000$ **we find that**
$S_{july} = 3$**,** $S_{august} = 4$**, and** $S_{july+august} = \dfrac{3+4}{\sqrt{2}} \approx 4.95$**.**

Fisher's method gives: $p_{july+august} = 0.00000077$**;**
$S^{Fisher}_{july+august} = 4.80$

Weighted combination

For Poisson distribution with $n_{obs} \gg 1$ **and** $\lambda \gg 1$ **(when we can approximate this distribution by Gaussian) we can define significance as**

$$S_1 = \frac{n_{obs_1} - n_{0_1}}{\sigma_1},\ S_2 = \frac{n_{obs_2} - n_{0_2}}{\sigma_2}.$$

Here $S_i > 0$ **corresponds for excess of events,** $S_i < 0$ **corresponds for shortage of events and** $\sigma_i = \sqrt{\lambda_i}, \quad i = 1, 2$**.**

Usualy people use $S = |\frac{n_{obs} - n_0}{\sigma}|$**.**

The rule for significances combining is

$$S(S_1, S_2) = \frac{S_1\ \sigma_1 + S_2\ \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}.$$

For general case

$$S(S_1, S_2, \ldots, S_n) = \frac{S_1\ \sigma_1 + \ldots + S_n\ \sigma_2}{\sqrt{\sigma_1^2 + \ldots + \sigma_2^2}}.$$

The case of small statistics

Consider now the case of small statistics. Namely, as an example, consider CMS experiment with $n_{obs,july} = n_{obs,august} = 1$ **and** $\lambda = \lambda_{august} = \lambda \ll 1$**. Then the probability to observe** $n \geq 1$ **events is determined by**

$$Pois(n \geq 1, \lambda) = \sum_{i=1}^{\infty} Pois(i, \lambda) \approx \lambda.$$

Correspondingly,

$$Pois(n \geq 2, \lambda) = \sum_{i=2}^{\infty} Pois(i, \lambda) \approx \frac{\lambda^2}{2}$$

and we find that

$$P_{july} = \lambda_{july},$$

$$P_{august} = \lambda_{august},$$

$$P_{july+august} = \frac{(\lambda_{july}+\lambda_{august})^2}{2} = \frac{(P_{july}+P_{august})^2}{2}.$$

A. For $P_{july} = P_{august} = 0.023 \quad (2\sigma : one - side)$

$$P_{july+august} = 0.001 \quad (3.07\sigma).$$

Compare with Fisher's formula

$$P^{Fisher}_{july+august} = P_{july} \cdot P_{august}(1 - ln(P_{july} \cdot P_{august})).$$

Fisher's method gives the value

$$P^{Fisher}_{july+august} = 0.0044 \quad (2.62\sigma).$$

The case of small statistics

B. For the case $S_{july} = 3\sigma$**,** $S_{august} = 4\sigma$**,** $n_{july} = 1$ **and** $n_{august} = 1$

$P_{july+august} = 0.95 \cdot 10^{-6} \quad (4.76\sigma)$
$P^{Fisher}_{july+august} = 0.77 \cdot 10^{-6} \quad (4.81\sigma)$

C. For the case $S_{july} = 3\sigma$**,** $S_{august} = 3\sigma$**,** $n_{july} = 1$ **and** $n_{august} = 1$

$P_{july+august} = 0.36 \cdot 10^{-5} \quad (4.49\sigma)$
$P^{Fisher}_{july+august} = 0.26 \cdot 10^{-4} \quad (4.05\sigma)$

In general for $n_{obs} \gg \lambda$

$$\sum_{n=n_1}^{\infty} P(n, \lambda_1) \approx \frac{\lambda_1^{n_1}}{n_1!} e^{-\lambda_1};$$

$$\sum_{n=n_2}^{\infty} P(n, \lambda_2) \approx \frac{\lambda_2^{n_2}}{n_2!} e^{-\lambda_2}$$

and

$$\sum_{n=n_1+n_2}^{\infty} P(n, \lambda_1 + \lambda_2) \approx \frac{(\lambda_1 + \lambda_2)^{n_1+n_2}}{n_1 + n_2!} e^{-(\lambda_1+\lambda_2)}.$$

Systematics

Let us consider the influence of systematic effects related with nonexact knowledge of parameter λ in Poisson formula.

Suppose $n_{obs,july} = n_{obs,august} = 600$, $\lambda_{july} = \lambda_{august} = 300$, and $\epsilon = \frac{1}{3}$ (uncertainty in the parameter λ determination).

For such case the significance is determined by approximate formula

$$S = \frac{n_{obs} - \lambda}{\sqrt{\lambda + (\epsilon\lambda)^2}}.$$

According to this formula

$$S_{july} = S_{august} = \frac{300}{\sqrt{300 + (100)^2}} \approx 3.$$

For july+august

$$S_{july+august} = \frac{600}{\sqrt{600 + (2 \cdot 100)^2}} \approx 3.$$

So we find that july+august combining does not help to increase significance, since the systematic error $\sim (\epsilon\lambda)^2$ dominates.

Conclusions

To conclude we think that in many cases for Poisson statistics the most natural rules of significances combining is the use of the fact that the sum of Poisson processes is the Poisson process.

$$\begin{cases} Pois(n_1, \lambda_1) \\ Pois(n_2, \lambda_2) \\ \dots \\ Pois(n_n, \lambda_n) \end{cases} \Longrightarrow Pois(\Sigma\, n_i, \Sigma\, \lambda_i)$$

In fact it is natural generalization of the original Stouffer method.

References

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

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

[3] F. James, *Statistical Methods in Experimental Physics*, 2nd edition, World Scientific, 2006.

[4] I.J. Good, *On the weighted combination of significance tests*, Journal of the Royal Statistical Society. Series B (Methodological) 17(2):264-265.

[5] H. Lancaster, *The combination of probabilities: an application of orthinormal functions*. Australian Journal of Statistics, **3**, 20-33, 1961.

[6] 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.

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

[8] 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.

[9] F. Mosteller and R. Bush, *Selected quantitative techniques*, in ed. G. Lindzey, *Handbookof Social Psychology* **vol. I**, 289-334, Addison-Wesley, Cambridge, Mass., 1954.

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

[11] S.Bityukov, N.Krasnikov and A.Nikitenko, On the combining significances, arXiv:physics/0612178(2006); S.Bityukov, N.Krasnikov, A.Nikitenko and V.Smirnova, Two approaches to combining significances, PoS ACAT08:118(2008).