UCRL-11530

University of California

Ernest O. Lawrence Radiation Laboratory

METHOD OF EFFECTIVE COUNTS FOR COMBINING EXPERIMENTAL RESULTS

TWO-WEEK LOAN COPY

This is a Library Circulating Copy which may be borrowed for two weeks. For a personal retention copy, call Tech. Info. Division, Ext. 5545

Berkeley, California

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Submitted to Nuclear Instruments
and Methods

UCRL-11530

UNIVERSITY OF CALIFORNIA
Lawrence Radiation Laboratory
Berkeley, California

AEC Contract No. W-7405-eng-48

METHOD OF EFFECTIVE COUNTS FOR COMBINING EXPERIMENTAL RESULTS

Frank S. Crawford, Jr.

June 26, 1964

Method of Effective Counts for Combining Experimental Results*

Frank S. Crawford, Jr.

Lawrence Radiation Laboratory
University of California
Berkeley, California

June 26, 1964

ABSTRACT

The usual way of combining experimental data by the method of least squares gives a biased result when the errors from the individual experiments are based on Poisson fluctuations. Instead, one would like to combine the data into one large experiment. This is usually impossible because most experimenters do not publish likelihood functions. The suggested method of effective counts provides a way to treat data from several experiments as though they came from an effective single experiment when only the physical quantity and its standard deviation are published.

INTRODUCTION

It is often desirable to combine the results of two (or more) experiments to obtain a weighted-average result for some quantity. If the experiments disagree they should not be combined. The compiler nevertheless usually insists on combining them, and of course if the disagreement is due to statistical fluctuations they should be combined. Suppose that the compiler has available from each experiment no information other than the experimental result and its standard deviation. This is often the case. How should he form the weighted average? The customary procedure is to use the method of least squares. But this biases the average towards the smaller of the two results. At least this is true when the experimental result is proportional to a counting rate and there are not many counts.

We illustrate the bias with a simple numerical example, from which we are also led to suggest a simple alternative method of averaging. Suppose that in experiment No. 1 nine counts are obtained in 2 minutes. The "published" result and standard deviation for the counting rate $R \pm \delta R$ is $R_1 = [n_1 \pm (n_1)^{1/2}]/t_1 = (9 \pm 3)/2 = 4.5 \pm 1.5$. In another experiment, No. 2, involving identical apparatus, 25 counts are obtained in 2 minutes. Thus $R_2 = (25 \pm 5)/2 = 12.5 \pm 2.5$. The two results are in poor agreement, but there is no way of choosing <u>a priori</u> between them, and perhaps it will be 2 years before another result is available. Therefore we may be sure that the results will be combined. The method of least squares prescribes $R = (w_1 R_1 + w_2 R_2)/(w_1 + w_2)$, with $w_i = 1/(\delta R_i)^2$. The standard deviation on R is given by $\delta R = (w_1 + w_2)^{-1/2}$. In our example we then have $R = [(1.5)^{-2}(4.5) + (2.5)^{-2}(12.5)]/[(1.5)^{-2} + (2.5)^{-2}]$, or $R = 6.62 \pm 1.29$. Why do we call this answer biased? Suppose that not only the results

$R_1 \pm \delta R_1$ and $R_2 \pm \delta R_2$ were available, but also the data itself--n_1, t_1, n_2, and t_2. Then if the discrepancy were believed to be due to statistical fluctuations, the most reasonable procedure would be to combine the data as though it were from one large experiment. Then we would have for the combined result

$$R = \frac{(n_1 + n_2) \pm (n_1 + n_2)^{1/2}}{t_1 + t_2} \quad . \tag{1}$$

In our example we would then find $R = [34 \pm (34)^{1/2}]/4 = 8.50 \pm 1.45$. We call this the unbiased average. It is exactly what is prescribed by the method of maximum likelihood (see Appendix A). The least-squares result is too small by 1.5 of its standard deviations! Of course it would be unusual for two determinations of R to disagree as violently as they did in our example (unless there are systematic errors), but it is in just those circumstances (poor agreement) that it matters how we combine the two results.

METHOD OF EFFECTIVE COUNTS

How can we obtain the "unbiased" average if the experimenters do not publish all of their data, but only the result and its standard deviation? Our suggestion consists of defining "effective counts" and "effective running time" (or track length, or some other denominator) for each experiment. We then use Eq. (1). For example, suppose $R_1 \pm \delta R_1$ and $R_2 \pm \delta R_2$ are two published values of a cross section (or of a small branching ratio) for which we believe, although the experimenter may not have told us, that the quoted errors were obtained essentially from the square root of the counts (or counts in the numerator if it is a small branching fraction).

We then define the effective counts as

$$n_1 = (R_1/\delta R_1)^2 \text{ and } n_2 = (R_2/\delta R_2)^2 \tag{2}$$

and the effective running times (whose dimension and meaning will not actually be time in most cases) as

$$t_1 = n_1/R_1 \text{ and } t_2 = n_2/R_2. \tag{3}$$

We then combine the experiments according to Eq. (1). Clearly in our numerical example we get the "unbiased" result, since the effective counts turn out to be the actual counts.

In experiments based on entirely different techniques and counting efficiencies, but for which the published uncertainties are based only on Poisson fluctuations in the counts, our method agrees with the maximum-likelihood method (see Appendix A), because then the effective counts will be the actual counts.

If the experimenter arrived at his quoted error in a more complicated manner, then our method need not be equivalent to combining the data as though they came from one large experiment. But our method is still probably preferable to least squares, and is perhaps as close as one can come to the maximum-likelihood result without access to the data or likelihood functions.[1]

FURTHER APPLICATIONS

As a second example, consider a branching fraction that is not small (but is still of course ≤ 1.0), so that fluctuations in the denominator can not be neglected. Suppose we have $f = n_a/N$, where $N = n_a + n_b$. Then we expect that the published standard deviation was based on the well-known relation $(\delta f)^2 = f(1-f)/N$. In any case we define the effective counts $N_1 = f_1(1-f_1)/(\delta f_1)^2$ and similarly for N_2. Then the effective counts of

type a are $n_{a1} = f_1 N_1$ and $n_{a2} = f_2 N_2$. The weighted average is thus

$$f = \frac{N_1 f_1 + N_2 f_2}{N_1 + N_2} \pm \left(\frac{f(1-f)}{N_1 + N_2} \right)^{1/2} . \tag{4}$$

Again, this is not the prescription of least squares, but Eq. (4) is equivalent to Eq. (1) if we think of the "denominator" t_1 and t_2 in Eq. (1) becoming the total counts N_1 and N_2 in the present case. For a small branching ratio, N_1 and N_2 are essentially completely known and the fluctuations are all due to the small number of counts $N_1 f_1$ and $N_2 f_2$ of the numerators.

Instead of the branching function f, the published quantity may be the branching ratio

$$r = n_a / n_b .$$

We then assume that the published standard deviation δr was based on $(\delta r/r)^2 = n_a^{-1} + n_b^{-1}$. We solve r and $(\delta r)^2$ for the effective n_a and n_b for experiment No. 1. This gives

$$(n_a)_1 = r_1^2 (1+r_1)/(\delta r_1)^2 ,$$

and

$$(n_b)_1 = (n_a)_1 / r_1 ;$$

and similarly for No. 2. The combined result is then

$$r = \frac{(n_a)_1 + (n_a)_2}{(n_b)_1 + (n_b)_2} \equiv \frac{n_a}{n_b} , \tag{5}$$

with $(\delta r/r)^2 = n_a^{-1} + n_b^{-1}$.

As a third example, consider the up-down decay asymmetry for a parity-nonconserving decay of a hyperon of spin 1/2. One encounters expressions like

$$a = (3/N) \sum_{i=1}^{N} x_i \pm [(3-a^2)/N]^{1/2} ,$$

where a is the decay asymmetry and x_i is the cosine of the decay particle in the hyperon rest frame. Therefore if two published values are $a_1 \pm \delta a_1$ and $a_2 \pm \delta a_2$, we define $N_i = (3-a_i^2)/(\delta a_i^2)$ and combine data from the two experiments into data from an equivalent single experiment

$$a = \frac{(N_1 a_1 + N_2 a_2)}{N_1 + N_2} \pm \left(\frac{3 - a^2}{N_1 + N_2}\right)^{1/2} . \tag{6}$$

Sometimes a decay asymmetry is calculated with only the "up" counts U and "down" counts D. Then, with $U + D \equiv N$, one has

$$a = \frac{2(U - D)}{N} \pm \left(\frac{4 - a^2}{N}\right)^{1/2} .$$

We can then define $N_i = (4 - a_i^2)/(\delta a_i)^2$ and obtain the combined value

$$a = \frac{N_1 a_1 + N_2 a_2}{N_1 + N_2} \pm \left(\frac{4 - a^2}{N_1 + N_2}\right)^{1/2} . \tag{7}$$

As a last example, consider a mean decay lifetime. For an experiment with no cutoffs or corrections we would have

$$\tau = (1/N) \sum_{i=1}^{N} t_i$$

with standard deviation $\delta\tau = \tau/(N)^{1/2}$. Therefore we <u>define</u> the effective counts $N_1 = (\tau_1/\delta\tau_1)^2$ and similarly for N_2. The "large combined experiment" then gives

$$\tau = \frac{N_1 \tau_1 + N_2 \tau_2}{N_1 + N_2} \pm \frac{\tau}{(N_1 + N_2)^{1/2}} . \tag{8}$$

This is a very different prescription from that of least squares and is equivalent to the maximum-likelihood method when we have $\tau = (1/N)\Sigma\tau_i$. Notice that if we set $R_1 \equiv \tau_1^{-1}$ as the "effective counting rate" for experiment No. 1, and $N_1 \tau_1 \equiv t_1$ as the "effective running time", and

similarly for experiment No. 2, then our prescription for τ is identical with Eqs. (1), (2), and (3).

In all of our examples we would probably use our prescription, no matter how the experimenters arrived at their quoted standard deviations. (Of course it might not be a bad idea to also read their paper.) If the experimenters publish likelihood functions, then there is a correct way to compile, and that is to multiply the likelihood functions. That is equivalent to combining the data into one large experiment. In most cases this is not practical (for instance, usually the likelihood functions are not published.) The method of effective counting rates is then a reasonably good substitute, to be preferred to the method of least squares.

There are of course many types of measurements for which the quoted error is <u>not</u> based on Poisson's fluctuations -- for example, a mass measurement. In those cases our suggestion does not apply.

It is a pleasure to acknowledge stimulating conversations with Arthur H. Rosenfeld and Frank T. Solmitz.

APPENDICES

A. Maximum Likelihood Derivation of Eq. (1)

We show here that Eq. (1) or its equivalent is prescribed by the method of maximum likelihood. Suppose that quantity R is measured in two experiments that differ entirely in such things as technique and detection efficiencies, but have one feature in common, namely, that the expected number of counts is proportional to R. For experiment No. 1 we have $\bar{n}_1 = a_1 R$, where $\bar{n}_1$ is the expected number of counts for a given value of R, and a_1 depends on the experimental method. Similarly $\bar{n}_2 = a_2 R$. The factors a_1 and a_2 are assumed to be completely known (no systematic errors). If R were known, $\bar{n}_1$ and $\bar{n}_2$ would also be known. Then the chance that experiment No. 1 would find exactly n_1 counts and No. 2 exactly n_2 counts would be $L = L_1 \; L_2 = \left[e^{-\bar{n}_1} (\bar{n}_1)^{n_1}/n_1!\right]\left[e^{-\bar{n}_2}(\bar{n}_2)^{n_2}/n_2!\right]$. Instead n_1 and n_2 are known and R is not known; L is called the likelihood function for R. If we consider experiment No. 1 alone and write $\bar{n}_1 = a_1 R$ in L_1, define $w_1 = -\ln L_1$, and drop into "const" all terms from w_1 that do not contain R, we then have $w_1 = \bar{n}_1 - n_1 \ln \bar{n}_1 + \text{const} = a_1 R - n_1 \times \ln R + \text{const}$. The maximum-likelihood result for R_1 is obtained by setting the derivative of w_1 with respect to R equal to zero, i.e., $0 = a_1 - n_1/R$, so that $R_1 = n_1/a_1$, and similarly, of course, $R_2 = n_2/a_2$. When we combine the experiments we have $w = w_1 + w_2 = (a_1 + a_2)R - (n_1 + n_2) \ln R + \text{const}$. Setting $dw/dR = 0$, we obtain $R = (n_1 + n_2)/(a_1 + a_2)$. This is just Eq. (1), with t_1 and t_2 generalized to a_1 and a_2. Thus we see that the maximum-likelihood prescription is to combine the two experiments into an equivalent single experiment by adding the counts with equal weight and similarly combining the effective running times. The difference between the prescriptions of the likelihood

method and the method of least squares is of course entirely due to the difference in properties of a Poisson distribution and a Gaussian distribution.

B. Average Bias in Introductory Example

In our introductory example of the bias that results from weighting by least squares we used a rather extreme example. Now we calculate the average bias. Suppose that $\bar{n}$ is the "true" result for the average number of counts per minute that one gets by counting for a very long time and dividing by the time. The probability that in experiment No. 1 (which lasts 1 minute), exactly n_1 counts are obtained in $P_1 = e^{-\bar{n}}(\bar{n})^{n_1}/[(n_1)!]$. For the similar experiment No. 2, the probability that we will get n_2 counts is $P_2 = e^{-\bar{n}}(\bar{n})^{n_2}/[(n_2)!]$. If we combine data from the two experiments as prescribed by the maximum-likelihood method we get

$$n \equiv \frac{n_1 + n_2}{2}$$

as the best estimate that can be obtained for $\bar{n}$.

If we perform this type of weighting (of two experiments) many times, we have for the average of n

$$\langle n \rangle = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} P_1 P_2 \left(\frac{n_1 + n_2}{2}\right) = (\bar{n} + \bar{n})/2 = \bar{n} .$$

This is "unbiased."

Instead we combine data from the two experiments by least squares, using

$$n = (w_1 n_1 + w_2 n_2)/(w_1 + w_2), \text{ with } w_i = 1(\delta n_i)^2 = 1/n_i. \text{ Thus}$$

$$n = \frac{2}{\frac{1}{n_1} + \frac{1}{n_2}} .$$

If we perform this kind of weighting many times we would find

$$\langle n \rangle = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} P_1 P_2 \left(\frac{2}{\frac{1}{n_1} + \frac{1}{n_2}} \right),$$

which is easily shown to give

$$\langle n \rangle = \bar{n} - \frac{1}{2} + \frac{e^{-2\bar{n}}}{2} ,$$

instead of the true value $\bar{n}$. In the limit of large $\bar{n}$, the least-squares result is unbiased.

FOOTNOTE AND REFERENCE

* Work performed under the auspices of the U. S. Atomic Energy Commission.

1. It seems to us highly unlikely that our suggestion is new. This article is therefore to be taken as largely pedagogical in nature.

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.