DD/EXP/62/11
22.3.1962

Determination of Scanning Efficiency and True Numbers of Events

from K.D. TOLSTOV, DUBNA 1962
Report No. P-864

CERN translation by T. HOFMOKL and D. SHAW*

Abstract (see original)

Text To evaluate the efficiency of scanning it is necessary to complete two scans and it is generally assumed that the probability is constant for finding in the second scan an event which was found in the first scan. Thus: p = const. Although this is tacitly assumed, it should be verified for each experiment. Let us proceed as follows:

Let the true number of events be n and the scanning efficiencies for the 1st, 2nd ...... ith scans be ϵ_1, ϵ_2 ..... ϵ_i respectively. Then the average number of events, m_i, found in the ith scan is

$$m_i = n \cdot \epsilon_i \qquad (1),$$

and the number of events missed in this particular scan in $n(1 - \epsilon_i)$. If the assumption "$p = \epsilon = \text{const}$" is valid, then the total number, m_{i+k}, of events included in two scans i and k is

$$m_{i+k} = n \cdot \epsilon_i + n(1 - \epsilon_i)\ \epsilon_k .$$

---

* This is not a literal translation but an attempt to express in English the essential content of the original Russian paper.

Now, let us define the efficiency, ϵ_{i+k}, for this combination of two scans as $\epsilon_{i+k} = \frac{m_{i+k}}{n}$ (2),

then, it follows that $\epsilon_{i+k} = \epsilon_i + \epsilon_k - \epsilon_i \cdot \epsilon_k$ (3).

In an analogous way, we find for three scans,

$$\epsilon_{1+2+3} = \epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_1 \cdot \epsilon_2 \cdot \epsilon_3 - [\epsilon_1 \cdot \epsilon_2 + \epsilon_1 \cdot \epsilon_3 + \epsilon_2 \cdot \epsilon_3] \quad (3a)$$

The probability ϵ_{ik} that an event will be found in both the ith and the kth scan is

$$\epsilon_{ik} = \epsilon_i \cdot \epsilon_k \quad (4).$$

Similarly, the probability ϵ_{123} for finding an event in all three scans 1, 2 and 3 is

$$\epsilon_{123} = \epsilon_1 \cdot \epsilon_2 \cdot \epsilon_3 \quad (4a).$$

If more than two scans are completed, equations such as (4) and (4a) relating the ϵ's are overdetermined and hence there may be more than one set of solutions. Let ϵ_i^k denote the efficiency for the ith scanning deduced from the numbers of events observed in the ith and kth scans. Then

$$m_{ik} = \epsilon_k \cdot \epsilon_i^k \cdot n$$

or $$\epsilon_i^k = \frac{m_{ik}}{m_k} \quad (5)$$

where m_{ik} denotes the number of events found both in the ith and kth scans. The weighted mean value for the efficiency of the ith scan, $\bar{\epsilon}_i$, may then be computed from the quantities ϵ_i^k. Hence the weighted mean number of events, $\bar{n}$, may be deduced from the individual numbers n_i found from m_i and $\bar{\epsilon}_i$ using the equation

$$n_i = \frac{m_i}{\bar{\epsilon}_i} \quad (6).$$

Alternatively, the value of n may be deduced from the combined scanning statistics,

$$n = \frac{m_{i+k}}{\epsilon_{i+k}} \qquad (7)$$

$$n = \frac{m_{1+2+3}}{\epsilon_{1+2+3}} \qquad (8)$$

where m_{1+2+3} is the total number of events included in three scans.

Let us now see under what conditions our basic assumption $p = \epsilon$ = constant is fulfilled.

According to (4)

$$\epsilon_{ik} = \frac{m_{ik}}{n} = \epsilon_i \cdot \epsilon_k \; ; \quad \epsilon_{123} = \frac{m_{123}}{n} = \epsilon_1 \cdot \epsilon_2 \cdot \epsilon_3 \qquad (9)$$

where m_{123} is the number of common events in the three scans. If condition (9) is satisfied within the limits of statistical error then $p = \epsilon$ = const. and the analysis is internally consistent. An additional criterion arises from a comparison between the observed and predicted frequency of events found only in one scan. Suppose an event was missed in both scans i and k; then the probability of finding it in the jth scan is

$$\epsilon_j^{i+k} = (1 - \epsilon_{i+k}) \cdot \epsilon_j \qquad (10).$$

(Transliterators' note: Tolstov states that the following is the most sensitive test.)

If p = const, the fact that some events are found and others are missed does not alter the constant probability for observing each and every event. For this reason, the scanning efficiencies deduced (a) from the true number n and (b) from the observed number m_i, must be equal. Let us apply this procedure to evaluate the efficiency for the 'j' and 'k' scans from a knowledge of m_i and the following quantities:

(i) m_{ik} = number of events common to scans i and k

(ii) m_{ij} = number of events common to scans i and j

(iii) m_{ijk} = number of events common to the three scans i, j and k.

Denoting the efficiency for scan j, determined from m_{ijk} and m_{ik} by ϵ_j^{ik}, we have:

$$\epsilon_j^{ik} = \frac{m_{ijk}}{m_{ik}} \quad (11).$$

However, since $m_{ik} = m_{ki}$, then $\epsilon_j^{ik} = \epsilon_j^{ki}$. Also, the calculated number of events found in scan j, $\hat{m}_j$, is

$$\hat{m}_j = \frac{m_{ij}}{\epsilon_j^{ik}} = \frac{m_{ij} \cdot m_{ik}}{m_{ijk}} \quad (12)$$

Then we compare, for different j, the following:

(a) m_j with $\hat{m}_j$

and (b) $\bar{\epsilon}_j$ with ϵ_j^{ik} (c.f. equation (6) which defines $\bar{\epsilon}_j$ as the weighted average value of the efficiency for the jth scan).

If there is any systematic disagreement between these values, for example, if $\epsilon_j^{ik} - \bar{\epsilon}_j > 0$, (and hence $\hat{m}_j < m_j$), it follows that $p \neq$ const.

Effort is usually concentrated on obtaining a scanning efficiency near to unity. However, if conditions (9) and (10) are not satisfied or $\hat{m}_j \neq m_j$, then <u>even a high efficiency</u> will not guarantee an accurate value for n. But, if these conditions are fulfilled, even <u>quite low efficiencies</u> determined with good statistical accuracy may ensure an accurate value for n.

It should be emphasized that in order to maintain p = const, it is necessary that ϵ_i should be maintained constant throughout the whole of the time spent on the ith scan. It is obviously not sufficient that the types of events observed have the same features.

Determination of Errors

1. Known efficiency of observation, $p = \epsilon = const$

The above method may be used to test whether or not this condition is fulfilled.

If m_i is the number of events, then probability $p(m_i)$ for the occurrence of this number may be expressed by means of the binomial distribution:

$$p(m_i) = \frac{n\,!}{m_i\,!\,(n - m_i)\,!}\,(\epsilon_i)^{m_i}\,(1 - \epsilon_i)^{n - m_i} \qquad (13)$$

where the true number of events n equals the number of 'tries'. The variance of this distribution is

$$\sigma^2_{m_i} = m_i\,(1 - \epsilon_i) \qquad (14)$$

σ_{m_i} can be expressed, (conditionally, since ϵ is given), by the standard deviation of the efficiency σ_{ϵ_i}, thus:

$$\sigma_{m_i} = n \cdot \sigma_{\epsilon_i} = \sqrt{m_i\,(1 - \epsilon_i)} = \sqrt{n \cdot \epsilon_i\,(1 - \epsilon_i)}$$

Hence the fractional standard deviation of ϵ_i is

$$\frac{\sigma_{\epsilon i}}{\epsilon_i} = \sqrt{\frac{1 - \epsilon_i}{n_i}} \qquad (15)$$

This formula is only valid if n is known whereas in practice the efficiency is not known a priori. The efficiency is derived from the experimental data and its error will be greater than given by equation 15. (See 2 below).

2. Constant efficiency for each scan but $\epsilon_i \neq \epsilon_k$

Let m_k be the number of events found in scan k and m_{ik} the number of common events to scans i and k. Obviously, equation 4 is valid since the ϵ's are constants. Hence, $m_{ik} = m_k \cdot \epsilon_i - n \cdot \epsilon_i \cdot \epsilon_k = n \cdot \epsilon_{ik}$.

Therefore, the probability of m_{ik} events appearing is

$$p(m_{ik}) = \frac{n\,!}{m_{ik}\,!\,(n - m_{ik})\,!}\,(\epsilon_{ik})^{m_{ik}}\,(1 - \epsilon_{ik})^{n - m_{ik}} \tag{16}$$

By analogy with equation 14, the variance $\sigma^2_{m_{ik}}$ will be given by

$$\sigma^2_{m_{ik}} = m_{ik}\,(1 - \epsilon_i \cdot \epsilon_k) = m_{ik}\,(1 - \epsilon_{ik}) \tag{17}$$

According to (5), the efficiency ϵ_i^k of the ith scan is a function of m_k amd m_{ik} and the fractional standard deviation is

$$\frac{\sigma_{\epsilon_i^k}}{\epsilon_i^k} = \sqrt{\frac{\sigma^2_{m_{ik}}}{m^2_{ik}} + \frac{\sigma^2_{m_k}}{m^2_k} - 2r\,\frac{\sigma_{m_{ik}}}{m_{ik}} \cdot \frac{\sigma_{m_k}}{m_k}} \tag{18}$$

where r is the correlation coefficient between m_k and m_{ik} thus:

$$r = \sqrt{\frac{\epsilon_i\,(1 - \epsilon_k)}{1 - \epsilon_i \cdot \epsilon_k}} \tag{19)*$$

Substituting in equation 18 the values of σ_{m_k}, $\sigma_{m_{ik}}$ and r, given respectively by equations (14), (17) and (19), we find

$$\frac{\sigma_{\epsilon_i^k}}{\epsilon_i^k} = \sqrt{\frac{1 - \epsilon_i}{m_{ik}}} \tag{20}$$

---

* Note: Formula (19) is proved in a supplement to the original Russian text.

3531/p

3. The efficiencies are small and not constant

The Poisson distribution is applicable for m_{ik} : $\sigma^2_{m_{ik}} = m_{ik}$.

$$\frac{\sigma_{\epsilon_i^k}}{\epsilon_i^k} = \frac{1}{\sqrt{m_{ik}}} \qquad (21)$$

when $\epsilon << 1$, equations 20 and 21 are practically equivalent.

4. Efficiencies are not small and vary with time and between different observers

This case cannot be represented in either the Binomial or Poisson form. Ultimately, for a large sample, it goes into a normal distribution. If the sample is sufficient then one may deduce errors in efficiencies by grouping the data. Then one must determine the quantity $\epsilon_{i\Upsilon}$, that is the efficiency of the 'Υ' group in the ith scan. After this the mean efficiency and its standard deviation must be deduced.

Practical Procedure

The best approximation to the true number of events n is obtained if formulae (9) and (10) are valid and especially if $\hat{m}_j = m_j$. This can only be tried if at least three scans have been carried out.

(a) If condition p = const. is true, then the formulae (6), (7b) and (8) are equivalent within experimental errors. The standard deviation of n may be calculated from the corresponding standard deviation of ϵ (equation 20).

(b) Usually $p \neq$ const. Then it is much better to use equation (6) since in applying this formula the assumption p = const. is only made once whereas in applying (7) and (2) this assumption is made twice. For this reason, it is necessary first of all to calculate $\bar{\epsilon}_i$, the weighted mean of the ϵ_i^k's and, since $p \neq$ const, equation (20) only gives a lower limit for

$\frac{\sigma_\epsilon}{\epsilon}$. The upper limit, for a given value of m_{ik}, may be deduced from equation 21. Then

$$\sqrt{\frac{1 - \epsilon_i}{m_{ik}}} < \frac{\sigma_{\epsilon_i}}{i} < \frac{1}{\sqrt{m_{ik}}} \qquad (22).$$

Thence, we may deduce n_i using equation (6)

$$\left[n_i = \frac{m_i}{\epsilon_i} \right]$$

and subsequently calculate n.