Class 1 Chapter 14
Nuclear Counting Statistics

I. Introduction

The statistical nature of radioactive decay was recognized soon after the discovery of radioactivity. In fact, the law of radioactive decay (Chapter III) can be deduced strictly from statistical considerations, proving it a statistical relationship subject to the laws of chance. Hence, in any sample containing a large number of radioactive atoms, some average number will disintegrate per unit time. But the exact number which disintegrate in any given unit of time fluctuates around the average value. In counting applications, it is important to estimate this fluctuation because it indicates the repeatability of results of a measurement.

II. Frequency Distributions

If one plots the frequency of occurrence of values against the values themselves for a series of identical measurements of a statistical process, a curve will result—the ***frequency distribution*** curve. Many statistical phenomena conform to certain standard frequency distributions. If this distribution is known, certain inferences about a population may be made by observing a small sample of the population. In nuclear counting statistics, frequency distributions of interest are the normal, and ***Poisson*** (pronounced *'pwah-sohn'*), and the chi-square distributions.

A. Normal Distribution

The normal distribution describes most statistical processes having a continuously varying magnitude. If one plots the frequency distribution curve for a large number of measurements on a quantity which conforms to the normal distribution, a familiar bell-shaped curve (similar to the one shown in Figure 1.14.1) will result. This is the normal distribution curve. It is characterized by two independent parameters: the mean (m) and the standard deviation (σ).

1. Mean

The mean is the average value of the quantity under observation. For the standard normal distribution (i.e., symmetrical about the mean), the mean value is the one that occurs with the highest frequency. Since in reality we observe only a portion of the population, we estimate the mean by a numerical average ($\bar{x}$).

$$\bar{x} = \frac{\sum x_i}{n} \qquad (1)$$

where x_i is the value of the i^{th} measurement, n is the total number of observations.

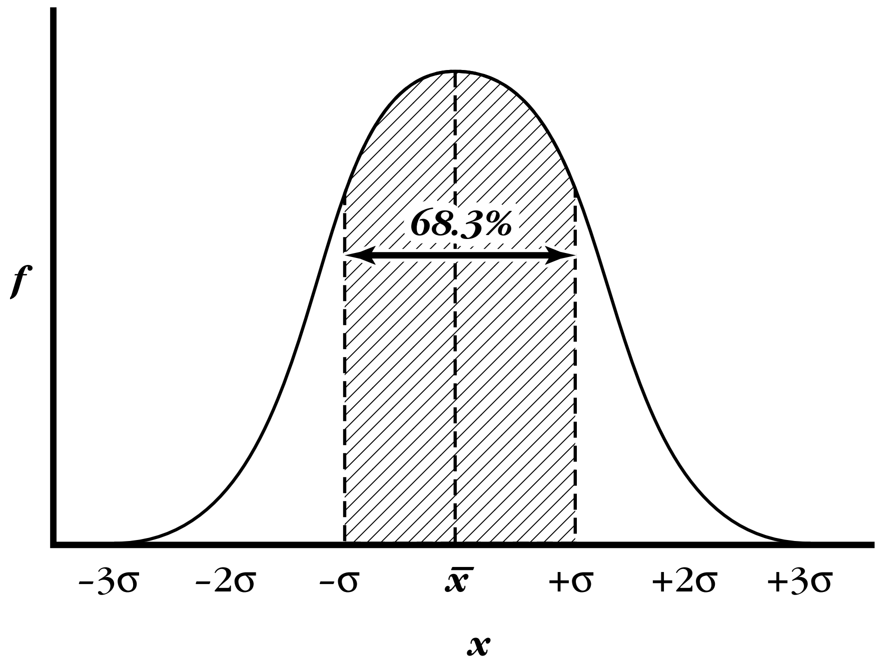


Figure 1.14.1 - Normal Distribution Curve

2. Standard deviation

The standard deviation is defined as the square root of the average of the squares of the individual deviations from the mean. Expressed mathematically, this is:

$$\sigma = \sqrt{\frac{\sum (x_i - m)^2}{n}} \qquad (2)$$

where m is the mean value, x_i is the value of the i^{th} measurement, and n is the total number of observations. As with the mean, we must estimate σ from a finite number of observations. The best estimate of σ is called s_x which is given by:

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \qquad (3)$$

The use of n–1 in the denominator results from the use of $\bar{x}$ instead of m in the numerator. In other words, we lose one degree of freedom by estimating m with $\bar{x}$. Since m and σ are ***independent*** parameters which characterize the normal distribution, it is possible to have an infinite number of values of σ for a given mean. However, the smaller the standard deviation, the greater the reproducibility of the measurements. If one covers one standard deviation on each side of the mean of a standard normal distribution curve, approximately 68% of the total area under the curve will be included. (See

Figure 1.14.1) Two standard deviations on each side of the mean include approximately 95% of the total area, three standard deviations–99.7%, etc. The practical significance of this is: If one estimates the mean of a series of normally distributed measurements by $\bar{x}$, and estimates the standard deviation by s_x , it can be said $\bar{x} = s_x$. Likewise, it can be said with 95% confidence that the true mean is somewhere between $\bar{x} \pm 2s_x$, etc.

B. Poisson Distribution

The Poisson distribution equation is:

$$P(x) = \frac{e^{-n}\, m^x}{x!} \qquad (4)$$

where **P(*x*)** is the probability that a given value in a series of observations will occur ***x*** times. The Poison distribution adequately predicts the frequency distribution resulting from the observation of a large number of events which, taken singly, have a small but constant likelihood of occurrence. The Poisson distribution is characterized by only one parameter–the mean. If the definition of standard deviation (II.A.2) is applied to the Poisson distribution equation, the following expression results:

$$\sigma = \sqrt{m} \qquad (5)$$

Thus, the standard deviation for the Poisson distribution depends on the mean. It can have only one value for a given mean value. Another difference between the normal and Poisson distributions is that ***m*** and ***x*** must be integers in the Poisson distribution, while the normal distribution is a continuous one.

It can be shown that radioactive decay obeys the Poisson distribution law. If one observes a sample containing a large number of radioactive atoms for a period of time which is short compared to the half-life, then the probability that a single atom will decay during the observation time is small but constant for equal time intervals. Also, only integer numbers of atoms decay in any time period. Hence, nuclear disintegrations obey Poisson statistics. If, however, the mean number of observed events is moderately large, say 100 or more, the Poisson distribution is adequately approximated by a special normal distribution for which $\sigma = \sqrt{m}$, or in terms of our estimates of these parameters:

$$S_x = \sqrt{\bar{x}} \qquad (6)$$

The approximation is usually considered acceptable if the mean value is 20 or greater. This is the preferred method for handling nuclear counting data, since it is less complex than working with the Poisson distribution directly.

III. APPLICATION TO NUCLEAR COUNTING DATA

The following symbols will be used throughout the remainder of this chapter:

N = total counts

t = counting period

n = **N/*t*** = count rate

The subscript, ***g***, refers to the sample plus background count (gross count), ***b*** refers to the background count alone, and s refers to the net sample count. Naturally, ***s*** must be obtained by subtraction, since it is impossible to observe directly the sample activity apart from the ever present background.

A. Standard Deviation in Total Count (N)

From Equation (6) the standard deviation in the total sample plus background count, s_{N_g}, is given by:

$$S_{N_g} = \sqrt{N_g} \qquad (7)$$

and the standard deviation in the total background count, s_{N_b}, is calculated from:

$$S_{N_b} = \sqrt{N_b} \qquad (8)$$

B. Standard Deviation in Count Rate (*n*)

To obtain the standard deviation in the gross count rate, divide both sides of Equation (7) by the counting period. Then:

$$S_{N_g} = \frac{S_{N_g}}{t_g} = \sqrt{\frac{N_g}{t_g^2}} = \sqrt{\frac{N_g}{t_g} \times \frac{1}{t_g}} \qquad (9)$$

But, $\frac{N_g}{t_g} = n_g$

Therefore, the standard deviation in the gross count rate, s_{N_g} , is calculated as follows:

$$S_{N_g} = \sqrt{\frac{n_g}{t_g}} \qquad (10)$$

Similarly, the standard deviation in the background count rate, s_{N_b} , is given by:

$$S_{N_b} = \sqrt{\frac{n_b}{t_b}} \qquad (11)$$

C. Standard Deviation in Net Count Rate (n_s)

The net count rate, n_s, is given by:

$$n_s = n_g - n_b$$

Notice the count ***rates*** are used, because total counts cannot be subtracted unless the counting period is the same for sample and background. This is rarely the case. The problem here is one of combining the standard deviations in the gross count rate and the background count rate. The method of combining standard deviations is $s_{x_1} + s_{x_2}$ to obtain the standard deviation in the sum or difference of two measurements $x_1 \pm x_2$ is:

$$S_{x_1 \pm x_1} = \sqrt{S_{x_1}^{2} + S_{x_2}^{2}} \tag{13}$$

Thus, by combining Equations (10) and (11), one obtains the following expression for INSERT, the standard deviation in the net count rate:

$$S_{N_g} = \sqrt{\left(\frac{N_g}{t_g}\right)^2 + \left(\frac{N_b}{t_b}\right)^2} = \sqrt{\frac{N_g}{t_g} + \frac{N_b}{t_b}} \tag{14}$$

Example:

Given the following data, find the standard deviations in:

(a) the total gross and background counts

(b) the gross background count rates

(c) the net count rate.

Report (c) at the 95% confidence level.

N_g = 40,000 counts

t_g = 10 minutes

N_b = 3,600 counts

t_b = 20 minutes

(a) $s_{N_g} = \sqrt{40,000} = 200$ counts

$s_{N_b} = \sqrt{3,600} = 60$ counts

(b) n_g = 40,000/ 10 = 4,000 cpm

n_b = 3,600 / 20 = 180 cpm

$s_{N_g} = \sqrt{4,000/10} = 20$ cpm

$s_{N_b} = \sqrt{180/20} = 3$ cpm

(c) $s_{N_g} = \sqrt{(20)^2 + (3)^2} = \sqrt{409} = 20.3$ cpm

Therefore, $n_s = 3,820 \pm 40.6$ cpm at the 95% confidence level.

D. Fractional Standard Deviation and Per Cent Uncertainty

It is sometimes convenience to express the uncertainty in a measurement as a fraction of the quantity itself. This is called the ***fractional standard deviation*** (FSD) and is determined as follows for the net count rate:

$$(FSD)_{n_g} = \frac{ks_{n_g}}{n_g} = \frac{k}{n_g}\sqrt{\frac{n_g}{t_g} + \frac{n_b}{t_b}}$$

where k is the number of standard deviations required to give the desired confidence level. The uncertainty expressed as a percentage of the total is obtained by multiplying the FSD by 100. In the previous example,

$$(FSD)\ Ns = 40.6 \div 3{,}820 = 0.0106$$

The uncertainty in the determination is 1.06% of the net count rate at the 95% confidence level.

E. Other Useful Parameters

1. Most probable error

The quantity commonly referred to as the most probable error is the deviation which corresponds to that value which will probably be exceeded on repeat measurements. In other words, it corresponds to the value which, if taken on each side of the mean, will include 50% of the area under the normal distribution curve. Expressed in terms of the standard deviation, the most probable error is equal to 0.6754σ.

2. Nine-tenths error

The nine-tenths error is the uncertainty which is expected to be exceeded 10% of the time on repeat measurements; it is 1.64σ.

F. The Chi-Square Test of Goodness of Fit

One of the most important applications of statistics to measurements is the investigation of whether or not a particular set of measurements fit an assumed distribution. The test most often used for this purpose on nuclear counting data is Pearson's chi-square test.

The quantity x^2 is defined as follows:

$$\chi^2 = \sum \frac{\left[(\textit{observed value})_i - (\textit{expected value})_i\right]^2}{(\textit{expected value})_i}$$

where the summation is over the total number of independent observations. The expected values are computed from any assumed frequency distribution. For nuclear counting data the assumed distribution is the Poisson, hence the expected value is equal to $\bar{x}$, the average number of counts recorded per interval. Thus:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\bar{x}} \qquad (15)$$

where n values of ***x*** are observed.

The data should be subdivided into at least five classifications, each containing at least five counts.

The steps in applying Pearson's chi-square test to counting data are as follows:

1. Compute $\bar{x} = \sum \frac{\Sigma x_i}{n}$

2. Compute χ^2 from Equation (15)
3. Determine the degrees of freedom (F).

   F is the number of ways the observed distribution may differ from the assumed.

   In our application, F = n – 1.
4. From Table X-1, with the values of χ^2 and F, determine P.

P is the probability that ***larger*** deviations than those observed would be expected due solely to chance if the observed distribution is actually identical to the assumed distribution. From this definition, it is obvious that too little deviation is possible as well as too much. The closer P is to 0.5, the better the observed distribution fits the assumed, for larger deviations than those observed are just as likely as not. The interpretation of P is for $0.1 \le P \le 0.9$, the observed and assumed distributions are very likely the same. If $P > 0.02$ or if $P > 0.98$ the equality of the distributions is very unlikely. Any other value of P would call for additional data to better define the observed distribution.

An example of the use of Pearson's test is as follows: The data in the following table are from a series of ten 2 minute counts of a standard source made with a G-M laboratory counter. We wish to determine whether these data reflect proper instrument operation. The chi-square test is applied as shown on page 113.

Table of Chi-Square Values*

Number of Determinations	Probability						
	0.99	0.95	0.90	0.50	0.10	0.05	0.01
3	0.020	0.103	0.211	1.386	4.605	5.991	9.210
4	0.115	0.352	0.584	2.366	6.251	7.815	11.354
5	0.297	0.711	1.064	3.357	7.779	9.488	13.277
6	0.554	1.145	1.610	4.351	9.236	11.070	15.086
7	0.872	1.635	2.204	5.348	10.645	12.592	16.812
8	1.239	2.167	2.833	6.346	12.017	14.067	18.475
9	1.646	2.733	3.490	7.344	13.362	15.507	20.090
10	2.088	3.325	4.168	8.343	14.684	16.919	21.666
11	2.558	3.940	4.865	9.342	15.987	18.307	23.209
12	3.053	4.575	5.578	10.341	17.275	19.675	24.725
13	3.571	5.226	6.304	11.340	18.549	21.026	26.217
14	4.107	5.892	7.042	12.340	19.812	22.362	27.688
15	4.660	6.571	7.790	13.339	21.064	23.685	29.141
16	5.229	7.261	8.547	14.339	22.307	24.996	30.578
17	5.812	7.962	9.312	15.338	23.542	26.296	32.000
18	6.408	8.672	10.085	16.338	24.769	27.587	33.409
19	7.015	9.390	10.865	17.338	25.989	28.869	34.805
20	7.633	10.117	11.651	18.338	27.204	30.144	36.191
21	8.260	10.851	12.443	19.337	28.412	31.410	37.566
22	8.897	11.591	13.240	20.337	29.615	32.671	38.932
23	9.542	12.338	14.041	21.337	30.813	33.924	40.289
24	10.196	13.091	14.848	22.337	32.007	35.172	41.638
25	10.856	13.848	15.659	23.337	33.196	36.415	42.980
26	11.524	14.611	16.473	24.337	34.382	37.382	44.314
27	12.198	15.379	17.292	25.336	35.563	38.885	45.642
28	12.879	16.151	18.114	26.336	36.741	40.113	46.963
29	13.565	16.928	18.939	27.336	37.916	41.337	48.278
30	14.256	17.708	19.768	28.336	39.087	42.557	49.588

* Usually tables in statistical texts give the probability of obtaining a value of $\chi 2$ as a function of df, the number of degrees of freedom, rather than of n, the number of replicate determinations. In using such texts, the value of df should be taken as n-1.

Figure 1.14.2 - Table of Chi-Square Values

x_i	$(x_i - \bar{x})^2$
264	144
267	225
242	100
261	81
233	361
247	25
237	225
263	121
243	81
263	121
2,520	1,484

$\bar{x}$ = 2,520 / 10 = 252

χ^2 = 1,484 / 252 = 5.9

F = 10 – 1 = 9

From the Table of Chi-Square Values, P ≈ 0.72. Thus, we conclude that the data reflect proper instrument operation.

G. Minimum Detectable Activity

Minimum detectable activity (MDA) is defined as the activity (usually in microcuries) which will result in a count rate significantly different from background for a given counting time. If the sample counting time is to equal the background counting time:

$$MDA = (k \div f)\sqrt{N_b \div t_b}$$

where ***k*** is the number of standard deviations corresponding to the chosen confidence level, and ***f*** is the calibration factor for the instrument (***f*** has units of cpm/µCi).

H. Application of Statistics to Ratemeter Readings

$$S_r = \sqrt{\frac{r}{2\,RC}}$$

where ***r*** is the ratemeter reading in counts per minute or counts per second, and RC is the time constant of the ratemeter in appropriate time units.

IV. Statistical Control Charts

One way to continually check the accuracy and precision of a nuclear counting instrument is with statistical control charts. A statistical control chart permits a periodic check to see if the observed fluctuation in the counting rate from a constant source of radioactivity is consistent with that predicted from statistical considerations. to construct such a chart, it is necessary first to make 20 or 30 independent measurements of the same source, keeping the counting time constant. Then a chi-square test must be performed on this data to insure proper instrument operation at the outset. Once proper operation is established, the mean counting rate and the standard deviation are calculated from the data. Next, a graph is constructed as shown in Figure 1.14.3, and daily counting rates are plotted.

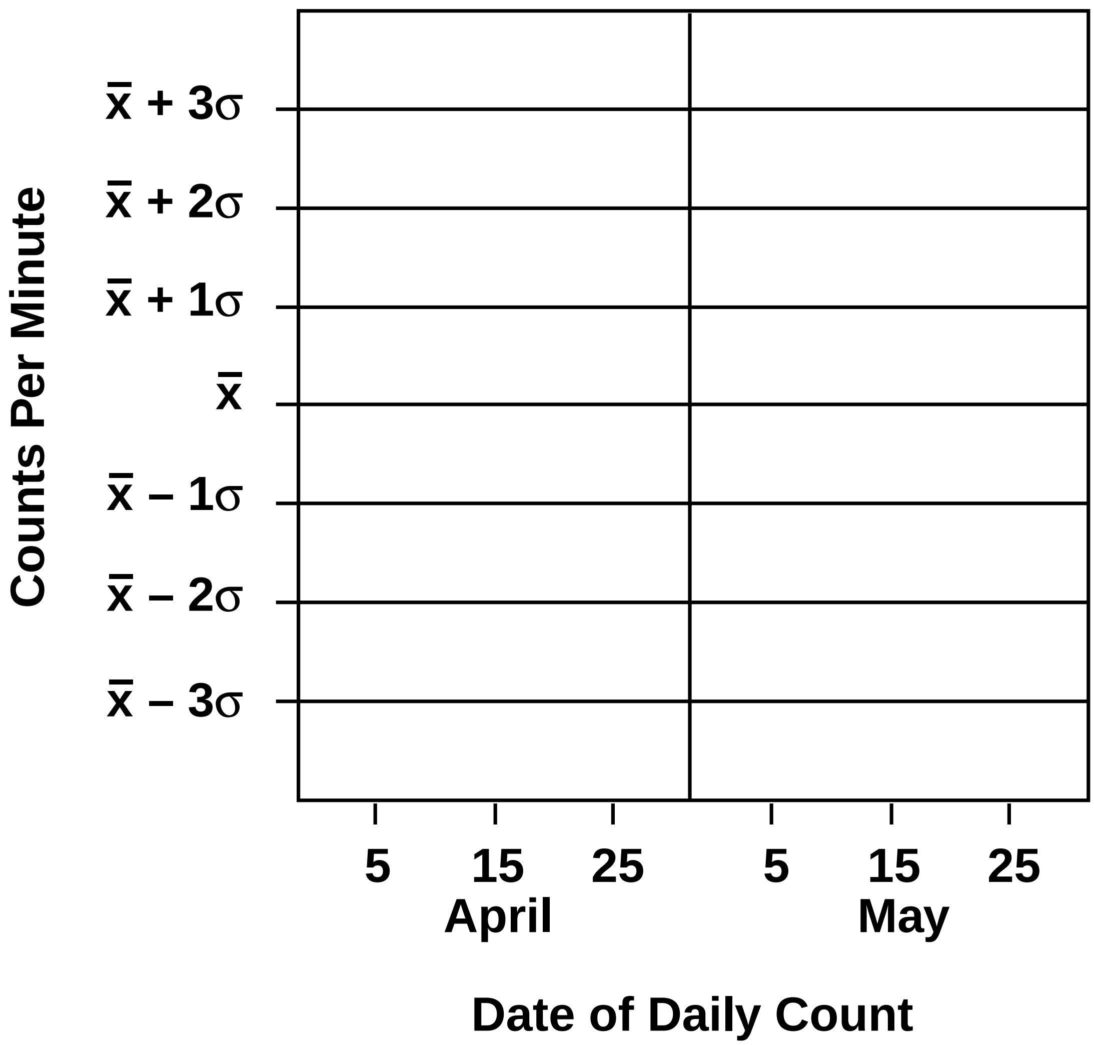


Figure 1.14.3 - Statistical Control Chart

It is wise to label the abscissa in calendar units to facilitate retrospective analyses. Individual deviations from the mean should not exceed one standard deviation more than 33% of the time, two standard deviations more than 5% of the time, etc. If a measurement falls outside the 95% (2σ) line it should be repeated. Since there is one chance in 20 that a single value will fall outside these limits by chance alone, the chances are one in 400 that such will be the case on two successive observations. Hence two successive measurements outside the 95% control limit is sufficient cause to suspect anomalous data. One should also watch for trends in the data, i.e., gradual but consistent changes in either direction.

One readily available source for constructing statistical control charts is the ever present background radiation. it is important to measure background at least daily on a routinely used instrument, since certain causes of erroneous data (e.g., contamination and external sources in the area) are reflected in changing background count rates.

V. Summary

The application of statistics to nuclear counting data is mandatory to the precision with which measurements are made. it should be emphasized that only the uncertainty due to the random nature of the decay process is considered in this chapter. If other significant sources of uncertainty are present, such as timing, they must be dealt with separately and included in the overall estimate of the accuracy.

Suggestions For Further Reading

1. Chase, G. D., and Rabinowitz, J. L., ***Principles of Radioisotope Methodology***, Burgess Publishing Co. (1965), chap. 4.
2. Wagner, H. N., ***Principles of Nuclear Medicine***, W. B. Saunders Co. (1968), pp. 36-44.
3. Quimby, E. H., and Feitelberg, S., ***Radioactive Isotopes in Medicine and Biology***, Lea and Febiger, Vol. 1 (1965), chap. 14.
4. Blahd, W. H., ***Nuclear Medicine***, McGraw-Hill Book Co. (1965), pp. 74-80.

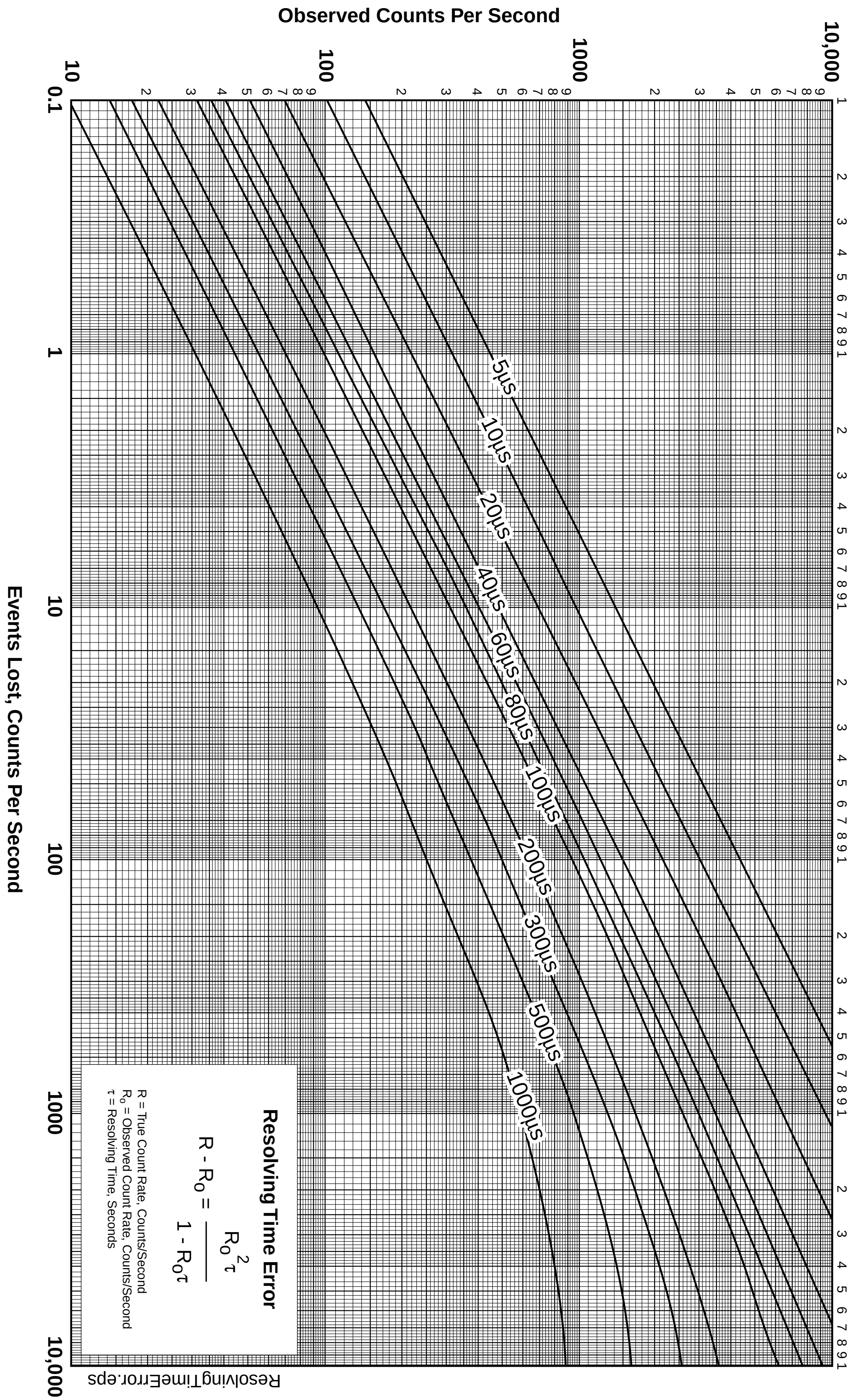


Figure 1.14.4 - Resolving Time Error

Error in Counts Per Minute as a Function of Total Count and Length of Count (95% Confidence Level)

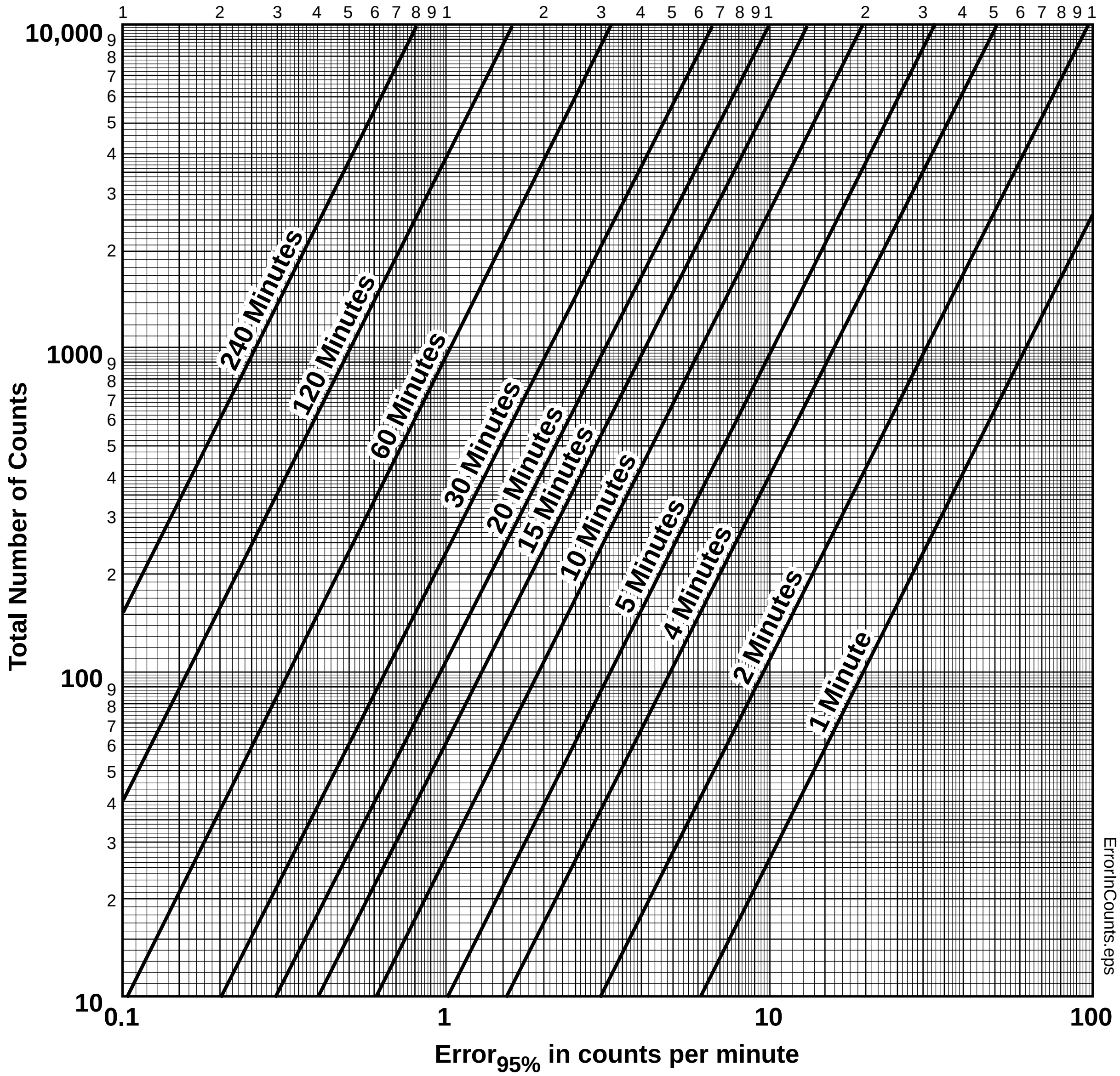


Figure 1.14.5 - Error in Counts per Minute

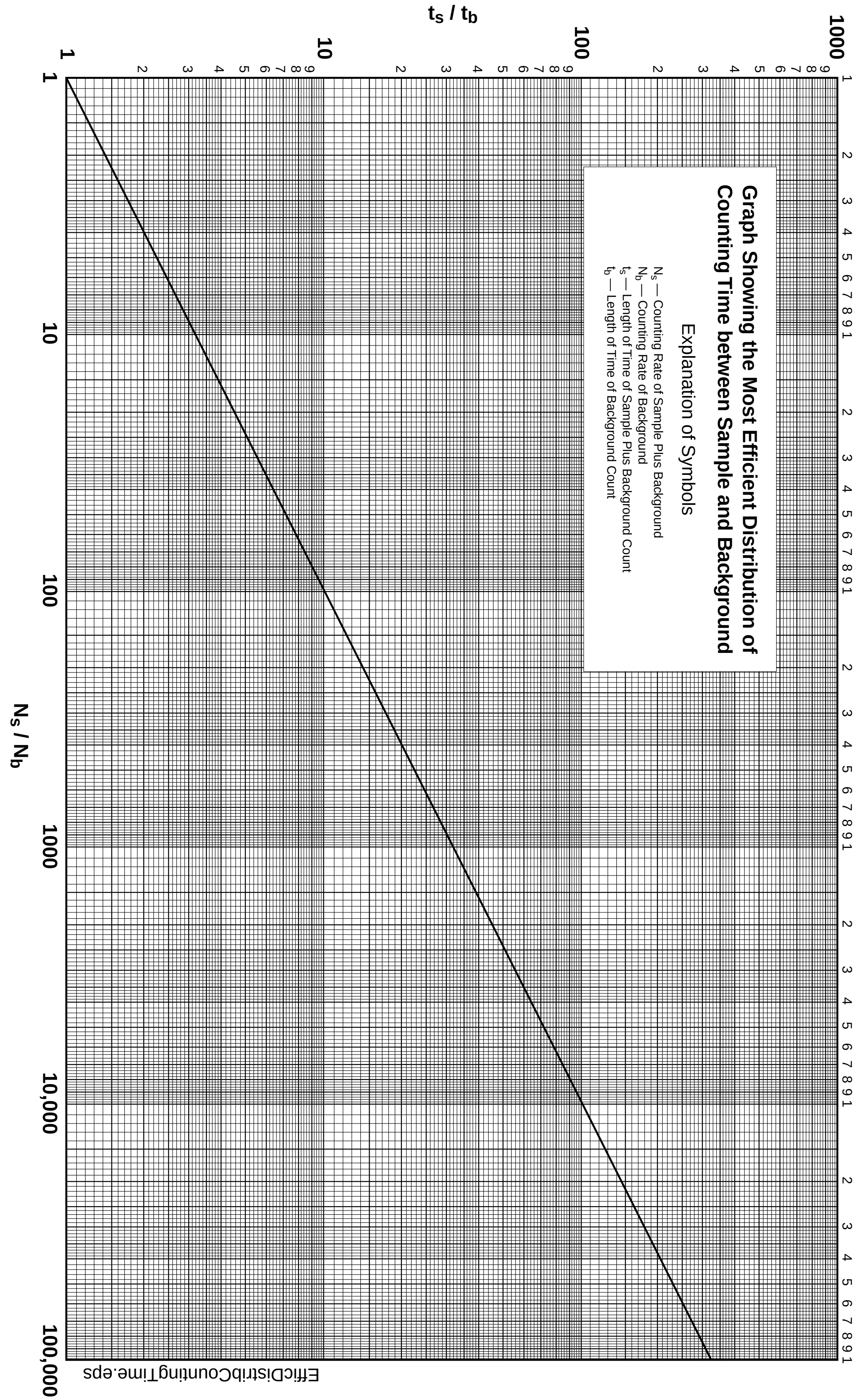


Figure 1.14.6 - Efficient distribution of counting time between sample and background

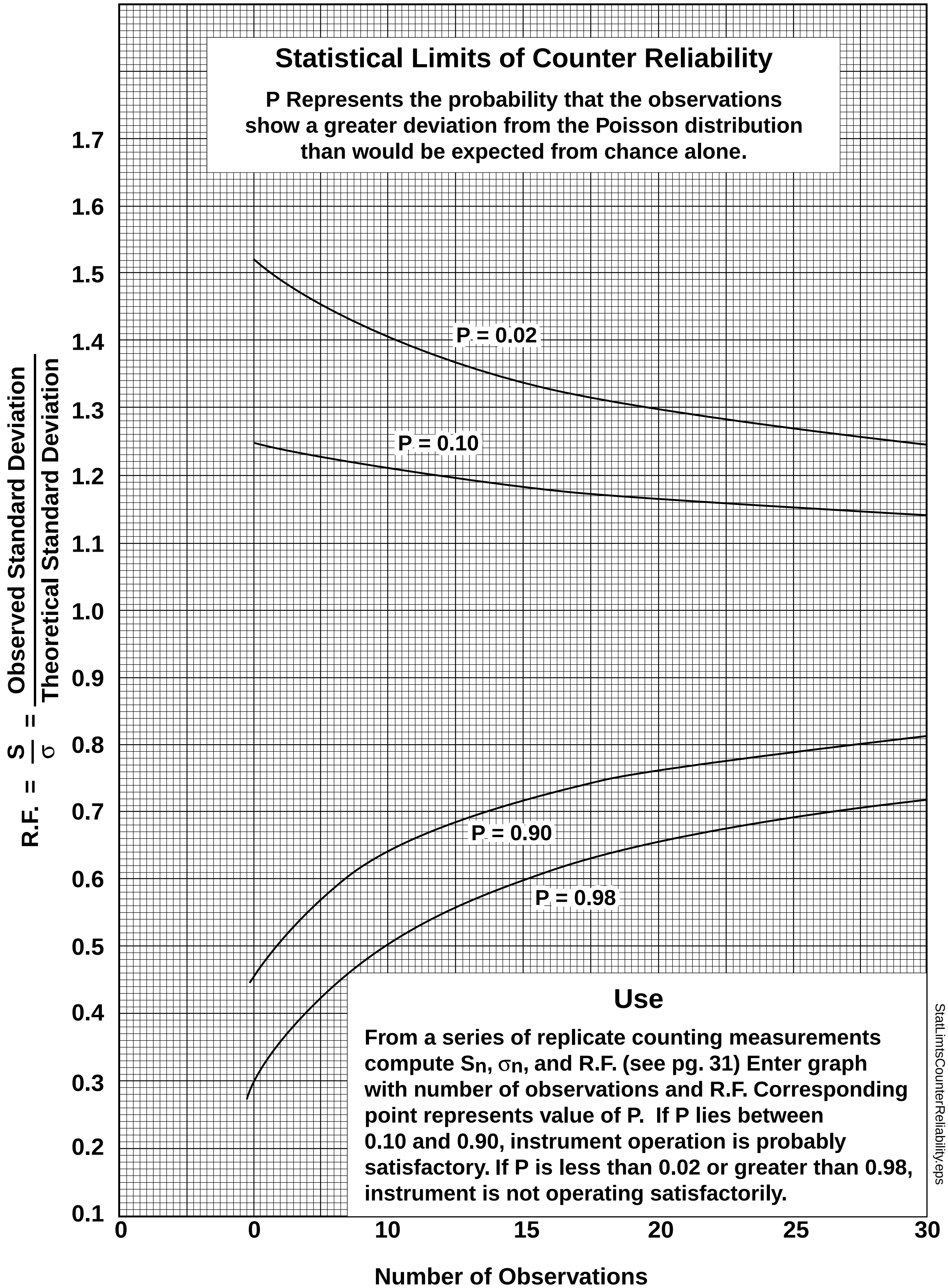
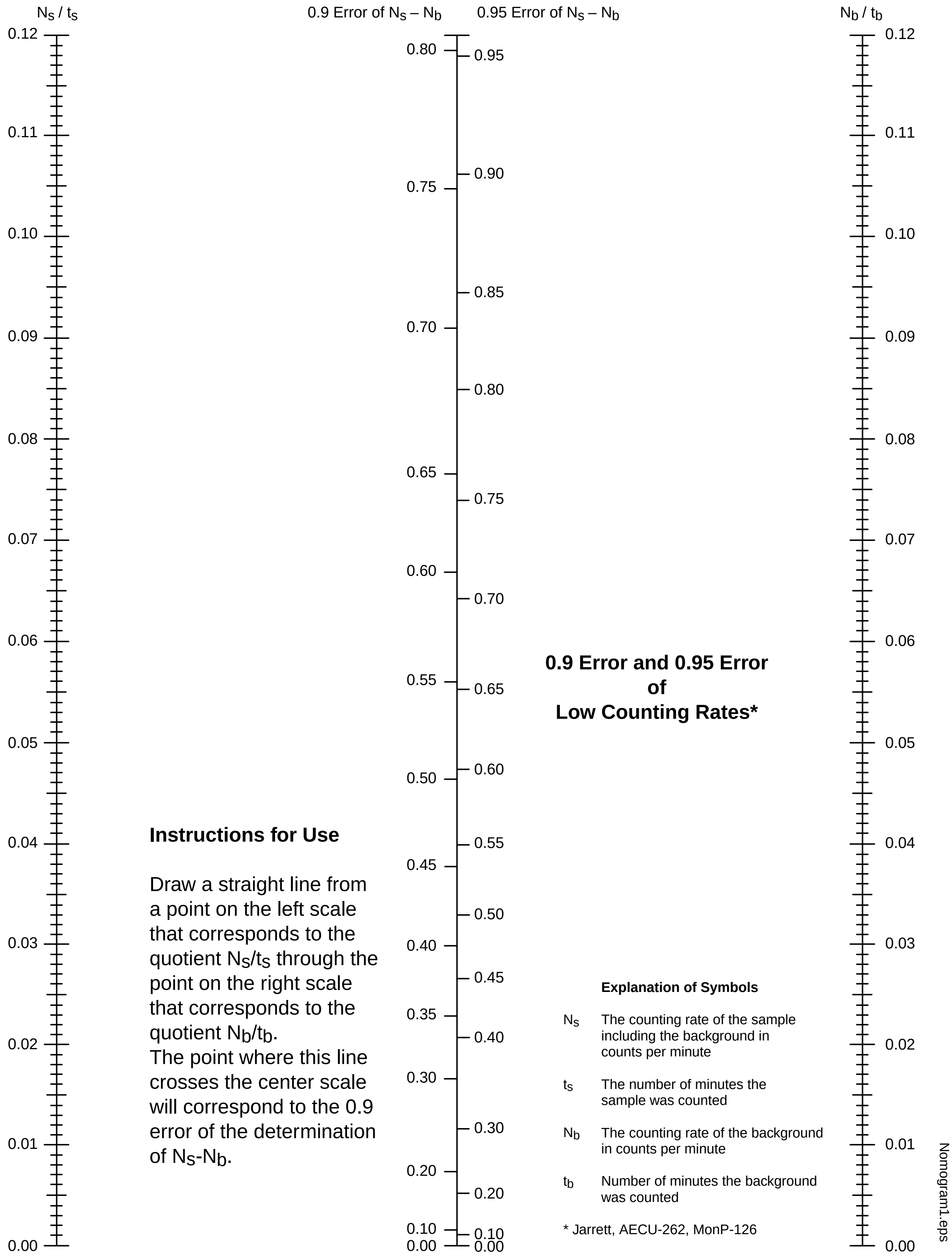


Figure 1.14.7 - Statistical Limits of Counter Reliability



1.14.8 - Nomogram for Error in Low Counting Rates

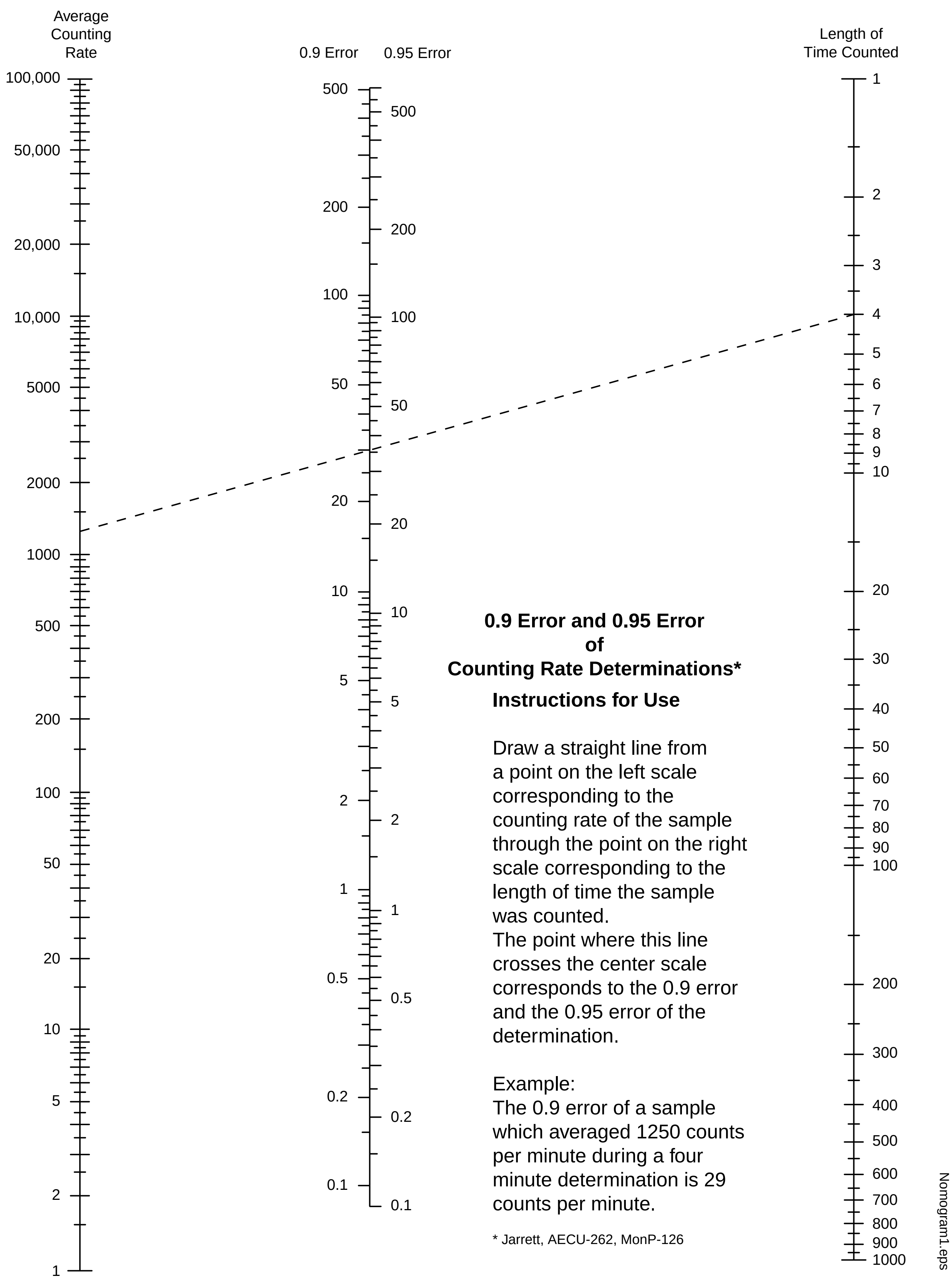
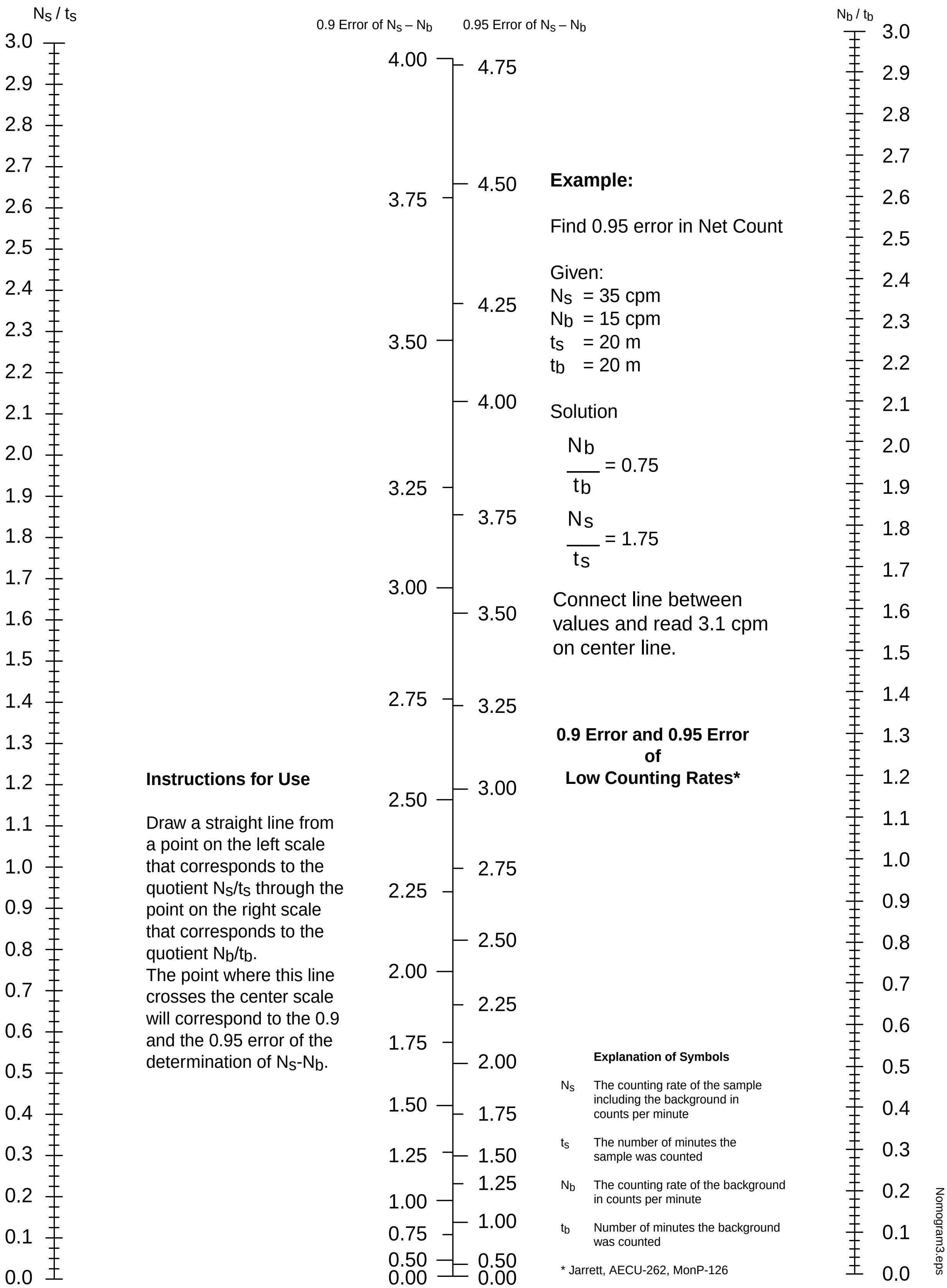


Figure 1.14.9 - Nomogram for Error in Counting Rate Determinations



Example:

Find 0.95 error in Net Count

Given:
N_s = 35 cpm
N_b = 15 cpm
t_s = 20 m
t_b = 20 m

Solution

$$\frac{N_b}{t_b} = 0.75$$

$$\frac{N_s}{t_s} = 1.75$$

Connect line between values and read 3.1 cpm on center line.

0.9 Error and 0.95 Error of Low Counting Rates*

Instructions for Use

Draw a straight line from a point on the left scale that corresponds to the quotient N_s/t_s through the point on the right scale that corresponds to the quotient N_b/t_b.
The point where this line crosses the center scale will correspond to the 0.9 and the 0.95 error of the determination of N_s-N_b.

Explanation of Symbols

N_s The counting rate of the sample including the background in counts per minute

t_s The number of minutes the sample was counted

N_b The counting rate of the background in counts per minute

t_b Number of minutes the background was counted

* Jarrett, AECU-262, MonP-126

Figure 1.14.10 - Nomogram for Error in Counting Rate Determinations