STATISTICS

To reduce the size of this section's PostScript file, we have divided it into three PostScript files. We present the following index:

PART 1

PART 2

PART 3

29.6. Errors and confidence intervals

We measure a mass, lifetime, or other physical quantity under the assumption that a "true answer" α exists. The conditions of the measurement introduce a random element, and our measurement (or combination of measurements) $\widehat{\alpha}_{\rm exp}$ samples a distribution with p.d.f. $f(\widehat{\alpha};\alpha)$. The unknown constant α appears as a parameter. We suppose that for every value of α we can find two values $\gamma_1(\alpha,\varepsilon)$ and $\gamma_2(\alpha,\varepsilon)$ such that repeated experiments would produce results in the interval $\gamma_1 < \widehat{\alpha} < \gamma_2$ a fraction $1-\varepsilon$ of the time, where

$$1-\varepsilon = \int_{\gamma_1}^{\gamma_2} f(\widehat{\alpha};\alpha)\,d\widehat{\alpha}\ . \tag{29.33}$$

This situation is shown in Fig. 29.1 (ignore the "unphysical region" part of the graph for now), where the region between the curves $\gamma_1(\alpha,\varepsilon)$ and $\gamma_2(\alpha,\varepsilon)$ is indicated by the domain $D(\varepsilon)$. It can be argued that since the point $(\alpha_{\rm actual},\widehat{\alpha}_{\rm exp})$ belongs to D, then our statement that repeated experiments would produce values of $\widehat{\alpha}$ in the interval $\gamma_1 < \widehat{\alpha} < \gamma_2$ is equivalent to the statement that the *confidence interval* $c_1 < \alpha < c_2$ includes $\alpha_{\rm actual}$ with probability $1-\varepsilon$ [1,6]. (We will call ε the *confidence coefficient.*) In this "confidence interval" or frequentist approach, α is a parameter, not a statistical variable. Instead, c_1 and c_2 vary from experiment to experiment and *are* statistical variables. It is very different to say that a lifetime τ is to be found in the interval $\tau_0 \pm \sigma_\tau$ with 68% probability than to say that the interval $\tau_0 \pm \sigma_\tau$ (which can vary from experiment to experiment) includes the actual, fixed, value of the lifetime with 68% probability. Any method which gives confidence intervals that contain the true value with probability $1-\varepsilon$ (no matter what the true value of α is) is said to have *coverage.* The frequentist intervals as constructed above have *coverage* by construction. Coverage is considered the most important property of confidence intervals [12].

The actual choice of γ_1 and γ_2, such that $\int_{\gamma_1}^{\gamma_2} f(\widehat{\alpha};\alpha)\,d\widehat{\alpha} = 1-\varepsilon$, can be made in an infinite number of ways, but in practical situations there are usually additional criteria. The most common is to choose *central intervals* such that the area of the excluded tail on either side is $\varepsilon/2$.

When the observable is restricted to integer values (as in the case of Poisson and binomial distributions), it is not generally possible to construct confidence intervals with exact coverage for all values of α. In these cases one constructs intervals with exact coverage for some values of α and unavoidable over-coverage for all other values. This is the best that can be done and still guarantee coverage for any true value. For a Poisson distribution negative values cannot occur, so $\gamma(\widehat{\alpha},\alpha)$ (with $\widehat{\alpha}$ an integer and α the Poisson mean) might be taken as the curve below which ε of the area under the distribution lies. (In this case the curve really consists of discrete points, since $\widehat{\alpha}$ can have only discrete values.) For $\varepsilon = 0.05$ the curve starts at $(\alpha,\widehat{\alpha}) = (3.0,0)$. If in a given experiment no decays to a certain final state are seen, we might then conclude that $\alpha < 3.0$ excludes the actual value of α with 95% probability. This statement can be converted to a similar statement about the branching fraction.

In Sec. 28 we discussed such confidence limits for a χ^2 distribution (where ε was called CL). Here we discuss confidence intervals for the Gaussian and Student's t-distribution,

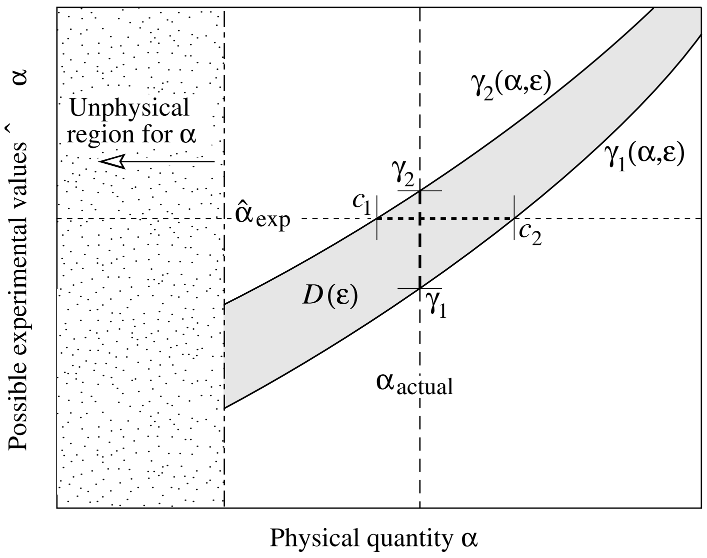

Figure 29.1: Confidence intervals for a single unknown parameter α. One might think of the p.d.f. $f(\widehat{\alpha};\alpha)$ as being plotted out of the paper as a function of $\widehat{\alpha}$ along each vertical line of constant α. The domain $D(\varepsilon)$ contains a fraction $1-\varepsilon$ of the area under each of these functions.

and confidence limits for the Poisson case. We then discuss the much more contentious situation in which the horizontal line at ordinate $\widehat{\alpha}$ in Fig. 29.1 enters $D(\varepsilon)$ at a boundary for unphysical values of α, so that at least c_1 is undefined—for example, if we find $\widehat{m}^2 = -30 \pm 50$ eV^2.

Extensive tables and graphs were once used to find confidence intervals and limits, but by now their main function is to confirm that software is working. FORTRAN mathematical libraries (IMSL, NAG, CERNLIB) are readily available, and a wide range of distributions are available in personal computer spreadsheet applications such as Microsoft® Excel [13]. Its built-in functions CHIDIST, NORMDIST, and TDIST (Student's t-distribution), along with "Solver," were used to produce or check the numbers given in this section.

29.6.1. *Gaussian errors*:

If the data are such that the distribution of the estimator(s) satisfies the central limit theorem discussed in Sec. 28.3.3, the Gaussian distribution is the basis of the error analysis. If there is more than one parameter being estimated, the multivariate Gaussian is used. For the univariate case with known σ,

$$1-\varepsilon = \int_{\widehat{\mu}-\delta}^{\widehat{\mu}+\delta} f(x;\,\widehat{\mu},\,\sigma^2)\,dx = \mathrm{erf}\left(\frac{\delta}{\sqrt{2}\,\sigma}\right) \tag{29.34}$$

is the probability that the true value of μ will fall within $\pm\delta$ $(\delta > 0)$ of the measured $\widehat{\mu}$. This interval will cover μ in a fraction $1-\varepsilon$ of all

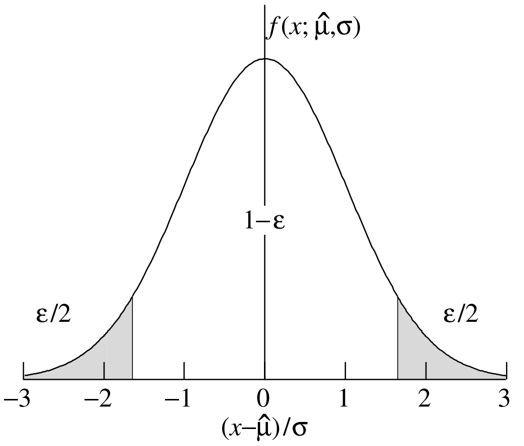

Figure 29.2: Illustration of a symetric 90% confidence interval (unshaded) for a measurement of a single quantity with Gaussian errors. Integrated probabilities, defined by ε, are as shown.

similar measurements. Fig. 29.2 shows a $\delta = 1.64\sigma$ confidence interval unshaded. The choice $\delta = \sqrt{\mathrm{Var}(\,\widehat{\mu}\,)} \equiv \sigma$ gives an interval called the *standard error* which has $1 - \varepsilon = 68.27\%$ if σ is known. Confidence coefficients ε for other frequently used choices of δ are given in Table 29.1. For other δ, find ε as the ordinate of Fig. 28.1 on the $n = 1$ curve at $\chi^2 = (\delta/\sigma)^2$. We can set a one-sided (upper or lower) limit by excluding above $\widehat{\mu} + \delta$ (or below $\widehat{\mu} - \delta$); ε's for such limits are 1/2 the values in the table above.

We have increased confidence that the interval covers the true value as $1 - \varepsilon$ increases, or χ^2 increases. We must be careful to distinguish this case from the other major use of Fig. 28.1, evaluation of goodness-of-fit (Sec. 29.5.0). In that case we have increased confidence in the fit as χ^2 decreases. In an attempt to reduce possible confusion in this discussion, we will use the ε notation (which corresponds to notation used in hypothesis testing [4]) when discussing confidence intervals and CL notation when discussing goodness-of-fit. Elsewhere in this *Review*, where the confusion between fit confidence level and interval (usually an upper or lower limit) confidence level does not arise, we follow the common practice of using "CL" to refer to the confidence level of the interval. This CL is understood to represent $1 - \varepsilon$.

If the variance σ^2 of the estimator is not known, but must be estimated from the data, then we need to incorporate the error in $\widehat{\sigma}$ into our confidence interval using Student's t distribution. If we have N data points with which we estimate k parameters, the Gaussian approximation is adequate for $N - k \gg 1$. Otherwise replace δ by a factor $T\widehat{\sigma}$, T being defined by

$$1 - \varepsilon = \int_{-T}^{T} f(t; N-k)\, dt \ , \tag{29.35}$$

where f for the Student's t-distribution is defined in Table 28.1. T is tabulated in Ref. 14 and in Table 29.2.

Table 29.1: Area of the tails ε outside $\pm\delta$ from the mean of a Gaussian distribution.

ε (%)	δ	ε (%)	δ
31.73	1σ	20	1.28σ
4.55	2σ	10	1.64σ
0.27	3σ	5	1.96σ
6.3×10^{-3}	4σ	1	2.58σ
5.7×10^{-5}	5σ	0.1	3.29σ
2.0×10^{-7}	6σ	0.01	3.89σ

Table 29.2: t limits containing $1-\varepsilon$ of the area of Student's t-distribution $f(t; N-k)$.

	ε (%)					
$N-k$	31.73	10.00	5.00	4.55	1.00	0.27
1	1.84	6.31	12.71	13.97	63.66	235.8
2	1.32	2.92	4.30	4.53	9.92	19.21
3	1.20	2.35	3.18	3.31	5.84	9.22
4	1.14	2.13	2.78	2.87	4.60	6.62
5	1.11	2.01	2.57	2.65	4.03	5.51
10	1.05	1.81	2.23	2.28	3.17	3.96
20	1.03	1.72	2.09	2.13	2.85	3.42
∞	1.00	1.64	1.96	2.00	2.58	3.00

For multivariate $\boldsymbol{\alpha}$ we must consider pairwise correlations. Assuming a multivariate Gaussian, Eq. (28.22), and subsequent discussion the standard error ellipse for the pair $(\widehat{\alpha}_m, \widehat{\alpha}_n)$ may be drawn as in Fig. 29.3.

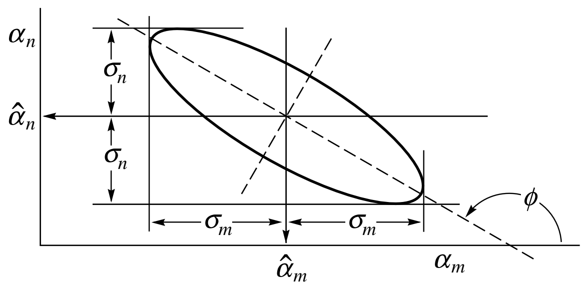

Figure 29.3: Standard error ellipse for the estimators $\widehat{\alpha}_m$ and $\widehat{\alpha}_n$. In this case the correlation is negative.

The minimum χ^2 or maximum likelihood solution is at $(\widehat{\alpha}_m, \widehat{\alpha}_n)$. The standard errors σ_m and σ_n are defined as shown, where the ellipse is at a constant value of $\chi^2 = \chi^2_{\rm min} + 1$ or $\ln \mathscr{L} = \ln \mathscr{L}_{\rm max} - 1/2$. The angle of the major axis of the ellipse is given by

$$\tan 2\phi = \frac{2\rho_{mn}\ \sigma_m\ \sigma_n}{\sigma_m^2 - \sigma_n^2}\ . \tag{29.36}$$

For non-Gaussian or nonlinear cases, one may construct an analogous contour from the same χ^2 or $\ln \mathscr{L}$ relations. Any other parameters $\widehat{\alpha}_\ell, \ell \neq m, n$ must be allowed freely to find their optimum values for every trial point.

For any unbiased procedure (*e.g.*, least squares or maximum likelihood) being used to estimate k parameters α_i, $i = 1, \ldots, k$, the probability $1 - \varepsilon$ that the true values of all k lie within the s-standard deviation ellipsoid may be found from Fig. 28.1. Read the ordinate as ε; the correct value of ε occurs on the $n = k$ curve at $\chi^2 = s^2$. For example, for $k = 2$, the probability that the true values of α_1 and α_2 simultaneously lie within the one-standard-deviation error ellipse ($s = 1$), centered on $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, is 39%. This probability only assumes Gaussian errors, unbiased estimators, and that the model describing the data in terms of the α_i is correct.

29.6.2. *Poisson processes—upper limits*:

Because the outcome of a Poisson process is an integral number of events, n_0, it is usually not possible to set confidence intervals for the true Poisson parameter μ at a certain exact ε. For large n_0 an approximate interval can be set using the Gaussian approximation, in our section on Probability, Sec. 28.3.2, and the techniques of Sec. 29.6.1.

For small n_0 we can define an upper limit N for μ as being that value of μ such that it would be at least $1 - \varepsilon$ (*e.g.*, 90% or 95%) probable that a random observation of n would then lie above the observed n_0. Thus

$$1 - \varepsilon = \sum_{n=n_0+1}^{\infty} f(n; N)\ ; \qquad \varepsilon = \sum_{n=0}^{n_0} f(n; N)\ . \tag{29.37}$$

Fig. 29.4 illustrates the case with $n_0 = 2$ and $1 - \varepsilon = 90\%$, for which it may be shown that $N = 5.32$. For any given n_0 and desired ε we can obtain N from the χ^2 Confidence Level figure because of a relation between the Poisson and the χ^2 distributions: read the ordinate as ε, find χ^2 on the curve for $n = 2(n_0 + 1)$; then $N = \chi^2/2$. Some useful values are given in Table 29.3.

The meaning of these upper limits is that, for a given true μ, the probability is *at least* $1 - \varepsilon$ that one will observe n_0 which will result in N which is $\geq \mu$. The probability for that to occur may be higher than $1 - \varepsilon$; for example, if $\mu \leq 2.30$ a "90%" upper limit will actually exceed μ 100% of the time. Note from Eq. (29.37) that for $n_0 = 0$, $N = -\ln \varepsilon$.

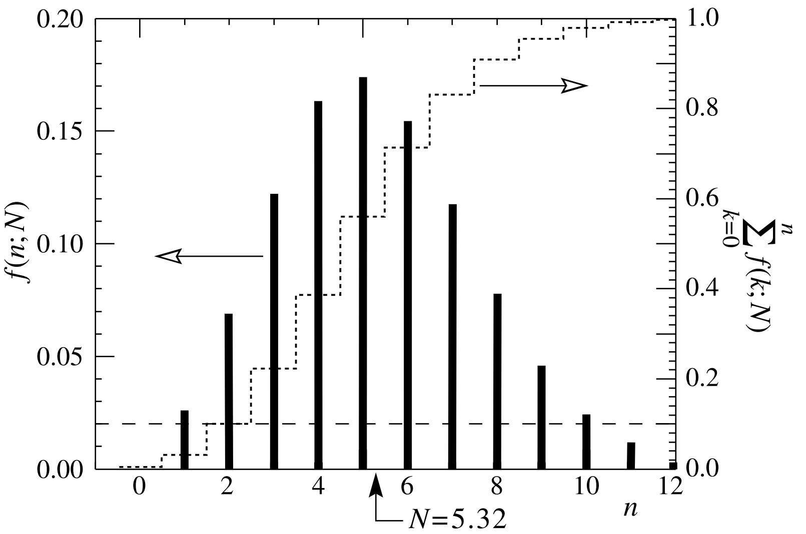

Figure 29.4: Illustration of Eq. (29.37) Poisson probabilities for an assumed mean of N. With an observed count $n_0 = 2$, $N = 5.32$ as shown gives summed probability $\varepsilon = 10\%$. The dotted summed probability curve (scale on right) has been displaced by -0.5 for clarity.

Table 29.3: Poisson upper limits N for n_0 observed events.

n_0	$\varepsilon =$ 10%	$\varepsilon =$ 5%	n_0	$\varepsilon =$ 10%	$\varepsilon =$ 5%
0	2.30	3.00	6	10.53	11.84
1	3.89	4.74	7	11.77	13.15
2	5.32	6.30	8	13.00	14.44
3	6.68	7.75	9	14.21	15.71
4	7.99	9.15	10	15.41	16.96
5	9.27	10.51	11	16.60	18.21