For example, in counting experiments one often groups the data in bins in order to associate a Poisson error with each bin. In this case y_i is the bin height and the error depends on the expectation value of the theory in each bin, as estimated by the best fit of the model. Since the variances are functions of the fitting parameters, neither the conditions of the Gauss-Markov theorem nor the assumption of Gaussian distributions (with or without fixed variances) in the context of the maximum likelihood approach are valid without a large-number approximation, and so an unbiased or efficient least-squares fit is not guaranteed. In such cases it seems more sensible to use one of the approaches discussed in Sec. 17.3.3.

17.5. 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

$$1-\varepsilon = \int_{\gamma_1}^{\gamma_2} f(\widehat{\alpha};\alpha)\, d\widehat{\alpha}\ . \tag{17.28}$$

This situation is shown in Fig. 17.1 (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.

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. For a Gaussian distribution, for example, choosing the limits symmetric about the mean minimizes the length of the interval. The area of the excluded tail on either side is then $\varepsilon/2$. 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. 16 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, 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. 17.1 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 [12]. 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.

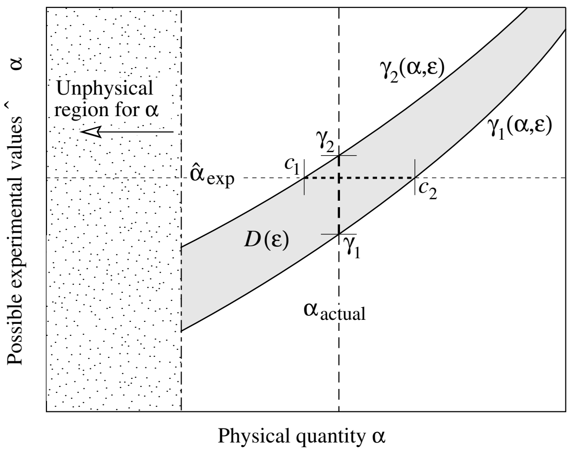

Figure 17.1: Confidence intervals for a single unknown parameter α. One might think of the p.d.f. $f(\widehat{\alpha};\alpha)$ as being plotted out of the paper as a function of $\widehat{\alpha}$ along each vertical line of constant α. The domain $D(\varepsilon)$ contains a fraction $1-\varepsilon$ of the area under each of these functions.

17.5.1. *Gaussian errors*:

If the data are such that the distribution of the estimator(s) satisfies the central limit theorem discussed in Sec. 16.3.3, 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,

$$1-\varepsilon = \int_{\widehat{\mu}-\delta}^{\widehat{\mu}+\delta} f(x;\widehat{\mu},\sigma^2)\,dx = \mathrm{erf}\left(\frac{x-\widehat{\mu}}{\sqrt{2}\,\sigma}\right) \tag{17.29}$$

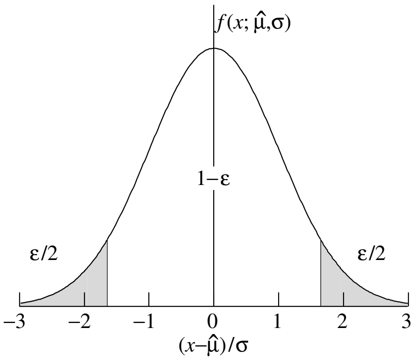

Figure 17.2: Illustration of a symetric 90% confidence interval (unshaded) for a measurement of a single quantity with Gaussian errors. Integrated probabilities, defined by ε, are as shown.

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 similar measurements. Fig. 17.2 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 17.1.

For other δ, find ε as the ordinate of Fig. 16.1 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. 16.1, evaluation of goodness-of-fit (Sec. 17.4.0). In that case we have increased confidence