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

28.6.3. *Bounded physical region**:

The measurement of a physical constant α results in an estimator $\widehat{\alpha}$, together with some knowledge of experimental error and therefore knowledge of $f(\widehat{\alpha};\alpha)$, the parameterized p.d.f. that allows us to state the probability with which repeated experiments would produce results in a given range. It does *not* permit us to comment about α itself, which in this language is a constant, not a statistical variable. At the beginning of this section we introduced the confidence interval, or frequentist, approach to the problem, and were able to say that with a given probability the unknown parameter could be found between (statistical) limits c_1 and c_2. But what if a physical boundary exists? Although polarization should be less than one and mass or its square should be greater than zero, experimental results do not always fall inside such a physical boundary because of statistical fluctuations.

However one might set a limit, there is little question about how to report and combine data [14]. A given experiment finds an unbiased estimator $\widehat{\alpha} = -5 \pm 10$ for a physical constant (*e.g.* the square of the mass of a neutrino, in eV^2). This value should be reported as the primary result. In case the true value is zero, for example, this "unphysical" result would not be unlikely. It can be combined with the results of other such experiments by forming the appropriately weighted average of unbiased results, including negative ones, to find an unbiased estimator which expresses our best knowledge of the parameter.

What if we wish to extend our concept of confidence limit to such a situation? The question of how to calculate an upper limit in the vicinity of a physical boundary is one of the most divisive in high-energy physics. We present two main approaches: The confidence interval, or frequentist, method, and the Bayesian method. "Classical method" is applied to one or the other by various writers, so we avoid the term.

1. The method of confidence intervals [1,15]. This is the approach described in the introduction, and requires little further explanation. It is presently the method in favor [1,6,14]. For a Gaussian distribution it gives the same result as the Bayesian approach with a flat prior distribution (see below) if the region containing α with the stated probability is far from an unphysical region, as in Fig. 28.1. Two cases in which this is untrue are shown in Fig. 28.5, where as a matter of convenience we assume that α must be positive. As before, we can define limits γ_1 and γ_2 for each value of the unknown parameter α, such that we can expect that a fraction $1-\varepsilon$ of repeated experiments to produce results between these limits. Since this can be done for each value of α, the limits are described by the functions $\gamma_1(\alpha,\varepsilon)$ and $\gamma_2(\alpha,\varepsilon)$. However, these *cannot* be extended into a region in which α makes no sense. Experimental result $\widehat{\alpha}_{\exp A}$, indicated in Fig. 28.5, is positive, but if the true value is α_{actual} a significant fraction of repetitions of the experiment would produce negative $\widehat{\alpha}$. In these cases there is no horizontal intercept c_1, so without further assumptions we cannot make a statement about the region which would cover α_{actual} in a given fraction of experiments. Experimental result $\widehat{\alpha}_{\exp B}$ presents a more serious problem, since it is so negative that there is no physical α for which the point $(\widehat{\alpha},\alpha)$ lies in the domain $D(\varepsilon)$. The reason why the frequentist method gives no confidence interval is clear: This measured value of $\widehat{\alpha}$ would be unlikely *no matter what the true value of α was.*

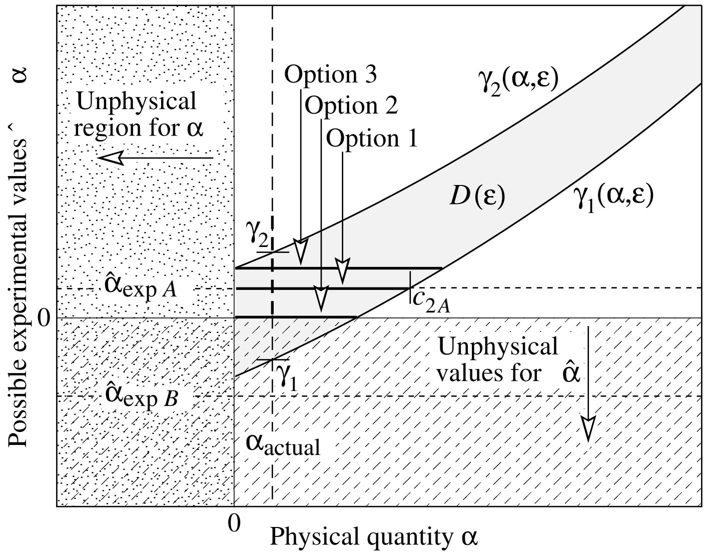

Figure 28.5: The situation near a physical boundary. In Fig. 28.1 the horizontal line for a given $\widehat{\alpha}_{\rm exp}$ crossed the domain $D(\varepsilon)$, bounded by $\gamma_1(\alpha,\epsilon)$ and $\gamma_2(\alpha,\epsilon)$ entirely in the physical region, entering at c_1 and leaving at c_2. The limits γ_1 and γ_2 cannot be defined in a region where α is not defined, so the functions cannot be continued into the unphysical region. As a result c_1 (for experiment A) or c_1 and c_2 (for experiment B) cannot be defined. Options 1, 2, and 3 label the ways one might define confidence intervals, as described in the text.

There are several *ad hoc* ways to set confidence limits in such cases, although many frequentists would prefer to stop with the weighted average of unbiased results—if the outcome is exceedingly unlikely, one should look to the experiment, not to the statistics. The methods we list below all involve placing c_1 on the physical boundary, which in our example is at $\alpha = 0$.

1. If $\widehat{\alpha}_{\rm exp} > \gamma_1(0,\varepsilon)$, as in Experiment A, c_2 is defined. Use it for the upper limit, whether or not $\widehat{\alpha}_{\rm exp} > 0$.
2. If $\widehat{\alpha}_{\rm exp} < 0$, as in Experiment B, use the c_2 corresponding to $\widehat{\alpha}_{\rm exp} = 0$.
3. If c_1 is not defined, "lift up" $\widehat{\alpha}$ to $\gamma_2(0,\varepsilon)$, where $c_1 = 0$. Use the corresponding c_2 as the upper limit.

These three options are shown in in Fig. 28.5; note that there are regions where more than one of them can be used, with different results. The third option is certainly the most conservative. For Gaussian $f(\widehat{\alpha};\alpha)$ the upper limit c_2 is a one-sided Gaussian confidence limit; read the tables for a 90% two-sided limit to obtain 95% one-sided limit. Alternatively, read the intercepts of the dotted lines in Fig. 28.7. (The horizontal axis is incorrectly labeled for this application.)

2. The Bayesian approach [3]. This is the approach favored in the older literature, and has (unfortunately and incorrectly) been referred to as the "PDG method" in certain

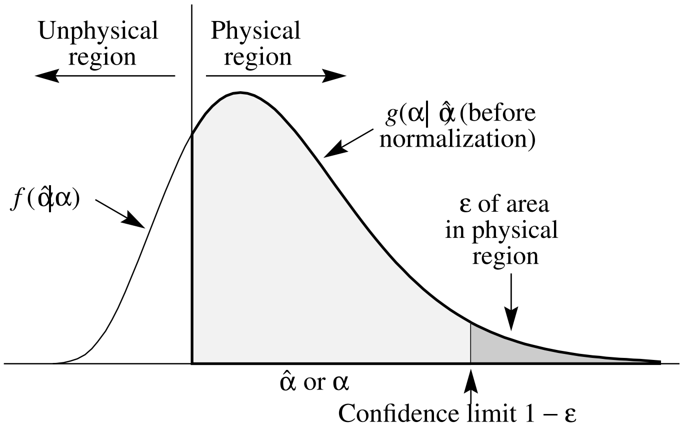

Figure 28.6: An example of a bounded physical region, in which a measurement $\widehat{\alpha}$ can fall in an unphysical region with significant probability. If we assume that α, the quantity we are trying to measure, cannot lie in the unphysical region (0 probability) but can lie anywhere in the physical region ("no prior knowledge"), then Bayes' theorem says that our new knowledge of the distribution of α, given our measurement $\widehat{\alpha}$, is given by the shaded function after appropriate renormalization.

papers. To begin with, it is argued that while α is not a statistical variable, our *knowledge* of α is less than complete, and it is fair to describe our uncertainty by treating α as a statistical variable. The parameterized p.d.f. $f(\widehat{\alpha};\alpha)$ is replaced by the conditional p.d.f. $f(\widehat{\alpha}|\alpha)$. The confidence limit question can then be rephrased: Our measurements provide $f(\widehat{\alpha}|\alpha)$, that is, information about $\widehat{\alpha}$ for a fixed and unknown value of α, while we really want to know $g(\alpha|\widehat{\alpha})$, which tells us that, given our measurement $\widehat{\alpha}$, the "true answer" α lies between α and $\alpha + d\alpha$ with probability $g(\alpha|\widehat{\alpha})\,d\alpha$. The connection is provided by Bayes' theorem (Eq. (27.7):

$$g(\alpha|\widehat{\alpha}) = \frac{f(\widehat{\alpha}|\alpha)\,\pi(\alpha)}{\int f(\widehat{\alpha}|\alpha)\,\pi(\alpha)\,d\alpha}\ . \qquad (28.38)$$

Here $\pi(\alpha)$ represents our "advance knowledge" of the value of α. In the usual case we claim no prior knowledge, so that before the experiment all physically reasonable values of α are equally probable: $\pi(\alpha)$ is a constant over the region of interest and zero in the unphysical region. This assumption leads to the conclusion that

$$g(\alpha|\widehat{\alpha}) = \begin{cases} f(\widehat{\alpha}|\alpha)/\int f(\widehat{\alpha}|\alpha)\,d\alpha & \text{if } \alpha \text{ is in the physical region;} \\ 0 & \text{otherwise;} \end{cases} \qquad (28.39)$$

where this time the integral is over the physical region. In Fig. 28.6 we assume that an ensemble of experiments would produce values for $\widehat{\alpha}$ which distribute as shown, with a significant probability of obtaining results with unphysical values. With our assumed step function $\pi(\alpha)$, the effect of Eq. (28.38) or (28.39) is to replace this distribution with the function shown by the shaded region, except that it is renormalized to unit area. By stating our confidence at the 90% level that α lies below the beginning of the dark shaded region, we mean that 90% of the area in the physical region is in the light shaded region.

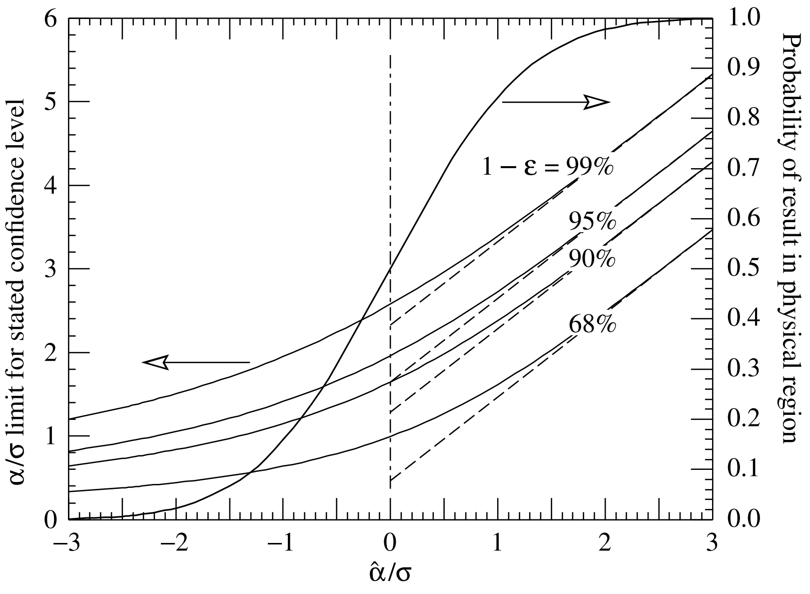

Figure 28.7: Application of the Bayesian scheme shown in Fig. 28.6 to the case of Gaussian $f(\widehat{\alpha}|\alpha)$. For example, if our measurement $\widehat{\alpha}$ is 1.0 standard deviations negative, then we conclude that $\alpha < 1.15\sigma$ with 90% probability—however, there is only a 31% probability that an experimental result as low as this would occur. Note that these are upper limits, so that the asymptote for large $\widehat{\alpha}/\sigma$ corresponds to a one-sided confidence interval, *e.g.*, the asymptote for a 95% confidence level is $\alpha < \widehat{\alpha} + 1.64\sigma$, corresponding to a 90% confidence interval for a two-sided distribution. The dashed lines show the frequentist limit; if Option 3 is used, these are extended horizontally to the right for negative $\widehat{\alpha}/\sigma$.

In most cases of interest in this Review, $\widehat{\alpha}$ is assumed to be a random value from a Gaussian distribution. Application of the procedure sketched in Fig. 28.6 then leads to the family of curves shown in Fig. 28.7. The confidence limit set by this method is always greater than the [one-sided] confidence interval set without the restriction of an unphysical region, and approaches it from above as the tail in the unphysical region becomes unimportant. It is also greater than any of the limits shown in Fig. 28.5. With a small modification (exclusion of that portion of the negative tail inside the physical region in the confidence interval definition), it smoothly approaches the usual two-sided confidence interval for Gaussian distributions.

Even so, it is not a valid confidence limit. If it were, the interval would include the true value of α with *exactly* $1 - \varepsilon$ probability no matter what the true value was. If the true answer is zero, our procedure, by guaranteeing a limit greater than zero for any experiment, also guarantees that the confidence interval for any ε includes α_{actual} with 100% probability. Only as α increases does the probability decrease toward the α-independent Gaussian result.

The error function corresponding to the right axis of Fig. 28.7 shows the probability that $\widehat{\alpha}/\sigma$ at or below the given value should occur. If the experimental value is exceeding improbable, then the formal confidence limit obtained by this or any other method means very little.

What about the arbitrariness of $\pi(\alpha)$? If the square of the neutrino mass is measured ($\alpha = m_\nu^2$), then should we not take the prior knowledge distribution as proportional to $\sqrt{\alpha}$?[†] There are other attractive options. Jeffreys points out that if $\pi(\alpha)\,d\alpha = d\alpha/\alpha$, then the distributions for α and α^n are proportional [16], but there are practical difficulties with this approach. Lynch has investigated prior distributions that are constant in α, α^2, and $\sqrt{\alpha}$ in the context of Gaussian $f(\widehat{\alpha}|\alpha)$, and has observed that assuming a prior distribution that is flat in α gives results that are much more satisfactory than one gets from the others [17]: All three methods have the property that the probability that the calculated limit contains the correct answer is 100% when $\alpha = 0$ and approaches the proper value when $\alpha \gg \sigma$, but the approach to the proper value as α increases is much faster when the prior distribution is taken to be flat in α. In this case the approach is also monotonic, giving it the "conservative" property that for no value of α will the method produce a limit that has a probability of being correct that is less than the stated confidence limit. Although there is nothing unique about the limit calculated with a constant $\pi(\alpha)$, it has desirable features and no obvious replacement.

Summary: If there is a significant probability of obtaining an estimator corresponding to an unphysical value for a parameter, there is no universally accepted way way to make a statement of the sort "α is less that c_2 with probability $1-\varepsilon$." A variety of upper limits can be defined, but no method is entirely satisfactory. The Bayesian method with a flat prior distribution gives a reasonable upper limit which combines everything we know about the unknown quantity α into a physically reasonable value, but it does not give a complete summary of the information contained in the experiment.

28.6.4. *Poisson processes with background* [18] :

If we observe n_0 events in a Poisson process which has two components, signal and background, estimating a limit on the signal is more complicated. Let μ_S be the unknown mean (the Poisson parameter) for the signal and μ_B be the mean for the sum of all backgrounds. Assume μ_B is known with negligible error; however we don't know n_B, the actual number of events resulting from the background. We do know that $n_B \le n_0$. If $\mu_B + \mu_S$ is large, the Gaussian approximation to the Poisson distribution (see Sec. 27.3.2) is usually adequate, and one can define confidence intervals or limits as above, assuming $\widehat{n}_B \approx \mu_B$ and therefore $\widehat{\mu}_S = n_0 - \mu_B$ with variance equal to n_0 (larger than $\widehat{\mu}_S$ to allow for the error in $\widehat{n}_B$).

Otherwise an upper limit can be defined by extension of the argument of the preceding section. Let N be the desired upper limit on μ_S with confidence coefficient ε. Set N to be that value of μ_S such that any random repeat of the current experiment with $\mu_S = N$ and the same μ_B would observe *more* than n_0 events in total *and* would have $n_B \le n_0$, all with probability $1-\varepsilon$. For any assumed N and μ_B we can calculate this probability:

$$1-\varepsilon = 1 - \frac{e^{-(\mu_B+N)} \displaystyle\sum_{n=0}^{n_0} \frac{(\mu_B+N)^n}{n!}}{e^{-\mu_B} \displaystyle\sum_{n=0}^{n_0} \frac{\mu_B^n}{n!}} \,. \tag{28.40}$$

We adjust N to obtain a desired ε. For $\mu_B = 0$ this converges to Eq. (28.37). As in that case (see the last paragraph of Section 28.6.2) this gives a *conservative* upper limit in that for any given true μ_S we get a true probability $\geq 1-\varepsilon$ that $N \geq \mu_S$, averaged over a large set of identically performed experiments. For $\varepsilon = 0.10$, Fig. 28.8 shows N as a function of n_0 and μ_B.

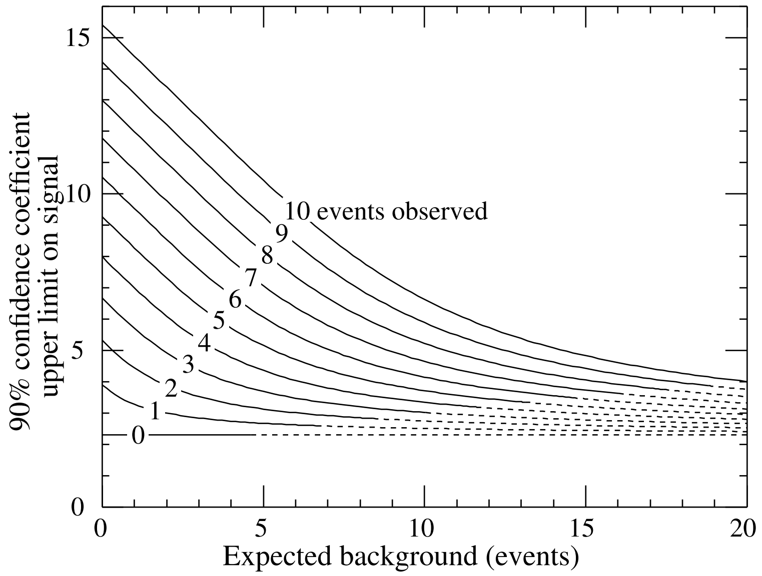

Figure 28.8: 90% confidence coefficient upper limit on the number of signal events as a function of the expected number of background events. For example, if the expected background is 8 events and 5 events are observed, then the signal is 4.0 (approximately) or less with 90% confidence. Dashed portions indicate regions where it is to be expected that the number observed would exceed the number actually observed $\geq$ 99% of the time, even in the complete absence of signal.

Averaging of experiments and other comparisons require that n_0 and μ_B be quoted and the technique used for upper limit extraction be given.

If $\mu_B \gg n_0$ the experimenter should question the probability of observing n_B as that n_0. If this is very small the background, μ_B, may not have been calculated properly and the upper limit for μ_S obtained under those assumptions may be too low. For example, in Fig. 28.8, the dashed portions of the curves lie in the region where n_0 is expected to exceed the observed value 99% of the time (or more), even in the complete absence of signal. In these regions one should be cautious about accepting the results of the measurement.

As in the Gaussian case (Sec. 28.6.3), whenever $n_0 < \mu_B$ some experimenters may prefer to use N calculated as if $n_0 \approx \mu_B$ rather than the smaller value obtained from the observed n_0.

*In addition to the references cited, communications with R.D. Cousins, F. James, G. Lynch, and B. Roe have been invaluable in formulating this section.

†There is an additional problem: Even if we set a confidence limit on m_ν^2 by some particular recipe, it translates into a different confidence limit on $\sqrt{m_\nu^2}$ except when a Bayesian procedure with Jeffreys' $\pi(\alpha) \propto 1/\alpha$ prior distribution is used.

References:

1. H. Cramér, *Mathematical Methods of Statistics,* Princeton Univ. Press, New Jersey (1958).
2. J. Orear, "Notes on Statistics for Physicists," Univ. of Calif. Radiation Lab. Report UCRL-8417 (1958), unpublished; "Notes on Statistics for Physicists, Revised," Cornell Univ. Report CLNS 82/511 (1982), unpublished.
3. F.T. Solmitz, Annu. Rev. Nucl. Sci. **14**, 375–402 (1964).
4. W.T. Eadie, D. Drijard, F.E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North Holland, Amsterdam and London, 1971).
5. L. Lyons, *Statistics for Nuclear and Particle Physicists,* (Cambridge University Press, New York (1986)).
6. B.P. Roe, *Probability and Statistics in Experimental Physics,* (Springer-Verlag, New York, 208 pp., 1992).
7. S. Baker and R. Cousins, Nucl. Instr. and Meth. **221**, 437 (1984).
8. W.H. Press *et al.*, *Numerical Recipes* (Cambridge University Press, New York, 1986).
9. W.H. Maindonald *et al.*, *Statistical Computation* (John Wiley and Sons, Inc., New York, 1984).
10. A. Basilevsky *et al.*, *Applied Matrix Algebra in the Statistical Sciences* (North Holland, New York, 1983).
11. F. James and M. Roos, "MINUIT, Function Minimization and Error Analysis," CERN D506 (Long Writeup). Available from the CERN Program Library Office, CERN-DD Division, CERN, CH-1211, Geneva 21, Switzerland.
12. Microsoft® is a registered trademark of Microsoft Corporation.
13. M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions* (Dover, New York, 1972).
14. F. James and M. Roos, Phys. Rev. **D44**, 299 (1991).

15. J. Neyman, *Lectures and Conferences on Mathematical Statistics,*" Washington (1938).
16. H. Jeffreys, *Theory of Probability,* 3rd ed. (Oxford Univ. Press, Oxford, England, 447 pp., 1961).
17. G. Lynch, "Calculations of Upper Limits," Particle Data Group Note PDG-93-03 (Sept. 1993) (unpublished).
18. O. Helene, Nucl. Instr. and Meth. **212**, 319 (1983).