CHAPTER 9

FITTING COMPOSITE CURVES

9.1 LORENTZIAN PEAK ON QUADRATIC BACKGROUND

Many fitting problems involve determining the parameters of a resonant peak or peaks, superimposed upon a background signal. Examples may be found in various types of spectroscopic studies where the objective is to determine the properties of one or more resonant states.

EXAMPLE 9.1 We consider a problem from nuclear or particle physics illustrated by the 4000-event histogram of Figure 9.1, which shows a large peak on a smoothly varying background. We shall assume that the data have been drawn from a distribution that includes a resonant state described by the Lorentzian distribution, and that the background can be described by a second-degree polynomial in the energy E.[1] We shall attempt to fit Equation (9.1) to the data to determine the amplitude A_0, the resonant energy E_0, and the full width at half maximum Γ.

$$y(E) = a_1 + a_2E + a_3E^2 + A_0\frac{\Gamma/(2\pi)}{(E - E_0)^2 + (\Gamma/2)^2} \tag{9.1}$$

We note that Equation (9.1) is linear in the parameters a_1, a_2, and a_3, but not in the parameters E_0 and Γ.

[1]These "data" were actually generated by the Monte Carlo method described in Chapter 5. The parameters used in the generation are listed in the second column of Table 9.1.

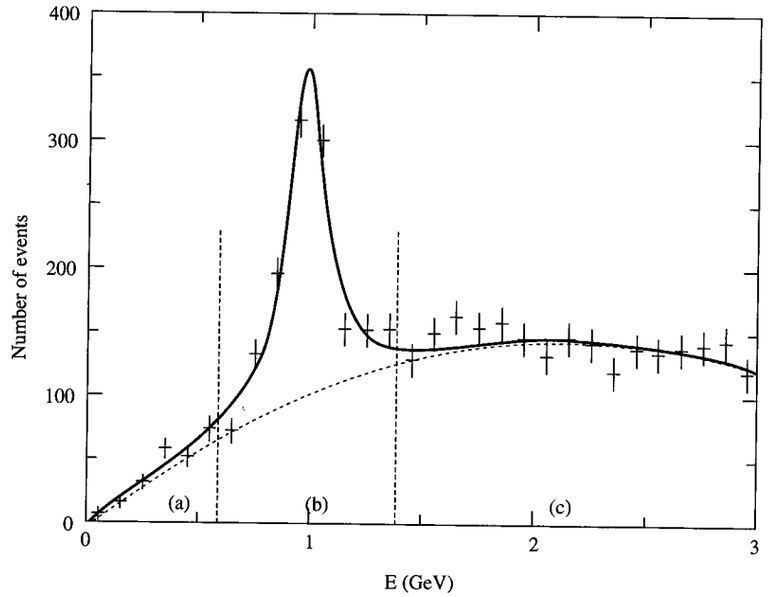

FIGURE 9.1
Histogram data in bins of 0.10 GeV of the 4000 simulated events generated from Equation (9.13) with parameters listed in column 2 of Table 9.1. The solid curve illustrates a fit of Equation (9.1) to the data. The dashed curve indicates the polynomial background.

We used the Marquardt method with numerical derivatives to fit Equation (9.1) to the histogram of Figure 9.1, because this is clearly the most flexible and convenient of the four methods considered in Chapter 8. The amplitudes of the polynomial function (a_1 through a_3), the amplitude of the Lorentzian peak ($a_4 = A_0$), and the mean E_0 and half-width Γ of the Lorentzian function (a_5 and a_6) were treated as free parameters of the fit. Starting values for a_5 and a_6 were obtained by inspecting the histogram of Figure 9.1; starting values for the other parameters, the coefficients of the various terms, were obtained by trial and error. Because the Marquardt method is exact for a function that is linear in the parameters, convergence of the fit is relatively insensitive to starting values of a_1 through a_4. The method is more sensitive to starting values for the Lorentzian parameters (E_0 and Γ). If starting values were too far from the obvious parameters of the peak, the program would coast to a halt in a shallow local minimum with obviously incorrect values for the parameters, and with a higher than expected value of χ^2. Starting values for all fits are listed in column 3 of Table 9.1.

Results of this six-parameter fit to the distribution in Figure 9.1 are summarized in column 4 of Table 9.1 and the curve calculated from Equation (9.1) with the parameters found in the fit is plotted on the histogram of Figure 9.1. The dashed curve shows the contribution of background under the peak. The χ^2 probability of the fit (7.9%) is low, but acceptable.

Because one of the objectives of the analysis of Example 9.1 is to determine E_0, the mean of the peak function of Equation (9.1), we must be careful in the

TABLE 9.1
Results of least-squares fits of Equations (9.1) and (9.13) to data displayed in Figures 9.1 and 9.2

	Values used to generate data	Starting values for fit	Six-parameter fit (Figure 9.1)	Six-parameter fit (Figure 9.2-inset)	Nine-parameter fit (Figure 9.2)
dof			24	54	51
χ^2			34.3	72.9	56.0
$P_{\chi 2}$			7.9%	4.4%	29.4
Num	4000		3944	3927	3994
a_1	1.0	1	2.2 ± 2.6	−2.2 ± 1.3	−2.1 ± 1.1
a_2	45.0	1	136.0 ± 8.1	73.9 ± 3.7	73.7 ± 3.6
a_3	−10.	1	−31.6 ± 3.1	−18.0 ± 1.4	−18.0 ± 1.4
	Peak 1				
a_4 (A_0)	20.0	1	79.8 ± 7.0	33.9 ± 2.7	28.8 ± 3.0
a_5 (E_0)	1.0	1	0.9838 ± 0.0068	0.9912 ± 0.0050	0.9968 ± 0.0044
a_6 (Γ_0)	0.1	0.1	0.197 ± 0.024	0.139 ± 0.015	0.108 ± 0.017
	Peak 2				
a_7 (A_1)	3.5	1	—	—	5.3 ± 2.2
a_8 (E_1)	0.8	0.825	—	—	0.824 ± 0.017
a_9 (Γ_1)	0.12	0.05	—	—	0.083 ± 0.034

choice of the value of the independent variable that we use in the fit. On the histogram of Figure 9.1, the value of E_i at the left-hand edge of selected bins is indicated, but for the fit we used the value of E at the center of each bin. If we had used values of E_i from the left-hand edge of the histogram bins, the value for E_0 from the fit would have been too low by half a bin width. For wide bins and a rapidly varying fitting function, it might be advisable to select the value of E_i for each bin by weighting according to the steepness of the function.

Note that the problem of selecting the absolute value of the abscissa corresponding to the ordinate value was not important in the determination of the mean lifetimes in Example 8.1 because lifetimes are determined effectively from differences, rather than absolute values, of the independent variable. We must, however, always take care when we plot results of a fit that the curve is not displaced half a bin width from the data.

Program 9.1 LORINFIT (Appendix E) illustrates use of the Marquardt method to fit a Lorentzian peak on a quadratic background.

9.2 AREA DETERMINATION

When dealing with problems of peaks and backgrounds, we may wish to determine not only the position and width of a peak in a spectrum, but also the number of

events or area of the peak, which may measure the intensity of a transition or the strength of a reaction. When peaks are not well separated, or when the contribution from background is substantial, least-squares fitting can provide a consistent method of extracting such information from the data.

The importance of consistency should not be underestimated. Whether or not the method chosen is the best possible method, as long as it involves a well-understood and clearly specified procedure, other experimenters will be able to check and compare the results safe in the knowledge that their comparisons are justified and meaningful. The method of least squares is considered to be an unbiased estimator of the fitting parameters and all parameters are presumed to be estimated as well as possible. This assumption is based on the validity of both the fitting function in describing the data and the least-squares method. If we try to fit the data with an incorrect fitting function, or try to fit data with uncertainties that do not follow the Gaussian distribution, then the fitting procedure may not yield optimum results.

Although we refer to the number of events as the area of a peak or plot, the true area is, of course, the number of events multiplied by the data interval or histogram bin width. Thus, to find the area A_p of the peak from the results of the fit in Example 9.1, we calculate

$$A_P = \int_{-\infty}^{\infty} A_0 \frac{\Gamma/(2\pi)}{(E - E_0)^2 + (\Gamma/2)^2} \tag{9.2}$$

Because we used the normalized form of the Lorentzian function, the integral is just the coefficient a_4 obtained in the search $A_P = A_0 = a_4$. The area of the peak on the histogram is the product of the number of events N_P in the peak and the width ΔE of the histogram bin

$$A_P = N_P \times \Delta E,$$

so the number of events in the peak is given by

$$N_P = A_P/\Delta E \tag{9.3}$$

The result from Example 9.1 is $N_P = (79.8 \pm 7.0)/0.1 = (798 \pm 70)$ events.

Alternatively, we might plot the background curve on the graph

$$y_b(E) = a_1 + a_2 E + a_3 E^2 \tag{9.4}$$

and count the number of events in the peak above the background in a selected range encompassing the peak. We have indicated such a range by vertical dotted lines at $E_0 - 2\Gamma$ and $E_0 + 2\Gamma$ in Figure 9.1. With this method we should be obliged to estimate and correct for events outside the selected region.

Uncertainties in Areas under Peaks

If we calculate the area of the peak from Equations (9.2) and (9.3), then the uncertainty should be estimated from the uncertainties in the parameters by the error propagation equation. We have used this method to obtain the uncertainty in the number of events of the peak of Figure 9.1 in the calculation that follows Equation (9.3).

The uncertainty σ_A in the area under a peak can also be estimated by considering the uncertainty in the parent distribution. If the data are distributed according to the Poisson distribution, the uncertainty in the area A_p is given by $\sigma_a^2 \simeq A_p$. If we obtain the area by counting the number of events above background, then the variance of the difference will be the sum *(not the difference)* of the variance of the total area under the peak and the variance of the subtracted background A_b:

$$\sigma_p^2 = \sigma_t^2 + \sigma_b^2 = A_t + A_b$$

where the subscripts p, b, and t correspond to peak, background, and total (= peak + background). In order to keep $s_t = A_t$ as small as possible, we should count events only in that region where the peak-to-background ratio is large and make corrections for the tails of the distribution.

Area under a Curve with Poisson Statistics

Curiously enough, if the data are distributed around each data point according to the Poisson distribution, as in a counting experiment, the method of least squares consistently *underestimates* the area under a fitted curve by an amount approximately equal to the value of χ^2. To show this, let us consider fitting such data with an arbitrary peak, represented by $bf_p(x; \mu, \sigma)$ plus a polynomial background similar to Example 9.1:

$$y(x) = a + bf(x; \mu, \sigma) \tag{9.5}$$

where we have simplified the background to a single term a for clarity.

Using the method of least squares, we define χ^2 to be the weighted sum of the squares of deviations of the data from the fitted curve

$$\chi^2 = \sum \left[\frac{1}{\sigma_i^2} (y_i - a - bf(x; \mu, \sigma))^2 \right] \tag{9.6}$$

and obtain the solution by minimizing χ^2 simultaneously with respect to each of the parameters. The required derivatives with respect to the two parameters a and b, in which the function is linear, are

$$\begin{aligned} \frac{\partial \chi^2}{\partial b} &= -2 \sum \left[\frac{1}{\sigma_i^2} (y_i - a - bf(x; \mu, \sigma)) f(x; \mu, \sigma) \right] = 0 \\ \frac{\partial \chi^2}{\partial a} &= -2 \sum \left[\frac{1}{\sigma_i^2} (y_i - a - bf(x; \mu, \sigma)) \right] = 0 \end{aligned} \tag{9.7}$$

We can write χ^2 in terms of the derivatives of Equation (9.7) as

$$\chi^2 = \sum \left[\frac{y_i}{\sigma_i^2} (y_i - a - bf(x; \mu, \sigma)) \right] + \frac{1}{2} \left(a \frac{\partial \chi^2}{\partial a} + b \frac{\partial \chi^2}{\partial b} \right) \tag{9.8}$$

and setting the derivatives to 0 gives

$$\chi^2 = \sum \left[\frac{y_i}{\sigma_i^2} (y_i - a - bf(x; \mu, \Gamma)) \right] \tag{9.9}$$

If the data represent the number of counts per unit time in a detector, then they are distributed according to the Poisson distribution and we can approximate $\sigma_i^2 \simeq y_i$. Equation (9.9) becomes

$$\begin{aligned} \chi^2_{\min} &\simeq \Sigma [y_i (a + bf(x; \mu, \Gamma))] \\ &= \text{area(data)} - \text{area(fit)} \end{aligned} \tag{9.10}$$

Thus, we observe that the area under the total fit is underestimated by an amount equal to $\chi^2_{\min}$.

For this derivation we require only that the fitting function consist of a sum of terms, each one of which is multiplied by a coefficient

$$y(x) = \sum_{j=1}^{m} a_j f_j(x) \tag{9.11}$$

The function $f_j(x)$ can contain any number of other parameters in nonlinear form, but may not contain any of the coefficients a_j. Even reparameterizing the function of Equation (9.5) [or Equation (9.1)] and minimizing χ^2 with respect to the area explicitly would not affect the discrepancy between the actual and estimated areas.

Note that for data that are distributed with a constant uncertainty $\sigma_i = \sigma$, the second equation of Equations (9.7) is sufficient to ensure that $\Sigma y(x_i) = \Sigma y_i$. It is the assumption of a Poisson distribution for the data $\sigma_i^2 = y_i$ that yields the discrepancy between the actual and estimated areas.

If the agreement between the fit and the data should be exact, $\chi^2 = 0$, then the estimated and actual areas would be equal. For a fitting function that is a good representation of the data, the value of χ^2 will approximately equal the number of degrees of freedom, so that if there are many bins and a few parameters to be determined, the average discrepancy will be about 1 per bin. Thus, the correction may be negligible for distributions with large numbers of events.

We would like to find ways to reduce the discrepancy. The fact that we know the approximate value of the discrepancy in the total histogram is, in itself, not very helpful because we do not know how to allocate the discrepancy between peak and background. We might find the ratio of the integral A_p of the peak [Equation (9.2)] to the integral A of the complete function Equation (9.1) and scale to the total number of events in the plot to estimate the number of events in the peak. This method assumes that the correction is proportional to the area. Another possibility is to make separate fits to the peak and background regions of the plot, so that we can try to assign the estimated correction separately to the two regions of the plot.

One obvious way of reducing the discrepancy between the area of the measured and fitted data is to reduce the value of χ^2 at the minimum so that the correction is small. A method of accomplishing this reduction, which is not universally accepted but which can be justified by practical considerations, is the technique of smoothing the data, averaging in some mathematically acceptable way over adjacent bins. (See Appendix A.5). Under any smoothing process there can be no overall gain in information, and a net improvement of the fit to the area must be offset by an increased uncertainty in the estimation of other parameters, such as the width and position of the peak. But smoothing will decrease the value of χ^2 at the minimum and thereby reduce the bias in the estimation of the area.

Referring to Table 9.1, we observe that the areas under the three fitted curves differ from the area under the data sample (4000 events), although the differences do not agree with the predicted values ($\chi^2_{\min}$), perhaps because of the complexity of the nonlinear fitting process. Linear least squares polynomial fits to appropriate data, such as the background distributions in Example 9.1, yield the expected differences between the area of the data and the fitted curves. See Exercise 9.1.

9.3 COMPOSITE PLOTS

Single Peak and Background

For a fitting function $y(x)$ that is separable into a peak $y_p(x)$ plus a background $y_b(x)$, such as Equation (9.1), it may be convenient to consider at least some facets of the fitting procedure separately. The least-squares procedure for minimizing χ^2 with respect to each of the parameters a_j,

$$\frac{\partial}{\partial a_j}\sum\left\{\frac{1}{\sigma_i^2}[y_i - y_b(x_i) - y_P(x_i)]^2\right\}^{=0} \tag{9.12}$$

can be considered equally well in terms of fitting the sum of the curves $y(x)$ to the total yield y_i or of fitting one function $y_p(x)$ to the difference spectrum $y_i' = y_i - y_b(x_i)$. The only provision is that the uncertainties in the data points of $\sigma_i' = \sigma_i$ must be the same in both calculations.

If the background curve can be assumed to be a slowly varying function under the peak, as in Figure 9.1, and may reasonably be interpolated under the peak from fitting on both sides, it may be preferable to fit the background curve $y_b(x)$ outside the region of the peak and to fit the peak function $y_p(x)$ only in the region of the peak.

Such a procedure might help isolate special problems that result from fitting with an incorrect peak or background. The χ^2 function measures not only the deviations of the parameters from an ideal fit, but also the discrepancy between the form chosen for the shape of the fitting function $y(x)$ and the parent distribution of the data. If the shape of the fitting function does not represent that of the parent distribution exactly, the value of χ^2 may have large contributions from local data regions. By fitting separate regions of a plot, it may be possible to discover whether the disagreement is in the background or the peak region. In the histogram of Figure 9.1, our interest is in the properties of the peak function, and not in the background, which we parameterize with a simple power series in E. However, the value of χ^2 for the fit is calculated for the entire plot and includes contributions from discrepancies between the background and the fitted curve, as well as between the peak and curve. We may be able to isolate problems to one or the other region by separating the fit into two parts.

Another reason for making separate fits to regions of a plot is to search for starting values for an overall fit. For example, when fitting a function that consists of peak functions plus background function, it may be useful first to fit the regions outside the peaks to get starting values for the background parameters and then to fit separately the region close to each peak, to find starting values for the peak parameters.

As an example, assume that we wish to find starting values for the fit of Equation (9.1) to the data of Figure 9.1. The following procedure could be used:

1. Separate the curve into three regions (*a*), (*b*), and (*c*) as indicated by the two vertical lines on Figure 9.1.
2. Fit the background polynomial $y_1(x) = a_1 + a_2E + a_3E^2$ simultaneously to regions below and above the peak to obtain provisional values for the parameters a_1 through a_3.
3. Fit the entire function of Equation (9.1) to the central region, with the fixed values of a_1 through a_3 obtained in step 2 to obtain values for the parameters a_4, a_5, and a_6.
4. Fit the entire function of Equation (9.1) simultaneously to regions (*a*) and (*c*), with the starting values of the parameters a_4 through a_6 set to the values obtained in steps 2 and 3 to obtain new values of the parameters a_1 through a_3.

If the parameters continue to change significantly on each iteration, the process can be repeated from step 2 as required. Alternatively, it may be sufficient to skip step 3 and to fit for all parameters after step 2.

In fitting the peak and background functions over different parts of the spectrum, it is important to note that the complete function $y(E)$ of Equation (9.1) must be fitted to both regions; that is, in the region outside the peak where the background is being fitted, the calculation of the tail of the peak must be included, and underneath the peak, the background terms must be included.

Multiple Peaks

Separation of closely spaced peaks is an important problem in many research fields. Although we should not attempt to extract information from our data by sorting in bins that are smaller than the uncertainties in our measurements, and should not use bin widths that are so narrow that the numbers of events in the bins are too small to satisfy Gaussian statistics, we also should not err in the other direction and risk suppressing important details. Selecting optimum bin sizes is critical. For some data samples, different bin widths for different regions of the data sample may be appropriate.

EXAMPLE 9.2 We have noted that, although the 4.4% probability for the fit to the data of Example 9.1 is rather low, it could be acceptable. However, because the data were plotted in rather coarse bins ($\Delta E = 0.1$ GeV), some information may have been suppressed. To check this possibility, we plotted the data in smaller bins ($\Delta E = 0.05$ GeV) as illustrated in Figure 9.2. (Note that in plotting Figure 9.2 we have eliminated some bins from the lower and upper edges of the histogram in order to enhance the display; all 60 bins are included in the fits.)

Plotted in smaller bins, the large peak near $E = 1.00$ GeV appears to be considerably narrower than indicated in Figure 9.1. There is also a suggestion of a possible excess of events in the bin centered at $E = 0.825$ GeV on the low-energy side of the main peak. As illustrated by the curve on Figure 9.2, a fit of the two-peak Equation (9.13) to the narrow-bin data, seems to confirm the existence of a second peak. To

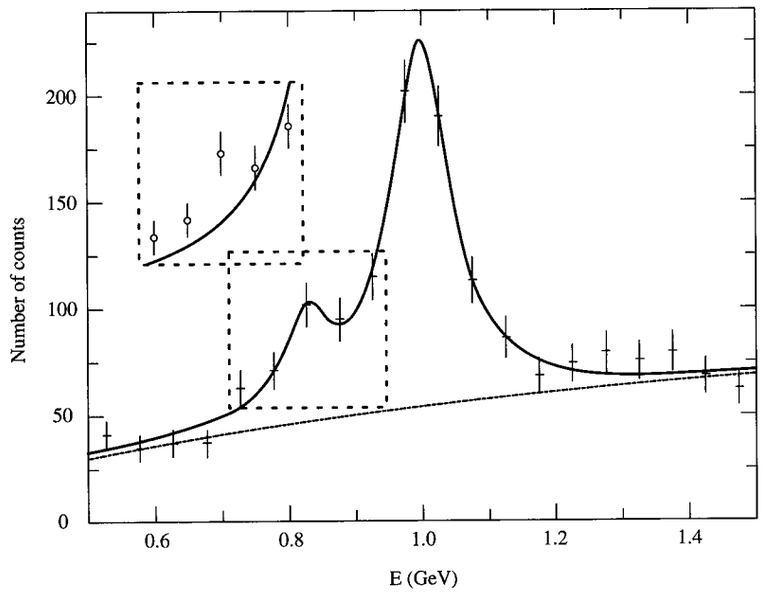

FIGURE 9.2
Histogram data in bins of 0.05 GeV of the 4000 simulated events shown in Figure 9.1. The solid curve illustrates a fit of Equation (9.13) to the data. The inset illustrates, in the region of the smaller peak, a fit of the single-peak Equation (9.1) to the entire data sample.

obtain this fit, we chose as starting values for the mass and width of the second peak, 0.825 and 0.05 GeV, respectively,

$$y(E) = a_1 + a_2E + a_3E^2 + A_0\frac{\Gamma_0/(2\pi)}{(E-E_0)^2 + (\Gamma_0/2)^2} + A_1\frac{\Gamma/(2\pi)}{(E-E_1)^2 + (\Gamma/2)^2} \tag{9.13}$$

suggested by examination of Figure 9.2.

Results of the fit are listed in column 6 of Table 9.1. The 29.4% chi-squared probability for this fit is a marked improvement from 4.4% for the single-peak fit. The inset on Figure 9.2 shows the region of the smaller peak with a curve calculated by fitting the single-peak Equation (9.1) to the entire data sample of Figure 9.2. Parameters determined in this fit are listed in column 5 of Table 9.1.

We can estimate the statistical significance of the smaller peak in Example 9.2 by counting the total number of events above the single-peak background (shown in the inset) and considering whether or not the excess is consistent with a statistical fluctuation. There are 102 events in the peak bin over a background of 69.5 events, corresponding to a fluctuation of $(102 - 69.5)\backslash\sqrt{69.5} = 3.9$ standard deviations in the background signal. Referring to Table C.2, we infer that there is a $(1 - 0.99990) = 0.00010$, or 0.01% probability that we should obtain a result

this large, or larger, from a statistical fluctuation. Thus, the smaller peak appears to be very well established.

But we should wait before rushing into publication; our analysis is not finished. We calculated the probability of finding a 3.9 standard deviation fluctuation in a particular bin. However, there are 60 bins in this data sample, and the fluctuation could have appeared in any of them. The probability that a 3.9 standard deviation would *not* appear in any of the 60 pairs is 0.9999^{60}, so the probability of observing the fluctuation in any of the bin pairs is $1 - 0.9999^{60} \sim 0.6\%$. This probability is low enough to give us considerable confidence that the smaller peak is not a fluctuation. If we had some a priori reason, such as a theoretical prediction or evidence from another experiment, to believe that the smaller peak should be located in the particular energy region where it appears, then the argument against a statistical fluctuation would be even more compelling.

While there appears to be firm statistical support for a second peak in the data of Example 9.2, that support depends strongly on our understanding of the contributions in the region of the second peak from the smooth background distribution and the tail of the large peak. If, for example, background counts were 10% higher, decreasing the excess by 10%, the fluctuation would decrease from 3.9 to 2.9 standard deviations and the probability of a fluctuation of this magnitude in any bin would increase from about 1% to 20%, a considerably less compelling number.

Are there further tests we can make on our data sample to help us understand the significance of our result? For problems such as this, where the statistical significance of a result may be in question, the Monte Carlo method (Chapter 5) provides a powerful tool for more detailed examination. We shall use this technique in Chapter 11 to make a simple statistical test of these data. A full Monte Carlo program, which incorporates all the known or estimated details involved in the creation of the data sample, is invaluable in the planning and analysis of a serious experiment.

SUMMARY

Background subtraction:

$$y_P(x) = y(x) - y_b(x) \qquad (y_P \longrightarrow \text{peak};\ y_b \longrightarrow \text{background})$$

Uncertainty in area of peak:

$$\sigma^2_{A_P} = \sigma^2_A + \sigma^2_{A_b} \qquad (\simeq A + A_B \text{ for Poisson statistics})$$

Area under fitted peak curve:

$$A_P = \int_{-\infty}^{+\infty} y_P(x)\,dx$$

Discrepancy in area under a curve with Poisson statistics:

$$\chi^2_{\min} = \sum\left[\frac{y_i}{\sigma_i^2}(y_i - y(x_i))\right] \simeq \text{area(data)} - \text{area(fit)}$$

EXERCISES

9.1 The following data are drawn from the background distribution illustrated by the dashed curve in Figure 9.1 The data points correspond to the numbers of counts in 15 histogram bins, which are 0.2 GeV wide, each *centered on* the indicated value of E.

E	0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7
N	4	30	49	71	87	91	120	136	147	133	130	118	142	122

Plot a of the data.

Use a linear-fitting technique, such as those described in Chapter 7, to fit a second-order polynomial to these data. Assume statistical uncertainties in the counts. Compare the number of events in the histogram to the number determined by the fit. Is the difference consistent with the prediction of Equation (9.10)?

9.2 Find the area of the peak in Figure 9.1 by counting the area between the vertical dotted lines and subtracting the estimated background. Refer to the data in Exercise 9.4. Estimate the correction for the tails. Estimate the uncertainty in your determination of the area.

9.3 Refer to the data of Exercise 8.6. Fit the histogram by the method outlined in Section 9.3 with separate fits of the background second-order polynomial to the regions outside the peak and of the Gaussian function to the region of the peak.

9.4 The accompanying table lists the numbers of events in the histogram bins of Example 9.1 from $E = 0.0$ to 3.0 GeV in steps of 0.05 GeV.

(*a*) Fit Equation (9.1) to the data to obtain the parameters for this distribution. Compare to the values of the parameters listed in column 5 of Table 9.1.

(*b*) Repeat the fit with adjacent bins merged (i.e., combine bins 1 and 2, bins 3 and 4, etc.) and observe the effect on the value of χ^2, the determination of the area of the peak, and the determination of the mean and half-width of the peak. Assume statistical uncertainties.

7	2	6	12	15	18	31	29	27	27	41	35	37	37	63	71	102	95	115	202
190	113	86	68	74	79	75	79	68	62	69	81	79	85	87	68	70	89	77	70
71	62	85	62	73	70	59	61	77	61	62	73	67	71	75	66	73	71	71	49

CHAPTER 10

DIRECT APPLICATION OF THE MAXIMUM-LIKELIHOOD METHOD

The least-squares method is a powerful tool for extracting parameters from experimental data. However, before a least-squares fit can be made to a data set that consists of individual measurements or events, the events must be sorted into a histogram, which may obscure some detailed structure in the data. Because the least-squares method was derived from the principle of maximum likelihood, it might be better in some instances to use the maximum-likelihood method directly to compare experimental data to theoretical predictions, without the necessity of *binning* data into histograms with the corresponding loss of information.

We have already used the method in Chapter 4 to find estimates for the mean and standard deviation of data obtained in repeated measurements of a single variable, where we have assumed that the measurements were distributed according to Gaussian probability. Now, we extend the method to other distribution functions and to multiparameter fits. Maximum-likelihood methods can be applied directly to many "curve fitting" problems, and such fitting is almost as easy to use as the least-squares method, and considerably more flexible. However, the direct maximum-likelihood method requires computations for each *measured event,* rather than for each *histogram bin* as in least-squares fitting, and therefore the technique may be too slow for very large data samples.

Direct maximum-likelihood calculations have an advantage over the least-squares method for two particular types of problems: (1) low-statistics experiments

with insufficient data to satisfy the requirement of Gaussian statistics for individual histogram bins and (2) experiments in which the fitting function corresponds to a different probability density function for each measured event so that binning the data leads to a reduction in information and a loss of sensitivity in determining the parameters. If the data set is sufficiently large, then the least-squares method can be applied to problems of either type, and that method is generally preferred in view of its smaller computing requirement. At any rate, it is not possible to extract more than minimal information from a very small data set, so we should expect the direct maximum-likelihood method to be most useful for intermediate problems with modest data samples.

10.1 INTRODUCTION TO MAXIMUM LIKELIHOOD

The basic maximum-likelihood procedure is relatively simple. Assume that we have a collection of N events corresponding to the measurement of an independent variable x_i and a dependent variable y_i, where i runs from 1 to N. We wish to obtain the parameters, $a_1, a_2, \ldots, a_m$, of a fitting function $y(x_i) \equiv y(x_i; a_1, a_2, \ldots, a_m)$ from these data. For each event, we convert $y(x_i)$ to a normalized probability density function

$$P_i \equiv P(x_i; a_1, a_2, \ldots, a_m) \tag{10.1}$$

evaluated at the observed value x_i. The likelihood function $\mathscr{L}(a_1, a_2, \ldots, a_m)$ is the product of the individual probability densities

$$\mathscr{L}(a_1, a_2, \ldots, a_m) = \prod_{i=1}^{N} P_i \tag{10.2}$$

and the maximum-likelihood values of the parameters are obtained by maximizing $\mathscr{L}(a_1, a_2, \ldots, a_m)$ with respect to the parameters.

In many experiments, the probability density function P_i will be made up of two components: a theoretical factor corresponding to the underlying principle being tested and an experimental factor corresponding to the biases introduced by experimental conditions.

EXAMPLE 10.1 In Example 5.7 we presented a Monte Carlo program for studying biases that could arise in an experiment to measure the mean life of the short-lived K_s^0 meson (or kaon). The example includes details of the experiment and Figure 5.4 illustrates schematically the experimental apparatus.

In brief, the experiment involves measuring the distance between the point of production and point of decay of the kaon, determining the meson's velocity, and calculating the meson's time of flight from production to decay. After correction for bias introduced by the finite size of the experimental apparatus, the mean life of the kaon could be determined from measurements of many such events.

The dashed rectangle on Figure 5.4 indicates the region in which events are col-

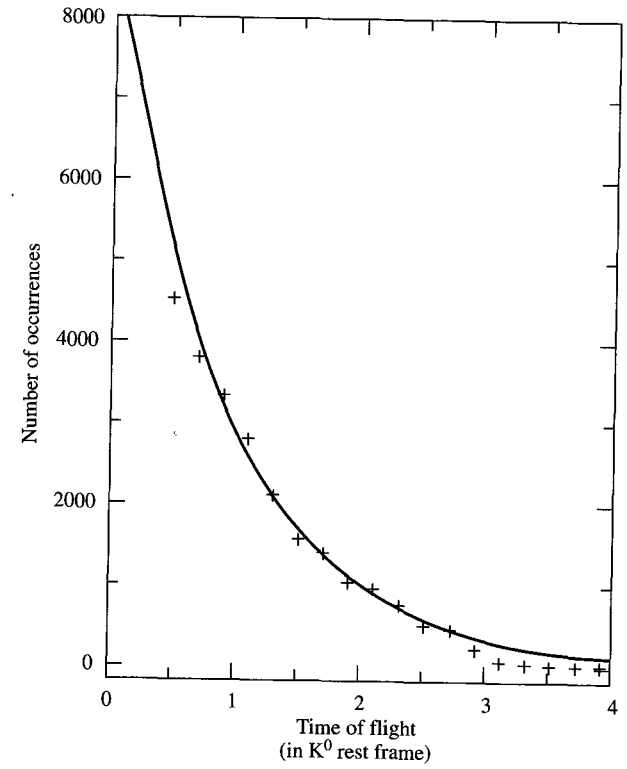

FIGURE 10.1
Frequency distribution of times of flight for 23,565 events that survived fiducial cuts in a 40,000-event Monte Carlo generation, as a function of the proper time (in units of 10^{-10} s). The exponential curve was calculated from the nominal value $\tau_K = 0.894 \times 10x^{-10}$ s to represent the expected distribution of the 40,000 generated events.

fall within the fiducial region bias the final calculation of the mean life and therefore we must understand the biases and make corrections.

In the following examples, we assume that the coordinates of the two vertices and the magnitude of the momentum of the decaying kaon have been determined.

We used the Monte Carlo program of Example 5.7, with the mean life of the kaon set to its nominal value of $\tau_K = 0.894 \times 10^{-10}$s, to generate 40,000 events in order to study the efficiency of the detector with reasonably high precision. It is important that the statistical uncertainties introduced in the determination of the efficiency function be negligible compared to the statistical and other uncertainties in the actual experiment. The distribution of the 23,565 generated events that survived fiducial cuts is shown as crosses in Figure 10.1 with the expected exponential distribution of the total 40,000-event sample shown as a smooth curve.

In Figure 10.2 we have plotted the resulting efficiency as a function of the times of flight of the kaons (the *proper time*) in their individual rest frames, with the effi-

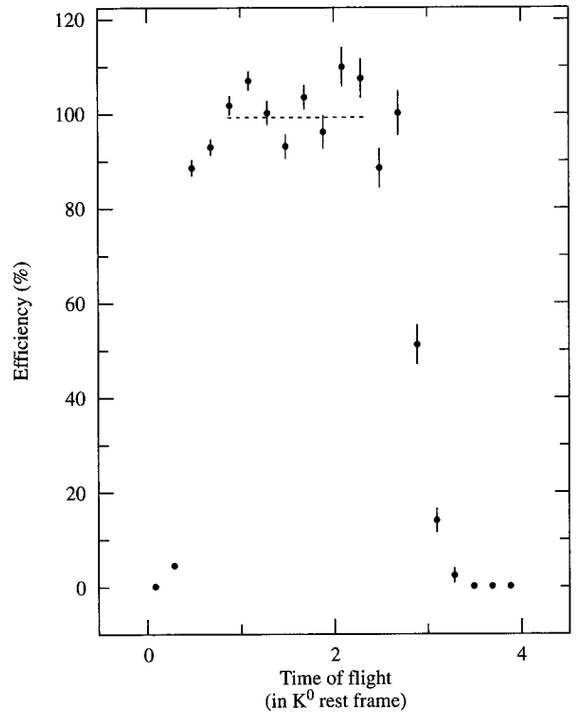

FIGURE 10.2
Efficiency function $\epsilon(T) = N(T)/N(T)$, calculated from the ratio of observed events (crosses) to expected events (smooth curve in Figure 10.1). The dotted line illustrates the region over which the efficiency reasonably may be assumed to be 100%.

We also used the Monte Carlo program, with different random-number seeds and the same nominal value of τ_K, to generate a small "data set" of 1000 events, of which 598 survived the fiducial cut, to use in testing our analysis procedures.

We shall discuss several aspects of the analysis of such data in the following examples.

EXAMPLE 10.1a: Least-squares Method Figure 10.3 shows on a semilogarithmic plot the distribution, as crosses (x), of the 598 events that survived the fiducial cuts from the total sample of 1000 events generated in Example 10.1. The straight line shows the expected distribution if there had been no efficiency losses. In order to extract the mean life of the kaon from these data, we apply the efficiency function illustrated in Figure 10.2 to correct for losses. The corrected data points are plotted in Figure 10.3 as data points with vertical error bars corresponding to the statistical uncertainties in the data, scaled by the efficiency factor. (Uncertainties in the correction factor were negligible.) The efficiency was assumed to be 100% in the region indicated by the horizontal dotted line in Figure 10.2. The very large error bars on "corrected" points at the two ends of the plot result from scaling low-statistics data points and illustrate the problem of using data in regions of low efficiency. Generally, it is

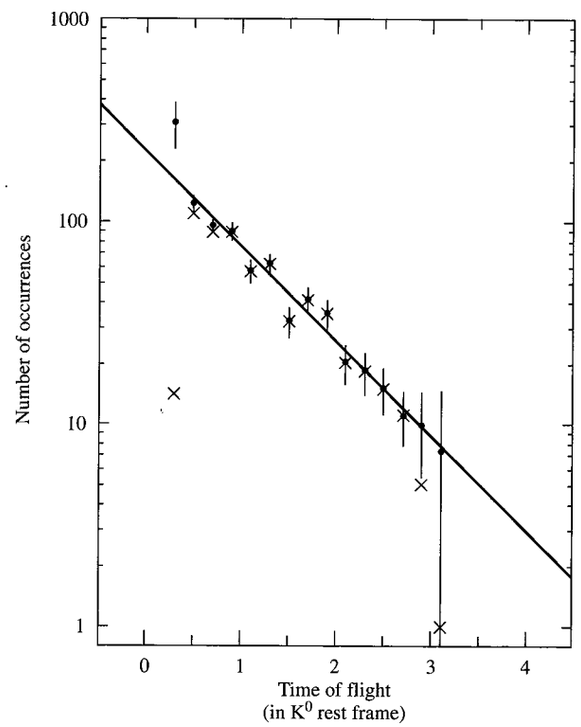

FIGURE 10.3
Semilogarithmic plot of the frequency distribution of 598 events that survived fiducial cuts from a 1000-event (Monte Carlo) data sample. The uncorrected data are shown as crosses; the data corrected for efficiency losses are shown as data points with error bars. The straight line shows the result of a linear least-squares fit to the corrected semilogarithmic data.

wise to eliminate points that require such large corrections from the sample, because they contribute little to the overall result and depend heavily on the corrections.

From the linear slope of the logarithmic plot, illustrated by the straight line through the data points, we obtain an "experimental" mean life $\tau = (0.925 \pm 0.058)$. Alternatively, we could have used a nonlinear least-squares fitting technique to determine τ directly from a linear plot of the data.

Direct Maximum Likelihood

Most actual experiments are more complex and have efficiency functions that are considerably more complicated than the one illustrated by our example. For such problems, application of direct maximum likelihood may be the preferable method for finding the best estimate of the parameters. To apply this method, we must define a probability function for each recorded event.

The probability of observing a single event that survives for a time t_i is

$$P_i = A_i p(t_i; \tau) \tag{10.3}$$

The first factor A_i represents the *detection efficiency,* or probability that the particle will decay within a predefined *fiducial volume* within our apparatus, so that a satisfactory measurement can be made of its flight time. This factor depends upon the coordinates of the production and decay vertices of the decaying particle, its momentum vector, and the geometry of the fiducial volume. The second factor $p(t_i; \tau)$ is proportional to the probability that a particle of mean lifetime τ will decay between time t_i and $t_i + \mathrm{dt}$ and is therefore proportional to $e^{-t_i/\tau}$. Equation (10.3) becomes

$$P_i = A_i e^{-t_i/\tau} \tag{10.4}$$

It might appear that the two factors in Equation (10.3) are independent, so that the detection efficiency factor is independent of the decay probability, but, as we have observed in the previous example, this is not generally true. Because of the finite size of our measuring apparatus, we may preferentially lose events that survive for very short times so that we can't make precise measurements of their flight paths, as well as those that survive for very long times and therefore decay outside the acceptable limits of our detectors. Losses of both types depend upon the mean life that we are attempting to determine, the "τ" in the second factor of Equation (10.3). For each particle that is observed to decay within the apparatus, we can define a *potential path length* as the distance it would travel if it had not decayed. Because each decaying particle has a different potential path length, we must calculate geometric factors to correct for those particles that decay outside the detector. The correction factors will depend on the parameters and will be a function of the production and decay coordinates and the momentum vectors of each decaying particle. Clearly, one element of good experiment design should be to minimize the dependence of these geometric correction factors on the parameters sought in the experiment.

Normalization for Maximum Likelihood

The factor A_i in Equation (10.4) corresponds to a normalization for each measurement to assure unit probability for observing in this experiment *any* event that has the mean life, coordinates, and kinematics of the observed decaying particle. To determine the normalizing factor A_i we refer to Figure 5.4 and consider the fiducial volume of our apparatus, indicated by the dashed rectangle. From each particle's production coordinates and momentum vector, we can determine the minimum distance d_1 that the particle must travel to enter the region and the maximum distance d_2 it can travel before leaving the region. (We can, of course, observe some events outside the fiducial volume, but we reject them because they cannot be measured precisely.) These minimum and maximum distances d_1 and d_2 must be converted to times of flight t_1 and t_2 in the rest frame of the decaying particles, and the normalizing factors A_i can then be determined from the condition

$$\int_{t_1}^{t_2} P_i dt_i = A_i \int_{t_1}^{t_2} e^{-t_i/\tau} dt_i = 1 \tag{10.5}$$

probability or the likelihood function for observing N such events in our experiment is just the product of the individual probability functions:

$$\mathscr{L}(\tau) = \prod_{i=1}^{N} P_i = \prod_{i=1}^{N} A_i e^{-t_i/\tau} \tag{10.6}$$

Parameter Search

Our object is to find the value of the parameter τ that maximizes this likelihood function. Because the probability of observing any particular event is less than 1, the product of a large number of such probabilities (one for each measured event) may be a very small number, and may, in fact, be too small for the computer to handle. To avoid problems, it is usually preferable to maximize the logarithm of the likelihood function

$$M = \ln \mathscr{L} \tag{10.7}$$

rather than the likelihood function itself, so that the product of Equation (10.6) becomes a sum. The logarithms should be reasonable, negative numbers. For our particular example, the logarithm of the likelihood function of Equation (10.6) is given by

$$M(\tau) = \ln[\mathscr{L}(\tau)] = \sum \left[\ln A_i - \frac{t_i}{\tau} \right] \tag{10.8}$$

with A_i defined by Equation (10.5). Note that A_i is a function of the unknown parameter τ, as well as of the production coordinates, momentum vector, and fiducial volume, and must be calculated separately for each event, *and for every trial value of* τ.

In general, this problem, like the corresponding nonlinear least-squares fitting problem, cannot be solved in closed form. However, either the grid- or gradient-search method of minimizing the χ^2 function discussed in Chapter 8 can be adopted directly. It is only necessary to search for a maximum of M (or a minimum value of $-M$) with the same routines we used in Chapter 8 to find a minimum of χ^2.

We may note a correspondence between the quantity $M(\tau)$, determined in Equation (10.7) from the likelihood function for *individual events,* and the goodness-of-fit parameter χ^2, determined by Equation (8.7) from the likelihood function $P(a)$ for *binned data:*

$$\chi^2 = -2\ln[\mathscr{L}(\tau)] + \text{constant} \tag{10.9}$$

In the limit of a large number of events, the two methods must yield the same value τ' for the maximum-likelihood estimate of the parameter τ. In both cases the likelihood function will be a Gaussian function of the parameter near the optimum value

$$\mathscr{L}(\tau) \propto \exp\left(-\frac{(\tau - \tau')^2}{2\sigma^2}\right) \tag{10.10}$$

EXAMPLE 10.1b Let us consider the simplest form of this problem. Assume that the unknown mean lifetime is sufficiently short so that our apparatus is large enough to include many lifetimes and, therefore, the loss of particles that decay at very long times is negligible. Let us also assume that our equipment can detect particles at very short as well as very long times. Then the limits on the normalization integral of Equation (10.5) become $t_1 = 0$ and $t_2 = \infty$ and A_i is the same for every event and is given by $A_i = 1/t$. The likelihood function becomes

$$\mathscr{L}(\tau) = \prod A_i e^{-t_i/\tau} = \prod \frac{e^{-t_i/\tau}}{\tau} \tag{10.11}$$

with logarithm

$$M(\tau) = \ln[\mathscr{L}(\tau)] = -\frac{1}{\tau}\sum t_i - N \ln \tau \tag{10.12}$$

We can obtain the maximum of Equation (10.12) by taking the derivative of $M(\tau)$ with respect to τ and setting it to 0:

$$\begin{aligned} \frac{dM(t)}{d\tau} &= \frac{d}{d\tau}\left\{-\frac{1}{\tau}\sum t_i - N \ln \tau\right\} \\ &= \frac{1}{\tau^2}\sum t_i - \frac{N}{\tau} = 0 \end{aligned} \tag{10.13}$$

The solution is $\tau = \Sigma t_i/N$; that is, the maximum-likelihood estimate of the mean life is just the mean of the individual lifetime measurements. We should have reached the same result if we had found the maximum of $\mathscr{L}(t)$ from Equation (10.11).

EXAMPLE 10.1c Suppose that we repeat the experiment, but with poorer experimental resolution so that we cannot distinguish the decay vertex (x_2, y_2, z_2) from the creation vertex (x_1, y_1, z_1) unless they are separated by a distance d_1. For simplicity, we assume that the decaying particles are all produced with the same velocity, so that the lower cutoff distance d_1 translates into the same lower cutoff in time t_1 for all events. (In an actual experiment, of course, the decaying particles would be produced with various velocities, so that the calculated lower cutoff time t_1 would vary from event to event.)

For this example, the normalization integral of Equation (10.5) becomes

$$A_i \int_{t_1}^{\infty} e^{-t_i/\tau} dt_i = 1 \tag{10.14}$$

which gives

$$A_i = \frac{e^{t_1/\tau}}{\tau} \tag{10.15}$$

The likelihood function becomes

so that

$$M = \ln \mathscr{L} = \sum \frac{[t_1 - t_i]}{\tau} - \sum \ln \tau \tag{10.17}$$

Setting

$$\frac{dM(\tau)}{d\tau} = 0 \tag{10.18}$$

gives

$$\frac{d}{d\tau}\sum\left\{\frac{[t_1 - t_i]}{\tau} - \ln \tau\right\} = -\sum\left\{\frac{t_1 - t_i}{\tau^2}\right\} - \frac{N}{\tau} = 0 \tag{10.19}$$

or

$$\tau = \frac{\Sigma[t_i - t_1]}{N} = \frac{\Sigma t_i}{N} - t_1 \tag{10.20}$$

As we should expect, the lifetime τ would have been overestimated if we had neglected to take account of the cutoff at short times.

EXAMPLE 10.1d Let us consider a more realistic problem in which we have both short and long cutoffs on the observable path. We also assume that the unstable particles are produced at various locations within the target and with various momentum vectors **p**.

For this example, we must calculate the normalization integral, Equation (10.5), separately for each event with individual values for t_1 and t_2 determined from the minimum and maximum distance cutoffs, d_1 and d_2, respectively. The resulting expression for the likelihood function is

$$\mathscr{L}(\tau) = \prod_{i=1}^{N} A_i e^{-t_i/\tau} = \prod_{i=1}^{N}\left[\frac{e^{-t_i/\tau}}{\tau[e^{-t_1/\tau} - e^{-t_2/\tau}]}\right] \tag{10.21}$$

with

$$M(\tau) = \ln[\mathscr{L}(\tau)]$$

Setting to zero the derivative of $M(\tau)$ with respect to τ gives us the equation for the maximum-likelihood value of τ. However, the resulting equation cannot be solved analytically for τ although it could be solved by interpolation (see Appendix A). We choose, rather, to maximize $M(\tau)$ by a one-dimensional grid-search method because search methods are more generally applicable to maximum-likelihood problems and can readily be extended to multiple parameter problems.

10.2 COMPUTER EXAMPLE

Sample Maximum Likelihood Fit

Example 10.1a. The events were generated with $\tau_K = 0.894 \times 10^{-10}$ s and the distribution of the selected events is illustrated by the crosses in Figure 10.3.

Program 10.1 MAXLIKE (Appendix E) A grid-search method to maximize the logarithm of the likelihood function of Equation (10.21). The routines have been written specifically for Example 10.1d.
STARTUP sets the range of the parameter TAU for the search.
FETCHDATA assigns the input data file, reads the limits of the fiducial region (d_1 and d_2), reads data for individual events.
SEARCH sets and increments TAU and calls LOGLIKE, which returns the logarithm of the likelihood function M. Compares each calculated value of M to the preceding value. Terminates the search when M stops increasing and starts to decrease, indicating that M has passed through a local maximum. At termination, fits a parabola to the last three points to find a better estimate of TAU at the maximum.
LOGLIKE calls LOGPROB to find the logarithm of the probability density for each event; sums to calculate the logarithm of the likelihood function.
LOGPROB calculates the logarithm of the probability density for an event.
ERROR calculates the uncertainty SIGTAU in TAUATMIN, the maximum likelihood value of the parameter TAU, by finding the change in TAU needed to decrease M by $\Delta M = 1/2$.
PLOTLIKECURVE (Not listed) calculates and plots the shape of the likelihood function in the region of the maximum. Plots a Gaussian curve with mean and standard deviation equal to TAUMIN and DTAU.

Grid-Search Solution

At each step the program increments τ by a preset amount $\Delta\tau$ and repeats the calculation until $M(\tau)$ has passed through a maximum and has started to decrease. The program fits a parabola to the three points that bracket the maximum to find the value τ' at the maximum of $M(\tau)$. For a more detailed problem, the program could be written to repeat the calculation with smaller values of $\Delta\tau$ to find a better estimate of τ', as in the fitting examples in Chapter 8. Either the grid- or gradient-search method of Chapter 8 could be adapted to solve multiparameter problems.

Results of the Fit

We analyzed the data set twice: first with data selected in the nominal fiducial region (10 cm to 40 cm), which gave $\tau' = (0.943 \pm 0.059) \times 10^{-10}$s for the 598 events that survived the cut, and then, to test the sensitivity of the calculation to our choice of fiducial region, with data selected in the less-appropriate fiducial region with $d_1 = 10$ cm and $d_2 = 20$ cm, which gave $\tau' = (0.78 \pm 0.14) \times 10^{-10}$s for the 373 events that survived this cut. Plots of the relative values of the likelihood function versus trial values of the parameter τ are shown as crosses in Figure 10.4a for the data selected in the nominal fiducial region and in Figure 10.4b for data selected in the less-

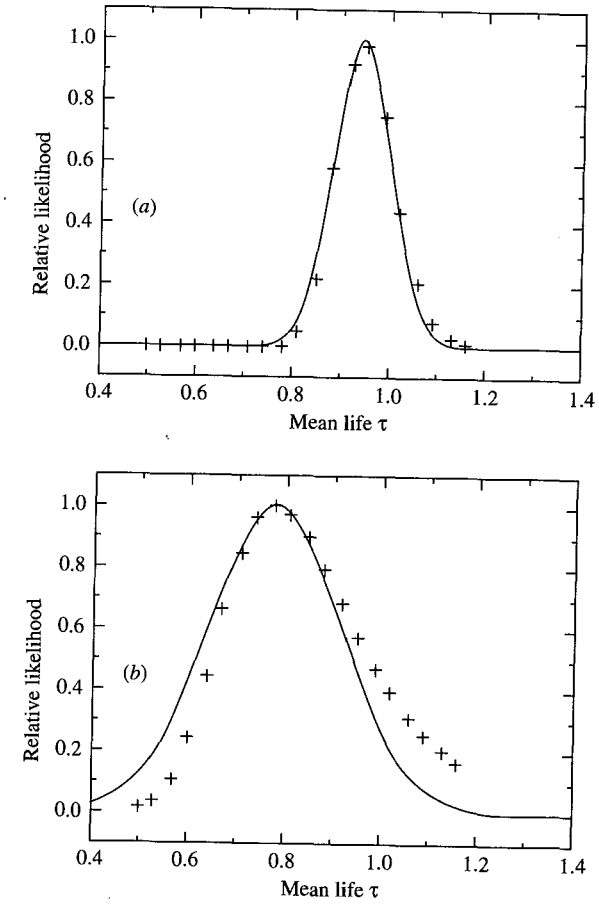

FIGURE 10.4
Relative values of the likelihood function versus trial values of the parameter for events that passed the fiducial cuts for the decay vertex. The data points are indicated by crosses; the smooth Gaussian curves were calculated from Equation (10.10) with the values of the means and standard deviations obtained in the two fits. (*a*) Nominal fiducial cuts: 10 − 40 cm; 598 events survived; $\tau' = 0.943 \times 10^{-10}$ s, $\sigma = 0.059 \times 10^{-10}$ s. (*b*) Incorrect fiducial cuts: 10 − 20 cm; 373 events survived; $\tau' = 0.78 \times 10^{-10}$ s, $\sigma = 0.14 \times 10^{-10}$ s.

fewer events and, therefore, gives a less-precise result. In an actual experiment, we should have to consider a trade-off between the number of surviving events in the sample, and the precision with which those surviving events could be measured, and choose our fiducial region to maximize the overall quality of the result.

We observed that, for a sufficiently large event sample, the likelihood function should become Gaussian in the parameters in the vicinity of a χ^2 minimum (a

maximum of the likelihood function) according to Equation (10.10), where τ' is the value of the parameter τ that maximizes the likelihood function. We show on Figures 10.4a and 10.4b Gaussian curves calculated from Equation (10.10), with τ' and σ determined by the respective fits. Both the data points and the Gaussian curves have been scaled to unit height at $\tau = \tau'$. The data points of Figure 10.4a closely follow the curve; in the lower statistics example in Figure 10.4b, the data points depart from the curve considerably.

Uncertainties

To estimate the uncertainty σ in our determination of τ', we found the change in τ necessary to decrease M by $\Delta M = 1/2$ from its value at the maximum τ' (corresponding to an increase of χ^2 by 1 or a change of $e^{-1/2}$ in the likelihood function $\mathscr{L}$). Because the likelihood function for the larger sample (Figure 10.4a) closely followed the Gaussian form, our estimate of the uncertainty should be satisfactory. However, the smaller sample (Figure 10.4b) was skewed from the Gaussian, so that our estimate of the standard deviation might be somewhat low. For multiparameter fits it is often useful to plot contours of χ^2 (or of M) as a function of pairs of the parameters to study the uncertainties. (See Chapter 11.)

There are several other ways to estimate the uncertainty in a parameter after performing a maximum-likelihood fit. If the distribution of the likelihood function is sufficiently close to a Gaussian, we can find σ_τ from Equation (8.11):

$$\sigma_\tau^2 = \left(\frac{\partial^2 M(\tau)}{\partial \tau^2} \right)^{-1} \tag{10.22}$$

If it is not possible to calculate Equation (10.22) exactly (although it is possible for our example), we can find the second derivative by taking finite differences as discussed in Appendix A.

If the likelihood function does not follow the Gaussian distribution, we can try a numerical integration of the likelihood function to find limiting values that include ~68.3% of the total area, corresponding to the 1 standard deviation limit. Alternatively, we may use a method suggested by Orear (1958) who points out that, for small event samples, where the likelihood function may not be very Gaussianlike, it may be preferable to calculate an average value of the second derivative through the equation

$$\overline{\frac{\partial^2 M}{\partial a^2}} = \frac{\int [\partial^2 M / \partial a^2] \, \mathscr{L}(a) \, da}{\int \mathscr{L}(a) \, da} \tag{10.23}$$

where a is the unknown parameter and the integrals are over the allowable range of the parameter. This procedure has the advantage over the method of Equation (10.22) of giving more weight to the tails of the distribution in cases where they drop off more slowly than those of a Gaussian curve.

Another method of determining the uncertainties in the parameters is to use a Monte Carlo calculation to produce simulated data sets, comparable to our measured

data, and to use the method outlined in Chapter 11 for determining confidence levels for our results. This method has the advantage that it depends only on the assumptions made in the Monte Carlo generation, and not on any statistical expectations about the shape of the likelihood function. In many experiments, especially those with low statistics, it provides the most reliable estimate of parameter uncertainties.

Goodness of Fit

One disadvantage of the direct maximum-likelihood method is that it does not provide a convenient test of the quality of the fit. The value at the peak of the likelihood function itself is not useful because it represents only the maximized probability for obtaining our particular experimental result and we have no way of predicting the expected probability.

An estimate of the goodness of fit can be obtained by making a histogram of the data and comparing it to a prediction based on our best estimate of the parameters. A Monte Carlo simulation of the experiment may be required to calculate the predicted distribution, with a χ^2 test to compare the data to the prediction.

It is not always clear just which data variable should be histogrammed for this purpose. We would like to find that variable on which the parameters depend most strongly. For our sample problem, the lifetime τ in the rest frames of the particles is an obvious choice, because that is the variable we would choose if we were to solve the problem by the least-squares method. However, it might be wise to try plots of several variables to be sure that the fit is satisfactory. To test, we could generate with our Monte Carlo program a large sample of events based on the parameters discovered in each search, apply the fiducial cuts, and calculate χ^2 from the agreement between the Monte Carlo results and our data sample. We should be aware that, because we did not actually minimize χ^2 for the experimental distribution with respect to the parameters, a satisfactory value of χ^2 may be at best an indication that nothing is drastically wrong with the solution.

SUMMARY

Normalized probability density function:

$$P_i \equiv P(x_i, a_1, a_2, \ldots, a_m)$$

Likelihood function:

$$\mathscr{L}(a_1, a_2, \ldots, a_m) = \prod_{i=1}^{N} P_i$$

Single-event probability density: $P_i = A_i \cdot p(x_i; a)$ where A_i is the detection efficiency and $p(x_i; a)$ is proportional to the interaction probability
Logarithm of likelihood function: $M = \ln \mathscr{L} = \Sigma \ln P_i$
Maximization of $\mathscr{L}$ *or of* M: $\partial \mathscr{L} / \partial a_j = 0$ or $\partial M / \partial a_j = 0$ for all a_j

Gaussian form of likelihood function for large data sample:

$$\mathscr{L}(a_j) \propto \exp\left(-\frac{(a_j - a_j')^2}{2\sigma^2}\right)$$

Uncertainties in parameters:

$$\sigma_j^2 = \left(\frac{\partial^2 M(a_j)}{\partial a_j^2}\right)^{-1}$$

Method for low statistics:

$$\overline{\frac{\partial^2 M}{\partial a^2}} = \frac{\int [\partial^2 M/\partial a^2]\, \mathscr{L}(a)\, da}{\int \mathscr{L}(a)\, da}$$

EXERCISES

10.1. In a scattering experiment, the angles of the scattered particles are measured and the cosines of the angles in the center-of-mass rest frame of the incident and target particles are calculated and recorded. Fifty such measurements, drawn from the distribution $y(x) = a_1 + a_2 \cos^2\theta$, are listed in the table. Use the direct maximum-likelihood method to determine the values of the parameters a_1 and a_2. Note that it is necessary to convert the distribution function $y(x_i)$ to a normalized probability function and that the normalization constant will be different for each pair of trial values of a_1 and a_2.

−0.999	−0.983	−0.956	−0.946	−0.933	−0.925	−0.916	−0.910
−0.881	−0.739	−0.734	−0.717	−0.715	−0.675	−0.665	−0.649
−0.621	−0.537	−0.522	−0.508	−0.499	−0.471	−0.460	−0.419
−0.403	−0.311	−0.305	−0.281	−0.170	−0.162	−0.063	0.214
0.438	0.444	0.508	0.586	0.638	0.677	0.721	0.730
0.768	0.785	0.790	0.793	0.877	0.896	0.931	0.938
0.948	0.993						

Because of the small amount of data, the uncertainties in the parameters a_1 and a_2 are so large that the values of the parameters are not very meaningful. Therefore, to complete the problem, you should use the Monte Carlo program written for Exercise 5.8 to generate 500 events and use your calculation to find the parameters from those data.

10.2. Students in an undergraduate physics laboratory determined the mass of the Λ hyperon by measuring graphically the energies and the momentum vectors of the proton and π meson into which the Λ hyperons decayed. Because of the large uncertainties in the measurements, the calculated square of the masses of the decaying particles forms a truncated Gaussian distribution that is limited on the low-mass side by $(M_p + M_\pi)^2 = 1.1617\ (\text{GeV}/c^2)^2$, but is not limited on the high-mass side. The following 50 numbers represent squares of the calculated masses in units of $(\text{GeV}/c^2)^2$.

1.2981	1.2618	1.2145	1.2539	1.4230	1.3963	1.3701	1.2303	1.3655	1.2042
1.3190	1.2086	1.2118	1.2078	1.2726	1.2438	1.1838	1.1666	1.1908	1.1922
1.2525	1.3615	1.1855	1.2697	1.2044	1.3397	1.4317	1.2713	1.2203	1.2817
1.2046	1.2856	1.1980	1.2595	1.1721	1.2608	1.1689	1.4838	1.1743	1.2954

Use the direct maximum-likelihood method to fit a truncated Gaussian to these data to determine the maximum-likelihood value of the mass of the squared particle. A search in two-parameter space will be required since neither the mean nor the width of the distribution is known.

Note that it is necessary to calculate numerically the normalization of the truncated Gaussian for each pair of trial values of the mean and standard deviation of the Gaussian function. It is advisable to set up a table of the integral of the standard Gaussian and to use interpolation to find the desired normalizations. A simple automatic or manual grid search will suffice for maximizing the likelihood function.

10.3. Use Program 5.4 (available on the website) to generate 1000 sample kaon decay events with nominal mean life $\tau = 0.894 \times 10^{-10}$ s.

(*a*) Plot a histogram of the times of flight of all the generated kaons in their own rest frames (proper times).

(*b*) Use Program 10.1 (available on the website), with nominal fiducial cuts on your data ($d_1 = 10.0$ cm and $d_2 = 40$ cm) to repeat the analysis of Example 10.1d to find the maximum likelihood solution τ' for the kaon mean life. Plot a histogram of the events that survive the cuts.

(*c*) With the value of τ', which you determined in part (*b*), and random number seeds that are different from those used in part (*a*), generate 20,000 events to serve as your estimate of the parent distribution. Apply the nominal fiducial cuts to these data and plot a histogram of the data in the same bins as you used in part (*b*).

(*d*) Calculate χ^2 for the agreement between your "experimental" histogram and the surviving events from the "parent" distribution. If the numbers of events in your bins of the parent distribution are large enough, their uncertainties can be ignored in this calculation. If they are not, you must use the combined statistical errors of the two distributions when calculating χ^2.

CHAPTER 11

TESTING THE FIT

11.1 χ^2 TEST FOR GOODNESS OF FIT

The method of least squares is based on the hypothesis that the optimum description of a set of data is one that minimizes the weighted sum of the squares of the deviation of the data y_i from the fitting function $y(x_i)$. The sum is characterized by the variance of the fit s^2, which is an *estimate* of the variance of the data σ^2. For a function $y(x_i)$, which is linear in m parameters and is fitted to N data points, we have

$$s^2 = \frac{1}{N-m}\frac{\Sigma\{(1/\sigma_i^2)[y_i - y(x_i)]^2\}}{(1/N)\Sigma(1/\sigma_i^2)} = \frac{1}{N-m}\sum w_i[y_i - y(x_i)]^2 \qquad (11.1)$$

where the factor $\nu = N - m$ is the number of degrees of freedom for fitting N data points (implied in the unlabeled sums) with m parameters and the weighting factor for each measurement is given by

$$w_i = \frac{1/\sigma_i^2}{(1/N)\Sigma(1/\sigma_i^2)}, \qquad (11.2)$$

the inverse of the variance σ_i^2 that describes the uncertainties in each point, normalized to the average of all the weighting factors.

The variance of the fit s^2 is also characterized by the statistic χ^2 defined in Equation (7.5) for polynomials:

with

$$y(x_i) = \sum_{k=1}^{m} a_k f_k(x_i)$$

The relationship between s^2 and χ^2 can be seen most easily by comparing s^2 with the reduced chi-square χ_ν^2,

$$\chi_\nu^2 = \frac{\chi^2}{\nu} = \frac{s^2}{\langle\sigma_i^2\rangle} \qquad (11.4)$$

where $\langle\sigma_i^2\rangle$ is the weighted average of the individual variances

$$\langle\sigma_i^2\rangle = \frac{(1/N)\Sigma((1/\sigma_i^2)\sigma_i^2)}{(1/N)\Sigma(1/\sigma_i^2)} = \left[\frac{1}{N}\Sigma\frac{1}{\sigma_i^2}\right]^{-1} \qquad (11.5)$$

and is equivalent to σ^2 if the uncertainties are all equal, $\sigma_i = \sigma$.

The parent variance of the data σ^2 is a characteristic of the dispersion of the data about the parent distribution and is not descriptive of the fit. The estimated variance of the fit s^2, however, is characteristic of both the spread of the data and the accuracy of the fit. The definition of χ^2, as the ratio of the estimated variance s^2 to the parent variance σ^2 times the number of degrees of freedom ν, makes it a convenient measure of the goodness of fit.

If the fitting function is a good approximation to the parent function, then the estimated variance s^2 should agree well with the parent variance σ^2, and the value of the reduced chi-square should be approximately unity, $\chi_\nu^2 = 1$. If the fitting function is not appropriate for describing the data, the deviations will be larger and the estimated variance will be too large, yielding a value of χ_ν^2 greater than 1. A value of χ_ν^2 less than 1 does not necessarily indicate a better fit, however; it is simply a consequence of the fact that there exists an uncertainty in the determination of s^2, and the observed values of χ_ν^2 will fluctuate from experiment to experiment. A value of χ_ν^2 that is very small may indicate an error in the assignment of the uncertainties in the measured variables.

Distribution of χ^2

The probability distribution function for χ^2 with ν degrees of freedom is given by

$$p_\chi(x^2; \nu) = \frac{(x^2)^{1/2(\nu-2)}e^{-x^2/2}}{2^{\nu/2}\Gamma(\nu/2)} \qquad (11.6)$$

The chi-square distribution of Equation (11.6) is derived in many texts on statistics[1] but we shall simply quote the results here.

The gamma function $\Gamma(n)$ is equivalent to the factorial function $n!$ extended to nonintegral arguments. It is defined for integral and half-integral arguments by the values at arguments of 1 and ½ and a recursion relation:

$$\left.\begin{aligned} &\Gamma(1) = 1 \qquad \Gamma(1/2) = \sqrt{\pi} \qquad \Gamma(n-1) = n\Gamma(n) \\ &\text{For integral values of } n \\ &\qquad \Gamma(n+1) = n! \qquad n = 0, 1, \ldots \\ &\text{For half-integral values of } n \\ &\qquad \Gamma(n+1) = n(n-1)(n-2)\cdots(3/2)(1/2\sqrt{\pi}) \\ &\qquad\qquad n = 1/2, 3/2, 5/2, \ldots \end{aligned}\right\} \quad ((11.7)$$

Calculating factorial functions can lead to computer overflow problems. For computational purposes it is convenient to replace the factorial form of the gamma function by a form of Stirling's approximation[2]:

$$\Gamma[n] = \sqrt{2\pi}e^{-n}n^{(n-1/2)}(1 + 0.0833/n) \tag{11.8}$$

This approximation, which is accurate to ~0.1% for all $n \geq 1/2$, avoids both the problems of overflow in calculating factorials and the necessity of testing and choosing the appropriate form for integral or half-integral argument. The trade-off is computer speed. Calculating exponentials may be slower than calculating factorials, but high speed usually is not required for nonrepetitive calculations.

If the function of the parent population is denoted by $y_0(x)$, the value of χ_0^2 determined from the parameters of the parent function

$$\chi_0^2 = \sum\left\{\frac{1}{\sigma_i^2}[y_i - y_0(x_i)]^2\right\} \tag{11.9}$$

is distributed according to Equation (11.6) with $\nu = N$ degrees of freedom. If the function $y(x)$ used in the determination of χ^2 contains m parameters, the value of χ^2 calculated from Equation (11.3) is distributed according to Equation (11.6) with $\nu = N - m$ degrees of freedom.

More useful for our purposes than the probability density distribution $p_\chi(x^2; \nu)$ of Equation (11.6) is the integral probability $P_\chi(\chi^2; \nu)$ between $x^2 = \chi^2$ and $\chi^2 = \infty$:

$$P_\chi(\chi^2; \nu) = \int_{\chi^2}^{\infty} P_\chi(x^2; \nu)\, dx^2 \tag{11.10}$$

Equation (11.10) describes the probability that a random set of n data points drawn from the parent distribution would yield a value of χ^2 equal to or greater than the tabulated value.

Program 11.1. CHI2PROB (Appendix E) χ^2-probability.
CHIPROBDENS computation of the function $p_\chi(\chi^2; \nu)$ [Equation (11.6)] using function GAMMA to approximate the gamma function.
CHIPROB Numerical calculation of the integral, Equation (11.10), by Simpson's rule. If variable overflow is a problem, double-precision variables could be employed.

The calculation returns the integral to an accuracy of about ±0.1%. The trade-off on accuracy versus speed of computation is controlled by the value of the constant DX, the integration step.

For the special case of 1 degree of freedom, $\nu = 1$, the χ^2-probability density function of Equation (11.6) takes the form

$$p_\chi(x^2; \nu) = e^{-x^2/2}/(2\pi x^2)^{1/2}$$

which is difficult to integrate numerically near $x = 0$. However, the integral is finite, and the function can be expanded in a Taylor series about $x = 0$ and integrated analytically. We use that technique for $\nu = 1$ and $\chi^2 < 2$.

Similarly, for $\nu = 2$, where the function takes the form

$$p_\chi(x^2; \nu) = e^{-x^2/2}/2$$

the analytic form of the integral is used.

For a fitting function that is a good approximation to the parent function, the experimental value of χ_ν^2 should be close to one and the probability from Equation (11.10) should be approximately 0.5. For poorer fits, the values of χ_ν^2 will be larger and the associated probability will be smaller. There is an ambiguity in interpreting the probability because χ_ν^2 is a function of the quality of the data as well as the choice of parent function, so that even correct fitting functions occasionally yield large values of χ_ν^2. However, the probability of Equation (11.10) is generally either reasonably close to 0.5, indicating a reasonable fit, or unreasonably small, indicating a bad fit. In fact, for most purposes, the reduced chi-square χ_ν^2 is an adequate measure of the probability directly. The probability will be reasonably close to 0.5 so long as χ_ν^2 is reasonably close to 1; that is, less than about 1.5.

Example 11.1. Consider the solution of the problem of fitting two exponential curves plus a linear background to the data from the radioactive silver decay of Example 8.1. The fit (see Table 8.5) gave $\chi^2 = 66.1$ for 54 degrees of freedom, or $\chi_\nu^2 = 1.22$, with $P_\chi(\chi^2; \nu) = 12.4\%$. We can interpret this result in the following way. Assume that the parameters we found are, indeed, the parameters of the parent distribution. Then, suppose that we were to repeat our experiment many times, drawing many different data samples from that parent distribution. Our result indicates that in 12.4% of those experiments we should expect to obtain fits that are no better than that listed in Table 8.5.

11.2 LINEAR-CORRELATION COEFFICIENT

Let us assume that we have made measurements of pairs of quantities x_i and y_i. We know from the previous chapters how to fit a function to these data by the least-squares method, but we should stop and ask whether the fitting procedure is justified and whether, indeed, there *exists* a physical relationship between the variables x and y. What we are asking here is whether or not the variations in the observed

For example, if, as in Example 6.1, we were to measure the potential difference across segments of a current-carrying wire as a function of the segment length, we should find a definite and reproducible correlation between the two quantities. But if we were to measure the potential of the wire as a function of time, even though there might be fluctuations in the observations, we should not find any significant reproducible long-term relationship between the pairs of measurements.

On the basis of our discussion in Chapter 6, we can develop a quantitative measure of the degree of correlation or the probability that a linear relationship exists between two observed quantities. We can construct a linear-correlation coefficient r that will indicate quantitatively whether or not we are justified in determining even the simplest linear correspondence between the two quantities.

Reciprocity in Fitting x Versus y

Our data consist of pairs of measurements (x_i, y_i). If we consider the quantity y to be the dependent variable, then we want to know if the data correspond to a straight line of the form

$$y = a + bx \tag{11.11}$$

We have already developed the analytical solution for the coefficient b, which represents the slope of the fitted line given in Equation (6.12):

$$b = \frac{N\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{N\Sigma x_i^2 - (\Sigma x_i)^2} \tag{11.12}$$

where the weighting factors in σ_i have been omitted for clarity. If there is no correlation between the quantities x and y, then there will be no tendency for the values of y to increase or decrease with increasing x, and, therefore, the least-squares fit must yield a horizontal straight line with a slope $b = 0$. But the value of b by itself cannot be a good measure of the degree of correlation because a relationship might exist that included a very small slope.

Because we are discussing the interrelationship between the variables x and y, we can equally well consider x as a function of y and ask if the data correspond to a straight-line form

$$x = a' + b'y \tag{11.13}$$

The values of the coefficients a' and b' will be different from the values of the coefficients a and b in Equation (11.11), but they are related if the variables x and y are correlated.

The analytical solution for the inverse slope b' is similar to that for b in Equation (11.12):

$$b' = \frac{N\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{N\Sigma y_i^2 - (\Sigma y_i)^2} \tag{11.14}$$

If there is a complete correlation between x and y, then there exists a relationship between the coefficients a and b of Equation (11.11) and between a' and b' of Equation (11.13). To see what this relationship is, we rewrite Equation (11.13):

$$y = -\frac{a'}{b'} + \frac{1}{b'}x = a + bx \tag{11.15}$$

and equate coefficients

$$a = -\frac{a'}{b'} \qquad b = \frac{1}{b'} \tag{11.16}$$

We see from Equation (11.16) that $bb' = 1$ for complete correlation. If there is no correlation, both b and b' are 0 and Equations (11.16) do not apply. We. therefore define, as a measure of the degree of linear correlation, the experimental linear-correlation coefficient $r \equiv \sqrt{bb'}$:

$$r \equiv \frac{N\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{[N\Sigma x_i^2 - (\Sigma x_i)^2]^{1/2}[N\Sigma y_i^2 - (\Sigma y_i)^2]^{1/2}} \tag{11.17}$$

The value of r ranges from 0, when there is no correlation, to ± 1, when there is complete correlation. The sign of r is the same as that of b (and b'), but only the absolute magnitude is important.

The correlation coefficient r cannot be used directly to indicate the degree of correlation. A probability distribution for r can be derived from the two-dimensional Gaussian distribution, but its evaluation requires a knowledge of the correlation coefficient ρ of the parent population. A more common test of r is to compare its value with the probability distribution for the parent population that is completely uncorrelated; that is, for which $\rho = 0$. Such a comparison will indicate whether or not it is probable that the data points could represent a sample derived from an uncorrelated parent population. If this probability is small, then it is more probable that the data points represent a sample from a parent population where the variables are correlated.

For a parent population with $\rho = 0$, the probability that any random sample of uncorrelated experimental data points would yield an experimental linear-correlation coefficient equal to r is given by[3]

$$p_r(r; \nu) = \frac{1}{\sqrt{\pi}} \frac{\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)} (1 - r^2)^{(\nu-2)/2} \tag{11.18}$$

where $\nu = N - 2$ is the number of degrees of freedom for an experimental sample of N data points. The gamma function for integral and half-integral values was defined in Equation (11.7).

Integral Probability

A more useful distribution than that of Equation (11.18) is the probability $P_c(r; N)$ that a random sample of N uncorrelated experimental data points would yield an

experimental linear-correlation coefficient as large as or larger than the observed value of $|r|$. This probability is the integral of $p_r(r; \nu)$ for $\nu = N - 2$:

$$P_c(r; N) = 2\int_{|r|}^{1} p_x(r; \nu)\, dx \qquad \nu = N - 2 \tag{11.19}$$

With this definition, $P_c(r; N)$ indicates the probability that the observed data could have come from an uncorrelated ($\rho = 0$) parent population. A small value of $P_c(r; N)$ implies that the observed variables are probably correlated.

Because Equation (11.19) cannot be integrated analytically, the function must be integrated either by making a series expansion of the argument and integrating term by term or by performing a numerical integration. With fast computers, the latter method is more convenient and generally applicable to such problems.

Program 11.2 LCORLATE (Appendix E) Correlation probability computations. LCORPROB computes the probability of Equation (11.19) by numerical integration. Input variables RCORR and NOBSERV correspond to the value of the experimental linear-correlation coefficient and the number of observations, respectively. (The number of degrees of freedom is the number of observations minus 2.) The program uses the following routines: LINCORREL computes the function $p_r(r; \nu)$ of Equation (11.18) using the approximation of Equation (11.8) for the gamma function (calculated by the function GAMMA in the program unit GENUTIL). Because LINCORREL is intended to be used as an argument to the integration routine SIMPSON, it can have only one argument. The parameter ν is passed in the global variable PSIMPS by the calling routine.
LINCORPROB computes $P_c(r; \nu)$ of Equation (11.19) by numerically integrating LINCORREL by Simpson's rule. The calculation returns the integral to an accuracy of about ± 0.01. The trade-off on accuracy versus speed of computation is controlled by the value of the constant DX, the integration step.

Example 11.2. For the data of Example 6.1, the linear-correlation coefficient r can be calculated from Equation (11.17) with the data of Table 6.1:

$$r = \frac{9 \times 779.3 - 450.0 \times 12.44}{\sqrt{(9 \times 28{,}500 - 450.0^2) \times (9 \times 21.32 - 12.44^2)}} = 0.9998$$

The probability for determining, from an uncorrelated population with $9 - 2 = 7$ degrees of freedom, a value of r equal to or larger than the observed value, can be calculated from Equation (11.19) (see Table C.3). The result $P_c(r; N) < 0.001\%$ indicates that it is extremely improbable that the variables x and V are linearly uncorrelated. Thus, the probability is high that the variables are correlated and the linear fit is justified.

Similarly, in the experiment of Example 6.2, the linear-correlation coefficient can be calculated from Equation (11.17) by including the weighting factors $\sigma_i^2 = y_i$ as in Table 6.2, so that, for example, N is replaced by Σw_i and Σx_i is replaced by $\Sigma w_i x_i$, and so forth:

$$0.03570 \times 81.02 - 0.1868 \times 10$$

Again, the probability $P_c(r; N)$ for $r = +0.9938$ with $\nu = 10 - 2 = 8$ degrees of freedom is very small ($< 0.001\%$), indicating that the change in counting rate C is linearly correlated to a high degree of probability with $x = 1/r^2$, the inverse square of the distance between the source and counter.

11.3 MULTIVARIABLE CORRELATIONS

If the dependent variable y_i is a function of more than one variable,

$$y_i = a + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \cdots \tag{11.20}$$

we might investigate the correlation between y_i and each of the independent variables x_{ij} or we might also enquire into the possibility of correlation between different variables x_{ij}. Here, we use the first subscript i to represent the observation, as in the previous discussions, and the second subscript j to represent the particular variable under investigation. The variables x_{ij} could be different variables, or they could be functions of x_i, $f(x_i)$, as in Chapter 7. We shall rewrite Equation (11.17) for the linear-correlation coefficient r in terms of another quantity s_{jk}^2.

We define the *sample covariance* s_{jk}^2:

$$s_{jk}^2 \equiv \frac{1}{N-1}\sum[(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \tag{11.21}$$

where the means $\bar{x}_j$ and $\bar{x}_k$ are given by

$$\bar{x}_j \equiv \frac{1}{N}\sum x_{ij} \quad \text{and} \quad \bar{x}_k = \frac{1}{N}\sum x_{ik} \tag{11.22}$$

and the sums are taken over the range of the subscript i from 1 to N. The weights have been omitted for clarity. With this definition, the sample variance for one variable s_j^2,

$$s_j^2 \equiv s_{jj}^2 = \frac{1}{N-1}\sum(x_{ij} - \bar{x}_j)^2 \tag{11.23}$$

is analogous to the sample variance s^2 defined in Equation (1.9):

$$s^2 = \frac{1}{N-1}\sum(x_i - \bar{x})^2 \tag{11.24}$$

It is important to note that the sample variances s_j^2 defined by Equation (11.23) are measures of the ranges of variation of the variables and not of the uncertainties in the variables.

Equation (11.21) can be rewritten for comparison with Equation (11.17) by substituting the definitions of Equation (11.22):

$$s_{jk}^2 \equiv \frac{1}{N-1}\sum[(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)] = \frac{1}{N-1}\sum(x_{ij}x_{ik} - \bar{x}_j\bar{x}_k) \tag{11.25}$$

If we substitute x_{ij} for x_i and x_{ik} for y_i in Equation (11.17), we can define the *sample linear-correlation coefficient* between any two variables x_i and x_k as

$$r_{jk} = \frac{s_{jk}^2}{s_j s_k} \tag{11.26}$$

with the covariances and variances s_{jk}^2, s_j^2, and s_k^2 given by Equations (11.23) and (11.25). Thus, the linear-correlation coefficient between the jth variable x_j and the dependent variable y is given by

$$r_{jy} = \frac{s_{jy}^2}{s_j s_y} \tag{11.27}$$

Similarly, the linear-correlation coefficient of the parent population of which the data are a sample is defined as

$$\rho_{jk} = \frac{\sigma_{jk}^2}{\sigma_j \sigma_k} \tag{11.28}$$

where σ_j^2, σ_k^2, and σ_{jk}^2 are the true variances and covariances of the parent population. These linear-correlation coefficients are also known as product-moment correlation coefficients.

With these definitions we can consider either the correlation between the dependent variable and any other variable r_{jy} or the correlation between any two variables r_{jk}.

Polynomials

In Chapter 7 we investigated functional relationships between y and x of the form

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \tag{11.29}$$

In a sense, this is a variation on the linear relationship of Equation (11.20) where the powers of the single independent variable x are considered to be various variables $x_j = x^j$. The correlation between the independent variable y and the mth term in the power series of Equation (11.29), therefore, can be expressed in terms of Equations (11.23) through (11.27):

$$\begin{aligned} r_{my} &= \frac{s_{my}^2}{s_m s_y} \\ s_m^2 &= \frac{1}{N-1}\left[\Sigma x_i^{2m} - \frac{1}{N}(\Sigma x_i^m)^2\right] \\ s_y^2 &= \frac{1}{N-1}\left[\Sigma y_i^2 - \frac{1}{N}(\Sigma y_i)^2\right] \end{aligned} \tag{11.30}$$

Weighted Fit

If the uncertainties in the data points are not all equal ($\sigma_i \neq \sigma$), we must include the individual standard deviations σ_i as weighting factors in the definition of variances, covariances, and correlation coefficients. From Chapter 6 the prescription for introducing weighting is to multiply each term in the sum by $1/\sigma_i^2$.

The formula for the correlation remains the same as Equations (11.26) and (11.27), but the formulas of Equations (11.21) and (11.23) for calculating the variances and covariances must be modified:

$$\begin{aligned} s_{jk}^2 &\equiv \frac{1/(N-1)\Sigma[(1/\sigma_i^2)(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)]}{(1/N)\Sigma(1/\sigma_i^2)} \\ s_j^2 \equiv s_{jj}^2 &= \frac{1/(N-1)\Sigma[(1/\sigma_i^2)(x_{ij} - \bar{x}_j)^2]}{(1/N)\Sigma(1/\sigma_i^2)} \end{aligned} \tag{11.31}$$

where the means $\bar{x}_j$ and $\bar{x}_k$ are also weighted means

$$\bar{x}_j = \frac{\Sigma x_{ij} w_i}{N} = \frac{\Sigma(x_{ij}/\sigma_i^2)}{\Sigma(1/\sigma_i^2)}$$

The weighting factors

$$w_i = \frac{1/\sigma_i^2}{(1/N)\Sigma(1/\sigma_i^2)} \tag{11.32}$$

for each data point are the inverse of the variances σ_i^2 that describe the uncertainties in each point, normalized to the average of all the weighting factors.

Multiple-Correlation Coefficient

We can extrapolate the concept of the linear-correlation coefficient, which characterizes the correlation between two variables at a time, to include multiple correlations between groups of variables taken simultaneously. The linear-correlation coefficient r of Equation (11.17) between y and x can be expressed in terms of the variances and covariances of Equation (11.31) and the slope b of a straight-line fit given in Equation (11.12):

$$r^2 = \frac{s_{xy}^4}{s_x^2 s_y^2} = b\frac{s_{xy}^2}{s_y^2} \tag{11.33}$$

In analogy with this definition of the linear-correlation coefficient, we define the *multiple-correlation coefficient* R to be the sum over similar terms for the variables of Equation (11.20):

$$R^2 \equiv \sum_{j=1}^{n}\left(b_j \frac{s_{jy}^2}{s_y^2}\right) = \sum_{j=1}^{n}\left(b_j \frac{s_j}{s_y} r_{jy}\right) \tag{11.34}$$

If we substitute x_{ij} for x_i and x_{ik} for y_i in Equation (11.17), we can define the *sample linear-correlation coefficient* between any two variables x_i and x_k as

$$r_{jk} = \frac{s_{jk}^2}{s_j s_k} \tag{11.26}$$

with the covariances and variances s_{jk}^2, s_j^2, and s_k^2 given by Equations (11.23) and (11.25). Thus, the linear-correlation coefficient between the jth variable x_j and the dependent variable y is given by

$$r_{jy} = \frac{s_{jy}^2}{s_j s_y} \tag{11.27}$$

Similarly, the linear-correlation coefficient of the parent population of which the data are a sample is defined as

$$\rho_{jk} = \frac{\sigma_{jk}^2}{\sigma_j \sigma_k} \tag{11.28}$$

where σ_j^2, σ_k^2, and σ_{jk}^2 are the true variances and covariances of the parent population. These linear-correlation coefficients are also known as product-moment correlation coefficients.

With these definitions we can consider either the correlation between the dependent variable and any other variable r_{jy} or the correlation between any two variables r_{jk}.

Polynomials

In Chapter 7 we investigated functional relationships between y and x of the form

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \tag{11.29}$$

In a sense, this is a variation on the linear relationship of Equation (11.20) where the powers of the single independent variable x are considered to be various variables $x_j = x^j$. The correlation between the independent variable y and the mth term in the power series of Equation (11.29), therefore, can be expressed in terms of Equations (11.23) through (11.27):

$$\begin{aligned} r_{my} &= \frac{s_{my}^2}{s_m s_y} \\ s_m^2 &= \frac{1}{N-1}\left[\Sigma x_i^{2m} - \frac{1}{N}(\Sigma x_i^m)^2\right] \\ s_y^2 &= \frac{1}{N-1}\left[\Sigma y_i^2 - \frac{1}{N}(\Sigma y_i)^2\right] \end{aligned} \tag{11.30}$$

Weighted Fit

If the uncertainties in the data points are not all equal ($\sigma_i \neq \sigma$), we must include the individual standard deviations σ_i as weighting factors in the definition of variances, covariances, and correlation coefficients. From Chapter 6 the prescription for introducing weighting is to multiply each term in the sum by $1/\sigma_i^2$.

The formula for the correlation remains the same as Equations (11.26) and (11.27), but the formulas of Equations (11.21) and (11.23) for calculating the variances and covariances must be modified:

$$\begin{aligned} s_{jk}^2 &\equiv \frac{1/(N-1)\Sigma[(1/\sigma_i^2)(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)]}{(1/N)\Sigma(1/\sigma_i^2)} \\ s_j^2 \equiv s_{jj}^2 &= \frac{1/(N-1)\Sigma[(1/\sigma_i^2)(x_{ij} - \bar{x}_j)^2]}{(1/N)\Sigma(1/\sigma_i^2)} \end{aligned} \tag{11.31}$$

where the means $\bar{x}_j$ and $\bar{x}_k$ are also weighted means

$$\bar{x}_j = \frac{\Sigma x_{ij} w_i}{N} = \frac{\Sigma(x_{ij}/\sigma_i^2)}{\Sigma(1/\sigma_i^2)}$$

The weighting factors

$$w_i = \frac{1/\sigma_i^2}{(1/N)\Sigma(1/\sigma_i^2)} \tag{11.32}$$

for each data point are the inverse of the variances σ_i^2 that describe the uncertainties in each point, normalized to the average of all the weighting factors.

Multiple-Correlation Coefficient

We can extrapolate the concept of the linear-correlation coefficient, which characterizes the correlation between two variables at a time, to include multiple correlations between groups of variables taken simultaneously. The linear-correlation coefficient r of Equation (11.17) between y and x can be expressed in terms of the variances and covariances of Equation (11.31) and the slope b of a straight-line fit given in Equation (11.12):

$$r^2 = \frac{s_{xy}^4}{s_x^2 s_y^2} = b\frac{s_{xy}^2}{s_y^2} \tag{11.33}$$

In analogy with this definition of the linear-correlation coefficient, we define the *multiple-correlation coefficient* R to be the sum over similar terms for the variables of Equation (11.20):

$$R^2 \equiv \sum_{j=1}^{n}\left(b_j \frac{s_{jy}^2}{s_y^2}\right) = \sum_{j=1}^{n}\left(b_j \frac{s_j}{s_y} r_{jy}\right) \tag{11.34}$$

multiple-correlation coefficient R characterizes the fit of the data to the entire function. A comparison of the multiple-correlation coefficient for different functions is therefore useful in optimizing the theoretical functional form.

We shall discuss in the following sections how to use these correlation coefficients to determine the validity of including each term in the polynomial of Equation (11.29) or the series of arbitrary functions of Equation (11.20).

11.4 *F* TEST

As noted in Section 11.1, the χ^2 test is somewhat ambiguous unless the form of the parent function is known, because the statistic χ^2 measures not only the discrepancy between the estimated function and the parent function, but also the deviations between the data and the parent function simultaneously. We would prefer a test that separates these two types of information so that we can concentrate on the former type. One such test is the F test, which combines two different methods of determining a χ^2 statistic and compares the results to see if their relation is reasonable.

F Distribution

If two statistic χ_1^2 and χ_2^2, which follow the χ^2 distribution, have been determined, the ratio of the reduced chi-squareds, $\chi_{\nu 1}^2$ and $\chi_{\nu 2}^2$, is distributed according to the F distribution[4]

$$f = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} \tag{11.35}$$

with probability density function

$$P_f(f; \nu_1, \nu_2) = \frac{\Gamma[(\nu_1 + \nu_2)/2]}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{f^{1/2(\nu_1-2)}}{(1 + f\nu_1/\nu_2)^{1/2(\nu_1+\nu_2)}} \tag{11.36}$$

where ν_1 and ν_2 are the numbers of degrees of freedom corresponding to χ_1^2 and χ_2^2. By the definition of χ_ν^2 [see Equation (11.4)], a ratio of ratios of variances

$$\frac{\chi_{\nu_1}^2}{\chi_{\nu_2}^2} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \tag{11.37}$$

is also distributed as F, where s_1 and s_2 are experimental estimates of standard deviations σ_1 and σ_2 pertaining to some characteristic of the same or different distributions.

As with our tests of χ^2 and the linear-correlation coefficient r, we shall be more interested in the integral probability

$$P_F(F; \nu_1, \nu_2) = \int_F^\infty p_f(f; \nu_1, \nu_2)\, df \tag{11.38}$$

which describes the probability of observing such a large value of F from a random set of data when compared to the correct fitting function. The integral function $P_F(F; \nu_1, \nu_2)$ is tabulated and graphed in Table C.5 for a wide range of F, ν_1, and ν_2.

A word of caution is in order concerning the use of these tables. Because the statistic F in Equation (11.35) is defined as the ratio of two determinations of χ^2 without specifying which must be in the numerator, we can define two statistics F_{12} and F_{21},

$$F_{12} = \frac{\chi_{\nu 1}^2}{\chi_{\nu 2}^2} \qquad F_{21} = \frac{\chi_{\nu 2}^2}{\chi_{\nu 1}^2} = \frac{1}{F_{12}} \tag{11.39}$$

which must both be distributed according to the F distribution.

If in some experiment our calculations yield a particular value of F_{12}, we can use Table C.5 to determine whether such a large value is less than 5% probable (Table C.6 and Figure C.6) or less than 1% probable (Table C.7 and Figure C.7). If the test value is less than the tabulated values, we must also make sure that it is not too small. To do this, we compare the value

$$F_{21} = 1/F_{12} \tag{11.40}$$

to the same tables and graphs, noting that the values of ν_1 and ν_2 are reversed. The values of ν_1 and ν_2 specified in Table C.5 correspond to the degrees of freedom for the numerator and denominator of Equation (11.39), respectively.

Example 11.3. Suppose that $F_{12} = 0.2$ with $\nu_1 = 2$ and $\nu_2 = 10$. For Table C.6, the observed value of F_{12} may be as high as 4.10 and still be exceeded by about 5% of random observations. Similarly, we compare $F_{21} = 1/F_{12} = 5.0$ with the 5% point for $\nu_1 = 10$ and $\nu_2 = 2$, which has a value of 19.4. Because the values of F_{12} and F_{21} are well within the 5% limits, we can have confidence in the fit.

What we are estimating in this example is the probability $P_F(F_{12}; \nu_1, \nu_2)$ that F_{12} is not too large and the probability $P_F(1/F_{12}; \nu_2, \nu_1)$ that F_{12} is not too small. It is tempting to simplify this procedure by assuming that

$$P_F(1/F_{12}; \nu_2, \nu_1) = P_F(F_{12}; \nu_1, \nu_2) \tag{11.41}$$

so that our test consists of determining F such that

$$P_F(F; \nu_1, \nu_2) = 0.05$$

with the requirement that

$$F > F_{12} > 1/F$$

This approximation is valid for reasonably large values of ν_1 and ν_2 but not for small values of either, as in the preceding example, where we have $4.10 > F_{12} > 1/19.4$.

Multiple-Correlation Coefficient

There are two types of F tests that are normally performed on least-squares fitting procedures. One is designed to test the entire fit and can be related to the multiple-correlation coefficient R. The other, to be discussed later, tests the inclusion of an additional term in the fitting function.

$$S_y^2 = \Sigma(y_i - \bar{y})^2 \tag{11.42}$$

this is a statistic that follows the χ^2 distribution with $N - 1$ degrees of freedom (only one parameter $\bar{y}$ must be determined from the N data points). It is a characteristic of quantities that follow the χ^2 distribution that they may be expressed as the sum of other quantities that also follow the χ^2 distribution such that the number of degrees of freedom of the original statistic is the sum of the numbers of degrees of freedom of the terms in the sum.

By suitable manipulation and rearrangement, it can be shown that S_y^2 can be expressed as the sum of the two terms,

$$\begin{aligned} S_y^2 = \sum(y_i - \bar{y})^2 &= \sum_{j=1}^{m}\left[(y_i - \bar{y})\sum_{j=1}^{m} a_j(f_j - \bar{f}_j)\right] + \sum_{j=1}^{m}(y_i - \sum a_j f_j)^2 \\ &= \sum_{j=1}^{m}[a_j\sum[(y_i - \bar{y})(f_j - \bar{f}_j)]] + \sum[y_i - y(x_i)]^2 \end{aligned} \tag{11.43}$$

where the fitting function is of the form

$$y(x_i) = \sum_{j=1}^{m} a_j f_j(x_i) \tag{11.44}$$

and we have

$$\bar{f}_j = \frac{1}{N}\sum f_j(x_i) \tag{11.45}$$

The left-hand side of Equation (11.43) is distributed as χ^2 with $N - 1$ degrees of freedom. The right-hand term is our definition of χ^2 from the Equation (11.3) and has $N - m$ degrees of freedom. Consequently, the middle term must be distributed according to the χ^2 distribution with $m - 1$ degrees of freedom.

By comparison with our definition of the multiple-correlation coefficient R in Equation (11.34), we can express this middle tern as a fraction R^2 of the statistic S_y^2:

$$\sum_{j=1}^{m} a_j\sum[(y_i - \bar{y})(f_j - \bar{f}_j)] = R^2\sum(y_i - \bar{y})^2 \tag{11.46}$$

Equation (11.43) becomes

$$\sum(y_i - \bar{y})^2 = R^2\sum(y_i - \bar{y})^2 + (1 - R^2)\sum(y_i - \bar{y})^2 \tag{11.47}$$

or

$$S_y^2 = R^2 S_y^2 + (1 - R^2)S_y^2 \tag{11.48}$$

where, as before, both terms on the right-hand side are distributed as χ^2, the first with $m - 1$ degrees of freedom and the second with $N - m$ degrees of freedom.

Thus, the physical meaning of the multiple-correlation coefficient becomes

From the definition of Equation (11.35), we can define a ratio F_R of the two terms in the right-hand side of Equation (11.47) that follow the F distribution with $\nu_1 = m - 1$ and with $\nu_2 = N - m$ degrees of freedom,

$$F_R = \frac{R^2/(m-1)}{(1-R^2)/(N-m)} = \frac{R^2}{(1-R^2)} \times \frac{(N-m)}{(m-1)} \tag{11.49}$$

From this definition of F_R in terms of the multiple-correlation coefficient R, it is clear that a large value of F_R corresponds to a good fit, where the multiple correlation is good and $R \simeq 1$. The F test for this statistic is actually a test that the coefficients are 0 ($a_j = 0$). So long as F_R exceeds the test value for F, we can be fairly confident that our coefficients are nonzero. If, on the other hand, $F_R < F$, we may conclude that at least one of the terms in the fitting function is not valid, is decreasing the multiple correlation by its inclusion, and should have a coefficient of 0.

Test of Additional Term

Because of the additive nature of functions that obey the χ^2 statistics, we can form a new χ^2 statistic by taking the difference of two other statistics that are distributed as χ^2. In particular, if we fit a set of data with a fitting function with m terms, the resulting value of chi-square associated with the deviations about the regression $\chi^2(m)$ has $N - m$ degrees of freedom. If we add another term to the fitting function, the corresponding value of chi-square $\chi^2(m + 1)$ has $N - m - 1$ degrees of freedom. The difference between these two must follow the χ^2 distribution for 1 degree of freedom.

If we form the ratio of the difference $\chi^2(m) - \chi^2(m + 1)$ to the new value $\chi_\nu^2(m + 1)$, we can form a statistic F_χ that follows the F distribution with $\nu_1 = 1$ and $\nu_2 = N - m - 1$:

$$F_\chi = \frac{\chi^2(m) - \chi^2(m+1)}{\chi^2(m+1)/(N-m-1)} = \frac{\Delta\chi^2}{\chi_\nu^2} \tag{11.50}$$

This ratio is a measure of how much the additional term has improved the value of the reduced chi-square and should be small when the function with $m + 1$ terms does not significantly improve the fit over the function with m terms. Thus, we can be confident in the relative merit of the new terms if the value of F_χ is large. As for F_R, this is really a test of whether the coefficient for the new term is 0 ($a_{m+1} = 0$). If F_χ exceeds the test value for F, we can be fairly confident that the coefficient should not be 0 and the term, therefore, should be included. Table C.5 and Figure C.5 are useful for testing F_χ. They give the value of F corresponding to various values of the probability $P_F(F; 1, \nu_2)$ and various values of ν_2 for the case where $\nu_1 = 1$. Thus, rather than evaluating F for critical values of the probability (for example, 5% or 1%), we can evaluate the probability corresponding to the observed value of F_χ.

A calculation of F_χ could be built into a linear regression program and the resulting value compared to a supplied test value F, to indicate whether or not the last term in the series is justified, and therefore, to determine how many terms in the

number of terms in the calculation manually. One should, however, be aware that the important figure of merit for added terms is the difference of the two values of χ^2 divided by the new value χ^2_ν of the *reduced* chi-square.

11.5 CONFIDENCE INTERVALS

The object of data fitting is to obtain values for the parameters of the fitted function, and the uncertainties in the parameters. The quality of the fit is indicated by χ^2 and its associated probability, and the uncertainties give the probabilities that our values of the fitted parameters are good estimates of the parent parameters. Whether we estimate our parameters by the least-squares method or by direct application of the maximum-likelihood method, as discussed in Chapter 10, we must always estimate the uncertainty in our parameters to indicate numerically our confidence in our results.

Generally, we assume Gaussian statistics and quote the standard deviation σ in a result, where σ appears in the Gaussian probability density function

$$p_G(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{11.51}$$

and determines the width of the distribution. As noted in Chapter 2, approximately 68.3% of the events of the Gaussian distribution fall within $\pm\sigma$ of the mean μ and approximately 95.4% fall within $\pm 2\sigma$.

Confidence Level for One-Parameter Fit

One way of looking at the 1 standard deviation limit is to consider that, in a series of repeated experiments, there is approximately a 68% chance of obtaining values within $\pm\sigma$ of the mean μ. Of course, we usually do not know μ, and perhaps not σ either, but have determined experimentally only $\bar{x}$ and s, our estimate of the parameters. However, as long as our experimental estimates $\bar{x}$ and s are reasonably close to the true values μ and σ, we can state that there is approximately a 68% probability that the true value of the measured parameter lies between $\bar{x}-s$ and $\bar{x}+s$, or that at the 68.3% *confidence level*, the true value of the parameter lies between these two limits.

We may wish to quote results in terms of other confidence levels. For example, we refer to the $\pm 2\sigma$ limit as the 95.4% confidence interval, or we may quote a 99% or 99.9% confidence level for a high-precision experiment. The conventional 1σ and 2σ limits are based on the Gaussian distribution, which may or may not apply to the data in question, and even an experimental distribution that nominally follows Gaussian statistics is apt to deviate in the tails.

For any distribution, represented by the normalized probability density function, $p_x(x;\mu)$, we determine the probability that a measurement of the parameter will fall between $\bar{x}-a$ and $\bar{x}+b$ by the integral

$$P_x=\int_{\bar{x}-a}^{\bar{x}+b} p_x(\bar{x};x)\,dx \tag{11.52}$$

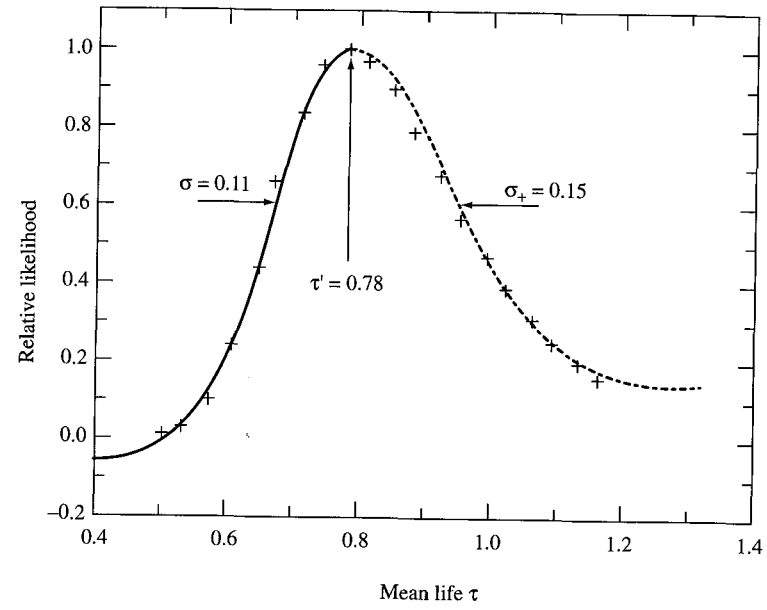

FIGURE 11.1
Relative values of the likelihood function versus trial values of the parameter for the 373-event sample of Example 10.1d. The data points (from Figure 10.4b) are indicated by crosses; the solid and dashed curves represent the results of fitting Gaussian curves separately to the two sides of the distribution. Parameters determined in the two fits are indicated on the graph. All measurements are in units of 10^{-10} s.

symmetrical about the mean. The uncertainties in our measurements may not be symmetrical, although the asymmetry may be hidden if we assume Gaussian statistics in our calculations. For example, the routines for finding uncertainties in parameters found by least-squares fitting (Chapters 7 and 8) generally assume a Gaussian distribution of the parameters and hence produce a single number for the uncertainties.

Example 11.4. As an example of an asymmetrical probability distribution, consider the 373-event data sample of Example 10.1d. In Figure 10.4b we plot as crosses the scaled values of the likelihood function for these data as a function of trial values of the parameter τ. The data points exhibit a marked asymmetry about the mean τ'. The dashed curve was calculated from Equation (10.10) with parameters obtained from the fit.

To make a better determination of σ from this curve, we considered the regions on each side of the mean separately and estimated two separate standard deviations, σ_L and σ_R, with the aid of Equation (1.11). To reduce the effect of the right-hand side tail on the value of σ_r, we imposed a cutoff at $\tau = 1.6$ and used only those data points below the cutoff in this calculation.

A composite curve formed of two Gaussians with the

the uncertainty in τ, so that we could report $\tau' = 0.78^{+0.15}_{-0.11}$, as indicated by the arrows on Figure 11.1 rather than $\tau' = 0.78 \pm 0.14$ as we did in Chapter 10. This is equivalent to finding the two positions at which the logarithm of the likelihood function has decreased by $\Delta M = ½$ as discussed in Section 10.2. Clearly this result is somewhat subjective if either side of the curve does not follow the Gaussian form. For this example, the value of σ_R depends on how much of the tail is included in the calculation.

Confidence Levels for Multiparameter Fits

The definition of the confidence level in a one-parameter experiment is generally straightforward. We can plot our data and observe if the distribution is Gaussian and estimate directly from the distribution of the probability that the true result lies between two specified values. When two or more variables have been determined and those variables exhibit some correlation, the definition of the confidence level becomes a little more difficult. Consider, for example, the determination of the mean lifetimes τ_1 and τ_2 of two unstable silver isotopes of Example 8.1. The problem was treated in Chapter 8 as a five-parameter problem, with parameters a_4 and a_5 corresponding to the two mean lifetimes, τ_1 and τ_2, respectively, and parameters a_1, a_2, and a_3 corresponding to the amplitudes of a uniform background and the two decaying states. The parameters of most interest in the experiment are a_4 and a_5, and we want to define a joint confidence interval for those two variables.

Figure 11.2 shows two sets of contours for the variation of χ^2 as a function of a_4 and a_5 from the least-squares fit by the Marquardt method discussed in Chapter 8. The small contours, drawn with solid lines, were calculated by holding the parameters a_1, a_2, and a_3 fixed at their optimum values (see Table 8.5) and varying a_4 and a_5 to obtain increases in χ^2 of 1, 2, and 3 from the minimum value. The large contours, shown as dashed lines, were calculated by allowing a_1, a_2, and a_3 to vary to minimize χ^2 for each pair of values of a_4 and a_5. The contour plots cover very different ranges because of the correlations of the displayed parameters, a_4 and a_5, with the remaining parameters a_1 through a_3. The tilt of the closed figures on each plot indicates the degree of correlation of parameters a_4 and a_5 with each other. In an ideal experiment, the contours are ellipses in the region of the χ^2 minimum and if a_4 and a_5 are not correlated, then, with suitable scaling of the axes, the ellipses are circles.

Which plot should we use? Additionally, how do we determine a confidence interval; that is, a region of the a_4-a_5 space in which we estimate there is, for example, a ~68% probability of finding the true values of the two parameters?

First, we should note that, because the fitting function, Equation (8.2), is not linear in the parameters, the methods of testing described in the previous sections strictly do not apply. However, we are much more likely to run into nonlinear fitting problems than the easier linear problems, so we shall continue with this example. At any rate, the function is linear in parameters a_1 through a_3, and we could make a linear expansion of it, over a limited region, in the parameters a_4 and a_5. In fact, this was the basis of a method of fitting nonlinear functions in Chapter 8.

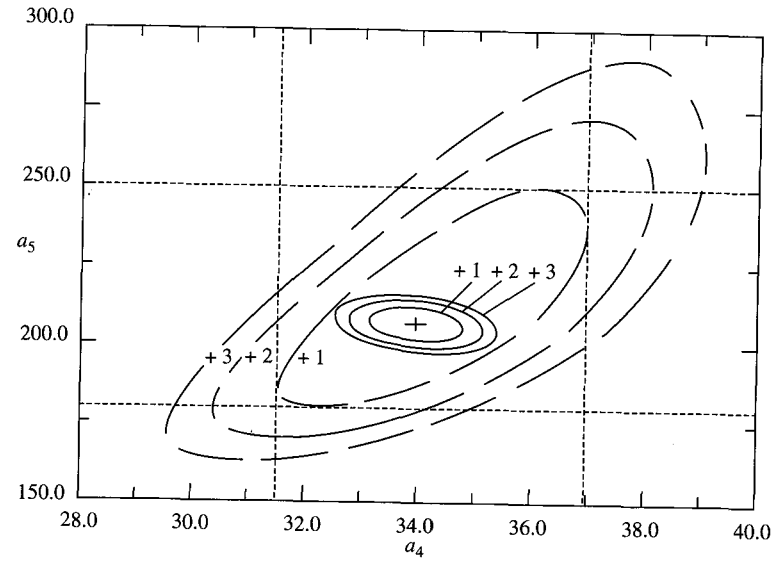

FIGURE 11.2
Two sets of contours for the variation of χ^2 with parameters a_4 and a_5 in the region of the χ^2 minimum. Data are from the least-squares fit by the Marquardt method discussed in Chapter 8. The small contours, drawn with solid lines, were calculated by holding parameters a_1 through a_3 fixed at their optimum while varying a_4 and a_5 to obtain increases in χ^2 of 1, 2, and 3 from the minimum values. The large contours, shown as dashed lines, were calculated by allowing a_1, a_2, and a_3 to vary to minimize χ^2 for each pair of values of a_4 and a_5.

from the contour plot, we should consider the full range of the outer limit of the $\Delta\chi^2 = 1$ contour, and not the intersection of that contour with the a_4 axis. This is equivalent to allowing a_5 to assume its best values for each chosen value of a_4, as we have already assumed for the parameters a_1 through a_3. The two dashed vertical lines indicate the two limits on a_4 that include the 1 standard deviation, or 68.3% of the probability, and the two horizontal lines indicate the 1 standard deviation limits for a_5.

How do we know that the vertical lines enclose 68.3% of the probability? By allowing the four parameters a_1, a_2, a_3, and a_5 to find their optimum values for each chosen value of a_4 and varying a_4, we have separated our χ^2 fitting problem into two parts: a fit of N data points to $m - 1$ parameters with $N - m - 1$ degrees of freedom and a variation of $\Delta\chi^2$ with a_4 about the minimum χ^2, with 1 degree of freedom. As we observed in the previous section, the two variations separately must follow their appropriate χ^2 distributions, so our variation of $\Delta\chi^2$ obeys the χ^2 probability distribution for 1 degree of freedom. If we look at the integrated probability distribution P_χ for 1 degree of freedom [Table C.4, or calculated from Equation (11.10)], we see that $\chi^2 \geq 1$ corresponds to 31.7% of the probability, or $\Delta\chi^2 < 1$ corresponds to 68.3%. Similarly, if we wish to find the limits for a_4

To find the 1 standard deviation region encompassed by the *joint variation* of two parameters, a_4 and a_5, with all other parameters optimized, we must draw the contour corresponding to that value of $\Delta\chi^2$ for 2 degrees of freedom that includes 68.3% of the probability. Referring again to Table C.4 or Equation (11.10), we find that we should draw the contour for $\Delta\chi^2 = 2.30$, and for the 2 standard deviation contour, we should choose $\Delta\chi^2 = 6.14$. Joint confidence intervals with more than two parameters are often of interest, but are difficult to display and are represented best by two-dimensional projections of contours for pairs of variables.

Confidence Level for a Predicted Value

Suppose the predicted value of a physical quantity is $\mu = 1000.0$, and we have made a measurement and obtained the value $\bar{x} = 999.4 \pm 2.0$. At what confidence level is the predicted value consistent with our measurement? The question could be rephrased as, "What is the probability of obtaining from the predicted parent distribution a distribution that is as bad as the one we got, or worse?" Because the shape of the parent distribution was not predicted, but only the value of the mean, we must use our value of the standard deviation, $\sigma \simeq 2.0$, as an estimate of that of the parent distribution. If the distribution is known to follow Gaussian statistics, then the required confidence is twice the integral of the standard Gaussian probability function from $x = \delta$ to ∞, where $\delta = |\mu - \bar{x}|/\sigma = |1000.0 - 999.4|/2.0$.

Now, suppose that the predicted value was necessarily positive—an intensity, for example. Then, we might again assume a Gaussian distribution, but only for positive values of the variable x, and therefore our confidence integral becomes the integral of the standard Gaussian from δ to ∞. However, because the total probability must be normalized to 1, we again multiply the integral by 2 so that the probability or confidence level is the same for both problems.

The method of determining the confidence level thus depends on the type of problem as well as the probability function that is applicable to the problem. For distributions that are symmetrical about their means, such as the Gaussian distribution, we generally consider the probability of obtaining a result that is the specified number of standard deviations from the mean, without regard to sign, unless a particular sign is excluded by the physical problem. For distributions such as the chi-square and Poisson distributions, which are only defined for positive values of their arguments, it is conventional to find a "one-sided" probability as in the case of the χ^2 distribution where we quote the probability of obtaining a value as large as or larger than the value we obtained for a given number of degrees of freedom.

11.6 MONTE CARLO TESTS

A Monte Carlo calculation can help us understand the statistical significance of our results and possibly obtain a better estimate of some of the parameters of the experiment. As a by-product, the Monte Carlo program may also help us identify biases

We want to find the probability of obtaining from a series of similar experiments a value $\bar{x}$ that differs from the predicted value μ by

$$\Delta x \geq |\mu - \bar{x}| \tag{11.53}$$

We can set up a Monte Carlo program to simulate our experiment and to generate events with the parameters predicted by the theoretical principle that we are testing and with the same cuts as those imposed by our experimental apparatus. Such a program can be quite complex, but it may already exist at the time of analysis, if, for example, a Monte Carlo program was written to help plan the experiment. Or it might be possible to use some geometric and kinematic quantities from the actual experiment and only generate those parts of each event that are affected by the parameters in question.

After the Monte Carlo program has been written and debugged, we can simulate repeated experiments with the same parent parameters and the same number of final measurements as in our real experiment. The data from each of these simulated experiments can be processed by our regular analysis program to obtain a group of "experimental" values of $\bar{x}$, and from the distribution of these values we can estimate the required probability.

Example 11.5. Let us use the Monte Carlo method to try to learn more about the significance of the small peak in our data of Example 9.2. Examination of Figure 9.2 leaves no doubt about the existence of a large peak at ~1.0 GeV. Without the fitted curve, the smaller peak near 0.8 GeV would be considerably less striking and further analysis might be helpful. (We note that, if the small peak were indeed spurious, we should have to refit the large peak to obtain a better estimate of its mean energy and width.) In Chapter 9, we estimated the probability to be about 0.01% that the smaller peak is just a fluctuation in a single bin above the single-peak background, with a probability of about 0.6% of such a fluctuation occurring in any one of the 60 bins into which the data were sorted. These are quite compelling numbers. Can we support them with a more detailed calculation by the Monte Carlo method?

We adapted to the study of this problem the Monte Carlo program and the least-squares fitting program, which were used to generate and analyze the data in Chapter 9. With the Monte Carlo program, we simulated the experiment according to Equation (9.1) to generate 4000 single-peak events in each of 1000 trial "experiments." The mean energy (E_0), half-width (Γ), and amplitude of the larger peak, and the amplitudes (a_1 through a_3) of the quadratic background, were set to the values obtained in the six-parameter fit, listed in Table 9.1.

To each set of trial data we fitted Equation (9.13), using identical procedures to those used in Chapter 9, with the exception that, starting values for the parameters of the smaller peak (a_7, a_8, and a_9) were set to the values obtained in the nine-parameter fits of Chapter 9, listed in column 6 of Table 9.1. We selected those fits that yielded parameters of the lower peak consistent with the values determined in Chapter 9 by imposing the following conditions: (1) We required that both the chi-square probability and the amplitude of the smaller peak (a_7) be equal to or greater than the corresponding fitted values listed for the nine-parameter fit in Table 9.1; (2) We required that the central energy of that peak be within plus or minus one histogram bin (0.05 GeV) of

TABLE 11.1
Results of generating 4000-event "experiments" from Equations (9.1) and (9.13) with parameters from fits listed in Table 9.1. We used several values of the amplitude A_1 of the smaller peak to test the sensitivities of our analysis to small and possibly spurious peaks.

A_1	Equation	Number of experiments	Number of successes
3.50	9.13	100	61
1.75	9.13	100	18
0.875	9.13	100	5
0.000	9.1	1000	~5

Chapter 9 for a single bin fluctuation. Tests made with other starting values and cuts for the smaller peak yielded similar numbers of survivors.

To check our procedure, we also generated and analyzed 100 two-peak trial "experiments" from Equation (9.13), with the parameters of the smaller peak set to the values from the nine-parameter fit listed in Table 9.1. From these 100 trials, 61, or 61%, survived the cuts. When we repeated the analysis with the amplitude of the smaller peak reduced by a factor of 2 (i.e., $a_4/2$), the success rate dropped to 18%, and a further reduction by another factor of 2 ($a_4/4$) reduced the success rate to 5%. The results of analyses are summarized in Table 11.1.

These results offer strong support for the existence of the smaller peak, and indicate that in a 4000-event experiment we might detect with reasonable probability a peak with only one-fourth the amplitude of the current smaller peak. Clearly, a Monte Carlo simulation should play an important role in planning this type of experiment. A carefully planned Monte Carlo program may be much better (and easier) than a detailed theoretical analysis for finding an answer to the question "How much data will be needed to establish (or disprove) the existence of a specified feature in a distribution."

We offer a final word of caution on using the Monte Carlo technique to study the statistical significance of experimental results. For Examples 9.2 and 11.5, we used a very simple problem to illustrate this technique. Yet, there are many opportunities for errors, which can lead to erroneous conclusions about the significance of our Chapter 9 data. In a larger study, it would be very easy to make a simple mistake that might lie undetected in the program and have a subtle effect on the results. It is important to test the program under a variety of conditions, and to examine results at intermediate stages before drawing conclusions from the result. In particular, if the results of the program lead to conclusions that violate intuition about the experiment, we should check and recheck the calculation. The Monte Carlo method is very powerful, and can enable us to solve very difficult

SUMMARY

Variance of the fit:

$$s^2 = \frac{1}{N-M}\frac{\Sigma\{(1/\sigma_i^2)[y_i - y(x_i)]^2\}}{(1/N)\Sigma(1/\sigma_i^2)} = \frac{1}{N-m}\sum w_i[y_i - y(x_i)]^2$$

Weighting factors:

$$w_i = \frac{1/\sigma_i^2}{(1/N)\Sigma(1/\sigma_i^2)}$$

Relationship between s^2 and χ^2:

$$\chi_\nu^2 = \frac{\chi^2}{\nu} = \frac{s^2}{\langle\sigma_i^2\rangle}$$

where

$$\langle\sigma_i^2\rangle = \left[\frac{1}{N}\sum\frac{1}{\sigma_i^2}\right]^{-1}$$

Probability $P_\chi(\chi^2; \nu)$ that any random set of N data points will yield a value of chi-square as large as or larger than χ^2:

$$P_\chi(\chi^2; \nu) = \int_{\chi^2}^{\infty}\frac{z^{1/2(\nu-2)}e^{-z/2}}{2^{\nu/2}\Gamma(\nu/2)}\,dz$$

Linear-correlation coefficient:

$$r \equiv \frac{N\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{[N\Sigma x_i^2 - (\Sigma x_i)^2]^{1/2}[N\Sigma y_i^2 - (\Sigma y_i)^2]^{1/2}}$$

Probability $P_c(r, N)$ that any random sample of uncorrelated experimental data points would yield an experimental linear-correlation coefficient as large as or larger than $|r|$:

$$P_c(r; \nu + 2) = \int_{|r|}^{1}\frac{1}{\sqrt{\pi}}\frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)}(1-r^2)^{(\nu-2)/2}$$

Sample covariance:

$$s_{jk}^2 \equiv \frac{1/(N-1)\Sigma[(1/\sigma_i^2)(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)]}{(1/N)\Sigma(1/\sigma_i^2)} \quad \text{with} \quad \bar{x}_j = \frac{\Sigma(x_{ij}/\sigma_i^2)}{\Sigma(1/\sigma_i^2)}$$

Sample variance: $\sigma_j^2 = \sigma_{jj}^2$

Sample linear-correlation coefficient:

$$r_{jk} = \frac{s_{jk}^2}{s_j s_k}$$

Multiple-correlation coefficient:

F test:

$$F = \frac{\chi^2_{\nu 1}}{\chi^2_{\nu 2}}$$

$$P_F(F; \nu_1, \nu_2) = \int_F^\infty p_f(f; \nu_1, \nu_2)\, df$$

F test for multiple-correlation coefficient R (for $\nu = N - m$):

$$F_R = \frac{R^2/(m-1)}{(1-R^2)/(N-m)} = \frac{R^2}{(1-R^2)} \times \frac{(N-m)}{(m-1)}$$

F test for χ^2 validity of adding $(m+1)$th term:

$$F_\chi = \frac{\chi^2(m) - \chi^2(m+1)}{\chi^2(m+1)/(N-m-1)} = \frac{\Delta\chi^2}{\chi^2_\nu}$$

Confidence limits: $1\sigma \longrightarrow 68.3\%$; $2\sigma \longrightarrow 95.4\%$; $3\sigma \longrightarrow 99.7\%$

EXERCISES

11.1. Discuss the meaning of χ^2 and justify the relationship between it and the sample variance $s^2 = \chi^2_\nu$.

11.2. Compare the exact calculation of the gamma function $\Gamma(n)$ of Equation (11.7) with the approximate calculation of Equation (11.8) for $n = ½, 1, 5½, 4, 9½, 10$.

11.3. From Equation (11.6), show that the χ^2-probability density for 1 degree of freedom can be written as

$$p(x^2) = \frac{e^{-x^2/2}}{\sqrt{2\pi x^2}}$$

Calculate to 1% the probability of obtaining a value of χ^2 that is less than 2.00 by expanding the function in a Taylor series and integrating term by term.

11.4. For a typical number of degrees of freedom ($\nu \simeq 10$), find, by numerically integrating Equation (11.6), the range of probability $P_\chi(\chi^2, \nu)$ for finding χ^2 as small as 0.5 or as large as 1.5. Use the approximation for the gamma function of Equation (11.8).

11.5. By numerically integrating Equation (11.6), find the probability of finding a value of $\chi^2_\nu = 1.5$ with $\nu = 100$ degrees of freedom. (Note that double-precision variables must be used.) Would you consider this to be a reasonably good fit?

11.6. Express the linear-correlation probability density of Equation (11.18) in terms of the approximation for the gamma function of Equation (11.8).

11.7. Work out the details of the calculation of the linear-correlation coefficients r for Examples 6.1 and 6.2.

11.8. If a set of data yields a zero slope $b = 0$ when fitted with Equation (11.11), what can you say about the linear-correlation coefficient r? Justify this value in terms of the correlation between x_i and y_i.

11.9. Find the linear-correlation coefficient r_1 between the independent variable T_i and the dependent variable V_i for the data of Example 7.1.

11.11. Express the multiple correlation R in terms of x_{ij}, y_i, and their averages.

11.12. Evaluate the multiple-correlation coefficient R for the data of Example 7.1.

11.13. Is a large value of F good or bad? Explain.

11.14. If we wish to set as an arbitrary criterion a probability of 0.01 for the F_χ test, what would be the reasonable average value for F test?

11.15. What different aspects of a fit do the F_R and F_χ tests represent?

11.16. Apply the F_χ test for the quadratic term to the data of Example 7.1 and state your conclusions. (Refer to Table 7.4.)

11.17. Show the intermediate steps in the derivation of Equation (11.43).

11.18. Estimate from Figure 11.2 the 90% confidence limit for each of the two mean lifetimes (a_4 and a_5) of Example 8.1 when all variables are allowed to find their optimum values.

APPENDIX A

NUMERICAL METHODS

There are several reasons why we might want to fit a function to a data sample, and several different techniques that we might use. If we wish to estimate parameters that describe the parent population from which the data are drawn, then the maximum-likelihood or least-squares method is best. If we wish to interpolate between entries in data tables to find values at intermediate points or to find numerically derivatives or integrals of tabulated data, then an interpolation technique will be more useful. Additionally, if we wish to obtain intermediate values between calculated coordinate pairs in order to plot a smooth curve on a graph, then we may wish to use a spline fitting method. In this appendix we shall summarize some standard methods for treating the latter two types of problems, as well as some methods of finding the roots of nonlinear functions, a different sort of interpolation problem.

A.1 POLYNOMIAL INTERPOLATION

With modern fast computers, the need for interpolating within tables to find intermediate values of tabulated functions has reduced markedly. Nevertheless, there are situations in which it may be convenient to represent a complicated function by a simple approximation over a limited range. For example, in a large Monte Carlo calculation, where computing time is a significant consideration, we may approximate a complex function by a simpler polynomial that can be calculated quickly. Alternatively, we may save time by creating a probability integral once at the beginning of the program. and interpolating to find values of x corresponding to the randomly chosen values of y.

For many purposes a linear or quadratic interpolation is satisfactory; that is, we fit a straight line to two coordinate pairs, or a parabola to three, and use the

is better and more convenient to represent a function over a limited region by a series of low-order approximations.

Lagrange's Interpolation Method

Here is a method that is easy to remember and can be used to expand a function to any order. We know it works because of the theorem that states that if you can find any nth-degree polynomial that passes exactly through $n + 1$ points, then you have found the one and only nth-degree polynomial that passes through those points. Think about it. It is obvious for $n = 1$ (2 points).

Let us start with an easy problem. Suppose we have two coordinate pairs (x_0, y_0) and (x_1, y_1), and we want to find the straight line that passes through both of them. We write a function of the form

$$P(x) = y_0 A_0(x) + y_1 A_1(x) \tag{A.1}$$

and search for a function $A_0(x)$ that is 1 when $x = x_0$ and 0 when $x = x_1$, and a function $A_1(x)$ that is 1 when $x = x_1$ and 0 when $x = x_0$. We can guess the form. If we write $A_0(x)$ as a fraction and set its numerator to $(x - x_1)$, then $A_0(x)$ will be 0 for $x = x_1$ and will be $(x_0 - x_1)$ for $x = x_0$. But we want $A_0(x) = 1$ for $x = x_0$, so the denominator of A_0 must be $(x_0 - x_1)$. We can make similar arguments for $A_2(x)$ and thus write as our interpolation equation

$$P(x) = y_0 \frac{(x - x_1)}{(x_0 - x_1)} + y_1 \frac{(x - x_0)}{(x_1 - x_0)} \tag{A.2}$$

Suppose we want a parabola that passes through three points. Then we simply write

$$P(x) = y_0 A_0(x) + y_1 A_1(x) + y_2 A_2(x) \tag{A.3}$$

and, following the previous arguments, write

$$P(x) = y_0 \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + y_1 \frac{(x - x_0)}{(x_1 - x_0)} \frac{(x - x_2)}{(x_1 - x_2)} + y_2 \frac{(x - x_0)}{(x_2 - x_0)} \frac{(x - x_1)}{(x_2 - x_1)} \tag{A.4}$$

The expansion to higher orders should be obvious. The kth term in an nth order expansion is given by the following product in which the $j = k$ term must be omitted:

$$\prod_{j=0}^{n} \frac{(x - x_j)}{(x_k - x_j)} y_k \quad (\text{excluding } j = k) \tag{A.5}$$

Note that the intervals in x need not be equally spaced. The interpolation for a well-behaved function $y = f(x)$ is completely general.

Newton's Divided Differences

not very convenient for repetitive calculations. It is not very convenient as an expansion either, because increasing the order of the expansion requires adding another factor to each term as well as adding another term. What we require is a more familiar form—a discrete analog of the Taylor expansion. For this we turn to Newton's method of divided differences.

There are several forms of the divided differences expansion, roughly characterized by the method we choose to define the differences, forward, backward, or about a central point. We shall restrict ourselves here to forward differences; that is, we calculate the variation of y with respect to x by taking increments in the positive x direction.

Again, consider a set of data points, (x_0, y_0), (x_1, y_1), (x_2, y_2), ... Let us assume that we wish to make a linear interpolation from x_0 to some point x with a first-degree polynomial. We define the zeroth divided difference as the function itself $f(x)$ evaluated at $x = x_0$:

$$f[x_0] \equiv f(x_0) = y_0 \tag{A.6}$$

The first divided difference is defined to be

$$f[x_0, x_1] \equiv \frac{f[x_1] - f[x_0]}{(x_1 - x_0)} \tag{A.7}$$

which is the slope of a linear function. Then, for a linear function,

$$f[x, x_0] = f[x_0, x_1] \tag{A.8}$$

or

$$\frac{f[x_0] - f[x]}{(x_0 - x)} = \frac{f[x_1] - f[x_0]}{(x_1 - x_0)} \tag{A.9}$$

which, on rearrangement of the terms, gives the first-order expansion

$$\begin{aligned} P_1(x) &= f[x_0] + (x - x_0)\frac{f[x_1] - f[x_0]}{(x_1 - x_0)} \\ &= f[x_0] + (x - x_0) f[x_0, x_1] \end{aligned} \tag{A.10}$$

where we have written $P_1(x)$ instead of $f(x)$ to indicate that the expansion is a polynomial approximation to the function $f(x)$.

To find the second-order expansion, we consider the second divided differences

$$f[x_0, x_1, x_2] \equiv \frac{f[x_2, x_1] - f[x_1, x_0]}{(x_2 - x_1)(x_1 - x_0)} \tag{A.11}$$

which corresponds to the slope of the slope, or the second derivative. This must be constant for a second-order function, so we have

$$f[x, x_0, x_1] = f[x_0, x_1, x_2] \tag{A.12}$$

Remainders

The extrapolation formula for an nth-order expansion is only exact when the function itself is an nth-degree polynomial. Otherwise, the *remainder* at x after n terms $R_n(x)$, defined as the difference between the original function $f(x)$ and the expansion $P_n(x)$, is given by

$$\begin{aligned} R_n(x) &= f(x) - P_n(x) \\ &= (x - x_0)(x - x_1) \cdots (x - x_n) f[x, x_0, x_1, \ldots, x_n] \end{aligned} \tag{A.14}$$

Calculation of the remainder requires the value of the function $f(x)$ at x, which is generally not available. (If it were, we might not be doing this expansion.) However, it may be possible to make an estimate of $f_n(x)$, or to use a nearby value, and thus find an estimate of $R_n(x)$. An expression for the remainder can also be obtained in terms of the $(n + 1)$th derivative of the function.[1]

Uniform Spacing

The divided difference expressions have a particular convenient form when the intervals in x are uniform; that is, if $x_2 - x_1 = x_3 - x_2 = x_i - x_{i-1} = h$. The divided difference of the previous discussion can be written

$$f[x_0, x_1] = \frac{f[x_1] - f[x_0]}{(x_1 - x_0)} = \frac{\Delta f(x_0)}{h}$$

or

$$\Delta f(x_0) \equiv f(x_1) - f(x_0) \quad \text{and} \quad h = x_1 - x_2 \tag{A.15}$$

and higher-order differences become

$$\Delta^2 f(x_0) \equiv \Delta[\Delta f(x_0)] = \Delta f(x_1) - \Delta f(x_0), \text{ etc.} \tag{A.16}$$

If we define the relative distance along the interval by

$$\alpha = (x - x_0)/h \tag{A.17}$$

we can write for the nth-order expansion,

$$\begin{aligned} P_n(x) = f(x_0) &+ \alpha \Delta f(x_0) + \alpha(\alpha - 1)\Delta^2 f(x_0)/2! + \cdots \\ &+ \alpha(\alpha - 1) \cdots (\alpha - n - 1)\Delta^n f(x_0)/n! \end{aligned} \tag{A.18}$$

Equation (A.18) is a finite difference analog of the familiar Taylor expansion with the important difference that the factors multiplying the coefficients $\Delta^k f(x_0)/n!$ are not successive powers of the relative distance from the starting point, but rather the product of relative distances from successive points used in the expansion, because $(\alpha - 1) = (x - x_0 - h)/h = (x - x_1)/h$, and so forth.

Extrapolation

Equations (A.15) through (A.18) are perfectly general for fitting exactly n sequen-

TABLE A.1
Uniform differences for cos θ

θ (degrees)	y	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5
0	1.0000	−0.0489	−0.0931	0.0139	0.0078	−0.0021
18	0.9511	−0.1420	−0.0792	0.0217	0.0056	
36	0.8090	−0.2212	−0.0575	0.0273		
54	0.5878	−0.2788	−0.0302			
72	0.3090	−0.3090				
90	−0.0000					

TABLE A.2
Extrapolation from 0 to 10° and from 0 to 75° in various orders

θ		Order				
(degrees)	cos θ	1	2	3	4	5
10	0.9848	0.9728	0.9843	0.9851	0.9848	0.9848
75	0.2588	0.7961	0.1819	0.2481	0.2589	0.2588

position of the first data point (x_0, y_0) can be anywhere, but for optimum interpolation, the values of x_0 and x_n should straddle the interpolation point x and be approximately equidistant from it.

The same formula can be used for extrapolating to values beyond the region of data, but the uncertainties in the validity of the approximation increase as x gets farther from the average of x_1 and x_n. The approximation is limited by both the degree of the interpolating polynomial and by uncertainties in the coefficients of the polynomial resulting from fluctuations in the data.

Example A.1. Table A.1 shows a uniform divided difference table for the cosine function for a range of the argument θ between 0 and 90°. Table A.2 shows values of cos θ for θ = 10 and 75° calculated from the divided difference table in orders 1 through 5. The interpolation starts at 0° so that only the top row of Table A.1 is used and thus, θ > 18°, the calculation is an extrapolation. The true value of cos θ is also listed. As we should expect, the large extrapolation to 75° is very poor in low order. Usually, an approximation can be improved by increasing the number of terms in the expansion. However, the better method would be to drop to a different line of the table; that is, to ensure that the calculation is an interpolation rather than an extrapolation.

A.2. BASIC CALCULUS: DIFFERENTIATION AND INTEGRATION

Differentiation

Let $f(x)$ be a function of the variable x. If x increases by an amount Δx, the function varies by an amount $\Delta f = f(x + \Delta x) - f(x)$. The ratio $\Delta f/\Delta x$ is a measure of the relative variation of $f(x)$ with x. In the limit, as Δx becomes infinitesimally small, the ratio $\Delta f/\Delta x$ for a continuous function $f(x)$ approaches an asymptotic value, the *derivative* df/dx of the function $f(x)$ with respect to x.

$$\frac{df(x)}{dx} \equiv \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{A.19}$$

The derivative of $f(x)$ at $x = x_0$ is written $\dfrac{df(x_0)}{dx}$ and corresponds to the slope of the function evaluated at x_0 or the tangent to the curve at that point.

Example A.2 To find the derivative of $f(x) = x^n$, we can expand the function $f(x + \Delta x)$ to first order in a Taylor series.

Thus, with $n = 4$, we have $f(x) = x^4$ and $df/dx = 4r^3$.

$$\frac{d(x^n)}{dx} = \lim_{\Delta x \to 0} \frac{(x^n + nx^{n-1}\Delta x) - x^h}{\Delta x}$$
$$= \frac{nx^{n-1}\Delta x}{\Delta x} = nx^{n-1}$$

Example A.3 For $f(x) = \sin x$, we can write

$$\sin(x + \Delta x) = (\sin x)(\cos \Delta x) + (\sin \Delta x)(\cos x)$$

and again expand $f(x)$ to obtain

$$\frac{d(\sin x)}{dx} = \lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin x}{\Delta x}$$
$$= \lim_{\Delta x \to 0} \frac{(\sin x)(\cos \Delta x) + (\sin \Delta x)(\cos x) - \sin x}{\Delta x}$$
$$= \frac{\sin x + (\Delta x)(\cos x) - \sin x}{\Delta x} = \cos x$$

Similarly, for $f(x) = \cos x$, we find $df/dx = -\sin x$.

SUMS AND PRODUCTS The derivative of a sum of functions is equal to the sum of the derivatives of the individual functions. Consider the function

$$f(x) = g(x) + h(x)$$

The derivative of this function is the sum of the derivatives of the individual terms.

$$\frac{df(x)}{dx} = \frac{dg(x)}{dx} + \frac{dh(x)}{dx}$$

$$f(x) = g(x) \times h(x)$$

We can rewrite Equation (A.19) as

$$\lim_{\Delta x \to 0} f(x + \Delta x) = \lim_{\Delta x \to 0} \left[f(x) = \Delta x \frac{df(x)}{dx} \right] \tag{A.20}$$

and show that

$$\begin{aligned} \frac{d[g(x) \times h(x)]}{dx} &= \lim_{\Delta x \to 0} \frac{g(x + \Delta x)h(x + \Delta x) - g(x)h(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{1}{\Delta x} \left\{ \left[g(x) + \Delta x \frac{dg(x)}{dx} \right] \left[h(x) + \Delta x \frac{dh(x)}{dx} \right] - g(x)h(x) \right\} \\ &= g(x) \frac{dh(x)}{dx} + h(x) \frac{dg(x)}{dx} \end{aligned}$$

FUNCTIONS OF FUNCTIONS If the function $f(x)$ can be expressed as a function of a function $g(x)$ of x,

$$f(x) = f[g(x)]$$

the derivative of $f(x)$ with respect to x can be expressed in terms of the derivative of $g(x)$ with respect to x. If we expand the definition of Equation (A.19) for the derivative, we can make use of the relationship of Equation (A.20) to expand still further.

$$\begin{aligned} \frac{df(x)}{dx} &= \lim_{\Delta x \to 0} \frac{f\left[g(x) + \Delta x \dfrac{dg(x)}{dx} \right] - f[g(x)]}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{f[g(x)] + \Delta x \dfrac{dg(x)}{dx} \dfrac{df(x)}{dg(x)} - f[g(x)]}{\Delta x} \\ &= \frac{df(x)}{dg(x)} \frac{dg(x)}{dx} \end{aligned} \tag{A.21}$$

Example A.4 If $f(x) = (a - bx^3)^2$, define $g(x) = a + bx^3$ so that $f(x) = [g(x)]^2$. The first factor in Equation (A.21) is the derivative of a square, and the second factor is the derivative of a cubic polynomial.

$$\frac{df(x)}{dg(x)} = 2g(x) = 2(a + bx^3) \qquad \frac{dg(x)}{dx} = 3bx^2$$

$$\frac{df(x)}{dx} = 2(a + bx^3)3bx^2 = 6bx^2(a + bx^3)$$

$$\frac{d^2f(x)}{dx^2} \equiv \frac{d}{dx}\left[\frac{df(x)}{dx} \right]$$

For the nth-order derivative $d^n f(x)/dx^n$, we simply take the derivative n times in succession. For example, if $f(x) = x^4$ as in Example A.2, the second derivative is $12x^2$. Similarly, the fourth derivative of either $\sin x$ or $\cos x$ is equal to itself.

PARTIAL DERIVATIVES If the function $f(x, y)$ is dependent on two variables x and y, we must define derivatives of the function with respect to each of the independent variables. To determine the *partial derivative* of f with respect to x, $\partial f/\partial x$, we consider that y is a constant and proceed as we would for an ordinary derivative. Similarly, to determine the partial derivative $\partial f/\partial y$ we consider that x is constant.

$$\frac{\partial f(x, y)}{\partial x} \equiv \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} = \frac{df(x)}{dx}$$

$$\frac{\partial f(x, y)}{\partial y} \equiv \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y} = \frac{df(y)}{dy}$$

Higher-order partial derivatives include not only higher-order derivatives with respect to one variable, but also cross-partial derivatives with respect to two or more variables simultaneously.

$$\frac{\partial^2 f(x, y)}{\partial x^2} \equiv \frac{\partial}{\partial x}\left[\frac{\partial f(x, y)}{\partial x} \right]$$

$$\frac{\partial^2 f(x, y)}{\partial x\, \partial y} \equiv \frac{\partial}{\partial x}\left[\frac{\partial f(x, y)}{\partial y} \right] = \frac{\partial}{\partial y}\left[\frac{\partial f(x, y)}{\partial x} \right] = \frac{\partial^2 f(x, y)}{\partial y\, \partial x}$$

MINIMA AND MAXIMA A function $f(x)$ is said to have a *local minimum* at $x = x_{\min}$ if the values of $f(x_{\min} \pm \Delta x)$ are larger than the value of $f(x_{\min})$ for infinitesimal changes Δx about $x_{\min}$. Similarly, the function has a *local maximum* if the values of $f(x_{\max} \pm \Delta x)$ are smaller than $f(x_{\max})$. At either a minimum or a maximum of a function, the derivative of the function is zero,

$$\frac{df(x_m)}{dx} = 0$$

corresponding to a tangent that is parallel to the x-axis.

The question of whether the function is a minimum or a maximum at x_m can be resolved by examining the second derivative. If the second derivative is positive, the curvature of the function is upward and $f(x_m)$ is a minimum. If the second derivative is negative, the $f(x_m)$ is a maximum.

FUNCTIONS OF MORE THAN ONE VARIABLE With functions of more than one variable, for example $f(x, y)$, we can still consider the function to have a minimum

Integration

Integration is the inverse of differentiation. To find the integral $F(x)$ of the function $f(x)$,

$$F(x) = \int f(x)dx$$

we must find a function $F(x)$ such that $\frac{dF(x)}{dx} = f(x)$.

However, this definition is not unique. An undetermined constant must be added to the solution to allow for the fact that the derivative of a constant is zero.

Example A.5 Consider the integral of the function $f(x) = x^3$. We observe that $F(x) = x^4/4$ is a solution:

$$\frac{dF(x)}{dx} = \frac{d(x^4/4)}{dx} = x^3 = f(x)$$

However, $F(x) = x^4/4 + C$ is also a solution, where C is any quantity that is not a function of x. Thus, the solution to an *indefinite integral* must include an added constant.

A *definite integral* is the integral of a function between two specific values of the independent variable, and is written

$$I = \int_a^b f(x)dx$$

To find the definite integral of a function, we integrate it, calculate the value of the integral at $x = b$ and at $x = a$, and find the difference between the two values. This is equivalent to calculating the area under the function $f(x)$ between the two limits a and b.

Example A.6 Consider the integral of the function $f(x) = x^3$ between the limits $x = 1.0$ and $x = 2.0$.

$$I = \int_{1.0}^{2.0} f(x)dx = \int_{1.0}^{2.0} x^3 dx = x^4/4\Big|_{1.0}^{2.0} = [2^4 - 1^4]/4 = 15/4$$

Note that a definite integral is not a function of variable of integration x.

From the results of Example A.3,

$$\int \sin x\, dx = -\cos x + C \qquad \text{and} \qquad \int \cos x\, dx = \sin x + C$$

A.3. NUMERICAL DIFFERENTIATION AND INTEGRATION

With the interpolation expressions discussed in Section A.1, it is relatively straight-

Differentiation

We can differentiate Equation (A.18) to find approximations for the derivatives of the function $f(x)$. We obtain

$$\frac{dP_n(x)}{dx} = \frac{1}{h}\frac{dP_n(x)}{d\alpha} = [\Delta f(x_0) + (2\alpha - 1)\Delta^2 f(x_0)/2! + (3\alpha^2 - 6\alpha + 2)\Delta^3 f(x_0)/3! + \cdots]/h \tag{A.22}$$

and

$$\frac{d^2P_n(x)}{dx^2} = \frac{1}{h^2}\frac{d}{d\alpha}\left[\frac{dP_n(x)}{d\alpha}\right] = [\Delta^2 f(x_0) + (\alpha - 1)\Delta^3 f(x_0) + \cdots]/h^2 \tag{A.23}$$

We should note that the use of forward differences introduces an asymmetry in the calculation. For a general solution, we could replace the forward differences by central differences, which are taken symmetrically about a central starting point. For a particular problem, we can usually arrange the expansion to provide reasonable symmetry of the differences about the point of interest. Thus, we can replace Equations (A.22) and (A.23) by

$$\frac{dP_n(x)}{dx} = \Delta f(x_0)/h = \frac{f(x + h/2) - f(x - h/2)}{h} \tag{A.24}$$

and

$$\frac{d^2P_n(x)}{dx^2} = \Delta^2 f(x_0)/h^2 = \frac{f(x + h) - 2f(x) + f(x - h)}{h^2} \tag{A.25}$$

Integration

Integrating Equation (A.18) leads to expressions for calculating the numerical integral in various orders, depending on the number of terms in the polynomial approximation. There are various forms for each order, depending on how we choose the limits of integration. We quote three of the most useful forms with the remainder estimates.

First-order, endpoint trapezoidal

$$\int_{x_0}^{x_1} f(x)\,dx = \frac{h}{2}[f(x_0) + f(x_1)] - \frac{h^3}{12}f^{(2)}(\xi)$$

(first-order closed-end trapezoidal)

$$\int_{x_0}^{x_2} f(x)\,dx = 2hf(x_1) + \frac{h^3}{3}f^{(2)}(\xi) \quad \text{(first-order open end)}$$

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] - \frac{h^5}{90}f^{(4)}(\xi)$$

(second-order closed-end Simpson's rule)

Note the large reduction on the error estimate in going from either of the first-order approximations to the second-order approximation.

For an integral over an extended range of x, it is usually advisable to employ a series of first- or second-order integrals over sections of the function, rather than to attempt to fit a large region with a higher-order function. In fact, it can be shown that the gain in accuracy in going from a second- to a third-order numerical integral is relatively small, and, for the same number of calculations of the ordinate y_j, the second-order Simpson rule may be more accurate than the third-order form. This relation applies in general to even and odd orders, so that, to make a significant improvement in the numerical integration of a function, one should advance to the next higher *even* order.

Thus, to find the integral by Simpson's rule of $f(x)$ over an extended range between $x = x_0$ and $x = x_n$, we divide the region into n equal intervals in x, with $nh = (x_n - x_0)$, to obtain

$$\int_{x_0}^{x_n} f(x)\,dx = \frac{h}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots + 4f(x_{n-1}) = f(x_n)] - \frac{nh^5}{180} f^{(4)}(\xi) \qquad \text{(A.26)}$$

where ξ is the value of x somewhere in the range of integration.

Program A.1 SIMPSON (Appendix E) calculates an extended integral by the second-order approximation of Equation (A.26). See Programs 11.1 and 11.2 for examples of the use of this routine.

The user supplies four arguments:

1. FUNCT: the name of the function to be integrated. The function must have one real argument. If other arguments are required, they must be made accessible to the function as global variables.
2. NINT: the number of *double* intervals. The interval is calculated as DX = (HILIM-LOLIM)\(2*NINT);
3. LOLIM and
4. HILIM: the integration limits.

A.4 CUBIC SPLINES

If we attempt to represent by an nth-degree polynomial a function that is tabulated at $n + 1$ points, we are apt to obtain disappointing results if n is large. The polynomial will necessarily coincide with the data points, but may exhibit large oscillations between points. In addition, if there are many data points, the calculations can become rather cumbersome. It is often better to make several low-order polynomial fits to separate regions of the function, and this procedure is usually satisfactory for simple interpolation in tables. However, if we want a smooth function, which passes through the data points, the results may not be satisfactory.

Suppose we have calculated a function at $n + 1$ points, and want to represent

the plot. It is unlikely that they will combine to form a smooth curve. What do we do now? Reach for our pencil and trusty drafting spline? No, we call up our spline fitting subroutine and let it join up the separate fits for us.

Spline fitting procedures have other uses besides plotting pretty curves on graphs, but the plotting function is of interest to us and is easily illustrated. Suppose we choose to make a series of cubic fits to successive groups of data points. What conditions do we need to produce a smooth curve that passes through the data points? We want the first and second derivatives, as well as the function itself, to be continuous at the data points. Suppose we consider a separate cubic polynomial for each interval on the graph, or a total of n polynomials for the $n + 1$ points. Then we write the polynomial equation, take derivatives, and, at each data point, equate the first and second derivatives of the left-side polynomial to those of the right-side polynomial.

Following the method discussed in Thompson (1984), we begin by writing the Taylor series for the cubic polynomial for interval i, expanded about the point x_i

$$y(x) = y(x_i) + (x - x_i)\frac{dy(x_i)}{dx} + (x - x_i)^2 \frac{d^2y(x_i)}{dx^3}\Big/2! + (x - x_i)^3 \frac{d^3y(x_i)}{dx^3}\Big/3! \qquad \text{(A.27)}$$

where the function and derivatives are evaluated at x_i. This can be written in a more concise form as

$$y(x) = y_i + (x - x_i)y_i' + (x - x_i)^2 y_i''/2 + (x - x_i)^3(y_{i+1}'' - y_i'')/6h \qquad \text{(A.28)}$$

where y_i' and y_i'' stand for the first and second derivatives evaluated at $x = x_i$ and the third derivative has been replaced by its divided difference form, which is exact for a cubic function. At $x = x_i$, we have $y = y_i$, as required. We can also set $x = x_{i+1} = x_i + h$ and solve the equation

$$y(x_{i+1}) = y_i + (x_{i+1} - x_i)y_i' + (x_{i+1} - x_i)^2 y_i''/2 + (x_{i+1} - x_i)^3(y_{i+1}'' + y_i'')/6h \qquad \text{(A.29)}$$

to obtain

$$y_{i+1} - y_i = hy_i' + h^2[2y_i'' + y_{i+1}'']/6 \qquad \text{(A.30)}$$

We repeat the calculation, using the equation for $y(x)$ in interval $i - 1$ [i.e., we replace i by $i - 1$ in Equation (A.29)],

$$y(x) = y_{i-1} + (x - x_{i-1})y_{i-1}' + (x - x_{i-1})^2 y_{i-1}''/2 + (x - x_{i-1})^3(y_i'' - y_{i-1}')/6h \qquad \text{(A.31)}$$

and again require that $y(x) = y(x_i)$ at the ith data point and obtain

$$y_i - y_{i-1} = hy_{i-1}' + h^2[2y_{i-1}'' + y_i'']/6 \qquad \text{(A.32)}$$

$$y'(x) = y'_i + (x - x_i)y''_i + (x - x_i)^2(y''_{i+1} - y''_i)/2h \tag{A.33}$$

which we equate to the first derivative in the interval $i - 1$ at the position $x = x_i$, to obtain

$$y'_i - y'_{i-1} = h[y''_i + y''_{i-1}]/2 \tag{A.34}$$

Similarly, equating the derivatives at the boundary $x = x_{i+1}$ gives

$$y'_{i+1} - y'_i = h[y''_{i+1}y''_i]/2 \tag{A.35}$$

(Repeating the procedure with the second derivative leads to an identity, because our use of the divided difference form for the third derivative assures continuity of the second derivative across the boundaries.) Eliminating the first derivatives from Equations (A.30), (A.32), (A.34), and (A.35) gives us the spline equation

$$y''_{i-1} + 4y''_i + y''_{i+1} = D_i \tag{A.36}$$

with

$$D_i = y[y_{i+1} - 2y_i + y_{i-1}]/h^2 \tag{A.37}$$

Note that the D_i are proportional to the second differences of the tabulated data and are all known. We can write Equation (A.36) as a set of linear equations relating the unknown variables y'', beginning with $i = 2$ and ending with $i = n - 1$:

$$y''_1 + 4y''_2 + y''_3 = D_2 \tag{A.38a}$$

$$y''_2 + 4y''_3 + y''_4 = D_3 \tag{A.38b}$$

$$\vdots$$

$$y''_{n-3} + 4y''_{n-2} + y''_{n-1} = D_{n-2} \tag{A.38c}$$

$$y''_{n-2} + 4y''_{n-1} + y''_n = D_{n-1} \tag{A.38d}$$

These equations can be solved for the second derivatives y''_i, as long as we know the values of y''_1 and y''_n. One possibility is to set the second derivatives to 0 to obtain *natural splines*. Alternatively, we may use the true second derivatives, if they are known, or a numerical approximation.

For example, suppose we have only four points to consider. Then, if we know y''_1 and y''_4, we can solve the simultaneous Equations (A.38a) and (A.38b) for y_2 and y_3. Similarly, if we have a full set of n equations, we can rewrite Equation (A.38a) to express $y''_2 = (D_2 - y''_1 - y''_3)/4$, and substitute this expression into Equation (A.38b) to eliminate y''_2. Then, we repeat the procedure to eliminate y''_3 from the next equation. We continue this procedure until we reach the last equation, which will contain only terms in y''_1, y''_{n-1}, and y''_n. Because y''_1 and y''_n are known, we can solve this equation for y''_{n-1}, and then work back down the chain determining successively y''_{n-2}, y''_{n-3}, and so forth, until we reach Equation (A.38a) from which we determine the last unknown y''_2. Once we have found the values of the y''_i, we can find the y'_i from Equation (A.30) or (A.32), and use Equation (A.28) to interpolate

An interesting alternative method of solving the set of simultaneous equations, Equations (A.38), is to set them up in a spreadsheet program. Then, when the boundary values y''_i and y''_n are supplied, the program will readjust the variables until they stabilize at the solutions to Equations (A.38). Although this method is not very practical for graphical applications where we want to build the solution into our plotting program, it does provide a quick way of finding the second derivatives and an interesting illustration of the solution.

As with all techniques, a certain amount of care must be exercised in using spline routines. The choice of a second derivative at the boundary may have an important effect on the interpolation at the ends of the function, and a wrong choice, for example, can produce undesirable shapes at the edges of a plot. Then too, although the spline routine assures a smooth variation between the data points, with continuity of the function and first and second derivative across the points, it cannot guarantee that there will be no peculiar oscillation between the points.

Program A.2 SPLINE INTERPOLATION (Appendix E)
SPLINEMAKE numerically calculates a table of second derivatives for a spline interpolation by the method discussed in the previous paragraphs.
SPLINEINT performs the interpolation. For simplicity, we have chosen to store only the second derivatives and to calculate the first and third derivatives as needed in functions D1YDX1 and D3YDX3. If speed is important, the derivatives could be computed and stored in arrays.

A spline interpolation routine is especially useful for plotting curves on graphs. The routine has been used to produce many of the graphs in this book.

A.5 ROOTS OF NONLINEAR EQUATIONS

Finding roots of nonlinear equations is essentially the reverse of an interpolation problem. When we interpolate a function, our object is to find a value of the dependent variable y at a specific value of the independent variable x. When we are searching for the root of a function, we are trying to find the value of x at a particular value, usually 0, of y. However, interchanging the variables completely changes the nature of the problem. Interpolation involves straightforward application of well-defined equations that are independent of the form of the original function: Finding roots of nonlinear equations may require different equations for different problems and almost always requires some sort of a search and iteration procedure.

The diffraction of light by a single slit provides an interesting example of a nonlinear equation. It is well known that the position of the interference maxima and minima from double slits and diffraction gratings can be determined analytically from consideration of the phase difference between the rays that pass through each slit, but only the minima of the diffraction pattern of a single slit can be found in this way. To find the position of a maximum, with the exception of the central one, we must differentiate the expression for the intensity with respect to

In Equation (A.39), I_0 is the intensity of the light at the central maximum ($\theta = 0$), I is the intensity at angle θ, λ is the wavelength of the light, and a is the slit width. The position of the maximum is given by solving

$$\frac{dI}{d\alpha} = 2I_0\left(\frac{\sin\alpha}{\alpha^3}\right)(\alpha\cos\alpha - \sin\alpha) = 0 \tag{A.40}$$

to obtain the value α_r at the root of the equation

$$f(\alpha) = \alpha_r - \tan\alpha_r = 0 \tag{A.41}$$

The first root is at $\alpha_r = 0$. The other roots cannot be calculated analytically and must be found by an iterative method. An approximate solution can be obtained by rewriting Equation (A.41) as

$$\alpha_r = \tan\alpha_r \tag{A.42}$$

and plotting separately the left and right sides to find the intersection of the straight line and the tangent curves. There are several mathematical ways to solve the problem, but making a plot of the function is always a good starting procedure.

Trial-and-Error: The Half-Interval Method

With a personal computer, trial-and-error may be a suitable method for solving the occasional root finding problem. An orderly approach is advisable and the half-interval method is convenient. The procedure is to write a little program that requests a trial value of the root and calculates the function and displays its value. The initial trial value might be obtained from a graph, or perhaps by mapping the function for various values of the independent variable x, until a reasonable estimate of the root has been obtained. Then, a second trial x is submitted, which produces a value of y on the other side of the root. The half-interval method begins at this point. The procedure is to select a third trial value that is midway between the two that bracket the root. For the fourth trial value, we use the mean of the most recent value, and whichever of the two previous trials was on the other side of the root. The process continues until the root is found to the desired accuracy.

This rather primitive method of root finding could be improved with a little programming to let the program decide which root to choose, to calculate the mean, and perform the next trial. The program could proceed in a loop until the root had been found to a predefined degree of accuracy, or the calculation could be stopped manually. However, if we are willing to program that little bit of logic, slightly more effort will produce a much faster root-finding program.

Secant Methods

The gain in speed comes from using the slope of the function in the calculation. We begin with two trial estimates of the root, x_k and x_{k+1}, preferably, but not necessar-

$$f(x) = y_k + (x - x_k)\frac{(y_{k+1} - y_k)}{(x_{k+1} - x_k)} \tag{A.43}$$

where we have written $y_k = f(x_k)$ and so forth. Setting $f(x) = 0$ and solving for x gives us an approximation to the value of x at the root:

$$x = x_k - y_k\frac{x_{k+1} - x_k}{y_{k+1} - y_k} = \frac{x_k y_{k+1} - x_{k+1}y_k}{y_{k+1} - y_k} \tag{A.44}$$

For the next trial, we replace x_{k+1} or x_{k+2} by the value x found in Equation (A.44) and repeat the calculation. The process can be repeated until the root is approximated as closely as desired. This is the first-order secant method.

There are various ways of choosing which of the previous values of x (x_k or x_{k+1}) to keep for the next iteration. The simplest is to keep the most recent value and discard the older value. Another way is to choose whichever is closer to the root [i.e., gives a smaller value of $f(x)$]. A third is to start the process with two values that straddle the root (i.e., give opposite signs for y_1 and y_2) and to continue to choose values that straddle the root after each iteration. This is the Regulo-Falsi method.

Clearly any method will find the root most quickly if the starting values are close to the root, but, in principle, the secant methods will almost always find a root of the function, eventually. With some functions, such as those that are antisymmetric about the root, there is the possibility that the search by the Regulo-Falsi method, for example, will jump back and forth across the root and never approach it. Additionally, for functions with several roots, we may not always find the one we want. Problems may also arise if two roots are very close together.

Newton-Raphson Method

Instead of calculating the slope by finite differences, as in the secant method, we could use the tangent, or derivative of the function, if it can be calculated. Then, we can replace Equation (A.43) by

$$f(x) = y_k + (x - x_k)\frac{df(x_k)}{dx} \tag{A.45}$$

where x_k and y_k are the values of x and $f(x)$ after the kth iteration. We find the next estimate x_{k+1} for the root, as before, by setting $f(x)$ in Equation (A.45) to zero to obtain

$$x_{k+1} = x_k - y_k \div \frac{df(x_k)}{dx} \tag{A.46}$$

Example A.7 Table A.3 shows steps in an iterative calculation of the second and third roots of Equation (A.41) by the secant and Newton-Raphson methods. Starting values were chosen by examining a plot of tan x versus x.

Simultaneous Nonlinear Equations

TABLE A.3
Determination of the first two nonzero roots of $\alpha = \tan \alpha$

Trial	First root		Second root	
	x	y	x	y
	(a) Newton's Method†			
0	4.40000	1.30368	7.70000	1.25713
1	4.53598	−1.07376	7.73028	−0.31270
2	4.50186	−0.17769	7.72545	−0.01188
3	4.49375	−0.00679	7.72525	−0.00002
4	4.49341	−0.00001	7.72525	−0.00000
5	4.49341	−0.00000		
	(b) Secant Method‡			
0	4.40000	1.30368	7.80000	−10.70682
0	4.50000	−0.13733	7.70000	1.25713
1	4.49047	0.05854	7.71051	0.78849
2	4.49332	0.00184	7.72819	−0.17931
3	4.49341	−000003	7.72491	0.02025
4	4.49341	0.00000	7.72524	0.00047
5			7.72525	−0.00000

†The calculation continues without assistance after the initial trial value has been selected

‡Two x, y pairs are required for each stage of the calculation. After the first trial, the most recently calculated x, y pair was used with whichever of the two previous pairs was closer to the root.

solve for the parameters a and b. We used the secant method to solve these equations.

Consider the two equations

$$f_a(u, v) = 0 \quad \text{and} \quad f_b(u, v) = 0 \tag{A.47}$$

which we wish to solve for u and v. We define the first *partial divided differences*,

$$\begin{aligned}
f_{au} &= f_a[u_0, v_0; u_1, v_0] \equiv \frac{f_a(u_1, v_0) - f_a(u_0, v_0)}{u_1 - u_0} \\
f_{av} &= f_a[u_0, v_0; u_0, v_1] \equiv \frac{f_a(u_0, v_1) - f_a(u_0, v_0)}{v_1 - v_0} \\
f_{bu} &= f_b[u_0, v_0; u_1, v_0] \equiv \frac{f_b(u_1, v_0) - f_b(u_0, v_0)}{u_1 - u_0} \\
f_{bv} &= f_b[u_0, v_0; u_0, v_1] \equiv \frac{f_b(u_0, v_1) - f_b(u_0, v_0)}{v_1 - v_0}
\end{aligned} \tag{A.48}$$

and, following Equation (A.43), write for a first-order expansion

If we assume that f_a and f_b are linear in u and v, we can find a first approximation to the roots by setting $f_a(u, v)$ and $f_b(u, v)$ to zero in Equation (A.49) and solving the two coupled linear equations for u and v:

$$\begin{aligned}
u f_{au} + v f_{av} - u_0 f_{au} - v_0 f_{av} + f_a(u_0, v_0) &= 0 \\
u f_{bu} + v f_{bv} - u_0 f_{bu} - v_0 f_{bv} + f_b(u_0, v_0) &= 0
\end{aligned} \tag{A.50}$$

Solution by the determinant method gives

$$\begin{aligned}
u_2 = u &= (A f_{bv} - B f_{av})/D \\
\text{and} \qquad v_2 = v &= (B f_{au} - A f_{bu})/D
\end{aligned} \tag{A.51}$$

with

$$\begin{aligned}
D &= f_{au} f_{bv} - f_{av} f_{bu} \\
A &= -u_0 f_{au} - v_0 f_{av} + f_a(u_0, v_0) \\
B &= -u_0 f_{bu} - v_0 f_{bv} + f_b(u_0, v_0)
\end{aligned} \tag{A.52}$$

We then repeat the procedure with coordinate pairs (u_1, v_1) and (u_2, v_2), to obtain the next approximation, until the roots have been found to the desired degree of accuracy.

A.6 DATA SMOOTHING

The concept of smoothing is not one that meets with universal approval. The discussion that follows should be considered with one caveat: For rigorously valid least-squares fitting, smoothing is neither desirable nor permissible; however, there are cases where smoothing can be beneficial, and, therefore, the techniques are introduced.

Consider, for example, the discussion of Section 9.2 of the determination of the area under a peak from a least-squares fit to a histogram of the data. Least-squares fitting techniques applied to data that are distributed according to Poisson distributions, rather than Gaussian distributions, underestimate the area of a peak by an amount equal to the value of χ^2. We have seen that we can improve the result by decreasing the value of χ^2 at its minimum. Similarly, if the shape of the fitting function does not exactly simulate that of the parent distribution, a better fit to the data by decreasing χ^2 can yield an improved estimate of the area under a peak.

Another example that might benefit from application of a smoothing algorithm is the parameterization of data for use in a Monte Carlo or other program. In preparing experimental proposals, it is often necessary to estimate yields and distributions based on currently available data. Such data are often sparse and generally must be expressed in parametric form for ease and speed of use in the Monte Carlo simulation program. Smoothing can be useful to average out fluctuations and allow the data to be expressed with a few parameters by a least-squares fit or an interpolation procedure.

In other words, if rigorously valid results are not required, but rather an averaged estimate of the distribution, smoothing may help obtain more reliable esti-

For example, an improved estimate of the area under a peak would be accompanied by an increased uncertainty in the estimates of the width and position of the peak.

Whatever smoothing or other manipulation is done must conserve the information pertaining to the desired parameters. The averaging techniques that we shall discuss, for example, conserve the area under a peak but not the width of the peak Similarly, this method would be useful for improving the estimate of the constant term of a polynomial but not the coefficients of the other terms.

Data smoothing is similar to the data "smearing" introduced in Chapter 5 to simulate measuring uncertainties in "measurements" generated by a Monte Carlo program. In the Monte Carlo program we used Gaussian smearing; that is, we allowed each event a Gaussian probability distribution about its mean.

In this section, we are dealing with binned data, and thus, for Gaussian smoothing, could consider a Gaussian integration that spreads each event over adjacent bins. Because our object here is to smooth the data, we are at liberty to choose the width of the smearing function to produce the desired degree of uniformity in the data, limited by the requirement that we do not damage the very variable we are trying to study.

The binomial distribution is a useful smoothing function. Suppose we want to smooth low statistics experimental data that follow a Gaussian peak in a way that preserves the area under the peak. Let us assume that the background slope is gentle enough that smoothing will not affect its determination drastically.

We can approximate the Gaussian peak with a binomial distribution with $p = \frac{1}{2}$ (see Section 2.1):

$$y(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \simeq \left(\frac{1}{2}\right)^n \frac{n!}{X!(n-X)!} \tag{A.53}$$

We can relate the widths σ and the means of the two distributions

$$\sigma_B^2 = np(1-p) = n/4 = \sigma^2 \quad \bar{X} = np = n/2 \quad \bar{x} = \mu \tag{A54}$$

to find the relationships among the parameters

$$n = 4\sigma^2 \quad X = x - \mu + n/2 = x - \mu + 2\sigma^2 \tag{A.55}$$

We can then express the binomial distribution of Equation (A.53) as

$$y(x) = \left(\frac{1}{2}\right)^n \frac{n!}{(n/2 + x - \mu)!(n/2 - x + \mu)!} \tag{A.56}$$

Let us smooth the data by averaging over adjacent channels with a binomial distribution spanning three channels:

$$y'(x) = 1/4y(x-1) + 1/2y(x) + 1/4y(x+1) \tag{A.57}$$

If we fold this averaging into the distribution of Equation (A.53), the result is also binomial:

The new distribution has the same mean $\bar{x} = \mu$ but a larger width $\sigma'^2 = n'/4 = (n+2)/4$ with the variance increased by $\frac{1}{2}$:

$$\sigma'^2 = \sigma^2 + \tfrac{1}{2} \tag{A.59}$$

Similarly, we could smooth over five channels by using a formula similar to Equation (A.57) but with five terms with coefficients given by the binomial expansion

$$\begin{aligned} y''(x) = {} & 1/16y(x-2) + 1/4y(x-1) + 3/8y(x) \\ & + 1/4y(x+1) + 1/16y(x+2) \end{aligned} \tag{A.60}$$

A five-channel smoothing is identical to two successive smoothings over three channels and yields a variance that is increased accordingly, $\sigma''^2 = \sigma^2 + 1$. Any such smoothings over $2n + 1$ adjacent channels is equivalent to n smoothings over three channels.

If we apply the smoothing of Equation (A.57) to a Gaussian distribution, the resulting distribution will also be nearly Gaussian because the shapes of the binomial and Gaussian distributions are nearly alike. In fact, if we are applying the smoothing because the original shape is not Gaussian enough, the averaging may make the shape more nearly Gaussian. If we apply binomial smoothing to a distribution that is not Gaussian, we should be aware that we are distorting the shape of the peak and making it more Gaussian.

If the width of the original Gaussian is not too small ($\sigma > 1$), the increase of Equation (A.59) should not be drastic because the addition is in quadrature. For a width $\sigma = 2$, for example, the new width $\sigma' = 2$ is only 5% larger. If the original width is very small ($\sigma < 1$), the approximation of Equation (A.53) is not valid because the Gaussian and binomial distributions are only similar in the limits of large n. A Gaussian fit to the data without smoothing would not be valid either, however, because the parameters of the fit are only meaningful if $\sigma \geq 1$. Because the averaging itself is a binomial distribution, the result is still expected to be a better approximation to a Gaussian distribution than the original data. For a smoothing over three channels, a Gaussian fit requires $\sigma \geq \sqrt{1/2}$ for the original data.

APPENDIX B

MATRICES

B.1 DETERMINANTS

In applying the method of least squares to both linear and nonlinear functions, we required the solution of a set of n simultaneous equations in n unknowns a_i similar to the following:

$$\begin{aligned} y_1 &= a_1 X_{11} + a_2 X_{12} + a_3 X_{13} \\ y_2 &= a_1 X_{21} + a_2 X_{22} + a_3 X_{23} \\ y_3 &= a_1 X_{31} + a_2 X_{32} + a_3 X_{33} \end{aligned} \tag{B.1}$$

where the constants y_i and X_{ij} are known quantities calculated from the data.

The symmetry of the right-hand side suggests that we write elements of the equations in a two-dimensional array

$$\boldsymbol{\alpha} = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix} \tag{B.2}$$

and separate the other terms and coefficients into one-dimensional arrays.

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{B.3}$$

Such arrays are called matrices, and we can write Equations (B.1) in matrix

Alternatively, because in our problems the matrix $\boldsymbol{\alpha}$ is always symmetric, that is, the element α_{ij} is equal to the element α_{ji}, we can write the matrices $\mathbf{a}$ and $\boldsymbol{\beta}$ as row matrices

$$a = [a_1 \quad a_2 \quad a_3] \quad \text{and} \quad b = [y_1 \quad y_2 \quad y_3] \tag{B.5}$$

and express Equation (B.1) as

$$\boldsymbol{\beta} = \mathbf{a} \cdot \boldsymbol{\alpha} \tag{B.6}$$

We shall be concerned primarily with linear one-dimensional matrices and with symmetric square two-dimensional matrices that have the same number of rows and columns and are mirror-symmetric about the diagonal. Consider a square matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2k} & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{j1} & A_{j2} & \cdots & A_{jk} & \cdots & A_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nk} & \cdots & A_{nn} \end{bmatrix} \tag{B.7}$$

The *degree* of the matrix $\mathbf{A}$ is the number n of rows and columns; the jkth *element* (or *component*) of the matrix is A_{jk}; the *diagonal terms* are A_{jj}. If the matrix is diagonally *symmetric*, $A_{jk} = A_{kj}$ and there are n^2 elements but only $n(n + 1)/2$ different elements.

Matrix Algebra

If $\mathbf{A}$ and $\mathbf{B}$ are two square symmetric matrices of degree n, then their sum $\mathbf{S}$ is a square symmetric matrix of degree n with elements that are the sums of the corresponding elements of the two matrices

$$\mathbf{A} + \mathbf{B} = \mathbf{S} \qquad S_{jk} = A_{jk} + B_{jk} \tag{B.8}$$

The product $\mathbf{P}$ of the matrices $\mathbf{A}$ and $\mathbf{B}$ is a square matrix of degree n, with elements determined in the following way:

$$\mathbf{AB} = \mathbf{P} \qquad P_{jk} = \sum_{m=1}^{n} (A_{jm} B_{mk}) \tag{B.9}$$

The elements of the jth row of $\mathbf{A}$ are multiplied by the elements of the kth column of $\mathbf{B}$ and the products are summed to obtain the jkth element of $\mathbf{P}$. In general, the matrix $\mathbf{P}$ will not be symmetric.

If $\mathbf{a}$ is a linear one-dimensional matrix, the product of $\mathbf{A}$ and $\mathbf{a}$ is only well de-

$$\begin{bmatrix} A_{11} & \cdots & & \cdots & A_{1n} \\ \vdots & \cdots & & \vdots & \vdots \\ A_{j1} & \cdots & A_{jk} & \cdots & A_{jn} \\ \vdots & \cdots & & \vdots & \vdots \\ A_{n1} & \cdots & & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_k \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \\ \vdots \\ c_n \end{bmatrix} \qquad c_j = \sum_{k=1}^{n} (A_{jk} a_k) \tag{B.10}$$

If **a** is a row matrix, it must multiply the square matrix on the left to yield another row matrix **r**.

$$[a_1 \quad \cdots \quad a_j \quad \cdots \quad a_n] \begin{bmatrix} A_{11} & \cdots & & \cdots & A_{1n} \\ \vdots & \cdots & & \cdots & \vdots \\ A_{j1} & \cdots & A_{jk} & \cdots & A_{jn} \\ \vdots & \cdots & & \cdots & \vdots \\ A_{n1} & \cdots & & \cdots & A_{nn} \end{bmatrix}$$

$$= [r_1 \quad \cdots \quad r_k \quad \cdots \quad r_n] \qquad r_j = \sum_{j=1}^{n} (a_j A_{jk}) \tag{B.11}$$

The product of two linear matrices depends on the order of multiplication. The product of a row matrix **a** times a column matrix **b** is a scalar. If the order is reversed, the result is a square matrix that is *diagonal;* that is, for which only the diagonal terms are nonzero:

$$[a_1 \quad \cdots \quad a_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum_{j=1}^{n} (a_j b_j)$$

$$\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} [a_1 \quad \cdots \quad a_n] = \begin{bmatrix} a_1 b_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & a_j b_j & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & \cdots & a_n b_n \end{bmatrix} \tag{B.12}$$

Determinants

The *determinant* of a square matrix is defined in terms of its algebra. The *order* of the determinant of a square matrix is equal to the degree n of the matrix. In this section, we shall mainly use determinants of order 3 as examples, although, unless otherwise specified, the comments apply to matrices of all orders. Manipulation of the rows may be substituted for columns throughout.

1. The determinant of the *unity matrix* is 1 where the unity matrix is defined as the diagonal matrix with all diagonal elements equal to 1:

2. If a column matrix of degree n is added to one column of a square matrix of degree n, the determinant of the result is the sum of the determinant of the original square matrix plus that of another square matrix obtained by substituting the column matrix for the modified column:

$$\begin{vmatrix} A_{11}+a_1 & A_{12} & A_{13} \\ A_{21}+a_2 & A_{22} & A_{23} \\ A_{31}+a_3 & A_{32} & A_{33} \end{vmatrix} = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} + \begin{vmatrix} a_1 & A_{12} & A_{13} \\ a_2 & A_{22} & A_{23} \\ a_3 & A_{32} & A_{33} \end{vmatrix} \tag{B.14}$$

3. If one column of a square matrix is multiplied by a scalar, the determinant of the result is the product of the scalar and the determinant of the original matrix:

$$\begin{vmatrix} cA_{11} & A_{12} & A_{13} \\ cA_{21} & A_{22} & A_{23} \\ cA_{31} & A_{32} & A_{33} \end{vmatrix} = c \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} \tag{B.15}$$

4. If two columns of a square matrix are interchanged, the determinant retains the same magnitude but changes sign:

$$\begin{vmatrix} A_{12} & A_{11} & A_{13} \\ A_{22} & A_{21} & A_{23} \\ A_{32} & A_{31} & A_{33} \end{vmatrix} = - \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} \tag{B.16}$$

5. The *minor* A^{jk} of an element A_{jk} of a square matrix of degree n is defined as the determinant of the square matrix of degree $n-1$ formed by removing the jth row and the kth column:

$$\mathbf{A} = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} \qquad A^{21} = \begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} \tag{B.17}$$

6. The *cofactor* $\mathrm{cof}(A_{jk})$ of an element A_{jk} of a square matrix of degree n is defined as the product of the minor and a phase factor:

$$\mathrm{cof}(A_{jk}) \equiv (-1)^{j+k} A^{jk} \tag{B.18}$$

7. With the preceding definitions 5 and 6, the determinant of a square matrix of degree n can be expressed in terms of cofactors of minors:

$$|\mathbf{A}| = \sum_{k=1}^{n} [A_{jk} \mathrm{cof}(A_{jk})] = \sum_{k=1}^{n} [(-1)^{j+k} A_{jk} A^{jk}] \tag{B.19}$$

Equation (B.19) is an iterative definition, because the cofactor is itself a determinant. The determinant of a matrix of degree 1, however, is equal to the single element of that matrix. The determinant of a square matrix of degree 2 is encountered often enough to make its explicit formula useful:

$$\begin{vmatrix} a & b \\ \end{vmatrix} = ad - bc \tag{B.20}$$

$$\begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = A_{11}\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{12}\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{13}\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix} \tag{B.21}$$

Computation

Matrix computation is generally simpler if we can manipulate matrices into diagonal form in which only the diagonal elements A_{jj} are nonzero. The determinant of a diagonal matrix is equal to the product of all the diagonal elements and the *trace* is their sum:

$$|\mathbf{A}_{\text{diag}}| = \prod_{j=1}^{n} A_{jj} \tag{B.22}$$

If we combine rules 2, 3, and 4 of the algebra for determinants, we can show that the determinant of a matrix is unchanged if the elements of any column, multiplied by an arbitrary scalar, are added to the elements of any other column. The determinant of the sum is equal to the sum of the two determinants, but one of these determinants has two identical columns except for a scalar factor that may be extracted, and is therefore equal to 0:

$$\begin{aligned} \begin{vmatrix} A_{11}+cA_{12} & A_{12} & A_{13} \\ A_{21}+cA_{22} & A_{22} & A_{23} \\ A_{31}+cA_{32} & A_{32} & A_{33} \end{vmatrix} &= \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} + c\begin{vmatrix} A_{12} & A_{12} & A_{13} \\ A_{22} & A_{22} & A_{23} \\ A_{32} & A_{32} & A_{33} \end{vmatrix} \\ &= |\mathbf{A}| \end{aligned} \tag{B.23}$$

Thus, it is possible to eliminate all elements except one from a row by successively subtracting one column, appropriately scaled, from each of the others. For example, if we perform the subtraction

$$A'_{jk} = A_{jk} - A_{1j}\frac{A_{j1}}{A_{11}} \tag{B.24}$$

on each row except the first, we eliminate all elements of the first column except A_{11} to obtain

$$\mathbf{A}' = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A'_{22} & A'_{23} \\ 0 & A'_{32} & A'_{33} \end{vmatrix} \tag{B.25}$$

Similarly, if we subsequently start with element A'_{22} and subtract an appropriately scaled second row from the rest of the rows,

$$A''_{jk} = A'_{jk} - A'_{2k}A'_{jk}/A'_{22} \tag{B.26}$$

all the elements of the second column vanish except A'_{22}:

$$\begin{vmatrix} A_{11} & 0 & A''_{13} \end{vmatrix}$$

Note that A'_{22} is not the original value A_{22}, but is modified as a result of the first subtraction.

By successively subtracting rows (or columns) scaled to their diagonal elements, we can produce a matrix that is diagonal. In practice, it is sufficient to eliminate only half of the nondiagonal elements so that all elements on one side of a diagonal are 0:

$$\begin{aligned} \mathbf{A} &= \begin{vmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{vmatrix} = \begin{vmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{vmatrix} \\ &= A_{11}A_{22}A_{33} \end{aligned} \tag{B.28}$$

B.2 SOLUTION OF SIMULTANEOUS EQUATIONS BY DETERMINANTS

Consider the following set of three equations in three coefficients a_1, a_2, and a_3. We shall consider the y_k and X_{jk} to be known quantities; that is, constants:

$$\begin{aligned} y_1 &= a_1X_{11} + a_2X_{12} + a_3X_{13} \\ y_2 &= a_1X_{21} + a_2X_{22} + a_3X_{23} \\ y_3 &= a_1X_{31} + a_2X_{32} + a_3X_{33} \end{aligned} \tag{B.29}$$

Let us consider the set of equations as if they were one matrix equation as in Equation (B.10):

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \tag{B.30}$$

with **a** and **y** represented by linear matrices and **X** represented by a square matrix. If we multiply the first equation of Equations (B.29) by the cofactor of X_{11} in the matrix of Equation (B.30), multiply the second equation by the cofactor of X_{21}, and multiply the third by the cofactor of X_{31}, then the sum of the three equations is an equation involving determinants according to Equation (B.18):

$$\begin{aligned} \begin{vmatrix} y_1 & X_{12} & X_{13} \\ y_2 & X_{22} & X_{23} \\ y_3 & X_{32} & X_{33} \end{vmatrix} &= a_1\begin{vmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{vmatrix} + a_2\begin{vmatrix} X_{12} & X_{12} & X_{13} \\ X_{22} & X_{22} & X_{23} \\ X_{32} & X_{32} & X_{33} \end{vmatrix} \\ &\quad + a_3\begin{vmatrix} X_{13} & X_{12} & X_{13} \\ X_{23} & X_{22} & X_{23} \\ X_{33} & X_{32} & X_{33} \end{vmatrix} \end{aligned} \tag{B.31}$$

The determinants in the two rightmost terms of Equation (B.31) both vanish

$$a_1 = \frac{\begin{vmatrix} y_1 & X_{12} & X_{13} \\ y_2 & X_{22} & X_{23} \\ y_3 & X_{32} & X_{33} \end{vmatrix}}{\begin{vmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{vmatrix}} \tag{B.32}$$

The denominator is the determinant of the square matrix **X** of Equation (B.30) and the numerator is the determinant of a matrix that is formed by substituting the column matrix y for the first column of the **X** matrix.

Similarly, *Cramér's rule* gives the solution for the jth coefficient a_j of a set of n simultaneous equations as the ratio of two determinants:

$$y_k = \sum_{j=1}^{n} (a_j X_{kj}) \qquad k = 1, n$$

$$a_j = \frac{|\mathbf{X}'(j)|}{|\mathbf{X}|} \tag{B.33}$$

The denominator is the determinant of the **X** matrix. The numerator $|\mathbf{X}'(j)|$ is the determinant of the matrix formed by substituting the **y** matrix for the jth column.

A matrix is singular if its determinant is 0. If the **X** matrix is singular, there is no solution for Equation (B.33). For example, if two of the n simultaneous equations are identical, except for a scale factor, there are really only $n - 1$ independent simultaneous equations, and therefore no solution for the n unknowns. In this case, the **X** matrix has two identical rows and therefore a 0 determinant.

Solution by Matrix Equations

Let us consider Equation (B.33) as if it were a matrix equation as in Equation (B.30). If the **X** matrix is square, we can consider the **y** and **a** linear matrices as either column matrices as in Equation (B.10) or row matrices as in Equation (B.11):

$$[y_k] = [a_j][X_{kj}] \tag{B.34}$$

If we could multiply this matrix by another matrix $\mathbf{X}'$ such that the right-hand side becomes just the linear matrix **a**, then we will have our solution for the coefficients a_j directly. The multiplication of matrices is associative; that is,

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} \tag{B.35}$$

Therefore, we require a matrix $\mathbf{X}'$ such that if it is multiplied by the matrix **X**, the result is the unity matrix:

$$[X_{kj}][X'_{kj}] = \mathbf{1} \tag{B.36}$$

The matrix $\mathbf{X}'$ that satisfies Equation (B.36) is called the inverse matrix $\mathbf{X}^{-1}$

$$[y_k][X_{jk}]^{-1} = [a_j]\mathbf{1} = [a_j] \tag{B.37}$$

We can express Equation (B.37) in more conventional form to give the solution for each of the coefficients a_j:

$$a_j = \sum_{k=1}^{n} (y_k X_{kj}^{-1}) \tag{B.38}$$

Thus, the solution for the n unknowns with n simultaneous equations is reduced to evaluating the elements of the inverse matrix $\mathbf{X}^{-1}$.

B.3 MATRIX INVERSION

The adjoint $\mathbf{A}^\dagger$ of a matrix **A** is defined as the matrix obtained by substituting for each element A_{jk} the cofactor of the transposed element A_{kj}:

$$A_{jk}^\dagger = \text{cof}(A_{kj}) \tag{B.39}$$

For a square symmetric matrix, the transposition makes no difference.

The inverse matrix $\mathbf{A}^{-1}$ defined in Equation (B.36) may be evaluated by dividing the adjoint matrix $\mathbf{A}^\dagger$ by the determinant of **A**:

$$A_{jk}^{-1} = \frac{A_{jk}^\dagger}{|\mathbf{A}|} \tag{B.40}$$

To show that this equality holds, we multiply both sides of Equation (B.40) by $|\mathbf{A}|\mathbf{A}$.

$$|\mathbf{A}|\mathbf{A}\mathbf{A}^{-1} = |\mathbf{A}|\mathbf{1} = \mathbf{A}\mathbf{A}^\dagger \tag{B.41}$$

Diagonal terms of the matrices in Equation (B.41) are equivalent to the formula of Equation (B.19) for evaluating the determinant:

$$|\mathbf{A}| = \sum_{k=1}^{n} (A_{jk} A_{jk}^\dagger) = \sum_{k=1}^{n} [A_{jk}\text{cof}(\mathrm{A}_{jk})] \tag{B.42}$$

Off-diagonal elements can be shown to vanish like those of the determinants of Equation (B.31). If the matrix **A** is singular (that is, if $|\mathbf{A}| = 0$), the inverse matrix $\mathbf{A}^{-1}$ does not exist and there is no solution to the matrix equation of Equation (B.34).

Gauss-Jordan Elimination

The formula of Equation (B.40) is generally too cumbersome for use in computing the inverse of a matrix. Instead, the Gauss-Jordan method of elimination is used to invert a matrix by building up the inverse matrix from a unity matrix while reducing the original matrix to unity.

Consider the inverse matrix $\mathbf{A}^{-1}$ as the ratio of the unity [illegible]

and adding the same rows scaled to the same constants), the ratio remains unchanged. If we perform the proper manipulation, we can change the denominator into the unity matrix; the numerator must then become equal to the inverse matrix $\mathbf{A}^{-1}$.

Let us write the 3×3 matrix $\mathbf{A}$ and the 3×3 unity matrix side by side and manipulate both to reduce the matrix $\mathbf{A}$ to the unity matrix. We start by using the formula of Equation (B.24) to eliminate the two off-diagonal elements of the first column:

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} - A_{12}\dfrac{A_{21}}{A_{11}} & A_{23} - A_{13}\dfrac{A_{21}}{A_{11}} \\ 0 & A_{32} - A_{12}\dfrac{A_{31}}{A_{11}} & A_{33} - A_{13}\dfrac{A_{31}}{A_{11}} \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ -\dfrac{A_{21}}{A_{11}} & 1 & 0 \\ -\dfrac{A_{21}}{A_{11}} & 0 & 1 \end{bmatrix} \tag{B.43}$$

Now, we divide the first row by A_{11} to get a diagonal element of

$$\begin{bmatrix} 1 & \dfrac{A_{12}}{A_{11}} & \dfrac{A_{13}}{A_{11}} \\ 0 & A_{22} - A_{12}\dfrac{A_{21}}{A_{11}} & A_{23} - A_{13}\dfrac{A_{21}}{A_{11}} \\ 0 & A_{32} - A_{12}\dfrac{A_{31}}{A_{11}} & A_{33} - A_{13}\dfrac{A_{31}}{A_{11}} \end{bmatrix} \begin{bmatrix} \dfrac{1}{A_{11}} & 0 & 0 \\ -\dfrac{A_{21}}{A_{11}} & 1 & 0 \\ -\dfrac{A_{21}}{A_{11}} & 0 & 1 \end{bmatrix} \tag{B.44}$$

The left matrix now has the proper first column. Let us relabel the matrices $\mathbf{B}$ (on the left) and $\mathbf{B}'$ (on the right) and perform the corresponding manipulations to obtain zeros in place of B_{12} and B_{32}, and then divide the second row by B_{22}:

$$\begin{bmatrix} 1 & 0 & B_{13} - B_{23}\dfrac{B_{12}}{B_{22}} \\ 0 & 1 & \dfrac{B_{23}}{B_{22}} \\ 0 & 0 & B_{33} - B_{23}\dfrac{B_{32}}{B_{22}} \end{bmatrix} \begin{bmatrix} B'_{11} - B'_{21}\dfrac{B_{12}}{B_{22}} & -\dfrac{B_{12}}{B_{22}} & 0 \\ \dfrac{B'_{21}}{B_{22}} & \dfrac{1}{B_{22}} & 0 \\ B'_{31} - B'_{21}\dfrac{B_{32}}{B_{22}} & -\dfrac{B_{32}}{B_{22}} & 1 \end{bmatrix} \tag{B.45}$$

After similar manipulation of the third column, the matrix on the left becomes the unity matrix and that on the right, therefore, must be the inverse matrix.

For computational purposes, even this method is somewhat inefficient in that two matrices must be manipulated throughout. Note, however, that at each stage of the reduction, there are only n (or three) useful columns of information in the two matrices. As each column is eliminated from the left matrix, the corresponding col-

(B.43), but instead of applying this formula to the first column, we divide the first column by $-A_{11}$ to get the first column on the right of Equation (B.43); the diagonal element must be divided twice to become $1/A_{11}$. Divide the rest of the first row by A_{11} to get the composite of the two matrices of Equation (B.44):

$$\begin{bmatrix} \dfrac{1}{A_{11}} & \dfrac{A_{12}}{A_{11}} & \dfrac{A_{13}}{A_{11}} \\ -\dfrac{A_{21}}{A_{11}} & A_{22} - A_{12}\dfrac{A_{21}}{A_{13}} & A_{23} - A_{13}\dfrac{A_{21}}{A_{11}} \\ -\dfrac{A_{21}}{A_{11}} & A_{32} - A_{12}\dfrac{A_{31}}{A_{11}} & A_{33} - A_{13}\dfrac{A_{31}}{A_{11}} \end{bmatrix} \tag{B.46}$$

A corresponding manipulation of the second column yields a matrix with the first two columns identical to those of the right side of Equation (B.45) whereas the last column is identical to that of the left side of Equation (B.45). Thus the inverse matrix is accumulated in the space vacated by the original matrix.

Computer Routine PROGRAM B.1 MATRIX (WEBSITE) includes two routines, MATINV and LINEARBYSQUARE. MATINV inverts a square matrix and calculates its determinant, substituting the inverted matrix into the same array as the original matrix.[1] Input variables are ARRAY, the matrix to be inverted, and NORDER, the order of its determinant.

The initial program loop iterates through the n columns of the matrix, reorganizing the matrix to get the largest element in the diagonal in order to reduce rounding errors and improve computational precision. The inversion procedure discussed above is then carried out and the determinant DET of the matrix is calculated from the diagonalized matrix. After inversion, the inverted matrix is stored back in ARRAY and the variable DET, the value of the determinant of the original matrix, is returned.

LINEARBYSQUARE multiplies a linear matrix (on the right) by a square matrix (on the left). For example, see Equation (B.30).

APPENDIX C

GRAPHS AND TABLES

The tables and graphs in this appendix are provided for easy reference. Computer routines for calculating several of the distributions and probability functions are listed in Appendix E. Routines are also available on the website for calculating probabilities.

C.1 GAUSSIAN PROBABILITY DISTRIBUTION

The probability density function $p_G(x; \mu, \sigma)$ for the Gaussian or normal error distribution is given by

$$p_G(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

If measurements of a quantity x are distributed in this manner around a mean μ with standard deviation σ, the probability $dP_G(x; \mu, \sigma)$ for observing a value of x, within an infinitesimally small interval dx, in a random sample measurement is given by

$$dP_G(x; \mu, \sigma) = p_G(x; \mu, \sigma)\, dx$$

Values of the probability density function $p_G(x; \mu, \sigma)$ are tabulated in Table

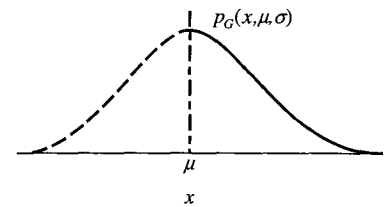

TABLE C.1
Gaussian probability density distribution. The Gaussian or normal error distribution $p_G(x; \mu, \sigma)$ versus $z = |x - \mu|/\sigma$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.39894	0.39892	0.39886	0.39876	0.39862	0.39844	0.39822	0.39797	0.39767	0.39733
0.1	0.39695	0.39654	0.39608	0.39559	0.39505	0.39448	0.39387	0.39322	0.39253	0.39181
0.2	0.39104	0.39024	0.38940	0.38853	0.38762	0.38667	0.38568	0.38466	0.38361	0.38251
0.3	0.38139	0.38023	0.37903	0.37780	0.37654	0.37524	0.37391	0.37255	0.37115	0.36973
0.4	0.36827	0.36678	0.36526	0.36371	0.36213	0.36053	0.35889	0.35723	0.35553	0.35381
0.5	0.35207	0.35029	0.34849	0.34667	0.34482	0.34294	0.34105	0.33912	0.33718	0.33521
0.6	0.33322	0.33121	0.32918	0.32713	0.32506	0.32297	0.32086	0.31874	0.31659	0.31443
0.7	0.31225	0.31006	0.30785	0.30563	0.30339	0.30114	0.29887	0.29659	0.29431	0.29200
0.8	0.28969	0.28737	0.28504	0.28269	0.28034	0.27799	0.27562	0.27324	0.27086	0.26848
0.9	0.26609	0.26369	0.26129	0.25888	0.25647	0.25406	0.25164	0.24923	0.24681	0.24439
1.0	0.24197	0.23995	0.23713	0.23471	0.23230	0.22988	0.22747	0.22506	0.22266	0.22025
1.1	0.21785	0.21546	0.21307	0.21069	0.20831	0.20594	0.20357	0.20122	0.19887	0.19652
1.2	0.19419	0.19186	0.18955	0.18724	0.18494	0.18265	0.18038	0.17811	0.17585	0.17361
1.3	0.17137	0.16915	0.16694	0.16475	0.16256	0.16039	0.15823	0.15609	0.15395	0.15184
1.4	0.14973	0.14764	0.14557	0.14351	0.14147	0.13944	0.13742	0.13543	0.13344	0.13148
1.5	0.12952	0.12759	0.12567	0.12377	0.12189	0.12002	0.11816	0.11633	0.11451	0.11271
1.6	0.11093	0.10916	0.10741	0.10568	0.10397	0.10227	0.10059	0.09893	0.09729	0.09567
1.7	0.09406	0.09247	0.09090	0.08934	0.08780	0.08629	0.08478	0.08330	0.08184	0.08039
1.8	0.07896	0.07755	0.07615	0.07477	0.07342	0.07207	0.07075	0.06944	0.06815	0.06688
1.9	0.06562	0.06439	0.06316	0.06196	0.06077	0.05960	0.05845	0.05731	0.05619	0.05509
2.0	0.05400	0.05293	0.05187	0.05083	0.04981	0.04880	0.04781	0.04683	0.04587	0.04492
2.1	0.04399	0.04307	0.04217	0.04129	0.04041	0.03956	0.03871	0.03788	0.03707	0.03627
2.2	0.03548	0.03471	0.03395	0.03320	0.03247	0.03175	0.03104	0.03034	0.02966	0.02899
2.3	0.02833	0.02769	0.02705	0.02643	0.02582	0.02522	0.02464	0.02406	0.02350	0.02294
2.4	0.02240	0.02187	0.02135	0.02083	0.02033	0.01984	0.01936	0.01889	0.01843	0.01798
2.5	0.01753	0.01710	0.01667	0.01626	0.01585	0.01545	0.01506	0.01468	0.01431	0.01394
2.6	0.01359	0.01324	0.01290	0.01256	0.01224	0.01192	0.01160	0.01130	0.01100	0.01071
2.7	0.01042	0.01015	0.00987	0.00961	0.00935	0.00910	0.00885	0.00861	0.00837	0.00814
2.8	0.00792	0.00770	0.00749	0.00728	0.00707	0.00688	0.00668	0.00649	0.00631	0.00613
2.9	0.00595	0.00578	0.00562	0.00546	0.00530	0.00514	0.00500	0.00485	0.00471	0.00457

	0.00	0.10	0.20	0.30	0.40
3.0	0.0044318	0.0032668	0.0023841	0.0017226	0.0012322
3.5	0.00087269	0.00061191	0.00042479	0.00029195	0.00019866
4.0	0.00013383	0.000089264	0.000058945	0.000038536	0.000024943
4.5	0.000015984	0.000010141	0.0000063701	0.0000039615	0.0000024391
5.0	0.0000014868	0.00000089730	0.00000053614	0.00000031716	0.00000018575
5.5	0.00000010771	0.00000006183	0.00000003514	0.00000001978	0.00000001102

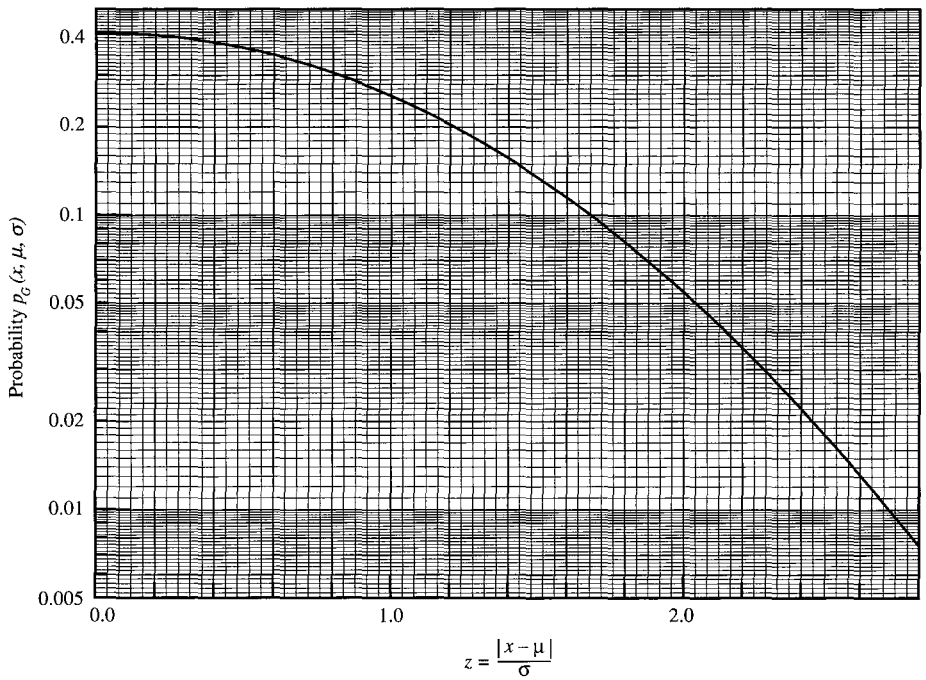

FIGURE C.1
Gaussian probability density distribution, $p_G(x; \mu, \sigma)$ versus $z = |x - \mu|/\sigma$

for z ranging from 0.0 to 3.0 in increments of 0.01 and up to 5.9 in increments of 0.1. This function is graphed on a semi-logarithmic scale as a function of z in Figure C.1.

C.2 INTEGRAL OF GAUSSIAN DISTRIBUTION

The integral $P_G(x; \mu, \sigma)$ of the probability density function $p_G(x; \mu, \sigma)$ for the Gaussian or normal error distribution is given by

$$P_G(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu - z\sigma}^{\mu + z\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

with

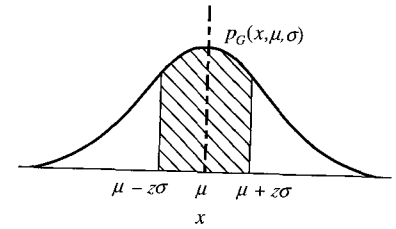

TABLE C.2
Integral of Gaussian distribution. The integral of the Gaussian probability density distribution. $P_G(x; \mu, \sigma)$ versus $z = |x - \mu|/\sigma$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0	0.00798	0.01596	0.02393	0.03191	0.03988	0.04784	0.05581	0.06376	0.07171
0.1	0.07966	0.08759	0.09552	0.10343	0.11134	0.11924	0.12712	0.13499	0.14285	0.15069
0.2	0.15852	0.16633	0.17413	0.18191	0.18967	0.19741	0.20514	0.21284	0.22052	0.22818
0.3	0.23582	0.24344	0.25103	0.25860	0.26614	0.27366	0.28115	0.28862	0.29605	0.30346
0.4	0.31084	0.31819	0.32551	0.33280	0.34006	0.34729	0.35448	0.36164	0.36877	0.37587
0.5	0.38292	0.38995	0.39694	0.40389	0.41080	0.41768	0.42452	0.43132	0.43809	0.44481
0.6	0.45149	0.45814	0.46474	0.47131	0.47783	0.48431	0.49075	0.49714	0.50350	0.50981
0.7	0.51607	0.52230	0.52847	0.53461	0.54070	0.54674	0.55274	0.55870	0.56461	0.57047
0.8	0.57629	0.58206	0.58778	0.59346	0.59909	0.60467	0.61021	0.61570	0.62114	0.62653
0.9	0.63188	0.63718	0.64243	0.64763	0.65278	0.65789	0.66294	0.66795	0.67291	0.67783
1.0	0.68269	0.68750	0.69227	0.69699	0.70166	0.70628	0.71085	0.71538	0.71985	0.72428
1.1	0.72866	0.73300	0.73728	0.74152	0.74571	0.74985	0.75395	0.75799	0.76199	0.76595
1.2	0.76985	0.77371	0.77753	0.78130	0.78502	0.78869	0.79232	0.79591	0.79945	0.80294
1.3	0.80639	0.80980	0.81316	0.81647	0.81975	0.82298	0.82616	0.82930	0.83240	0.83546
1.4	0.83848	0.84145	0.84438	0.84727	0.85012	0.85293	0.85570	0.85843	0.86112	0.86377
1.5	0.86638	0.86895	0.87148	0.87397	0.87643	0.87885	0.88123	0.88358	0.88588	0.88816
1.6	0.89039	0.89259	0.89476	0.89689	0.89898	0.90105	0.90308	0.90507	0.90703	0.90896
1.7	0.91086	0.91272	0.91456	0.91636	0.91813	0.91987	0.92158	0.92326	0.92491	0.92654
1.8	0.92813	0.92969	0.93123	0.93274	0.93422	0.93568	0.93711	0.93851	0.93988	0.94123
1.9	0.94256	0.94386	0.94513	0.94638	0.94761	0.94882	0.95000	0.95115	0.95229	0.95340
2.0	0.95449	0.95556	0.95661	0.95764	0.95864	0.95963	0.96059	0.96154	0.96247	0.96338
2.1	0.96426	0.96513	0.96599	0.96682	0.96764	0.96844	0.96922	0.96999	0.97074	0.97147
2.2	0.97219	0.97289	0.97358	0.97425	0.97490	0.97555	0.97617	0.97679	0.97739	0.97797
2.3	0.97855	0.97911	0.97965	0.98019	0.98071	0.98122	0.98172	0.98221	0.98268	0.98315
2.4	0.98360	0.98404	0.98448	0.98490	0.98531	0.98571	0.98610	0.98648	0.98686	0.98722
2.5	0.98758	0.98792	0.98826	0.98859	0.98891	0.98922	0.98953	0.98983	0.99012	0.99040
2.6	0.99067	0.99094	0.99120	0.99146	0.99171	0.99195	0.99218	0.99241	0.99264	0.99285
2.7	0.99306	0.99327	0.99347	0.99366	0.99385	0.99404	0.99422	0.99439	0.99456	0.99473
2.8	0.99489	0.99504	0.99520	0.99534	0.99549	0.99563	0.99576	0.99589	0.99602	0.99615
2.9	0.99627	0.99638	0.99650	0.99661	0.99672	0.99682	0.99692	0.99702	0.99712	0.99721

	0.00	0.10	0.20	0.30	0.40
3.0	0.9973002	0.9980648	0.9986257	0.99903315	0.99932614
3.5	0.99953474	0.99968178	0.99978440	0.99985530	0.999903805
4.0	0.999936656	0.999958684	0.999973308	0.999982920	0.999989174
4.5	0.9999932043	0.9999957748	0.9999973982	0.9999984132	0.99999904149
5.0	0.99999942657	0.99999966024			

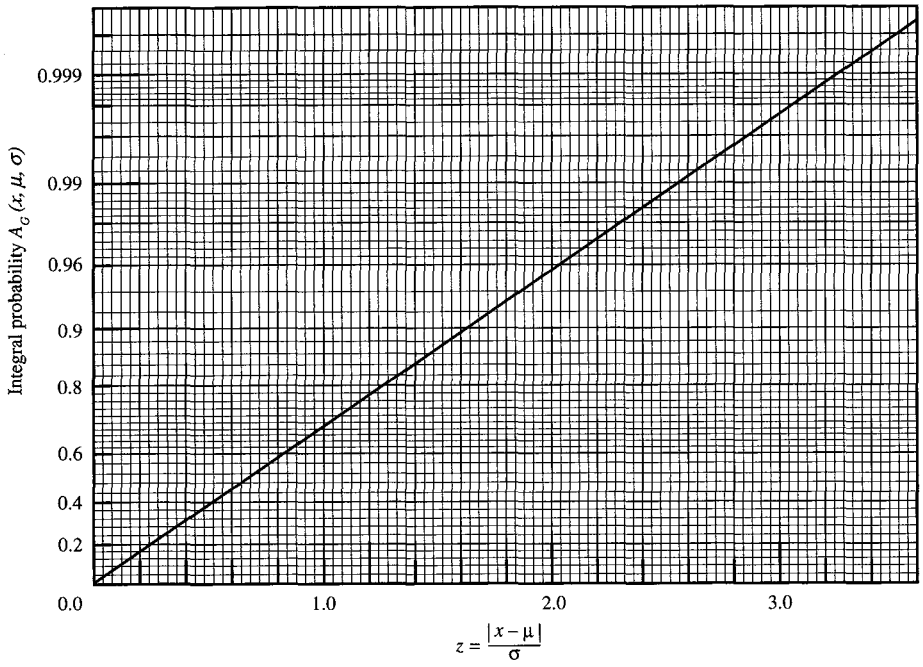

FIGURE C.2
Integral of the Gaussian probability density distribution, $p_G(x; \mu, \sigma)$ versus $z = |x - \mu|/\sigma$

If measurements of the quantity x are distributed according to the Gaussian distribution around a mean μ with standard deviation σ, $P_G(x; \mu, \sigma)$ is equal to the probability for observing a value of x in a random sample measurement that is between $\mu - z\sigma$ and $\mu + z\sigma$; that is, it is the probability that $|x - \mu| < z\sigma$.

Values of the integral $P_G(x; \mu, \sigma)$ are tabulated in Table C.2 as a function of z, for z ranging from 0.0 to 3.0 in increments of 0.01 and up to 5.9 in increments of 0.1. This function is graphed on a probability scale as a function of z in Figure C.2.

A related function is the *error function* erf Z:

$$\text{erf } Z = \frac{1}{\sqrt{\pi}} \int_{-Z}^{Z} e^{-z^2} dz = P_G(z\sqrt{2}; 0, 1)$$

The function that is tabulated and graphed is the shaded area between the limits $\mu \pm z\sigma$ as indicated.

C.3 LINEAR-CORRELATION COEFFICIENT

The probability distribution $p_r(r; \nu)$ for the linear correlation coefficient r for ν

$$p_r(r; \nu) = \frac{1}{\sqrt{\pi}} \frac{\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)} (1 - r^2)^{(\nu-2)/2}$$

The probability of observing a value of the correlation coefficient larger than r for a random sample of N observations with ν degrees of freedom is the integral of this probability $P_c(r; N)$:

$$P_c(r; N) = \frac{1}{\sqrt{\pi}} \frac{\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)} \int_{|r|}^{1} (1 - x^2)^{(\nu-2)/2} dx \qquad \nu = N - 2$$

If two variables of a parent population are uncorrelated, the probability that a random sample of N observations will yield a correlation coefficient for those two variables greater in magnitude than $|r|$ is given by $P_c(r; N)$.

Values of the coefficient $|r|$corresponding to various values of the probability $P_c(r; N)$ are tabulated in Table C.3 for N ranging from 3 to 100, and values of $P_c(r; N)$ ranging from 0.001 to 0.5. The functional dependence of r corresponding to representative values of $P_c(r, N)$ is graphed on a semi-logarithmic scale as a smooth variation with the number of observations N in Figure C.3.

The function that is tabulated and graphed is the shaded area under the tails of the probability curve for values larger than $|r|$ as indicated.

C.4 χ^2 DISTRIBUTION

The probability density distribution $p_\chi(\chi^2; \nu)$ for χ^2 is given by

$$p_\chi(\chi^2; \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} (\chi^2)^{(\nu-2)/2} e^{-\chi^2/2}$$

The probability of observing a value of χ^2 that is larger than a particular value for a random sample of N observations with ν degrees of freedom is the integral of this probability $P_\chi(\chi^2; \nu)$:

$$P_\chi(\chi^2; \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_{\chi^2}^{\infty} (x^2)^{(\nu-2)/2} e^{-x^2/2} d(x^2)$$

Values of the reduced chi-square $\chi^2_\nu = \chi^2/\nu$ corresponding to various values of the integral probability $P_\chi(\chi^2; \nu)$ of exceeding χ^2 in a measurement with ν degrees of freedom are tabulated in Table C.4 for ν ranging from 1 to 200. The functional dependence of $P_\chi(\chi^2; \nu)$ corresponding to representative values of ν is graphed in Figure C.4 as a smooth variation with the reduced chi-square χ^2_ν.

The function that is tabulated and

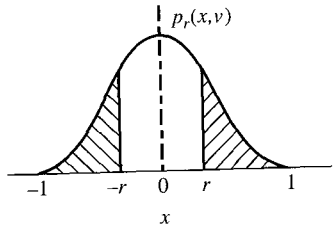

TABLE C.3
Linear-correlation coefficient. The linear-correlation coefficient r versus the number of observations N and the corresponding probability $P_c(r; N)$ of exceeding r in a random sample of observations taken from an uncorrelated parent population ($\rho = 0$)

	P								
N	**0.50**	**0.20**	**0.10**	**0.050**	**0.020**	**0.010**	**0.005**	**0.002**	**0.001**
3	0.707	0.951	0.988	0.997	1.000	1.000	1.000	1.000	1.000
4	0.500	0.800	0.900	0.950	0.980	0.990	0.995	0.998	0.999
5	0.404	0.687	0.805	0.878	0.934	0.959	0.974	0.986	0.991
6	0.347	0.608	0.729	0.811	0.882	0.917	0.942	0.963	0.974
7	0.309	0.551	0.669	0.754	0.833	0.875	0.906	0.935	0.951
8	0.281	0.507	0.621	0.707	0.789	0.834	0.870	0.905	0.925
9	0.260	0.472	0.582	0.666	0.750	0.798	0.836	0.875	0.898
10	0.242	0.443	0.549	0.632	0.715	0.765	0.805	0.847	0.872
11	0.228	0.419	0.521	0.602	0.685	0.735	0.776	0.820	0.847
12	0.216	0.398	0.497	0.576	0.658	0.708	0.750	0.795	0.823
13	0.206	0.380	0.476	0.553	0.634	0.684	0.726	0.772	0.801
14	0.197	0.365	0.458	0.532	0.612	0.661	0.703	0.750	0.780
15	0.189	0.351	0.441	0.514	0.592	0.641	0.683	0.730	0.760
16	0.182	0.338	0.426	0.497	0.574	0.623	0.664	0.711	0.742
17	0.176	0.327	0.412	0.482	0.558	0.606	0.647	0.694	0.725
18	0.170	0.317	0.400	0.468	0.543	0.590	0.631	0.678	0.708
19	0.165	0.308	0.389	0.456	0.529	0.575	0.616	0.662	0.693
20	0.160	0.299	0.378	0.444	0.516	0.561	0.602	0.648	0.679
22	0.152	0.284	0.360	0.423	0.492	0.537	0.576	0.622	0.652
24	0.145	0.271	0.344	0.404	0.472	0.515	0.554	0.599	0.629
26	0.138	0.260	0.330	0.388	0.453	0.496	0.534	0.578	0.607
28	0.133	0.250	0.317	0.374	0.437	0.479	0.515	0.559	0.588
30	0.128	0.241	0.306	0.361	0.423	0.463	0.499	0.541	0.570
32	0.124	0.233	0.296	0.349	0.409	0.449	0.484	0.526	0.554
34	0.120	0.225	0.287	0.339	0.397	0.436	0.470	0.511	0.539
36	0.116	0.219	0.279	0.329	0.386	0.424	0.458	0.498	0.525
38	0.113	0.213	0.271	0.320	0.376	0.413	0.446	0.486	0.513
40	0.110	0.207	0.264	0.312	0.367	0.403	0.435	0.474	0.501
42	0.107	0.202	0.257	0.304	0.358	0.393	0.425	0.463	0.490
44	0.104	0.197	0.251	0.297	0.350	0.384	0.416	0.453	0.479
46	0.102	0.192	0.246	0.291	0.342	0.376	0.407	0.444	0.469
48	0.100	0.188	0.240	0.285	0.335	0.368	0.399	0.435	0.460
50	0.098	0.184	0.235	0.279	0.328	0.361	0.391	0.427	0.451
60	0.089	0.168	0.214	0.254	0.300	0.330	0.358	0.391	0.414
70	0.082	0.155	0.198	0.235	0.278	0.306	0.332	0.363	0.385

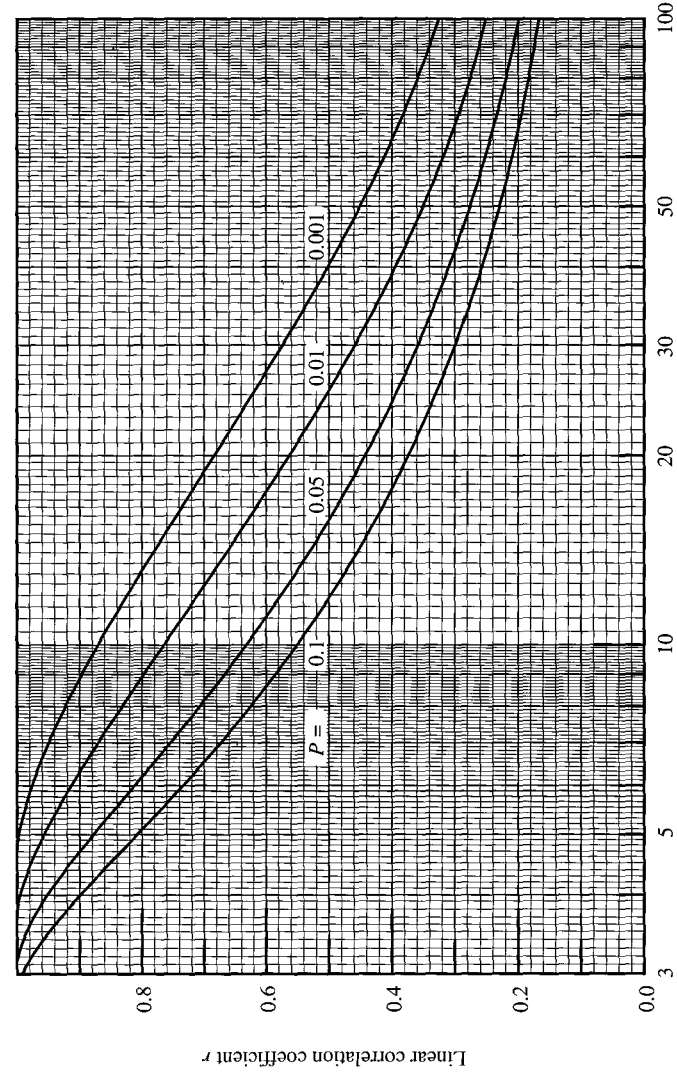

FIGURE C.3
The linear-correlation coefficient r versus the number of observations N and the corresponding probability $P_c(r; N)$ that the variables are not correlated.

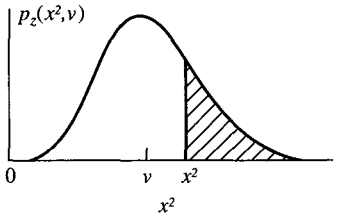

TABLE C.4

χ^2 distribution. Values of the reduced chi-square $\chi^2_\nu = \chi^2/\nu$ corresponding to the probability $P_\chi(\chi^2; \nu)$ of exceeding χ^2 versus the number of degrees of freedom ν

	P							
ν	0.99	0.98	0.95	0.90	0.80	0.70	0.60	0.50
1	0.00016	0.00063	0.00393	0.0158	0.0642	0.148	0.275	0.455
2	0.0100	0.0202	0.0515	0.105	0.223	0.357	0.511	0.693
3	0.0383	0.0617	0.117	0.195	0.335	0.475	0.623	0.789
4	0.0742	0.107	0.178	0.266	0.412	0.549	0.688	0.839
5	0.111	0.150	0.229	0.322	0.469	0.600	0.731	0.870
6	0.145	0.189	0.273	0.367	0.512	0.638	0.762	0.891
7	0.177	0.223	0.310	0.405	0.546	0.667	0.785	0.907
8	0.206	0.254	0.342	0.436	0.574	0.691	0.803	0.918
9	0.232	0.281	0.369	0.463	0.598	0.710	0.817	0.927
10	0.256	0.306	0.394	0.487	0.618	0.727	0.830	0.934
11	0.278	0.328	0.416	0.507	0.635	0.741	0.840	0.940
12	0.298	0.348	0.436	0.525	0.651	0.753	0.848	0.945
13	0.316	0.367	0.453	0.542	0.664	0.764	0.856	0.949
14	0.333	0.383	0.469	0.556	0.676	0.773	0.863	0.953
15	0.349	0.399	0.484	0.570	0.687	0.781	0.869	0.956
16	0.363	0.413	0.498	0.582	0.697	0.789	0.874	0.959
17	0.377	0.427	0.510	0.593	0.706	0.796	0.879	0.961
18	0.390	0.439	0.522	0.604	0.714	0.802	0.883	0.963
19	0.402	0.451	0.532	0.613	0.722	0.808	0.887	0.965
20	0.413	0.462	0.543	0.622	0.729	0.813	0.890	0.967
22	0.434	0.482	0.561	0.638	0.742	0.823	0.897	0.970
24	0.452	0.500	0.577	0.652	0.753	0.831	0.902	0.972
26	0.469	0.516	0.592	0.665	0.762	0.838	0.907	0.974
28	0.484	0.530	0.605	0.676	0.771	0.845	0.911	0.976
30	0.498	0.544	0.616	0.687	0.779	0.850	0.915	0.978
32	0.511	0.556	0.627	0.696	0.786	0.855	0.918	0.979
34	0.523	0.567	0.637	0.704	0.792	0.860	0.921	0.980
36	0.534	0.577	0.646	0.712	0.798	0.864	0.924	0.982
38	0.545	0.587	0.655	0.720	0.804	0.868	0.926	0.983
40	0.554	0.596	0.663	0.726	0.809	0.872	0.928	0.983
42	0.563	0.604	0.670	0.733	0.813	0.875	0.930	0.984
44	0.572	0.612	0.677	0.738	0.818	0.878	0.932	0.985
46	0.580	0.620	0.683	0.744	0.822	0.881	0.934	0.986
48	0.587	0.627	0.690	0.749	0.825	0.884	0.936	0.986
50	0.594	0.633	0.695	0.754	0.829	0.886	0.937	0.987
60	0.625	0.662	0.720	0.774	0.844	0.897	0.944	0.989
70	0.649	0.684	0.739	0.790	0.856	0.905	0.949	0.990
80	0.669	0.703	0.755	0.803	0.865	0.911	0.952	0.992
90	0.686	0.718	0.768	0.814	0.873	0.917	0.955	0.993
100	0.701	0.731	0.779	0.824	0.879	0.921	0.958	0.993
120	0.724	0.753	0.798	0.839	0.890	0.928	0.962	0.994
140	0.743	0.770	0.812	0.850	0.898	0.934	0.965	0.995

TABLE C.4
(continued)

	P							
ν	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	0.708	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	0.916	1.204	1.609	2.303	2.996	3.912	4.605	6.908
3	0.982	1.222	1.547	2.084	2.605	3.279	3.780	5.423
4	1.011	1.220	1.497	1.945	2.372	2.917	3.319	4.617
5	1.026	1.213	1.458	1.847	2.214	2.678	3.017	4.102
6	1.035	1.205	1.426	1.774	2.099	2.506	2.802	3.743
7	1.040	1.198	1.400	1.717	2.010	2.375	2.639	3.475
8	1.044	1.191	1.379	1.670	1.938	2.271	2.511	3.266
9	1.046	1.184	1.360	1.632	1.880	2.187	2.407	3.097
10	1.047	1.178	1.344	1.599	1.831	2.116	2.321	2.959
11	1.048	1.173	1.330	1.570	1.789	2.056	2.248	2.842
12	1.049	1.168	1.318	1.546	1.752	2.004	2.185	2.742
13	1.049	1.163	1.307	1.524	1.720	1.959	2.130	2.656
14	1.049	1.159	1.296	1.505	1.692	1.919	2.082	2.580
15	1.049	1.155	1.287	1.487	1.666	1.884	2.039	2.513
16	1.049	1.151	1.279	1.471	1.644	1.852	2.000	2.453
17	1.048	1.148	1.271	1.457	1.623	1.823	1.965	2.399
18	1.048	1.145	1.264	1.444	1.604	1.797	1.934	2.351
19	1.048	1.142	1.258	1.432	1.586	1.773	1.905	2.307
20	1.048	1.139	1.252	1.421	1.571	1.751	1.878	2.266
22	1.047	1.134	1.241	1.401	1.542	1.712	1.831	2.194
24	1.046	1.129	1.231	1.383	1.517	1.678	1.791	2.132
26	1.045	1.125	1.223	1.368	1.496	1.648	1.755	2.079
28	1.045	1.121	1.215	1.354	1.476	1.622	1.724	2.032
30	1.044	1.118	1.208	1.342	1.459	1.599	1.696	1.990
32	1.043	1.115	1.202	1.331	1.444	1.578	1.671	1.953
34	1.042	1.112	1.196	1.321	1.429	1.559	1.649	1.919
36	1.042	1.109	1.191	1.311	1.417	1.541	1.628	1.888
38	1.041	1.106	1.186	1.303	1.405	1.525	1.610	1.861
40	1.041	1.104	1.182	1.295	1.394	1.511	1.592	1.835
42	1.040	1.102	1.178	1.288	1.384	1.497	1.576	1.812
44	1.039	1.100	1.174	1.281	1.375	1.485	1.562	1.790
46	1.039	1.098	1.170	1.275	1.366	1.473	1.548	1.770
48	1.038	1.096	1.167	1.269	1.358	1.462	1.535	1.751
50	1.038	1.094	1.163	1.263	1.350	1.452	1.523	1.733
60	1.036	1.087	1.150	1.240	1.318	1.410	1.473	1.660
70	1.034	1.081	1.139	1.222	1.293	1.377	1.435	1.605
80	1.032	1.076	1.130	1.207	1.273	1.351	1.404	1.560
90	1.031	1.072	1.123	1.195	1.257	1.329	1.379	1.525
100	1.029	1.069	1.117	1.185	1.243	1.311	1.358	1.494
120	1.027	1.063	1.107	1.169	1.221	1.283	1.325	1.446
140	1.026	1.059	1.099	1.156	1.204	1.261	1.299	1.410
160	1.024	1.055	1.093	1.146	1.191	1.243	1.278	1.381
180	1.023	1.052	1.087	1.137	1.179	1.228	1.261	1.358
200	1.022	1.050	1.083	1.130	1.170	1.216	1.247	1.338

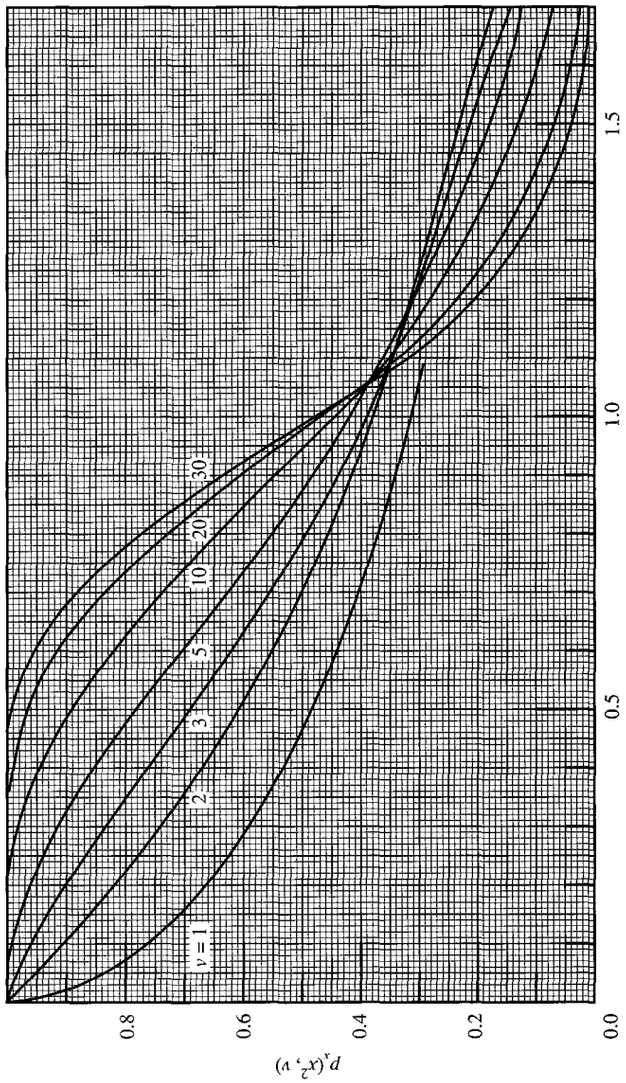

FIGURE C.4
The probability $P_\chi(\chi^2; \nu)$ of exceeding χ^2 versus the reduced chi-square $\chi^2_\nu = \chi^2/\nu$ and the number of degrees of freedom ν.

C.5 *F* DISTRIBUTION

The probability distribution for F is given by

$$p_f(f, \nu_1, \nu_2) = \frac{\Gamma[(\nu_1 + \nu_2)/2]}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{f^{(\nu_1-2)/2}}{(1 + f\nu_1/\nu_2)^{1/2(\nu_1+\nu_2)}}$$

The probability of observing a value of F that is larger than a particular value for a random sample with ν_1 and ν_2 degrees of freedom is the integral of this probability:

$$P_F(F; \nu_1, \nu_2) = \int_F^\infty p_f(f; \nu_1, \nu_2)\, df$$

Values of F corresponding to various values of the integral probability $P_F(F; \nu_1, \nu_2)$ of exceeding F in a measurement are tabulated in Table C.5 for $\nu_1 = 1$ and graphed in Figure C.5 as a smooth variation with the probability. Values of F corresponding to various values of ν_1 and ν_2 ranging from 1 to ∞ are listed in Table C.6 and graphed in Figure C.6 for $P_F(F; \nu_1, \nu_2) = 0.05$ and in Table C.7 and Figure C.7 for $P_F(F; \nu_1, \nu_2) = 0.01$. These values were adapted by permission from Dixon and Massey (1969).

The function that is tabulated and graphed is the shaded area under the tail of the probability curve for values larger than F as indicated.

C.6 STUDENT'S *t* DISTRIBUTION

The probability distribution for Student's t is given by[1]

$$f(t, \nu) = \frac{1}{\sqrt{(\nu\pi)}} \frac{\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

Student's t distribution describes, as a function of the number of degrees of freedom ν, the distribution of the parameter $t = |x - \bar{x}|/s_\mu$, where t is the number of standard deviations s_μ of the sample distribution by which x differs from $\bar{x}$. This distribution takes account of the fact that the sample standard deviation s_μ is an *estimate* of the parent standard error σ_μ and, as such, will vary for different samples drawn from the same parent distribution, just as the sample means vary. If $\bar{x}$ represents the mean of N numbers and x is not derived from the data, then $\nu = N - 1$. If both x and $\bar{x}$ are means, s_μ must be the joint standard deviation of both x and $\bar{x}$, and ν must be the total number of degrees of freedom. In the limit of large numbers of degrees of freedom, Student's t and Gaussian probability distributions agree; for small ν, that is, low-statistics experiments, the Gaussian distribution overestimates the probability and Student's t is preferred.

Table C.8 lists probabilities obtained by integrating Student's t distribution from $x = \bar{x} - ts_\mu$ to $x = \bar{x} + ts_\mu$ where $t = |\bar{x} - x|/s_\mu$. The integrals are listed as functions of t and of the number of degrees of freedom ν. The values corresponding to Gaussian probability (which are independent of ν) are listed in the last column.

[1] "Review of Particle Physics" *The European Physical Journal C,* vol. 15 (2000), p. 193.

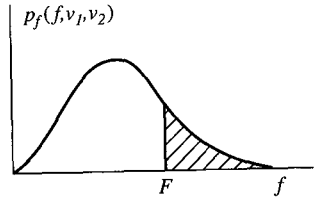

TABLE C.5
F distribution, $\nu = 1$. Values of F corresponding to the probability $P_F(F;1, \nu_2)$ of exceeding F (with $\nu_1 = 1$ degrees of freedom) versus the larger number of degrees of freedom ν_2.

Degrees of freedom	Probability (P) of exceeding F							
ν_2	**0.50**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
1	1.000	5.83	39.90	161.00	648.00	4050.00	16200.00	406000.0
2	0.667	2.57	8.53	18.50	38.50	98.50	198.00	998.0
3	0.585	2.02	5.54	10.10	17.40	34.10	55.60	167.0
4	0.549	1.81	4.54	7.71	12.20	21.20	31.30	74.1
5	0.528	1.69	4.06	6.61	10.00	16.30	22.80	47.2
6	0.515	1.62	3.78	5.99	8.81	13.70	18.60	35.5
7	0.506	1.57	3.59	5.59	8.07	12.20	16.20	29.2
8	0.499	1.54	3.46	5.32	7.57	11.30	14.70	25.4
9	0.494	1.51	3.36	5.12	7.21	10.60	13.60	22.9
10	0.490	1.49	3.28	4.96	6.94	10.00	12.80	21.0
11	0.486	1.47	3.23	4.84	6.72	9.65	12.20	19.7
12	0.484	1.46	3.18	4.75	6.55	9.33	11.80	18.6
15	0.478	1.43	3.07	4.54	6.20	8.68	10.80	16.6
20	0.472	1.40	2.97	4.35	5.87	8.10	9.94	14.8
24	0.469	1.39	2.93	4.26	5.72	7.82	9.55	14.0
30	0.466	1.38	2.88	4.17	5.57	7.56	9.18	13.3
40	0.463	1.36	2.84	4.08	5.42	7.31	8.83	12.6
60	0.461	1.35	2.79	4.00	5.29	7.08	8.49	12.0
120	0.458	1.34	2.75	3.92	5.15	6.85	8.18	11.4
∞	0.455	1.32	2.71	3.84	5.02	6.63	7.88	10.8

Note: For larger values of the probability P, the value of F is approximately $F \simeq [1.25(1 - P)]^2$.

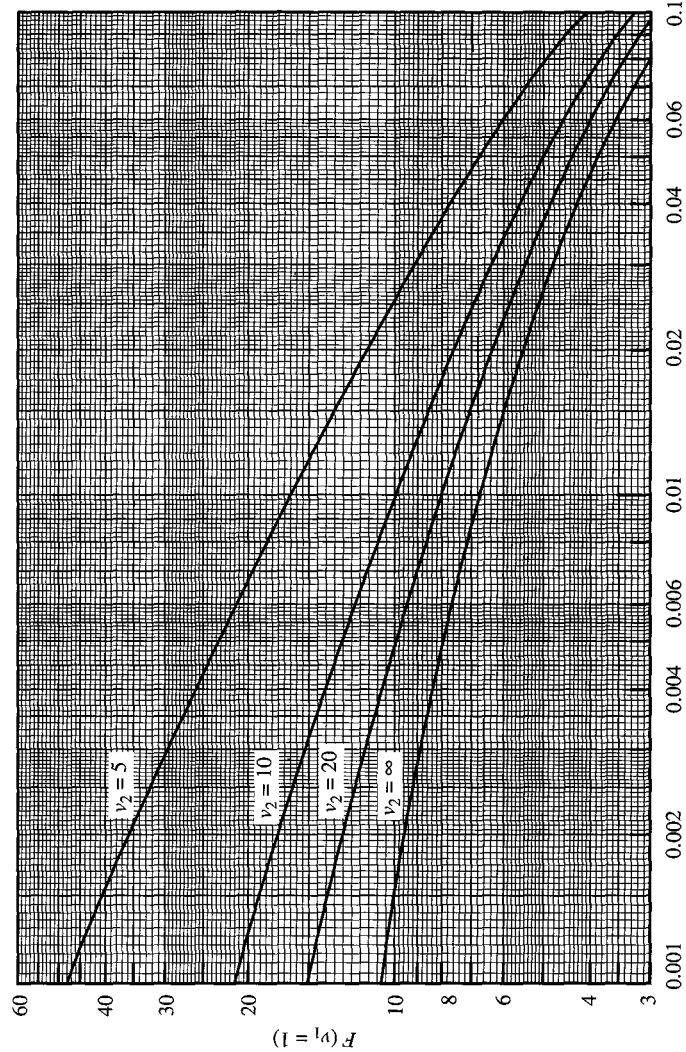

FIGURE C.5
The probability $P_F(F; 1, \nu_2)$ of exceeding F versus F and ν_2 for $\nu_1 = 1$.

TABLE C.6
F distribution, 5%. Values of F corresponding to the probability $P_F(F; \nu_1, \nu_2) = 0.05$ of exceeding F for ν_1 versus ν_2 degrees of freedom

Degrees of freedom ν_2	Degrees of freedom ν_1							
	2	4	6	8	10	15	20	100
1	200.00	225.00	234.00	239.00	242.00	246.00	248.00	253.00
2	19.00	19.20	19.30	19.40	19.40	19.40	19.40	19.50
3	9.55	9.12	8.94	8.85	8.79	8.70	8.66	8.55
4	6.94	6.39	6.16	6.04	5.96	5.86	5.80	5.66
5	5.79	5.19	4.95	4.82	4.73	4.62	4.56	4.41
6	5.14	4.53	4.28	4.15	4.60	3.94	3.87	3.71
7	4.74	4.12	3.87	3.73	3.64	3.51	3.44	3.27
8	4.46	3.84	3.58	3.44	3.35	3.22	3.15	2.97
9	4.26	3.63	3.37	3.23	3.14	3.01	2.94	2.76
10	4.10	3.48	3.22	3.07	2.98	2.85	2.77	2.59
11	3.98	3.36	3.09	2.95	2.85	2.72	2.65	2.46
12	3.89	3.26	3.00	2.85	2.75	2.62	2.54	2.35
15	3.68	3.06	2.79	2.64	2.54	2.40	2.33	2.12
20	3.49	2.87	2.60	2.45	2.35	2.20	2.12	1.91
24	3.40	2.78	2.51	2.36	2.25	2.11	2.03	1.80
30	3.32	2.69	2.42	2.27	2.16	2.01	1.93	1.70
40	3.23	2.61	2.34	2.18	1.08	1.92	1.84	1.59
60	3.15	2.53	2.25	2.10	1.99	1.84	1.75	1.48
120	3.07	2.45	2.18	2.02	1.91	1.75	1.66	1.37
∞	3.00	2.37	2.10	1.94	1.83	1.67	1.57	1.24

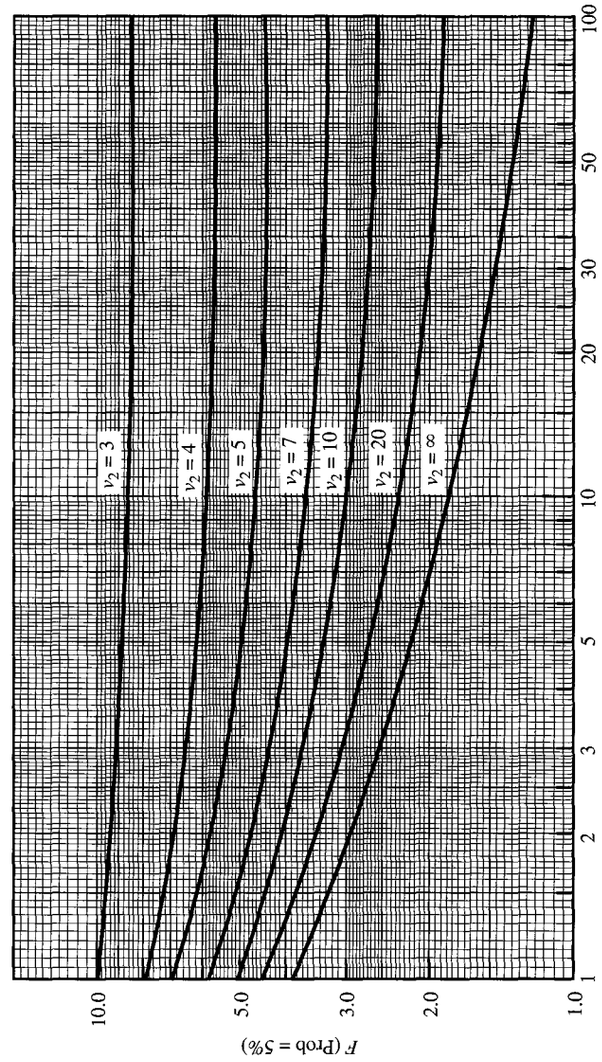

FIGURE C.6
Test values of $F(\nu_1, \nu_2)$ versus the numbers of degrees of freedom ν_1 and ν_2 for a probability $P_F(F; \nu_1, \nu_2) = 0.05$ of exceeding F.

TABLE C.7
F distribution, 1%. Values of F corresponding to the probability $P_F(F; \nu_1, \nu_2) = 0.01$ of exceeding F for ν_1 versus ν_2 degrees of freedom

Degrees of freedom ν_2	Degrees of freedom ν_1							
	2	4	6	8	10	15	20	100
1	5000.00	5620.00	5860.00	5980.00	6060.00	6160.00	6210.00	6330.00
2	99.00	99.20	99.30	99.40	99.40	99.40	99.40	99.50
3	30.80	28.70	27.90	27.50	27.20	26.90	26.70	26.20
4	18.00	16.00	15.20	14.80	14.50	14.20	14.00	13.60
5	13.30	11.40	10.70	10.30	10.10	9.72	9.55	9.13
6	10.90	9.15	8.47	8.10	7.87	7.56	7.40	6.99
7	9.55	7.85	7.19	6.84	6.62	6.31	6.16	5.75
8	8.65	7.01	6.37	6.03	5.81	5.52	5.36	4.96
9	8.02	6.42	5.80	5.47	5.26	4.96	4.81	4.42
10	7.56	5.99	5.39	5.06	4.85	4.56	4.41	4.01
11	7.21	5.67	5.07	4.74	4.54	4.25	4.10	3.71
12	6.93	5.41	4.82	4.50	4.30	4.01	3.86	3.47
15	6.36	4.89	4.32	4.00	3.80	3.52	3.37	2.98
20	5.85	4.43	3.87	3.56	3.37	3.09	2.94	2.54
24	5.61	4.22	3.67	3.36	3.17	2.89	2.74	2.33
30	5.39	4.02	3.47	3.17	2.98	2.70	2.55	2.13
40	5.18	3.83	3.29	2.99	2.80	2.52	2.37	1.94
60	4.98	3.65	3.12	2.82	2.63	2.35	2.20	1.75
120	4.79	3.48	2.96	2.66	2.47	2.19	2.03	1.56
∞	4.61	3.32	2.80	2.51	2.32	2.04	1.88	1.36

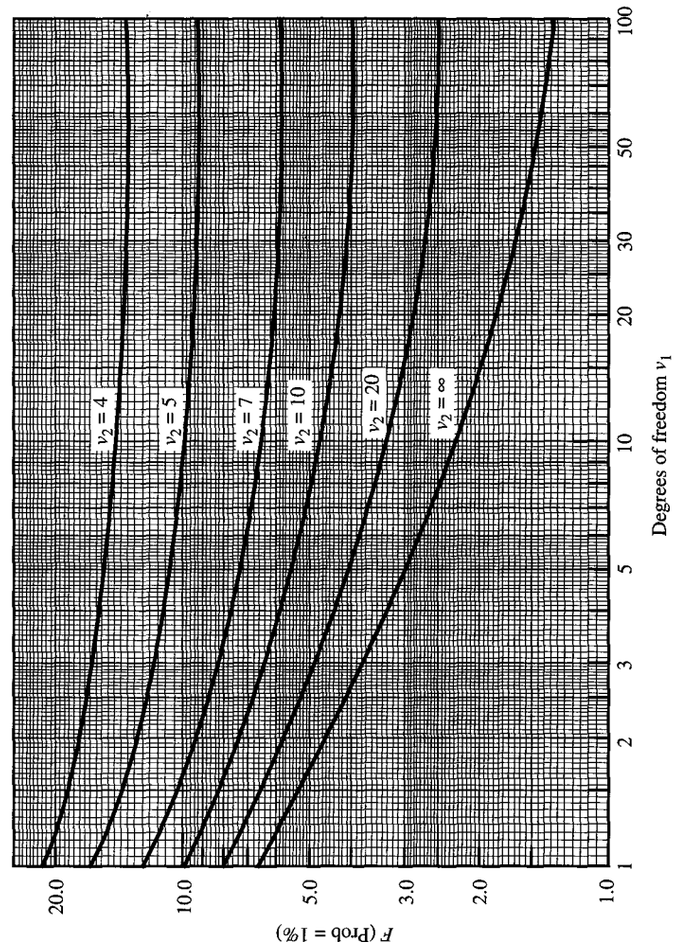

FIGURE C.7
Test values of $F(\nu_1, \nu_2)$ versus the numbers of degrees of freedom ν_1 and ν_2 for a probability $P_F(F; \nu_1, \nu_2) = 0.01$ of exceeding F.

TABLE C.8

$P_T(x; \mu, \sigma)$ versus $t = |x - \mu|/\sigma$; Integral of Student's t distribution between $x = \bar{x} - ts_\mu$ and $\bar{x} + ts_\mu$, expressed in percent.

	$\nu = N - 1$															
t	2	3	4	5	6	8	10	12	16	20	25	30	35	40	50	Gaussian probability
0.6	39.1	40.9	41.9	42.5	43.0	43.5	43.8	44.0	44.3	44.5	44.6	44.7	44.8	44.8	44.9	45.1
0.7	44.4	46.6	47.8	48.5	49.0	49.6	50.0	50.3	50.6	50.8	51.0	51.1	51.1	51.2	51.3	51.6
0.8	49.3	51.8	53.2	54.0	54.6	55.3	55.8	56.1	56.5	56.7	56.9	57.0	57.1	57.2	57.3	57.6
0.9	53.7	56.6	58.1	59.1	59.7	60.6	61.1	61.4	61.9	62.1	62.3	62.5	62.6	62.7	62.8	63.2
1.0	57.8	60.9	62.6	63.7	64.4	65.3	65.9	66.3	66.8	67.1	67.3	67.5	67.6	67.7	67.8	68.3
1.1	61.4	64.8	66.7	67.9	68.7	69.7	70.3	70.7	71.2	71.6	71.8	72.0	72.1	72.2	72.3	72.9
1.2	64.7	68.4	70.4	71.6	72.5	73.6	74.2	74.7	75.2	75.6	75.9	76.0	76.2	76.3	76.4	77.0
1.3	67.7	71.6	73.7	75.0	75.9	77.0	77.7	78.2	78.8	79.2	79.5	79.7	79.8	79.9	80.0	80.6
1.4	70.4	74.4	76.6	78.0	78.9	80.1	80.8	81.3	81.9	82.3	82.6	82.8	83.0	83.1	83.2	83.8
1.5	72.8	77.0	79.2	80.6	81.6	82.8	83.6	84.1	84.7	85.1	85.4	85.6	85.7	85.9	86.0	86.6
1.6	75.0	79.2	81.5	83.0	83.9	85.2	85.9	86.4	87.1	87.5	87.8	88.0	88.1	88.3	88.4	89.0
1.7	76.9	81.3	83.6	85.0	86.0	87.3	88.0	88.5	89.2	89.5	89.9	90.1	90.2	90.3	90.5	91.1
1.8	78.7	83.1	85.4	86.8	87.8	89.1	89.8	90.3	90.9	91.3	91.6	91.8	92.0	92.1	92.2	92.8
1.9	80.2	84.7	87.0	88.4	89.4	90.6	91.3	91.8	92.4	92.8	93.1	93.3	93.4	93.5	93.7	94.3
2.0	81.7	86.1	88.4	89.8	90.8	92.0	92.7	93.1	93.7	94.1	94.4	94.5	94.7	94.8	94.9	95.4
2.1	83.0	87.4	89.7	91.0	92.0	93.1	93.8	94.3	94.8	95.1	95.4	95.6	95.7	95.8	95.9	96.4
2.2	84.1	88.5	90.8	92.1	93.0	94.1	94.8	95.2	95.7	96.0	96.3	96.4	96.6	96.6	96.8	97.2
2.3	85.2	89.5	91.7	93.0	93.9	95.0	95.6	96.0	96.5	96.8	97.0	97.1	97.3	97.3	97.4	97.9
2.4	86.2	90.4	92.6	93.9	94.7	95.7	96.3	96.7	97.1	97.4	97.6	97.7	97.8	97.9	98.0	98.4
2.5	87.1	91.3	93.3	94.6	95.4	96.3	96.9	97.2	97.6	97.9	98.1	98.2	98.3	98.3	98.4	98.8
2.6	87.9	92.0	94.0	95.2	95.9	96.8	97.4	97.7	98.1	98.3	98.5	98.6	98.6	98.7	98.8	99.1
2.7	88.6	92.6	94.6	95.7	96.5	97.3	97.8	98.1	98.4	98.6	98.8	98.9	98.9	99.0	99.1	99.3
2.8	89.3	93.2	95.1	96.2	96.9	97.7	98.1	98.4	98.7	98.9	99.0	99.1	99.2	99.2	99.3	99.5
2.9	89.9	93.8	95.6	96.6	97.3	98.0	98.4	98.7	99.0	99.1	99.2	99.3	99.4	99.4	99.5	99.6
3.0	90.5	94.3	96.0	97.0	97.6	98.3	98.7	98.9	99.2	99.3	99.4	99.5	99.5	99.5	99.6	99.7
3.2	91.5	95.1	96.7	97.6	98.2	98.7	99.1	99.2	99.4	99.6	99.6	99.7	99.7	99.7	99.8	99.9
3.4	92.4	95.8	97.3	98.1	98.6	99.1	99.3	99.5	99.6	99.7	99.8	99.8	99.8	99.9	99.9	99.9
3.6	93.1	96.3	97.7	98.5	98.9	99.3	99.5	99.6	99.8	99.8	99.9	99.9	99.9	99.9	99.9	100.0
3.8	93.7	96.8	98.1	98.8	99.1	99.5	99.7	99.8	99.8	99.9	99.9	99.9	99.9	100.0	100.0	100.0
4.0	94.3	97.2	98.4	99.0	99.3	99.6	99.8	99.8	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0

Note: The Gaussian probability for each value of t is listed in the last column.

APPENDIX D

HISTOGRAMS AND GRAPHS

Graphs of experimental data and of theoretical predictions have always been important tools for scientists, in both the actual performance of research and in presentations of results. In recent years we have seen a proliferation of graphics displays as fast inexpensive computers and printers have facilitated the display-making process. Scientists have benefited from the new techniques and equipment, with many excellent commercial programs available for creating high-quality scientific graphics suitable for publication.

In science, the object is to present results in a straightforward manner so that relevant points are illustrated clearly and without bias. Graphs with suppressed zeros, which are common in advertisements, are not often seen in scientific papers. Bar graphs tend to be simple histograms rather than the multibar, brightly colored displays of magazines and newspapers. In fact, although the use of color is growing, especially in direct publication on the Internet, few scientific preprints and papers are printed in color, although discrete use can clarify graphical presentations significantly. Error bars, which are rare indeed in advertisements, are essential in a scientific presentation. Exaggerated perspective and distorted scales have very limited use in scientific work whereas semilogarithmic plots that are often used in science are not often seen in business publications.

It is often convenient to have graphics routines that are part of a simulation or an analysis program, rather than to use a separate graphing program. For example, in a Monte Carlo simulation, it is essential to be able to produce histograms and data graphs quickly at each stage of the study. Generations of scientists have made simple histograms on monitors or printers to make preliminary [illegible]

More elegant and detailed graphs can be created by using the graphics features of particular programming languages, and those provided by data analysis programs and spreadsheets. Such programs can produce high-quality graphs and charts suitable for presentations and publications. Many of the graphs in this book, such as those in Chapter 2, were created by programs written in Fortran and Pascal. Others, such as those in Chapter 11, were created in Origin, a very powerful data analysis program with strong graphing features.

D.1 MAKING A GRAPH

Whether a scientific graph is produced by hand or by computer, there are several basic principles that should be followed. The graph should be large enough to be read and understood easily, with appropriately proportioned abscissa and ordinate. Axes should be labeled with large, clean letters, and the axes scales should be clearly indicated. If more than a single function is displayed, or if both data and curves are displayed, a box, or legend, may be superimposed on the graph to indicate the meaning of different symbols. In scientific journals, a description of the graph is generally included as text below the abscissa label. In internal papers and preprints, these descriptions are often collected in a separate section of the paper. For visual presentation, some descriptive material may be included in a box on the graph, but it is important that text be large enough to be clearly legible. One should avoid scattering too much material over any graph, which gives a busy appearance. A properly made graph should not require many words of explanation.

It is generally advisable to plot the independent variable as the abscissa and the dependent variable as the ordinate. However, if the independent data have a high degree of uncertainty while the corresponding measurements of the dependent data can be made with high precision, then it might be wise to interchange the two axes to simplify least-squares fitting.

Reasonable, convenient values and intervals should be chosen for the scale marks on the two axes. For example, if abscissa values range from 0 to 400, it might be reasonable to divide the x-axis into eight parts and thus to mark the abscissa with major, labeled ticks at 0, 100, 200, 300, and 400, with minor ticks half-way between. Dividing the axis into six parts and putting ticks at 66.7, 133.3, and so forth, would make it very difficult for a reader to interpret.

In general, error bars should be included for ordinate variables except for simple histograms where the text clearly specifies that the uncertainties are statistical and therefore given by the square root of the value of the coordinate. Unless otherwise noted, error bars generally indicate the standard deviation. Error bars usually are not necessary for abscissa variables. However, if appropriate, they may be drawn to indicate the resolution of the measurement or setting, or they may simply indicate the range of the variable over which data have been collected or grouped, as in the case of the width of a histogram bin. The text must explain the meaning of such error bars. If no error bar is shown for the abscissa, then it is useful to draw a circle or other symbol at each data point to indicate the position of the central val-

D.2 GRAPHICAL ESTIMATION OF PARAMETERS

A graph of y versus x often provides a convenient way of estimating parameters of the relation $y = y(x)$. The simplest example is the straight line

$$y = A + Bx \tag{D.1}$$

where the slope and the intercept can be estimated by making a graph and drawing a straight line that relates y to x. Clearly the better way to handle this problem is by a least-squares fitting technique, but the graphical method can be useful in both research and instructional laboratories for obtaining quick preliminary estimates of experimental results.

If we wish to find from the graph the uncertainty in our estimate of the slope, then we should attempt to draw two lines through the data, corresponding to estimates of the largest and smallest *reasonable* slopes, s_1 and s_2. We should take account in the uncertainties in the data points, if they are available, and, because we are trying to estimate the uncertainty as a standard deviation, we should attempt to draw these two lines to bracket about two-thirds of the data points—not all the points. Making this estimate is often difficult and subjective, especially if there are few points and they exhibit a lot of scatter. The mean slope s is just the average of our two slopes,

$$s = (s_1 + s_2)/2 \tag{D.2}$$

and an approximate estimate of the uncertainty is the magnitude of half the difference

$$\sigma = |s_1 - s_2|/2 \tag{D.3}$$

To gain practice in determining parameters from a graph, it is a worthwhile exercise to estimate the parameters from the graph and to compare those estimates with the results of a least-squares fit to the data. We should note that the two lines selected to give a reasonable estimate of the uncertainty in the slope may not be the same two lines we might draw to obtain a reasonable estimate of the uncertainty in the intercept. Figure D.1 displays the data of Figure 1.1b, with lines bracketing the points to show (a) reasonable ranges for estimating the intercept, and (b) reasonable ranges for estimating the slope. These lines were actually calculated from the results of a least-squares fit of the equation $Y = A + Bx$ to the data, which yielded the parameters A and B and their uncertainties σ_A and σ_B. We calculated the two lines in Figure D.1a from the equations $Y = (A \pm \sigma_A) + Bx$ and those in Figure D.1b from the equations $Y = A + (B \pm \sigma_B)x$. We note that these lines are just particular examples of an infinite number of such lines corresponding to all combinations of the slope and intercept within one standard deviation ranges, and in any given graph, a decision must be made on which lines to draw. In particular, allowing the lines to intersect at the intercept as in Figure D.1b may not give the best evaluation of the

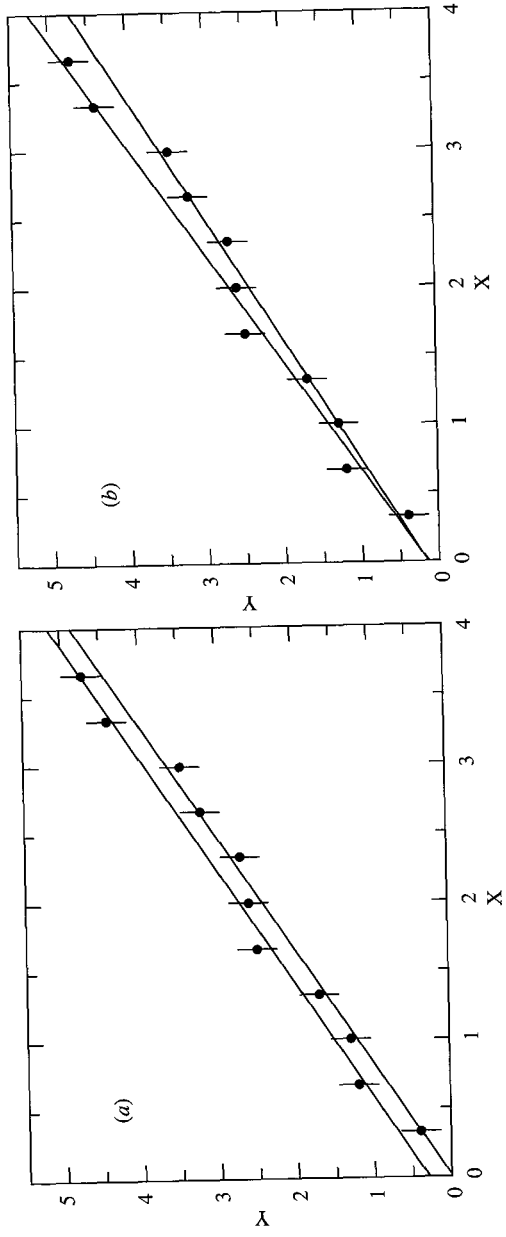

FIGURE D.1
Results of a least-squares fit of a straight line to the data of Figure 1.1b plotted to show separately the range of ±1 standard deviation (*a*) in the intercept, and (*b*) in the slope. Units on the axes are arbitrary.

Semilogarithmic Graphs

When dealing with an exponential decay function, it is convenient to display the activity as a function of time on a semilogarithmic graph. That is, if the relation is

$$y(t) = y_0 e^{-at} \tag{D.4}$$

we plot a graph of $\log(y)$ versus x. Fortunately semilogarithmic graph paper is readily available so that it is not necessary actually to calculate any logarithms to make this plot. We merely have to select paper with the appropriate number of powers of 10 for our plot, label the axes, and plot y versus x on the graph. Such a graph is illustrated in Figure 8.1 for Example 8.1.

Semilogarithmic graph paper comes in various *cycles,* corresponding to the number of *decades* or powers of 10 that can be plotted on a single sheet. Thus, for example, on three-cycle paper we can plot y values that range from 1 to 1000 (or from 0.01 to 10.0, etc.). Note that we can never plot y values that are zero or negative on semilogarithmic paper. This is a problem when dealing with subtracted distributions, such as the counting experiment of Example 8.1, where, if we wish to plot the number of counts remaining after we have subtracted the average background from cosmic rays, we discover that, at large times, some bins have negative net counts. Those points, of course, cannot be displayed on a semilogarithmic graph. A full, least-squares fit to the total, unsubtracted data sample is clearly the right way to solve this problem, but if we are to attempt a graphical solution, we should be aware of this limitation.

We can determine from our data the parameter a in Equation (D.4) by finding the slope of the straight line on the semilogarithmic graph just as we found the slope on ordinary graph paper for a simple linear plot. Note that when calculating the slope we must compute the logarithms of the y values. Thus, if the two ends of the straight line have coordinates (x_1, y_1) and (x_2, y_2), the slope is given by

$$s = \frac{\ln(y_2) - \ln(y_1)}{x_2 - x_1} = \frac{\ln(y_2/y_1)}{x_2 - x_1} \tag{D.5}$$

The uncertainty in the slope can again be determined by drawing two straight lines that bracket the mean slope, although the logarithmic form of the plot decreases the accuracy in this determination.

Full-Logarithmic Graphs

If we wish to display a power relation of the form $y = Ax^n$, we may make a plot of y versus x on full-logarithmic paper or *log-log* paper. The result will be a straight line with slope n and we can obtain the slope, and therefore the exponent n, from the graph. This technique could be used, for example, to check the $1/r^2$ law for radiation intensity as a function of distance, by plotting a graph of intensity versus distance on log-log paper.

introduced into the uncertainties in the process. Plotting on semilogarithmic or full-logarithmic paper is equivalent to such a variable change and we should attempt to compensate for these distortions, if necessary.

D.3 HISTOGRAMS AND FREQUENCY PLOTS

If we wish to display the frequency distribution of a measured variable x, then a histogram is generally the simplest and clearest form of presentation. For example, we may have observed particles emitted in the decay of an unstable state and wish to present the number detected in successive time intervals as in Example 2.4. Alternatively, we may have measured secondary particles in a scattering experiment and wish to display the distribution of their energies. In such cases, we can display the frequency distribution of the individual measurements, or *events,* as a histogram of $f(x)$ versus x, where $f(x_i)$ is the number of events that have values of x between x_i and $x_i + \Delta x$, and Δx is the histogram interval or bin width.

An alternate procedure for displaying binned data, which is especially useful for distributions with large numbers of bins, or for data with nonstatistical uncertainties, is to make a regular graph of frequency versus the measured variable, a *frequency plot,* with the data points indicated by crosses and uncertainties by error bars. This procedure is especially convenient when there are many bins or when error bars must be displayed, as illustrated in Figure 8.1.

A convenient procedure for finding the frequency distribution of (or *binning*) a continuous variable x is to label a bin with a tick mark at the lower limit x_1 of the bin and to count within a bin those events for which $x_i \leq x < x_i + \Delta x$. This is suitable for most, but not all, data sets. Choice of the bin width depends on a number of factors. In the ideal situation with a large quantity of high-precision data, the bin width could be chosen to be very small. However, in real experiments, the number of events may not be very large and each x coordinate will have some uncertainty. As a general rule, the bin width should not be less than the uncertainty in the measured variable x and one should be very wary of any data structure that is narrower than the uncertainty in x. If the number of events is relatively small, then even wider binning may be necessary. With such data, the competition between statistical significance and resolution of narrow effects in the histogram may become important. A histogram with less than ten events in its highest bin is not generally very informative, considering that the uncertainty in that bin will be over 30%.

A problem arises when the bin width of a histogram is close to or equal to the least count of the data. This can happen when the data are integral numbers or with data that have been collected by a digital device. The previous suggestions that the histogram bins be labeled with the lower limit at the left of the bin may not be reasonable for such data, and it may be better to place tick marks at the center of the bins.

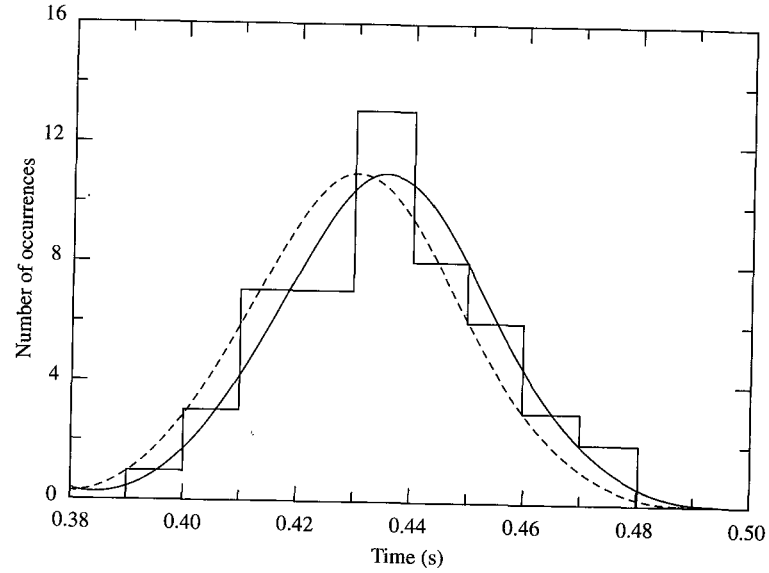

FIGURE D.2
Histogram of measured times plotted with the bin width equal to the least count of a digital clock. The numbers on the abscissa correspond to the lower time limit of the bin. The dashed Gaussian curve was calculated from the mean and standard deviation of the measurements. The solid curve was calculated with the mean increased by half the bin width to correct for the truncation of the data.

timer starts when the ball is released and stops when it hits the floor. Uncertainties in the measurements come mainly from variations in the starting and stopping times.

The student's measurements have been plotted in the histogram of Figure D.2 where the bin width is equal to the least count (0.01ts). We assume that the digital clock truncates the measured times so each time measurement corresponds to the left-hand edge of a bin and the actual value of the time is somewhere within the bin limits. Thus, in this case it is appropriate to indicate the lower value of the bin limit at the left-hand edge of the bin.

The dashed Gaussian curve was calculated from the mean ($\bar{t} = 0.431$ s) and standard deviation (s = 0.0184) of the measurements. The curve clearly is shifted to the left relative to the data. The discrepancy is caused by the fact that we neglected to correct for the truncation of the data by the digital clock. To correct the mean we must add to it half the width of a bin to obtain $\bar{t} = 0.431 + 0.005 = 0.436$ s. The Gaussian curve, calculated from the corrected mean, is shown as a solid line.

Normalized Curves on Histograms

When superimposing a theoretical curve on a data histogram, we often want to scale

to unit area, such as the Gaussian probability function of Equation (2.23). The area of one event on the histogram is equal to the bin width Δx multiplied by a unit interval on the ordinate. Thus, the total area of the histogram is equal to the bin width multiplied by the total number of events ($A = N\Delta x$). To scale the curve to the area of the histogram, we multiply the values $p(x_i)$, calculated from the equation of the probability distribution, by the product of the number of events on the plot and the bin width, so that the plotted curve becomes

$$y(x_i) = p(x_i) \times N\Delta x \tag{D.6}$$

D.4 GRAPHICS ROUTINES

We include source files on the website for routines which can be used to make simple graph and histograms. Most of the sample computer routines in this book make calls to these routines.

Program D.1 QUIKSCRP (website) accepts data that define graphs and histograms and writes a script file that can be read and interpreted by the executable program QDISPLAY.EXE (website) to produce displays on the monitor. Details of the calling procedures can be seen in the routine PLOTIT in the program unit \CHAPT-6\FITUTIL (website) called from the program \CHAPT-6\FITLINE (APPENDIX E). For this program QuikScrp writes an output file FITLINE.SCR.

Program D.2 QUIKHIST (website) collects data and presents a character-based histogram on the monitor. Printed output is also available. PROGRAM 5.2: \CHAPT-5\POISDCAY illustrates use of this program.

Program D.3 QDISPLAY.EXE (website) is an executable program that reads a script file written by QUIKSCRP and interprets the file to create a graphics display on the monitor. The command line instruction for running QDISPLAY with the script file produced by the program FITLINE is QDISPLAY FITLINE.

APPENDIX E

COMPUTER ROUTINES IN FORTRAN

This appendix lists several routines that illustrate the material of the text. The routines are listed in Fortran 77, an old, but quite readable version of that ever-popular programming language. All routines have been tested; however, most of them requires subsidiary routines and drivers that are not listed. Complete programs and routines are available on the Web in C++ as well as in Fortran. Readers are urged to log onto the website at www.mhhe.com/bevington to download these programs.

We have tried to keep the routines simple, trading efficiency for clarity where necessary. To make explicit which modules are required to form a complete program, and to avoid the need for command strings to link the object programs into an executable program, we have chosen to use the INCLUDE statement to present the compiler with a single source file from which to compile a single object module incorporating all required routines. We also use the INCLUDE statement to copy blocks of COMMON and other variable-defining statements into routines.

Because readers may not be familiar with Fortran, we list a few basic principles that should help in understanding the instructions and following their logic. This list includes only a selection of language elements that appear in the sample programs.

STATEMENTS

The format of Fortran statements was defined in terms of the 80-column Hollerith card:

columns 2–5: statement label (a number);

column 6: reserved for a single digit number to indicate a continuation of the statement from the previous line;

columns 7–72: program statements;

columns 73–80: not used.

Although it is not necessary to follow rigorously this scheme with a modern interactive compiler on a personal computer (for example, "tabs" can be used), the general order must be followed.

PROGRAM FLOW

Program flow can be controlled by IF statements, by IF THEN statements (with ELSEIF and ENDIF), by DO AND DOWHILE statements that may refer to a termination label (all statements labels are numerical) or to the DO terminator; ENDDO, and by GOTO statements. Excessive use of the GOTO statement can lead to very confusing programs. In order to facilitate following the program flow, we have indented groups of instructions that are accessed through a control statement, such as IF THEN, or DO.

Examples

```
DO 100I = 1 TO 20
  X = I
  Y(I) = SQRT(X)
100 CONTINUE
```

```
DO I = 1 TO 20
  X = I
  Y = SQRT(X)
ENDDO
```

```
X = 1
DO WHILE X .LE. 20
  Y = SQRT(X)
  X = X + 1
ENDDO
```

VARIABLE DEFINITIONS

Fortran does not require the rigorous variable typing of newer languages. As default typing, variables with names beginning with I, J, K, L, M, or N are defined as INTEGER; variable names beginning with other letters are identified as REAL. However, we have attempted to identify most of the variables in the routines and in some instances have violated the default typing for program clarity.

Examples

```
INTEGER S1, S3, N/10/
REAL X, T, TPRIME, SIGMAT/1.0/
LOGICAL NEXTVAR/.FALSE./
```

Note that the variables N, SIGMAT and, NEXTVAR in the preceding examples have been initialized to the values 10, 1.0, and FALSE, respectively. The DATA statement also can be used to initialize variables. DATA SQRTPI/1.7724539/

Other types include:

```
CHARACTER
COMPLEX
DOUBLE PRECISION
```

Variables defined in named COMMON statements are available to any routine that includes the statement. Local variables can be defined in DIMENSION STATEMENTS. Array sizes may be defined in PARAMETER statements or directly in a COMMON or DIMENSION statement.

Examples

```
PARAMETER(MAXPARAM 10)
COMMON/FITVARS/ NPTS, M, NFREE, MARRAY(MAXPARAM),
ZARRAY(200)
DIMENSION NPLAN(30).
```

Fortran has several types of subprograms that can be called from another routine: SUBROUTINE and FUNCTION are the most common. Data types defined in a subprogram must be consistent with the definitions in the calling routine. A function name must specify its own data type.

Examples

```
CALL SETRANDOMDEVIATESEED(S1, S2, S3)
TPRIME = GAUSSSMEAR(T,SIGMAT)
REAL FUNCTION GAUSSSMEAR(X,DX)
SUBROUTINE SETRANDOMDEVIATESEED(SA,SB,SC)
```

The INCLUDE statement copies the specified file into the body of the program.

Example

```
INCLUDE '\CHAPT-5\MONTEINC.FOR'
```

As well as comment statements that begin with a "C" in column 1, comments may appear in statement lines, preceded by the exclamation point (!).

E.1 Routines from Chapter 5

```
C PROGRAM 5.1: \CHAPT-5\HOTROD.FOR
C SIMULATED VARIATION OF TEMPERATURE ALONG A METAL ROD
C 10 CM ROD-TEMPERATURE IS ZERO AT ONE END, 100 DEGREES C AT OTHER.
C USES MONTELIB
      PROGRAM HOTROD
      INTEGER S1, S2, S3, N/10/          !--- GENERATE 10 POINTS AT 1 CM INTERVALS
      REAL  X, T, TPRIME, SIGMAT/1.0/ !--- WITH AN UNCERTAINTY OF +-1 DEGREE
      REAL GAUSSSMEAR
      S1 = 1171
      S2 = 343
      S3 = 1322
      CALL SETRANDOMDEVIATESEED(S1, S2, S3)
      PRINT *,'     HOT ROD TEST DATA,   SIGMA=',  SIGMAT
      X = -0.5
      DO 100 I = 1, N
         X     = X + 1.0            !--- POSITION ALONG ROD
         T     = 10.0*X             !--- CALCULATE MEAN TEMPERATURE AT POINT
         TPRIME = GAUSSSMEAR(T,SIGMAT)          !--- SMEAR IT
         PRINT *,I, X, T, TPRIME
  100 CONTINUE
      CALL EXIT
      END
      INCLUDE C:\CHAPT-5\MONTELIB.FOR

C PROGRAM 5.2: \CHAPT-5\POISDCAY.FOR
C SIMULATED DECAY OF AN UNSTABLE STATE.
C USES QUIKHIST, MONTELIB
      PROGRAM POISDCAY       !--- GENERATE A 200-EVENT POISSON  HISTOGRAM
      REAL LO/0/, INT/1/, HI/22/
      INTEGER NEVENTS/400/, POISSONDEVIATE
      REAL MU/8.4/
      INTEGER S1, S2, S3, I, K
      REAL   X
      S1 = 1171
      S2 = 343
      S3 = 1322
      CALL SETRANDOMDEVIATESEED(S1, S2, S3)
      CALL HISTINIT(' ')          !---OUTPUT FILE NAME OR ' ' FOR MONITOR OUTPUT
      CALL HISTSETUP(1,LO,INT,HI,'POISSON - COUNTS/10 SEC')
      K=POISSONDEVIATE(MU,.TRUE.)         !--- INITIALIZE - MAKE THE TABLE
      DO 100 I = 1, NEVENTS
         K = POISSONDEVIATE(MU,.FALSE.)
         X = K
         CALL HISTOGRAM(1,X)
  100 CONTINUE
      CALL HISTDISPLAYALL(.FALSE.) !DUMMY ARG-COMPAT. WITH QUIKSCRP
      CALL EXIT
      END
      INCLUDE \CHAPT-5\MONTELIB.FOR
```



```
C PROGRAM 5.3: \CHAPT-5\MONTELIB.FOR
C MONTE CARLO LIBRARY ROUTINES
      SUBROUTINE SETRANDOMDEVIATESEED(SA,SB,SC)
      INCLUDE '\CHAPT-5\MONTEINC.FOR'
      INTEGER SA, SB, SC
      SEED1 = SA
      SEED2 = SB
      SEED3 = SC
      RETURN
      END

      SUBROUTINE GETRANDOMDEVIATESEED(SA,SB,SC)
      INCLUDE '\CHAPT-5\MONTEINC.FOR'
      INTEGER SA, SB, SC
      SA = SEED1
      SB = SEED2
      SC = SEED3
      RETURN
      END

      REAL FUNCTION RANDOMDEVIATE()          !--- WICHMANN AND HILL
      INCLUDE '\CHAPT-5\MONTEINC.FOR'
      REAL  TEMP
      SEED1 = 171*MOD(SEED1,177) - 2*(SEED1 / 177)
      IF (SEED1 .LT. 0 ) SEED1 = SEED1 + 30269
      SEED2 = 172*MOD(SEED2,176) - 35*(SEED2 / 176)
      IF (SEED2 .LT. 0 ) SEED2 = SEED2 + 30307
      SEED3 = 170*MOD(SEED3,178) - 63*(SEED3 / 178)
      IF (SEED3 .LT. 0 ) SEED3 = SEED3 + 30323
      TEMP = SEED1/30269. + SEED2/30307. + SEED3/30323.
      RANDOMDEVIATE = TEMP-AINT(TEMP)
      RETURN
      END

C -FIND A RANDOM VARIABLE DRAWN FROM THE GAUSSIAN DISTRIBUTION-
      REAL FUNCTION RANDOMGAUSSDEVIATE()        !--- BOX-MUELLER
      INCLUDE '\CHAPT-5\MONTEINC.FOR'
      LOGICAL  NEXTVAR/.FALSE./
      REAL R, F, Z1, Z2, X1RANGAUSS, RANDOMDEVIATE
      IF (NEXTVAR) THEN
         NEXTVAR = .FALSE.
         RANDOMGAUSSDEVIATE = X2RANGAUSS
      ELSE
  100    Z1 = -1 + 2*RANDOMDEVIATE()
         Z2 = -1 + 2*RANDOMDEVIATE()
         R  = Z1*Z1 + Z2*Z2
         IF (R .GE. 1) GOTO 100
         F = SQRT(-2*ALOG(R)/R)
         X1RANGAUSS = Z1*F
         X2RANGAUSS = Z2*F
         RANDOMGAUSSDEVIATE = X1RANGAUSS
      ENDIF
      RETURN
      END
```

```
      NEXTVAR = .TRUE.
    ENDIF
    RETURN
    END

    REAL FUNCTION GAUSSSMEAR(X,DX)
    REAL X, DX
    REAL RANDOMGAUSSDEVIATE
    GAUSSSMEAR = X + RANDOMGAUSSDEVIATE() * DX
    RETURN
    END

C -RECURSION METHOD FOR POISSON PROBABILITY (P(N,M). TO FIND P(N,M) MUST
C   ALL WITH SUCCESSIVE ARGUMENTS J=0,1,..N. MAX MU=85, NO LIMIT ON X
    REAL FUNCTION POISSONRECUR(J, M)
    INCLUDE '\CHAPT-5\MONTEINC.FOR'
    INTEGER J
    REAL M
    IF (J.EQ.0 ) THEN
      POISS = EXP(-M)
    ELSE
      POISS = (POISS*M)/J        !--- POISS =  (M^J)EXP(-MU/J)
    ENDIF
    POISSONRECUR = POISS
    RETURN
    END

C -FIND A RANDOM VARIABLE DRAWN FROM THE POISSON DISTRIBUTION
    INTEGER FUNCTION POISSONDEVIATE(MU, INIT)
    INCLUDE '\CHAPT-5\MONTEINC.FOR'
    INTEGER  I, X, N
    REAL     MU, P , R, POISSONRECUR
    LOGICAL  INIT
    IF (INIT ) THEN                    ! --- MAKE TABLE OF SUMS ---
      N = AINT(MU + 8* SQRT(MU))       ! ---IE., 8*SIGMA
      IF (N .GT. MAXBINS ) THEN
        PRINT *, 'OVERFLOW ERROR IN ROUTINE POISSON DEVIATE'
        CALL EXIT
      ENDIF
      PTABLE(0) = POISSONRECUR(0,MU)
      DO 100 I = 1, N-1
        P = POISSONRECUR(I,MU)
        PTABLE(I) = PTABLE(I-1)+P
100   CONTINUE
      PTABLE(N) = 1                    ! --- ASSURE UNIT PROBABILITY ---
    ELSE                               ! --- GENERATE AN EVENT ---
      X = -1
      R = RANDOMDEVIATE()
200   X = 1 + X
      IF (PTABLE(X) .LE. R) GOTO 200 !— REPEAT UNTIL PTABLE(X) >= X
      POISSONDEVIATE = X
```



```
    ENDIF
    RETURN
    END

PROGRAM 5.4:  \CHAPT-5\KDECAY.FOR          (WEBSITE)
C ILLUSTRATION OF EXAMPLE 5.7

C PROGRAM 5.5:  \CHAPT-5\MONTEINC.FOR
C COMMON FOR MONTE CARLO LIBRARY
    COMMON/MC/ SEED1, SEED2, SEED3, X2RANGAUSS, POISS, PTABLE
    PARAMETER (MAXBINS = 100)
    INTEGER SEED1, SEED2, SEED3
    REAL X2RANGAUSS, PTABLE(0:MAXBINS)
    REAL*8 POISS
C--------------END MONTEINC ------------------------
```

E.2 Routines from Chapter 6

```
C PROGRAM 6.1: \CHAPT-6 FITLINE.FOR
C LEAST-SQUARES FIT TO A STRAIGHT LINE BY METHOD OF DETERMINANTS
C USES  FITUTIL
    PROGRAM FITLINE
C ------------------------MAIN  ROUTINE----------------------
    INCLUDE '\CHAPT-6 FITVARS.FOR'
    CHARACTER*40  TITLE
    CHARACTER*1 VORG, READCHAR
    INTEGER I
    REAL DET, CHI2, CALCCHISQ
    M = 2                                          !--- FIND 2 PARAMETERS
    PRINT *, '(V)OLTS OR (G)EIGER? '
    VORG = READCHAR()
    IF (VORG .EQ. 'V') THEN
      CALL FETCHDATA('\CHAPT-6\VOLTS.DAT' ,TITLE)      !--- EXAMPLE 6.1
    ELSEIF ((VORG .EQ. 'G') .OR. (VORG .EQ. 'G')) THEN
      CALL FETCHDATA('\CHAPT-6\GEIGER.DAT',TITLE)      !--- EXAMPLE 6.2
      DO 100 I = 1 , NPTS
      X(I)  = 1/X(I)**2                            !--- FITTING 1/R^2
100   CONTINUE
    ENDIF
    CALL LINEFIT(DET)
    CALL CALCULATEY        !---  FILL ARRAY YCALC FOR CALCCHISQ AND PLOTIT
    CHI2 = CALCCHISQ()
    CALL OUTPUT(.FALSE. , 'CON', CHI2, TITLE)  !--- FALSE FOR NO ERROR MATRIX
    IF (VORG .EQ. 'V') THEN
      CALL PLOTIT('FITLINE.SCR',.FALSE.,.FALSE., !---SCRPT FILE,LOG?,SPLINE?
     1   'C', ABS(X(2)-X(1))/20,          !--- DATA CIRCLE, RAD OF CIRCLE
     2   0.0, 0.0, 100.0, 3.0,            !---  X1,Y1,X2,Y2
```

```
     3    5, 6,                                !---  # X-DIV, # Y-DIV
     4    'X (CM)', 'POTENTIAL DIFF(ERENCE (VOLTS)')  !---  AXIS LABELS
          ELSEIF (VORG .EQ. 'G') THEN
             CALL PLOTIT('FITLINE.SCR',.FALSE. ,.FALSE.,
     1    'C', ABS(X(2)-X(1))/50,  0.0, 0.0, 30.0, 1000.0,  6, 5,
     2    'SQUARED INVERSE DISTANCE (1/M^2)', 'NUMBER OF COUNTS PER SEC')
          ENDIF
          READ *
          CALL CLOSEGRAPHICS
          END

          SUBROUTINE CALCULATEY        !---  FILLS ARRAY YCALC
          INCLUDE '\CHAPT-6\FITVARS.FOR'
          INTEGER I
          DO 100 I= 1 , NPTS
             YCALC(I) = A(1) + A(2)*X(I)
      100 CONTINUE
          RETURN
          END

          REAL FUNCTION CALCCHISQ() !---  ASSUMES ARRAY YCALC HAS BEEN FILLED
          INCLUDE '\CHAPT-6\FITVARS.FOR'
          INTEGER  I
          REAL     CHI2
          CHI2=0.
          DO 100 I = 1 , NPTS
             CHI2 = CHI2 + ( (Y(I)-YCALC(I))/SIGY(I))**2
      100 CONTINUE
          CALCCHISQ = CHI2
          RETURN
          END

          SUBROUTINE LINEFIT(DET)
          INCLUDE '\CHAPT-6\FITVARS.FOR'
          REAL DET
          INTEGER I
          REAL SUMWT, SUMX, SUMY, SUMX2, SUMY2, SUMXY, WEIGHT
          SUMWT = 0
          SUM   = 0
          SUMY  = 0
          SUMX2 = 0
          SUMY2 = 0
          SUMXY = 0
      C  --------- ACCUMULATE WEIGHTED SUMS ----------
          DO 100  I= 1 , NPTS
             WEIGHT = 1/SIGY(I)**2
             SUMWT  = SUMWT  + WEIGHT
             SUMX   = SUMX   + WEIGHT * X(I)
```



```
             SUMY   = SUMY   + WEIGHT * Y(I)
             SUMX2  = SUMX2  + WEIGHT * X(I)**2
             SUMY2  = SUMY2  + WEIGHT * Y(I)**2
             SUMXY  = SUMXY  + WEIGHT * X(I)*Y(I)
      100 CONTINUE
      C ---CALCULATE THE PARAMETERS - CUT OUT IF DETERMINANT IS NOT > 0 ---
          DET =  SUMWT * SUMX2 - SUMX * SUMX
          IF (DET .GT. 0 )  THEN
             A(1)    = (SUMX2*SUMY  - SUMX*SUMXY)/DET
             A(2)    = (SUMXY*SUMWT - SUMX*SUMY) /DET
             SIGA(1) = SQRT(SUMX2/DET)
             SIGA(2) = SQRT(SUMWT/DET)
          ELSE
             CALL ERRORABORT('DETERMINANT < OR = 0 IN LINEFIT')
          ENDIF
          RETURN
          END
          INCLUDE '\CHAPT-6\FITUTIL.FOR' ! FITUTIL INCLUDES QUIKSCRP.FOR


      C PROGRAM 6.2: \CHAPT-6\FITVARS.FOR        (WEBSITE)
      C INCLUDE FILE OF CONSTANTS, VARIABLES AND  ARRAYS FOR LEAST-SQUARES FITS
      C ALL GLOBAL TYPES, CONSTANTS AND VARIABLES ARE DECLARED HERE.
      C THE ARRAY LIMITS MAXDATA AND MAXPARAM CAN BE SET AS REQUIRED
                                                    FOR PARTICULAR PROBLEMS.

      C PROGRAM 6.3: \CHAPT-6\FITUTIL.FOR        (WEBSITE)
      C GENERAL UTILITY ROUTINES
```

E.3 Routines from Chapter 7

```
      C PROGRAM 7.1: \CHAPT-7\MULTREGR.FOR
      C LEAST-SQUARES FIT TO A POWER SERIES AND TO LEGENDRE POLYNOMIALS.
      C USES  FITFUNC7, MAKEAB7, MATRIX, FITUTIL

            PROGRAM MULTREGR
      C     M = NUM OF PARAMETERS, NPTS=NUMBER OF DATA PAIRS,
      C     DATA AND UNCERTAINTIES ARE IN ARRAYS  X, Y, DY.
            INCLUDE '\CHAPT-6\FITVARS.FOR'
            COMMON /FITVARS7/PAE
            CHARACTER * 1 PAE
            REAL    DET, CHI2, CALCCHISQ
            INTEGER I
            LOGICAL SPL
            CHARACTER*1 READCHAR
            CHARACTER*40 TITLE
            PRINT *,   '(P)OWER SERIES, (A)LL LEGENDRE TERMS TO L = 4,'
```

```
      PRINT *,   'OR (E)VEN LEGENDRE TERMS(L = 0,2,4).'
      PRINT *,   'TYPE P,A OR E '
      PAE = READCHAR()
 1000 FORMAT(A1)
      IF (PAE .EQ. 'P') THEN
         CALL FETCHDATA ('\CHAPT-7\THERMCOU.DAT', TITLE)
         PRINT *,   'TYPE NUMBER OF PARAMETERS '
         READ *, M
      ELSEIF (PAE .EQ. 'A') THEN
         CALL FETCHDATA('\CHAPT-7\LEGENDRE.DAT',TITLE)
         M = 5
      ELSEIF (PAE .EQ. 'E' ) THEN
         CALL FETCHDATA('\CHAPT-7\LEGENDRE.DAT',TITLE)
         M = 3
      ENDIF  !--- PAE
      CALL MAKEBETA                        !--- SET UP THE LINEAR BETA MATRIX
      CALL MAKEALPHA                       !--- SET UP THE SQUARE ALPHA MATRIX
      CALL MATINV(M, ALPHA, DET)           !--- INVERT ALPH TO GET EPSILON MATRIX
      CALL LINEARBYSQUARE(M,BETA,ALPHA,A) !--- BETA X EPS = PARAMETER MATRIX
      CALL CALCULATEY
      CHI2 = CALCCHISQ()
      DO 100 I = 1, M
         SIGA(I) = SQRT(ALPHA(I,I))
  100 CONTINUE
      CALL OUTPUT(.TRUE., 'CON', CHI2, TITLE) !---  TRUE TO PRINT ERROR MATRIX
      IF (M .GT. 2 ) THEN
         SPL = .TRUE.                      !---  PLOT A CURVE
      ELSE
         SPL = .FALSE.                     !---  PLOT A LINE
      ENDIF
      IF (PAE .EQ. 'P') THEN
         CALL PLOTIT('MULTREGR.SCR', .FALSE., SPL,    !--- FILE,LOG?,SPLINE
     1   'C', (X(2)-X(1))/12,      !--- DATA CIRCLES, RADIUS OF DATA CIR
     2   -10., -2., 110., 4.,          !--- X1,Y1, X2,Y2
     3    6, 6,                        !— X,Y GRID MARKS
     4   'TEMPERATURE (DEGREES CELSIUS)','VOLTAGE (MV)')
      ELSE IF ((PAE .EQ. 'A') .OR. (PAE .EQ. 'E')) THEN
         CALL PLOTIT('MULTREGR.SCR', .FALSE., .TRUE.,
     1   'C', (X(2)-X(1))/10,  0., 0.,180., 1500., 6, 6,
     2   'THETA(DEGREES)', 'NUMBER OF COUNTS')
      ENDIF                                  !--- PAE
      CALL CLOSEGRAPHICS
      END
      INCLUDE '\CHAPT-7\FITFUNC7.FOR'
      INCLUDE '\CHAPT-7\MAKEAB7.FOR'
      INCLUDE '\CHAPT-6\FITUTIL.FOR'
      INCLUDE '\APPEND-B\MATRIX.FOR'

C PROGRAM 7.2: \CHAPT-7\FITFUNC7.FOR
C FITTING FUNCTIONS FOR CHAPTER 7 EXAMPLES.
```



```
      REAL FUNCTION POWERFUNC(K, XX)
      INTEGER K
      REAL   XX
      REAL   YY
      INTEGER I
      YY = 1
      IF (K .GT. 1 ) THEN
         DO 100 I= 2, K
         YY = XX * YY
  100 CONTINUE
      ENDIF
      POWERFUNC = YY
      RETURN
      END

      REAL FUNCTION LEGFUNC(K, XX)
C DEFINE SEPARATE TERMS IN A SERIES, Y = A0*L0(X) + A1*L1(X) + ..
C NOTE K = 1 CORRESPONDS TO ZEROTH ORDER.
C VAR  PAE : CHAR   'P'-POWER SERIES,
C 'A'-ALL LEGENDRE TERMS TO ORDER M,
C 'E'-EVEN LEGENDRE TERMS)
C
      COMMON /FITVARS7/PAE
      CHARACTER *1 PAE
      INTEGER K
      REAL   XX
      INTEGER KK, I
      REAL  C, PI/3.14159/, LEGPOLY(11) !---  I.E., 0TH THRU 10TH ORDER
      IF (PAE .EQ. 'E') KK = 2*K - 1
      IF (PAE .EQ.'A')  KK = K
      C = COS(PI*XX/180)
      LEGPOLY(1) = 1 !---  FOR BETTER EFFICIENCY, COULD CALC ONCE AND SAVE
      IF (KK .GT. 1 ) THEN
         LEGPOLY(2) = C
         IF (KK .GT. 2 ) THEN
         DO 100 I = 3, KK
           LEGPOLY(I)=((2*I-1)*C*LEGPOLY(I-1)-(I-1)*LEGPOLY(I-2))/I
  100    CONTINUE
      ENDIF            !---  KK > 2
      ENDIF               !---  KK > 1
      LEGFUNC = LEGPOLY(KK)
      RETURN
      END

      REAL FUNCTION FUNCT(K, XX)
      INTEGER K
      REAL XX
      REAL LEGFUNC, POWERFUNC
      COMMON /FITVARS7/PAE
      CHARACTER * 1 PAE
```

```
      IF ((PAE .EQ. 'A') .OR. (PAE.EQ.'E')) FUNCT = LEGFUNC(K,XX)
      IF  (PAE .EQ. 'P')  FUNCT = POWERFUNC(K,XX)
      RETURN
      END

      SUBROUTINE CALCULATEY
      INTEGER  I, K
      REAL     YY, FUNCT
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      DO 100 I=1, NPTS
         YY = 0
         DO 200 K = 1, M
            YY = YY + A(K) * FUNCT(K,X(I))
200      CONTINUE
         YCALC(I) = YY
100   CONTINUE
      RETURN
      END

      REAL FUNCTION CALCCHISQ() !--- ASSUMES ARRAY YCALC HAS BEEN FILLED
      INTEGER I
      REAL    CHI2
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      CHI2=0.
      DO 100 I = 1, NPTS
         CHI2 = CHI2 + ( (Y(I)-YCALC(I)) / SIGY(I))**2
100   CONTINUE
      CALCCHISQ = CHI2
      RETURN
      END
```

```
C PROGRAM 7.3: \CHAPT-7\MAKEAB7.FOR
C ROUTINES TO SET UP THE BETA AND ALPHA MATRICES FOR LINEAR REGRESSION
C USES  MATRIX, FITFUNC7
C
      SUBROUTINE MAKEBETA            !--- MAKE THE BETA MATRICES
      INTEGER  I, K
      REAL FUNCT
      INCLUDE 'C:\CHAPT-6\FITVARS.FOR'
      DO 100  K=1, M
         BETA(K)=0
         DO 200 I=1, NPTS
         BETA(K)=BETA(K) + Y(I)*FUNCT(K, X(I))/SIGY(I)**2
200   CONTINUE
100   CONTINUE
      RETURN
      END

      SUBROUTINE MAKEALPHA           !--- MAKE THE ALPHA MATRICES
      INTEGER I,J,K
      REAL FUNCT
```



```
      INCLUDE 'C:\CHAPT-6\FITVARS.FOR'
      DO 100 J=1, M
         DO 200 K=1, M
            ALPHA(J,K)=0
               DO 300 I=1, NPTS
                  ALPHA(J,K) = ALPHA(J,K)+FUNCT(J, X(I))*FUNCT(K, X(I))/SIGY(I)**2
               300 CONTINUE
         200  CONTINUE
      100  CONTINUE
      RETURN
      END
```

E.4 Routines from Chapter 8

```
C PROGRAM 8.0: \CHAPT-8\NONLINFT.FOR
C MAIN CALLING ROUTINE FOR NON-LINEAR FITTING METHODS
C USES GRIDSEAR, GRADSEAR, EXPNDFIT, MARQFIT, FITFUNC8, MAKEAB8,
C      NUMDERIV, MATRIX, FITUTIL
       PROGRAM NONLINFT
       INTEGER  TRIAL, J, METHOD
       REAL STEPDOWN, LAMBDA, CHISQR, CALCCHISQ
       CHARACTER*40 TITLE
       REAL STEPSCALE(4)/0.49999, 0.99999, 0.001, 0.001/
       INCLUDE '\CHAPT-6\FITVARS.FOR'
       PRINT *,'    (1)GRID SEARCH,    (2)GRADIENT SEARCH'
       PRINT *,'    (3)CHISQ EXPANSION, (4)FUNCTION EXPANSION'
       PRINT *, 'TYPE 1, 2, 3, OR 4  --- '
       READ *, METHOD
       CHICUT   = 0.01
       STEPDOWN  = 0.1       !--- STEP DOWN THE GRADIENT IN GRADLS
       LAMBDA    = 0.001     !--- FOR MARQUARDT METHOD ONLY
       STEPSIZE  = STEPSCALE(METHOD)  !--- SCALES DELTAA(J)
       CALL FETCHDATA('\CHAPT-8\RADIODK.HST',TITLE)
       CALL FETCHPARAMETERS     !--- USES NPTS, MUST FOLLOW FETCHDATA
       TRIAL    = 0
       CHISQR  = CALCCHISQ()
       CHIOLD = CHISQR + CHICUT + 1
       DO WHILE (ABS(CHIOLD - CHISQR) .GE. CHICUT)
          CHIOLD = CHISQR
          PRINT 1000, TRIAL, CHISQR
1000   FORMAT(' TRIAL #', I4, ' CHISQ =', F10.1)
          PRINT 1100, (A(J), J = 1,M)
1100   FORMAT(6F12.4)
          PRINT *
          GOTO (110, 120, 130, 140) , METHOD
110       CALL GRIDLS(CHISQR)
          GOTO 150
120       CALL GRADLS(CHISQR, STEPDOWN)
          GOTO 150
130       CALL CHIFIT(CHISQR)
          GOTO 150
```

```
140     CALL MARQUARDT(CHISQR, CHICUT, LAMBDA)
150     TRIAL = TRIAL +1
      ENDDO
151   CALL CALCULATEY
      IF ((METHOD .EQ. 1) .OR. (METHOD .EQ. 2)) THEN
         DO 200 J = 1, M
            SIGA(J) = SIGPARAB(J)                    !--- DCHI2 = 1
200      CONTINUE
         CALL OUTPUT(.FALSE., 'CON' ,CHISQR, TITLE) !--- NO ERROR MATRIX
      ELSEIF ((METHOD .EQ. 3) .OR. (METHOD .EQ. 4)) THEN
         IF (METHOD .EQ. 4 ) THEN
            CALL MARQUARDT(CHISQR,CHICUT,0)          !--- GET ERROR MATRI
         ENDIF
         DO 300 J = 1, M
            SIGA(J) = SIGMATRX(J)                    !--- ERROR MATRIX
300      CONTINUE
         CALL OUTPUT(.TRUE., 'CON', CHISQR, TITLE) !--- WITH ERROR MATRIX
      ENDIF
      CALL PLOTIT('NONLIN.SCR', .TRUE., .TRUE., !--- SCRPT FILE, LOG?, SPLINE?
     1  'C', (X(2)-X(1))/5,    !--- DATA CIRCLES, RADIUS OF CIRCLES
     2  0., 1., 900., 1000.,   !--- RANGES-X1,Y1,X2,Y2
     3  6, 6,                  !--- NUM X-AXIS DIV, NUM Y-AXIS DIV
     4  'TIME (SEC)', 'NUMBER OF COUNTS')  !--- AXIS LABELS
      CALL CLOSEGRAPHICS
      END

C SAMPLE FITTING FUNCTION FOR NON-LINEAR FITS
C EXAMPLE IS SUM OF 2 EXPONENTIALS ON A CONSTANT BACKGROUND
      REAL FUNCTION  EXPF(A,X)
      REAL A,X
      REAL YY, ARG
      ARG = ABS(X/A)
      IF (ARG .GT. 60 ) THEN
         YY = 0
      ELSE
         YY = EXP(-ARG)
      ENDIF
      EXPF = YY
      RETURN
      END

      FUNCTION  YFUNCTION(XX) !--- REAL
      REAL YFUNCTION, XX, EXPF
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      YFUNCTION = A(1) + A(2)*EXPF(A(4),XX) + A(3)*EXPF(A(5),XX)
      RETURN
      END

      INCLUDE '\CHAPT-8\GRIDSEAR.FOR'    !--- 1-GRID SEARCH METHOD
      INCLUDE '\CHAPT-8\GRADSEAR.FOR'    !--- 2-GRADIENT SEARCH METHOD
      INCLUDE '\CHAPT-8\EXPNDFIT.FOR'    !--- 3-FUNCTION EXPANSION METHOD
```



```
      INCLUDE '\CHAPT-8\MARQFIT.FOR'   !--- 4-MARQUARDT METHOD
      INCLUDE '\CHAPT-6\FITUTIL.FOR'
      INCLUDE '\CHAPT-8\FITFUNC8.FOR'     !--- USED BY ALL METHODS
      INCLUDE '\CHAPT-8\MAKEAB8.FOR'      !--- USED BY METHODS 4 AND 5
      INCLUDE '\CHAPT-8\NUMDERIV.FOR'     !--- USED BY METHODS 4 AND 5
      INCLUDE '\APPEND-B\MATRIX.FOR'      !--- USED BY METHODS 4 AND 5

C  PROGRAM 8.1: \CHAPT-8\GRIDSEAR.FOR
C NON-LINEAR FIT BY THE GRID-SEARCH METHOD
C USES FITFUNC8, FITUTIL
      SUBROUTINE GRIDLS(CHISQR)
      REAL CHISQR
      REAL CALCCHISQ
      REAL SAVE,DELTA, DELTA1,DEL1,DEL2,AA,BB,CC,DISC,ALPH,X1,X2
      INTEGER J
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      CHISQ2 = CALCCHISQ()
C  -FIND LOCAL MINIMUM FOR EACH PARAMETER-
      DO 100 J = 1, M
         DELTA   = DELTAA(J)
         A(J)    = A(J) + DELTA
         CHISQ3  = CALCCHISQ()
         IF (CHISQ3 .GT. CHISQ2 ) THEN
            DELTA = -DELTA            !--- STARTED IN WRONG DIRECTION
            A(J)  = A(J) + DELTA
            SAVE  = CHISQ2            !--- INTERCHANGE 2 AND 3 SO 3 IS LOWER
            CHISQ2 = CHISQ3
            CHISQ3 = SAVE
         ENDIF                        !--- IF (CHISQ3 ...
C  -INCREMENT OR DECREMENT A(J) UNTIL CHI SQUARED INCREASES-
  110 CONTINUE
      CHISQ1 = CHISQ2              !--- MOVE BACK TO PREPARE FOR QUAD FIT
      CHISQ2 = CHISQ3
      A(J)   = A(J) + DELTA
      CHISQ3 = CALCCHISQ()
      IF (CHISQ3 .LE. CHISQ2) GOTO 110
C  -FIND MINIMUM OF PARABOLA DEFINED BY LAST THREE POINTS-
      DEL1 = CHISQ2 - CHISQ1
      DEL2 = CHISQ3 - 2*CHISQ2 + CHISQ1
      DELTA1 = DELTA * (DEL1/DEL2 + 1.5)
      A(J) = A(J)  - DELTA1
      CHISQ2 = CALCCHISQ()          !--- AT NEW LOCAL MINIMUM
C  -ADJUST DELTA FOR CHANGE OF 2 FROM CHISQ AT MINIMUM-
      AA = DEL2/2                   !--- CHISQ = AA*A(J)**2 + BB*A(J) + CC
      BB = DEL1 - DEL2/2
      CC = CHISQ1-CHISQ2
      DISC = BB**2 -4*AA*(CC-2)     !--- CHISQR DIFF(ERENCE) = 2
      IF (DISC .GT. 0 ) THEN     !--- IF NOT, THEN PROBABLY NOT PARABOLIC YET
         DISC =SQRT(DISC)
         ALPH = (-BB - DISC)/(2*AA)
```

```
      X1 = ALPH*DELTA +  A(1) - 2*DELTA  !--- A(J) AT CHISQ MINIMUM+2
      DISC = BB**2 - 4*AA*CC
      IF (DISC.GT.0 ) THEN
         DISC = SQRT(DISC)
      ELSE
         DISC = 0                        !--- ELIM ROUNDING ERR
      ENDIF
      ALPH = (-BB - DISC)/(2*AA)
      X2 = ALPH*DELTA + A(1) - 2*DELTA !--- A(J) AT CHISQ MINIMUM
      DELTA = X1 - X2
      DELTAA(J) = DELTA
    ENDIF                              !--- IF (DISC .GT. 0 ...
100 CONTINUE                           !--- DO J
    CHISQR = CHISQ2
    RETURN
    END

C PROGRAM 8.2: \CHAPT-8\GRADSEAR.FOR
C NON-LINEAR LEAST-SQUARES FIT BY GRADIENT SEARCH METHOD
C USES FITFUNC8, FITUTIL
      SUBROUTINE CALCGRAD
      INTEGER J
      REAL  SUM, DELTA, FRACT/0.001/, CALCCHISQ
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      SUM = 0
      DO 100 J = 1, M
         CHISQ2 = CALCCHISQ()
         DELTA  = FRACT * DELTAA(J) !--- DIFF(ERENTIAL ELEMENT FOR GRADENT
         A(J)   = A(J) + DELTA
         CHISQ1 = CALCCHISQ()
         A(J)   = A(J) - DELTA
         GRAD(J) = CHISQ2 - CHISQ1          !--- 2*DELTA*GRAD
         SUM    = SUM + GRAD(J)**2
100   CONTINUE
      DO 200 J = 1, M
         GRAD(J) =  DELTAA(J)*GRAD(J)/SQRT(SUM)    !--- STEP * GRAD
200   CONTINUE
      RETURN
      END

      SUBROUTINE GRADLS(CHISQR, STEPDOWN)
      REAL CHISQR, STEPDOWN
      REAL STEPSUM, STEP1, CALCCHISQ
      INTEGER J
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      CALL CALCGRAD          !--- CALCULATE THE GRADIENT
C -EVALUATE CHISQR AT NEW POINT AND MAKE SURE CHISQR DECREASES-
      CHISQ3 = CHISQ2 + 1
      DO WHILE (CHISQ3 .GT. CHISQ2)
         DO J = 1, M
            A(J) = A(J) + STEPDOWN * GRAD(J) ! SLIDE DOWN
         ENDDO
         CHISQ3 = CALCCHISQ()
         IF (CHISQ3 .GE. CHISQ2 ) THEN
            DO J = 1, M      ! MUST HAVE OVERSHOT MINIMUM
               A(J) = A(J) - STEPDOWN * GRAD(J) ! RESTORE
            ENDDO
            STEPDOWN = STEPDOWN/2          ! DECREASE STEPSIZE
         ENDIF
      ENDDO
      STEPSUM = 0
C -INCREMENT PARAMETERS UNTIL CHISQR STARTS TO INCREASE-
      DO WHILE (CHISQ3 .LT. CHISQ2)
         STEPSUM = STEPSUM + STEPDOWN   ! COUNTS TOTAL INCREMENT
         CHISQ1 = CHISQ2
         CHISQ2 = CHISQ3
         DO J = 1, M
            A(J) = A(J) + STEPDOWN * GRAD(J)
         ENDDO
         CHISQ3 = CALCCHISQ()
      ENDDO             !DOWHILE
C -FIND MINIMUM OF PARABOLA DEFINED BY LAST THREE POINTS-
      STEP1=STEPDOWN*((CHISQ3-CHISQ2)/(CHISQ1-2*CHISQ2+CHISQ3)+0.5)
      DO J = 1, M
         A(J) = A(J) - STEP1 * GRAD(J)    ! MOVE TO MINIMUM
      ENDDO
      CHISQR = CALCCHISQ()
      STEPDOWN = STEPSUM              ! START WITH THIS NEXT TIME
      RETURN
      END

C PROGRAM 8.3: \CHAPT-8\EXPNDFIT.FOR
C NON-LINEAR LEAST-SQUARES FIT BY EXPANSION OF THE FITTING FUNCTION
C USES  FITFUNC8, MAKEAB8, MATRIX
      SUBROUTINE CHIFIT(CHISQR)
      INTEGER J
      REAL DET, CALCCHISQ
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      CALL MAKEBETA
      CALL MAKEALPHA
      CALL MATINV(M, ALPHA, DET)      !--- INVERT MATRIX
      CALL LINEARBYSQUARE(M,BETA,ALPHA,DA)  !--- EVALULATE PARAM
INCREMENTS
      DO 100 J = 1, M
         A(J) = A(J) + DA(J)             !--- INCREMENT TO NEXT SOLUTION.
100   CONTINUE
      PRINT *,'A',(A(J),J=1,M)
      CHISQR =  CALCCHISQ()
      RETURN
      END
```

```
C PROGRAM 8.4: \CHAPT-8\MARQFIT.FOR
C NON-LINEAR FIT BY THE GRADIENT-EXPANSION (MARQUARDT) METHOD
C USES FITFUNC9, MAKEAB8, MATRIX
      SUBROUTINE MARQUARDT(CHISQR, XICUT, LAMBDA)
      INTEGER J
      REAL CHISQR, XICUT, LAMBDA
      REAL DET, CALCCHISQ
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      DO
         CALL MAKEBETA
         CALL MAKEALPHA
         DO 100 J = 1, M
            ALPHA(J,J) = (1 + LAMBDA) * ALPHA(J,J)
  100    CONTINUE
         CALL MATINV(M, ALPHA, DET)        !--- INVERT MATRIX
         IF (LAMBDA .LE. 0 ) RETURN !--- FINAL CALL TO GET THE ERROR MATRIX.
         CALL LINEARBYSQUARE(M,BETA,ALPHA,DA)!--- EVAL PARAM INCREMENTS
         CHISQ1 = CHISQR
         DO 200 J = 1, M
            A(J) = A(J) + DA(J)          !--- INCR TO NEXT SOLUTION
  200    CONTINUE
         CHISQR = CALCCHISQ()
         IF ( CHISQR .LE. CHISQ1 + XICUT ) RETURN
         DO 300 J = 1, M
            A(J) = A(J)-DA(J)    !--- RETURN TO PREV SOLUTION
  300    CONTINUE
         CHISQR = CALCCHISQ()
         LAMBDA = 10*LAMBDA  !--- AND REPEAT THE CALC, WITH LARGER LAMBDA
      END DO
      END

C PROGRAM 8.5: \CHAPT-8\FITFUNC8.FOR
C USES FITVARS
C -THE FOLLOWING ROUTINES ARE GENERAL FOR FITTING ANY FUNCTION-
      SUBROUTINE CALCULATEY
      REAL YFUNCTION
      INCLUDE '\CHAPT-6/FITVARS.FOR'
      DO 100 I = 1, NPTS
         YCALC(I) = YFUNCTION(X(I))
  100 CONTINUE
      RETURN
      END

      REAL FUNCTION CALCCHISQ()
      REAL  CHI2, YFUNCTION
      INCLUDE '\CHAPT-6/FITVARS.FOR'
      CHI2=0.
      DO 100 I = 1, NPTS
         CHI2 = CHI2 + ( (Y(I)-YFUNCTION(X(I)))/SIGY(I))**2
  100 CONTINUE
      CALCCHISQ = CHI2
```



```
      RETURN
      END

C -STANDARD DEVIATION CALC'D FROM CHISQ CHANGE OF 1 (PARABOLA FIT)
      REAL FUNCTION  SIGPARAB(J)
      INTEGER J
      REAL CALCCHISQ
      INCLUDE '\CHAPT-6/FITVARS.FOR'
      CHISQ2 = CALCCHISQ()
      A(J)   = A(J) + DELTAA(J)
      CHISQ3 = CALCCHISQ()
      A(J)   = A(J) - 2*DELTAA(J)
      CHISQ1 = CALCCHISQ()
      A(J)   = A(J) + DELTAA(J)
      SIGPARAB = DELTAA(J)*SQRT(2/(CHISQ1-2*CHISQ2+CHISQ3))
      RETURN
      END

C -STANDARD DEVIATION CALC'D FROM DIAGONAL TERMS IN ERROR MATRIX
      REAL FUNCTION SIGMATRX(J)
      INTEGER J
      REAL SIG
      INCLUDE '\CHAPT-6/FITVARS.FOR'
      SIG = SQRT(ABS(ALPHA(J,J)))
      IF (ALPHA(J,J) .LT. 0 ) SIG = - SIG  !--- NOTE- AN ERROR
      SIGMATRX = SIG
      RETURN
      END

C PROGRAM 8.6: \CHAPT-8\MAKEAB8.FOR
C MATRIX SET-UP FOR NON-LINEAR FITS
C USES FITFUNC8, NUMDERIV
C
      SUBROUTINE MAKEBETA    !---MAKE BETA MATRICES FOR NON-LINEAR FITTING
      INTEGER  J
      INCLUDE '\CHAPT-6/FITVARS.FOR'
      DO 100 J = 1, M
         BETA(J) = -0.5*DXISQ_DA(J)
  100 CONTINUE
      RETURN
      END

      SUBROUTINE MAKEALPHA  !--- ALPHA MATRICES FOR NON-LINEAR FITTING
      INTEGER  J, K
      INCLUDE '\CHAPT-6\FITVARS.FOR'
      DO 100 J = 1, M
         ALPHA(J,J) =  0.5 * D2XISQ_DA2(J)
         IF (ALPHA(J,J) .EQ. 0 ) THEN
            PRINT *, 'DIAGONAL ELEMENT IS ZERO, J =',J
         STOP
         ENDIF
```

```
      IF (J .GT. 1 ) THEN
         DO 200 K = 1, J-1
            ALPHA(J,K) = 0.5*D2XISQ_DAJK(J,K)
            ALPHA(K,J) =  ALPHA(J,K)
200      CONTINUE              !--- DO K
      ENDIF                    !--- IF J
100   CONTINUE                 !---  DO J
      DO 300 J = 1, M
         IF (ALPHA(J,J) .LT. 0 ) THEN
         ALPHA(J,J) = -ALPHA(J,J)
         IF (J .GT. 1 ) THEN
            DO 400 K = 1, J-1
               ALPHA(J,K) = 0
               ALPHA(K,J) = 0
400         CONTINUE             !--- DO K
         ENDIF                   !--- IF J
      ENDIF                      !--- IF ALPHA
300   CONTINUE                   !--- FOR J
      RETURN
      END
```

E.5 Routines from Chapter 9

```
C PROGRAM 9.1: \CHAPT-9\LORENFIT.FOR
C MAIN CALLING ROUTINE FOR FIT TO LORENTZIAN + POLYNOMIAL
C USES  FITFUNC9, MARQFIT, MATRIX,  NUMDERIV, MAKEAB8,  FITUTIL
      PROGRAM LORENFIT
      CHARACTER*40  TITLE
      INTEGER TRIAL, J
      REAL   XSHIFT, CHISQR, LAMBDA, YFUNCTION
      REAL  STEPSCALE(4)/ 0.49999, 0.99999, 0.001, 0.001/
      INCLUDE 'C:\F\CHAPT-6\FITVARS.FOR'
      CHICUT    = 0.01
      LAMBDA    = 0.001              ! FOR MARQUARDT METHOD ONLY
      STEPSIZE  = STEPSCALE(4)       ! SCALES DELTAA[J]
      CALL FETCHDATA('\F\CHAPT-9\SINGLE.HST',TITLE)
      XSHIFT = (X(2)- X(1))/2
      DO  J = 1, NPTS
         X(J) = X(J) + XSHIFT        ! MOVE TO BIN CENTER
      ENDDO
      CALL FETCHPARAMETERS           ! USES NPTS, MUST FOLLOW FETCHDATA
      TRIAL     = 0
      CHISQR    = CALCCHISQ()
      CHIOLD = CHISQR + CHICUT +1
      DO WHILE (ABS(CHIOLD - CHISQR) .GT. CHICUT)
         CHIOLD = CHISQR
         PRINT *,'TRIAL #',TRIAL,' CHISQ = ',CHISQR
         PRINT *, (A(J), J = 1, M)
         CALL MARQUARDT(CHISQR, CHICUT, LAMBDA)
         TRIAL  = 1 + TRIAL
```



```
         ENDDO
         CALL CALCULATEY
         CALL MARQUARDT(CHISQR,CHICUT,0)        ! GET ERROR MATRIX
         DO J = 1, M
            SIGA(J) = SIGMATRX(J)               ! ERROR MATRIX
         ENDDO
         CALL OUTPUT(.TRUE., 'CON', CHISQR,TITLE) ! WITH ERROR MATRIX
         DO J = 1, NPTS
            X(J) = X(J) - XSHIFT                ! RESTORE TO LEFT EDGE
         ENDDO
         CALL PLOTIT('LORENFIT.SCR',.FALSE.,.TRUE.,! SCRIPT FILE, LOG?, SPLINE?
     1   'H', 0.0,                  ! HIST, 0(NOT USED)
     2   0.0, 0.0, 3.0, 220.0,          ! X1, Y1, X2, Y2 FOR PLOT
     3   6, 6,                      ! NUM GRID MARKS X,Y
     4   'E (GEV)', 'NUMBER OF COUNTS')    ! LABELS
C -PLOT THE BACKGROUND-
         A(4) = 0.0
         A(7) = 0.0
         DO J = 1, NPTS
            YCALC(J) = YFUNCTION(X(J))
         ENDDO
         CALL SPLINEMAKE(NPTS,0,0,X,YCALC)
         CALL SCURVE(1, 40, 5, 0.025, X)    ! SPLINE CURVE
         CALL CLOSEGRAPHICS
         END

C LORENTZIAN PEAK ON A QUADRATIC BACKGROUND
         REAL FUNCTION YFUNCTION(XX) ! LORENTZIAN ON POLYNOMIAL
         REAL XX
         REAL YY, PI/3.1415927/
         INCLUDE '\F\CHAPT-6\FITVARS.FOR'
         YY = A(1) + A(2)*XX + A(3)*XX**2 + A(4)*A(6)/(2*PI)
     1   /((XX-A(5))**2 + A(6)**2/4)
         YFUNCTION = YY
         RETURN
         END
         INCLUDE '\F\CHAPT-6\FITUTIL.FOR'
         INCLUDE '\F\CHAPT-9\FITFUNC9.FOR'
         INCLUDE '\F\CHAPT-8\MARQFIT.FOR'       ! MARQUARDT METHOD
         INCLUDE '\F\CHAPT-8\MAKEAB8.FOR'       ! USED BY MARQFIT
         INCLUDE '\F\CHAPT-8\NUMDERIV.FOR'      ! USED BY MARQFIT
         INCLUDE '\F\APPEND-B\MATRIX.FOR'       ! USED BY MARQFIT
```

E.6 Routines from Chapter 10

```
C PROGRAM 10.1: \CHAPT-10\MAXLIKE.FOR
C DIRECT MAXIMUM LIKELIHOOD EXAMPLE
C  USES FITUTIL, QUIKSCRP
      PROGRAM MAXLIKE
      REAL    SIGTAU, TAUMAX, MAXM  !--- M IS LOG OF LIKELIHOOD FUNCTION
      INCLUDE '\CHAPT-10\MAXLINCL.FOR'
```

```
      CALL GETDATA('\CHAPT-10\TEST.DAT')     !--- WAS DA50
      CALL SEARCH(TAUMAX, MAXM)
      CALL WRITEOUTPUT(SIGTAU, TAUMAX, MAXM)
      CALL PLOTLIKECURVE(TAUMAX, SIGTAU, MAXM)
      CALL CLOSEGRAPHICS
      END

      SUBROUTINE GETDATA(INFILE)
      INTEGER  IEVNUM
      CHARACTER*(*) INFILE
      CHARACTER TITLE(80)
      INCLUDE '\CHAPT-10\MAXLINCL.FOR'
      C = 3.00
      LOSEARCH = 0.50
      HISEARCH = 1.5          !--- SEARCH RANGE
      TAUSTEP  = 0.01
      XLO      = 0.50         !---  PLOT RANGE
      XHI      = 1.2
      YLO      = 0.0
      YHI      = 1.2
      NTRIALS  = (HISEARCH - LOSEARCH)/TAUSTEP
      OPEN(5, INFILE)         !---  INPUT DATA FILE
      READ(5, *) TITLE
      PRINT *,' ',TITLE
      READ(5, *) NEVENTS, MASS, D1, D2
      IEVNUM = 1
      NEVENTS = 0
      DO WHILE (IEVNUM .GT.0)
         READ(5, *) IEVNUM, XPRODUCTION, PLAB, XDECAY
         IF (IEVNUM .GT.0) THEN
            IF ((XDECAY .GE. D1) .AND. (XDECAY .LT. D2)) THEN
               NEVENTS = 1 + NEVENTS
               LTOTSCALE =  MASS/(C*PLAB)  !---  = 1/(C*BETA*GAMMA)
               TIMES(NEVENTS)=(XDECAY - XPRODUCTION)*LTOTSCALE !---PROPER T
C  CONVERT D1 AND D2 TO TIME LIMITS, LOTLIM AND HITLIM,
C   I.E., INTEGRATION LIMITS IN PROPER TIME FROM THE PRODUCTION VERTEX.
               LOTLIM(NEVENTS) = (D1 - XPRODUCTION)*LTOTSCALE
               HITLIM(NEVENTS) = (D2 - XPRODUCTION)*LTOTSCALE
            ENDIF
         ENDIF
      ENDDO
      PRINT *, 'END OF FILE - ', IEVNUM, ' EVENTS READ'
      PAUSE
      RETURN
      END

      REAL FUNCTION LOGPROB(K, TAU)
      INTEGER K
      REAL TAU
      REAL A, B
      INCLUDE '\CHAPT-10\MAXLINCL.FOR'
```



```
C     D1 AND D2 ARE BEGINNING AND END OF THE FIDUCIAL REGION.
C     MUST CVT TO LOTLIM AND HITLIM WHICH ARE INTEGRATION LIMITS IN PROPER
      TIME,
C     MEASURED FROM PRODUCTION VERTEX.
C     NOW, CALC PROBABILITY-
      B = EXP(-HITLIM(K)/TAU)
      A = EXP(-LOTLIM(K)/TAU)
      PROB = EXP(-TIMES(K)/TAU)/(TAU*(A - B))
      LOGPROB = ALOG(PROB)
      RETURN
      END

      REAL FUNCTION LOGLIKE(T)
      REAL T, LOGPROB
      INTEGER I
      REAL    M, PROB
      INCLUDE '\CHAPT-10\MAXLINCL.FOR'
      M = 0.0
      DO 100 I = 1, NEVENTS
         PROB = LOGPROB(I,T)
         M = PROB + M
  100 CONTINUE
      LOGLIKE = M
      RETURN
      END

      SUBROUTINE SEARCH(TAUATMAX, MAXM)
      REAL TAUATMAX, MAXM
      INTEGER TRIAL
      REAL M1, M2, M3, DEL1, DEL2, DELTA1, TAU, MLIKELI, LOGLIKE
      INCLUDE '\CHAPT-10\MAXLINCL.FOR'
      M2   = -1000
      MAXM  = -1.0E20
      TAU   =  LOSEARCH
      DO 100 TRIAL = 0, NTRIALS
         MLIKELI = LOGLIKE(TAU)
         PRINT *,'TRIAL',TRIAL,' TAU=', TAU,' LOG LIKELIHOOD=',MLIKELI
         M3 = MLIKELI
         IF (M3 .GT. M2 ) THEN   !---  REMEMBER, THESE ARE NEGATIVE
            M1 = M2
            M2 = M3
         ELSE                    !---  LEAVING MAXIMUM
C     FIND MAXIMUM OF PARABOLA DEFINED BY LAST THREE POINTS-
            DEL1    = M2 - M1
            DEL2    = M3 - 2*M2 + M1
            DELTA1  = TAUSTEP * (DEL1/DEL2 + 1.5)
            TAU     = TAU  - DELTA1
            TAUATMAX = TAU
            MAXM    = LOGLIKE(TAU)   !---  AT  MAXIMUM OF PARABOLA
            RETURN
         ENDIF
```

```
      TAU = TAU + TAUSTEP
100 CONTINUE
    RETURN
    END

    REAL FUNCTION ERROR(T, DT)   !---  1/SQRT(-2ND DERIVATIVE OF LOG(L))
    REAL T, DT
    REAL  T1, T2, D2YDT2, ERR, LOGLIKE
    T1 = T - DT
    T2 = T + DT
    D2YDT2 = (LOGLIKE(T2) - 2*LOGLIKE(T) + LOGLIKE(T1))/DT**2
    ERR   = 1/SQRT(-D2YDT2)
    ERROR  = ERR
    RETURN
    END

C PROGRAM 10.2 \CHAPT-10\MAXLINCL.FOR    (WEBSITE)
C INCLUDE FILE FOR MAXLIKE
```

E.7 Routines from Chapter 11

```
C PROGRAM 11.1: \CHAPT-11\CHI2PROB.FOR
C CALCULATE CHI^2 PROB. DENS. & THE CHI^2 PROB. INTEGRAL
C USES CHIPROBDENS AND CHIPROB
    PROGRAM CHI2PROB
    REAL CHI2,  CHIPROB
    INTEGER NFREE
    PRINT *,'CALCULATE CHI2 PROBABILITY DENSITY FUNCTION & INTEGRAL',
   1  ' PROBABILITY'
    PRINT *, 'TYPE  NUM DEG OF FREEDOM AND  CHI2.  (EXIT ON ^C)'
    READ *,  NFREE, CHI2
    PRINT 1000, CHIPROBDENS(CHI2, NFREE), CHIPROB(NFREE, CHI2)
1000 FORMAT(' CHI^2 PROB. DENS. = ',F7.3,' , CHI^2 PROBABILITY=',F7.3)
    PRINT *,' ***** NOTE THAT TABLE C.4 REFERS TO CHI^2/NFREE****'
    END

C THE FOLLOWING THREE ROUTINES ARE INCLUDED
C   IN THE PROGRAM UNIT  C:\CHAPT-6\FITUTIL.FOR (WEBSITE)
    REAL FUNCTION CHIPROB(NFREE, CHI2)      !---  MAX NFREE =  56
    EXTERNAL CHIX
    COMMON/UTIL/ GLSIMPS
    REAL CHIX, SIMPSON, GLSIMPS
    INTEGER NFREE
    REAL PI, CHI2, CLIM, INTFROMLIM
    DATA   CLIM /2/,              !---  EXPANSION LIMIT FOR NFREE = 1
   1  INTFROMLIM /0.157/,         !---  INTEGRAL FROM CLIM TO INFINITY
   2     DX0 /0.2/                !---  DETERMINES ACCURACY OF INTEGRATION
   3     PI/3.14159/
    INTEGER NINT
    IF (CHI2 .GE. 1) THEN
```



```
      NINT = 5
    ENDIF
    IF (CHI2 .GT. 15*SQRT(NFREE) ) THEN !---  QUICK CUTOUT
      CHIPROB = 0
    ELSE
      GLSIMPS = FLOAT(NFREE)/2          !---  GLSIMPS IS GLOBAL FOR CHIX
     IF (NFREE .EQ. 1 )   THEN
        IF (CHI2 .LT. CLIM ) THEN
       CHIPROB = 1-SQRT(CHI2/2/PI)*
   1    (2 - CHI2*(1/3 - CHI2*(1/20 - CHI2*(1/168 - CHI2/1728))))
        ELSE
          CHIPROB = INTFROMLIM - SIMPSON(CHIX,NINT,CLIM,CHI2)
   1                /GAMMA(NFREE/2.0)/2.0**(NFREE/2.0)
        ENDIF                           !---  IF (CHI2 ...)
      ELSE IF (NFREE .EQ. 2 ) THEN
        CHIPROB = EXP(-CHI2/2)          !---  INTEGRABLE
      ELSE
        CHIPROB =  1 - SIMPSON(CHIX, NINT, 0, CHI2)
   1            /GAMMA(NFREE/2.0)/2.0**(NFREE/2.0)
      ENDIF                             !---  IF (NFREE ...)
    ENDIF
    RETURN                              !---  IF (NFREE ...)
    END

    REAL FUNCTION CHIPROBDENS(X,NFREE)
    REAL NUM, DEN, H, X
    INTEGER NFREE
    H = NFREE/2.0
    NUM = X**(H-1) * EXP(-X/2)
    DEN =  2**H * GAMMA(H)
    CHIPROBDENS = NUM/DEN
    RETURN
    END

C  USED BY CHIPROB (FOR SIMPSON WHICH ALLOWS ONLY 1 ARGUMENT.)
    REAL FUNCTION CHIX(X)
    COMMON/UTIL/ GLSIMPS
    REAL GLSIMPS
    REAL X
    IF (X.EQ.0) THEN
      CHIX = 0.0
    ELSE
      CHIX = X**(GLSIMPS-1)*EXP(-X/2)  !---  GLSIMPS = H = NFREE/2
    ENDIF
    RETURN
    END

C THIS FOLLOWING ROUTINE IS INCLUDED
C   IN THE PROGRAM UNIT \CHAPT-6\FITUTIL.FOR   (WEBSITE)
C APPROXIMATE GAMMA FUNCTION WITH H = NFREE/2
    REAL FUNCTION GAMMA(H)
```

```
      REAL H, PI/3.1415927/
      GAMMA = SQRT(2.0*PI) * EXP(-H)*(H**(H-0.5)) * (1.0 + 0.0833/H)
      RETURN
      END

C PROGRAM 11.2: \CHAPT-11\LCORPROB.FOR
C CALCULATE LINEAR CORRELATION PROBABILITY INTEGRAL
C USES LCORLATE
      PROGRAM LCORPROB
      INTEGER NOBSERV
      REAL LINCORPROB, RCORR
      PRINT *, 'TEST INTEGRAL OF LINEAR CORRELATION FUNCTION'
      PRINT *, 'TYPE-# OBSERVATIONS, LINEAR CORRELATION COEFFICIENT: '
      READ *, NOBSERV, RCORR
      PRINT *, 'INTEGRAL CORRELATION FUNCTION= ',
     1  LINCORPROB(NOBSERV-2, RCORR)
      END
      INCLUDE '\CHAPT-11\LCORLATE.FOR'

C LINEAR-CORRELATION PROBABILITY FUNCTION AND INTEGRAL
C USES FITUTIL
      REAL FUNCTION LINCORPROB(NFREE, HILIM)
      EXTERNAL LINCORREL    !--- FOR USE IN FUNCTION SIMPSON
      INTEGER NFREE
      REAL   HILIM
      REAL DX /0.01/, LOLIM/0.0/, LINCORREL, SIMPSON
      INTEGER  NINT
      COMMON/UTIL/GLSIMPS
      GLSIMPS = NFREE !--- GLOBAL FOR FUNCTION LINCORREL (FOR SIMPSON)
      NINT  = INT((HILIM - LOLIM)/DX)
      LINCORPROB = 1-2*SIMPSON(LINCORREL, NINT, LOLIM,  HILIM)
      RETURN
      END

      REAL FUNCTION LINCORREL(R)
      REAL R
      COMMON/UTIL/GLSIMPS      !--- GLSIMS = NFREE MUST BE GLOBAL FOR
      DATA SQRTPI/1.7724539/   !   FUNCT 'SIMPSONS' WHICH ALLOWS ONLY 1 ARG
      LINCORREL = GAMMA((GLSIMPS+1)/2)/GAMMA(GLSIMPS/2)
     1    *EXP( (GLSIMPS-2)/2 * ALOG(1 - R**2))/SQRTPI
      RETURN
      END
```

E.8 Routines from Appendix A

```
PROGRAM A.1 SIMPSON
C THE FOLLOWING ROUTINE IS INCLUDED
C   IN THE PROGRAM UNIT \CHAPT-6\FITUTIL    (WEBSITE)
C -SIMPSON'S RULE FOR "FUNCTX(X:REAL):REAL"
C  IF FUNCTX HAS OTHER PARAMETERS, THEY MUST BE GLOBAL, E.G., GLSIMPS
```



```
      REAL FUNCTION SIMPSON(FUNCTX, NINTS, LOLIM, HILIM) !--- 2 CALCS/INTERVAL
      EXTERNAL FUNCTX !--- THIS STATEMENT REQ'D IN CALLING PGM ALSO
      REAL  FUNCTX, SUM, X, DX, LOLIM, HILIM
      INTEGER  NINTS, I
       X = LOLIM
       DX = (HILIM - LOLIM)/(2*NINTS)
       SUM=FUNCTX(X)
       SUM= SUM - FUNCTX(HILIM)
       DO 100 I = 1, NINTS
       X=X+2*DX
         SUM=SUM + 4*FUNCTX(X-DX) + 2*FUNCTX(X)
  100   CONTINUE
      SUM = SUM
        SIMPSON = SUM*DX/3.0
        RETURN
       END

PROGRAM A.2 SPLINE INTERPOLATION
C PROGRAM A.1: \APPEND-A\SPLINTST.FOR
C TEST CUBIC SPLINE INTERPOLATION
      PROGRAM SPLINTST
      CHARACTER TITLE(80)
      REAL D2A, D2B, XS, X(100), Y(100), SPLINEINT
      INTEGER N, I
      OPEN(5,'\APPEND-A\SPLINE.DAT')     !--- TEST DATA FILE
      READ(5,1000) TITLE
      PRINT 1000, ' ',TITLE
 1000 FORMAT(80A1)
      READ(5,*) N, D2A, D2B !--- NO. OF POINTS, 2ND DERIVATIVES AT BOUNDARY
      PRINT *,'DATA TABLE: N=', N
      PRINT *,'    X              Y'
      DO 100 I = 1, N
         READ(5,*) X(I), Y(I)
         PRINT *, X(I), Y(I)
  100 CONTINUE
      CALL SPLINEMAKE(N, D2A, D2B, X, Y)
      CLOSE(5)
  200 PRINT *,'TYPE A VALUE OF X  (EXIT WITH ^C)'
      READ *, XS
      PRINT *, 'INTERPOLATED Y = ', SPLINEINT(XS)
      GOTO 200
      END

C ROUTINES FOR CUBIC SPLINE INTERPOLATION.
C CONSTANT INTERVALS IN THE INDEPENDENT VARIABLE ARE ASSUMED.
      SUBROUTINE SPLINEMAKE(NN, D2YDX2A, D2YDX2B, XIN, YIN)
      INTEGER NN
      REAL   D2YDX2A, D2YDX2B, XIN(100), YIN(100)
C -COMMON VARIABLES SET IN SPLINEMAKE, USED IN SPLINEINT-
      COMMON/SPLINES/N, H, XX(100), YY(100), D2YDX2(100)
      INTEGER N
```

```
      REAL H, XX, YY, D2YDX2
      INTEGER I
      REAL  A(100), DELT1(100), DELT2(100), B(100)
      N = NN        !--- USED BY SPLININT, THROUGH COMMON/SPLINES/
      H = (XIN(N) - XIN(1))/(N-1)
      DO 100 I = 1, N
         XX(I) = XIN(I)
         YY(I) = YIN(I)
100   CONTINUE
      D2YDX2(1) = D2YDX2A !---  END VALUES OF 2ND DERIVATIVES FROM INPUT
      D2YDX2(N) = D2YDX2B
      A(2) = 4
      DO 200 I =  3, N-1
         A(I) = 4-1/A(I-1)                !---  COEFFICIENTS
200   CONTINUE
      DO 300 I = 2, N
         DELT1(I) = YIN(I) - YIN(I-1)     !---  1ST DIFFERENCES
300   CONTINUE
      DO 400 I = 2, N-1                   !---  2ND DIFFERENCES X 6
         DELT2(I) = 6*(DELT1(I+1) - DELT1(I))/(H*H)
400   CONTINUE
      B(2) = DELT2(2) - D2YDX2(1)      !---  B COEFFICIENTS
      DO 500 I= 3, N-1
         B(I) = DELT2(I) - B(I-1)/A(I-1)
500   CONTINUE
      B(N-1) = B(N-1) - D2YDX2(N)
      D2YDX2(N-1) = B(N-1)/A(N-1)
      DO 600 I = N-2, 2, -1
         D2YDX2(I) = (B(I) - D2YDX2(I+1))/A(I) !---  2ND DERIVATIVES
600   CONTINUE
      RETURN
      END

      REAL FUNCTION DYDX(I) !---  FIRST DERIVATIVE (WEBSITE)
      INTEGER I
      COMMON/SPLINES/N, H, XX(100), YY(100), D2YDX2(100)
      INTEGER N
      REAL H, XX, YY, D2YDX2
      DYDX = (YY(I+1)-YY(I))/H - H*(D2YDX2(I)/3+D2YDX2(I+1)/6)
      RETURN
      END

      REAL FUNCTION D3YDX3(I) !---  THIRD DERIVATIVE (WEBSITE)
      INTEGER I
      COMMON/SPLINES/N, H, XX(100), YY(100), D2YDX2(100)
      INTEGER N
      REAL H, XX, YY, D2YDX2
      D3YDX3 = (D2YDX2(I+1) - D2YDX2(I))/H
      RETURN
      END
```



```
      REAL FUNCTION SPLINEINT(X) !--- INTERPOLATE IN TABLE (FROM SPLINEMAKE)
      REAL X
      COMMON/SPLINES/N, H, XX(100), YY(100), D2YDX2(100)
      INTEGER N
      REAL H, XX, YY, D2YDX2, DYDX, D3YDX3, DX
      INTEGER I
      I = INT((X-XX(1))/H)+1
      IF (I .LT. 1 )  I = 1
      IF (I .GT. N-1 ) I = N-1
      DX = X -XX(I)
C -INTERPOLATE
      IF (I .EQ. N ) THEN
         SPLINEINT = YY(I)
      ELSE
         SPLINEINT = YY(I) + (DYDX(I) + (D2YDX2(I)/2 +D3YDX3(I)/6*DX)*DX)*DX
      ENDIF
      RETURN
      END
```

E.9 Routines from Appendix B

```
C PROGRAM B.1: \APPEND-B\MATRIX.FOR
C INVERT A SQUARE MATRIX
C USES FITVARS
      SUBROUTINE MATINV(M, MARRAY, DET)
      INTEGER M
      REAL MARRAY(10,10), DET
      INTEGER IK(10), JK(10)
      INTEGER I, J, K, L
      REAL AMAX, SAVE
      DET=0
C -FIND LARGEST ELEMENT
      DO 100  K = 1, M
         AMAX=0
1500     DO 200  I = K, M
            DO 300 J = K , M
               IF ( ABS(MARRAY(I,J)) .GT. ABS(AMAX) )  THEN
                  AMAX = MARRAY(I,J)
                  IK(K) = I
                  JK(K) = J
               ENDIF
300         CONTINUE          !--- DO J
200      CONTINUE             !--- DO I
         IF (AMAX .EQ. 0 )  RETURN  !--- WITH 0 DETERMINANT AS SIGNAL
         DET = 1
C -INTERCHANGE ROWS AND COLUMNS TO PUT AMAX IN MARRAY(K,K)
         I = IK(K)
         IF (I .LT. K ) THEN
            GOTO 1500
         ELSEIF (I .GT. K ) THEN
            DO 400 J = 1, M
```

```
            SAVE = MARRAY(K,J)
            MARRAY(K,J) = MARRAY(I,J)
            MARRAY(I,J) = -SAVE
 400      CONTINUE                !--- DO J
        ENDIF                     !--- IF I
        J = JK(K)
        IF (J .LT. K ) THEN
           GOTO 100
        ELSEIF (J .GT. K ) THEN
           DO 500 I = 1, M
            SAVE = MARRAY(I,K)
            MARRAY(I,K) = MARRAY(I,J)
            MARRAY(I,J) = -SAVE
 500      CONTINUE  !--- DO I
        ENDIF  !--- IF J
C -ACCUMULATE ELEMENTS OF INVERSE MATRIX
        DO 600 I = 1, M
           IF (I .NE. K )
   1         MARRAY(I,K) = -MARRAY(I,K)/AMAX
 600    CONTINUE !--- DO I
        DO 700 I = 1, M
           DO 800 J = 1, M
              IF ((I .NE. K) .AND. (J .NE. K) )
   1          MARRAY(I,J) = MARRAY(I,J) + MARRAY(I,K)*MARRAY(K,J)
 800      CONTINUE                !--- DO J
 700    CONTINUE                  !--- DO I
        DO 900 J = 1, M
           IF (J .NE. K )
   1       MARRAY(K,J) = MARRAY(K,J)/AMAX
 900    CONTINUE                  !--- DO J
        MARRAY(K,K) = 1/AMAX
        DET = DET * AMAX
 100  CONTINUE                    !--- DO K
C -RESTORE ORDERING OF MATRIX
      DO 1000 L = 1, M
        K = M + 1 - L
        J = IK(K)
        IF (J .GT. K ) THEN
           DO 1100 I = 1, M
            SAVE = MARRAY(I,K)
            MARRAY(I,K) = -MARRAY(I,J)
            MARRAY(I,J) = SAVE
 1100     CONTINUE                !--- DO I
        ENDIF                     !--- IF J
        I = JK(K)
        IF (I .GT. K ) THEN
           DO 1200 J = 1, M
           SAVE = MARRAY(K,J)
           MARRAY(K,J) = -MARRAY(I,J)
           MARRAY(I,J) = SAVE
 1200   CONTINUE !--- DO J
```



```
        ENDIF                     !--- IF I
 1000 CONTINUE                    !--- DO L
      RETURN
      END

      SUBROUTINE LINEARBYSQUARE(M, A, B, C)  !--- MATRIX PRODUCT
      INTEGER M
      REAL A(10), B(10,10), C(10)
      INTEGER  I,J
      DO 100 I = 1, M
         C(I)=0
         DO 200 J = 1, M
         C(I)=C(I) +A(J)*B(I,J)
 200  CONTINUE
 100     CONTINUE
      RETURN
      END
```

E.10 Routines from Appendix C

```
C PROGRAM C.1: \APPEND-C\STUDENTST.FOR
C CALCULATES BOTH THE GAUSSIAN PROBABILITY
C   AND THE STUDENT'S T PROBABILITY FOR EXCEEDING  A GIVEN VALUE
C   OF (MU-X)/SIGMA, WHERE MU IS THE MEAN VALUE OF X AND SIGMA IS
C   THE UNCERTAINTY IN THE MEAN.
C FOR SPEED, AND TO REDUCE POSSIBILITY OF OVERFLOW, WE
C   CALCULATE THE RATIO OF THE GAMMA FUNCTIONS DIRECTLY
C   IN FUNCTION GAMMACONST.
C TO IMPROVE SPEED AND ACCURACY BY USING SIMPSON'S FOR INTEGRATION
C
      PROGRAM STUDENTS_T
      REAL GP,TP, T
      INTEGER NU
      PRINT *, 'TYPE NDOF AND T = |MU - X|/SIGMA '
      READ *, NU, T
      CALL GTPROB(GP, TP, NU, T)
      PRINT 1100, 100*TP, 100*(1-TP)
      PRINT 1200, 100*GP, 100*(1-GP)
 1100 FORMAT(' PROB (STUDENT''S T)  = ',F5.2,'%, 1-PROB = ',F5.2, '%')
 1200 FORMAT(' PROB (GAUSSIAN)     = ',F5.2, '%, 1-PROB = ',F5.2,'%')
      END

      REAL FUNCTION STUDENTST(NU, T, G)    !STUDENT'S T DISTRIBUTION
      INTEGER NU
      REAL T, G, X
C X = (1/SQRT(NU*PI)  * (GAMMA((NU+1)/2)/GAMMA(NU/2))*(1+T^2/NU)^(-(NU+1)/2)
      X  = G*EXP( (-(NU+1)/2)* ALOG(1+T*T/NU))
      STUDENTST = X
      END

      REAL FUNCTION GAUSS(X)
```

```
      REAL PI/3.14159/,X
      GAUSS = EXP(-X*X/2)/SQRT(2*PI)
      RETURN
      END

C  GAUSSIAN AND STUDENT'S T PROBABILITIES
      SUBROUTINE GTPROB(GPROB, TPROB, N, T) !INTEGRAL FROM -T TO +T
      REAL GPROB, TPROB,T
      INTEGER N
      REAL GAM, T1, SUMT, SUMG, DT
      GAM   = GAMMACONST(N)        !RATIO OF GAMMAS - FOR SPEED
      DT    = 0.0001                !INTEGRATION STEP
      T1    = 0
      SUMT = 0
      SUMG = 0
      DOWHILE ((T1 .LT. T) .AND. (SUMT*DT .LT. 0.5))  !SIMPLE INTEGRATION.
C   REPLACE BY SIMPSON'S RULE FOR BETTER SPEED AND ACCURACY
         SUMT  = SUMT + STUDENTST(N,T1,GAM)
         SUMG  = SUMG + GAUSS(T1)
         T1    = T1 + DT
      ENDDO
      TPROB   = 2*SUMT*DT
      GPROB   = 2*SUMG*DT
      RETURN
      END

      REAL FUNCTION GAMMACONST(N)
C   G  = GAMMA((H+1)/2)/GAMMA(H/2)/SQRT(H*PI)
C   PRE-CALCULATE RATIO FOR SPEED AND TO  AVOID OVERFLOW
      INTEGER N
      REAL PI/3.14159/
      REAL H, Y1,Y2, G
      H = N
      Y1 = -0.5*(H+1) + 0.5*(H)  *ALOG(0.5*(H+1))
      Y2 = -0.5*H     + 0.5*(H-1)*ALOG(0.5*H)
      G  = EXP(Y1-Y2)*(1+0.0833/(0.5*(H+1)))/((1+0.0833/(0.5*H))
     1  *SQRT(H*PI))
      GAMMACONST = G
      RETURN
      END
      END
```

E.11 Routines from Appendix D

```
C PROGRAM D.1: \APPEND-D\QUIKSCRP.FOR
C CREATE A SCRIPT FILE TO DISPLAY SIMPLE GRAPHS AND HISTOGRAMS
C THE FILE IS READ AND INTERPRETED BY \APPEND-D\QDISPLAY.EXE

C PROGRAM D.2: \APPEND-D\QUIKHIST.FOR
C ASSIGNS DATA TO HISTOGRAM BINS AND PLOTS HISTOGRAM EITHER
C   AS SCREEN CHARACTERS OR IN SCREEN GRAPHISC THROUGH QUIKSCRP
```

REFERENCES

Anderson, R. L. and E. E. Houseman, *Tables of Orthogonal Polynomial Values Extended to N = 104,* Research Bulletin 297, Agricultural Experimental Station, Iowa State University (April, 1942).

Arndt, R. A. and M. H. MacGregor, Nucleon-Nucleon Phase Shift Analysis by Chi-Squared Minimization, in *Methods in Computational Physics,* vol. 6, pp. 253–296, Academic Press, New York (1966).

Baird, D. C., *Experimentation: An Introduction to Measurement Theory and Experiment Design,* Prentice-Hall, Englewood Cliffs, N.J. (1988).

Bajpai, A. C., I. M. Calus, and J. A. Fairley, *Numerical Methods for Engineers and Scientists,* Wiley, Chichester (1977).

Beers, Y., *Introduction to the Theory of Error,* Addison-Wesley, Reading, Mass. (1957).

Box, G. E. P. and M. E. Müller, A Note on the Generation of Random Normal Deviates, *Ann. Math. Statist.,* vol. 29, pp. 610–611 (1958).

David, F. N., *Tables of the Correlation Coefficients,* Cambridge University Press, London (1938).

Dixon, W. J. and F. J. Massey, Jr., *Introduction to Statistical Analysis,* McGraw-Hill, New York (1969).

Eadie, W. T., D. Drijard, F. E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics,* North-Holland, Amsterdam (1971).

Hamilton, W. C., *Statistics in Physical Science,* Ronald Press, New York (1964).

Hamming, R. W., *Numerical Methods for Scientists and Engineers,* McGraw-Hill, New York (1962).

Handbook of Chemistry and Physics, Chemical Rubber Co., Cleveland, Ohio (1973).

Hildebrand, F. B., *Introduction to Numerical Analysis,* McGraw-Hill, New York (1956).

Hoel, P. G., *Introduction to Mathematical Statistics,* Wiley, New York (1954).

IBM, *System/360 Scientific Subroutine Package, Programmer's Manual* (360A-CM-03X).

Knuth, D. E., Seminumerical Algorithms, in *The Art of Computer Programming,* vol. 2, pp. 29ff., Addison-Wesley, Reading, Mass. (1981).

Marquardt, D. W., An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *J. Soc. Ind. Appl. Math.,* vol. II, no. 2, pp. 431–441 (1963).

Melkanoff, M. A., T. Sawada, and J. Raynal, Nuclear Optical Model Calculations, in *Methods in Computational Physics,* vol. 6, pp. 2–80, Academic Press, New York (1966).

Merrington, M. and C. M. Thompson, Tables of Percentage Points of the Inverted Beta (*F*) Distribution, *Biometrica,* vol. 33, pt. 1, pp. 74–87 (1943).

Orear, J., Notes on Statistics for Physicists, UCRL-8417, University of California Radiation Laboratory, Berkeley, Calif. (1958).

Ostle, B., *Statistics in Research,* Iowa State College Press, Ames, Iowa (1963).

Pearson, K, *Tables for Statisticians and Biometricians,* Cambridge University Press, London (1924).

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes, The Art of Scientific Computing,* Cambridge University Press, New York (1986).

Pugh, E. M. and G. H. Winslow. *The Analysis of Physical Measurements,* Addison-Wesley, Reading, Mass. (1966).
"Review of Particle Physics" The European Physical Journal C, vol. 15, p. 193 (2000).
"Review of Particle Properties" Physics Letters, vol. 170B, p. 53 (1986).
Taylor, J. R., *An Introduction to Error Analysis,* University Science Books, Mill Valley, Calif. (1982).
Thompson, W. J., *Computing in Applied Science,* Wiley, New York (1984).
Wichmann, B. and D. Hill, Building a Random Number Generator, *Byte Magazine,* March, p. 127 (1987). *Applied Statistics,* vol. 31, pp. 188–190 (1982).
Young, H. D., *Statistical Treatment of Experimental Data,* McGraw-Hill, New York (1962).
Zerby, C. D., Monte Carlo Calculation of the Response of Gamma-Ray Scintillation Counters, in *Methods in Computational Physics,* vol. 1, pp. 90–133, Academic Press, New York (1963).

ANSWERS TO SELECTED EXERCISES

Chapter 1

1.1. (*a*) 5 (*b*) 2 (*c*) 2 (*d*) 5 (*e*) 4
(*f*) 1 (*g*) 3 (*h*) 3 (*i*) 3 (*j*) 4

1.3. (*a*) 980. (*b*) 84,000 (*c*) 0.0094 (*d*) 3.0×10^2
(*e*) 4.0 (*f*) NA (*g*) 5300 (*h*) 4.0×10^2
(*i*) 4.0×10^2 (*j*) 3.0×10^4

1.5. Mean = 73.48; median = 73; most probable value = 70

1.7. Standard deviation = 15.52

Chapter 2

2.2. (*a*) 20 (*b*) 6 (*c*) 120 (*d*) 270, 725

2.3. For $p = 1/2$, 0.015625, 0.093750, 0.234375, 0.31250, 0.234375, 0.093750, 0.015625

2.6. 4.1 for one lemon; 37 for two lemons; 1000 for three lemons

2.9. (*a*) $2.3 \simeq 2$ students (*b*) 8%

2.13. (*a*) 0.0011 (*b*) $\sim 3 \times 10^{-20}$

2.15. Mean number hitting counter in the 200-ns time interval:

$\bar{x} = \sum_{x=0}^{\infty} x P_P(x; \mu) = \mu$; mean number recorded $= \sum_{x=1}^{\infty} 1 P_P(x; \mu) =$ $1 - P_P(0, \mu) = 1 - e^{-\mu}$; Efficiency $= (1 - e^{-\mu})/\mu$.

2.17. $\bar{r} = \int_0^\infty r P(r) dr = 6CR^4$; $\int_0^\infty P(r) dr = 1$, so $C = 1/(2R^3)$ and $\bar{r} = 3R$

Chapter 3

3.3. The relative uncertainty in r should be one-half the relative uncertainty in L.

3.5. 1.503 ± 0.024

3.7. (*a*) 15300 ± 6700 (*b*) 165 ± 11

3.9. $\bar{n} = 3.61$; $s = 1.88$

Chapter 4

4.1. $s = 2.18$; $\sigma_\mu = 0.44$
4.3. Fig. 2.3: $\chi^2 = 1.39$ for 5 bins; $\chi^2_\nu = 0.35$
Fig. 2.4: $\chi^2 = 4.88$ for 7 bins; $\chi^2_\nu = 0.81$
4.7. Mean total counts in 1-min interval = 123.2; $\sigma = 9.4$; $\sigma_\mu = 3.0$
(*a*) Background counts in 1-min interval = 11.6; $\sigma = 1.5$
(*b*) Difference = 111.6 ± 3.3 counts per minute from the source
4.9. 32.81 ± 0.46
4.11. (*a*) $1.96\sigma = 31.0$ or 3.1%
(*b*) $1.96s = 30.1$ or 3.0%.
4.13. (*c*) $\chi^2 = 14.7$ (calculated with σ)
(*d*) $\langle\chi^2\rangle = \nu = N - 1 = 12$

Chapter 5

5.10. For 6 rows: (*b*) 8, 48, 120, 160, 120, 48, 8 (*c*) $\sigma = 1.22$

Chapter 6

6.1. $a = 114.3 \pm 9.6$; $b = 9.58 + 0.89$, $\chi^2 = 10.1$
6.4. $b = 3.60 \pm 0.03$; $\chi^2 = 11.9$

Chapter 7

7.2. $a_1 = 512.0 \pm 45.9$; $a_2 = 348.3 \pm 21.8$; $\chi^2 = 13.2$
$\alpha_{11} = 21.09$; $\alpha_{12} = \alpha_{21} = -147.1$; $\alpha_{22} = 476.1$
7.4. All terms: $\chi^2 = 17.21$ for 12 degrees of freedom
Even terms: $\chi^2 = 17.59$ for 14 degrees of freedom
$a_1 = (849.6 \pm 15.4) - (335.5 \pm 85.7)x^2 + (847.3 \pm 87.8)x^4$ with $x = \cos(\theta)$
7.10. $a_1 = 0.0001 \pm 0.0009$; $a_2 = v_0 = 0.871 \pm 0.018$
$a_1 = g/2 = 4.870 \pm 0.057$ (after iterating)

Chapter 8

8.3. (*a*) $\mu = 1.8741 \pm 0.0005$; $\chi^2 = 13.70$
(*b*) $\mu = 1.8471 \pm 0.0005$; $\Gamma = 0.0555 \pm 0.0008$; $\chi^2 = 13.3$
8.4. $a_1 = 148.6 \pm 31.0$; $a_2 = 31.0 \pm 1.1$; $\chi^2 = 13.0$
$\epsilon_{11} = 65.6$; $\epsilon_{12} = \epsilon_{21} = -6.26$; $\epsilon_{22} = 1.156$

Chapter 9

9.4. (*b*) $\chi^2 = 34.2$ for 24 degrees of freedom
Fitted parameters a_1 through a_6:

−2.2	136.	−31.6	79.8	0.098	0.20

Uncertainties s_1 through s_6:

2.6	8.	3.1	7.0	0.007	0.02

Chapter 10

10.1. $a_1 = 4.16$; $a_2 = 22.8$ at the maximum of the likelihood function

Chapter 11

11.4. Approximately 10% probability
11.5. Approximately 0.1% probability; not a very good fit
11.9. 0.9985
11.10. 0.9729
11.12. 0.9997
11.14. $F \simeq 10$ for $\nu_1 = 1$; $F \simeq 5$ for $\nu_1 \leq \nu_2$
11.18. $\Delta\chi^2 = 2.7$; $a_4 = 3.4^{+4.5}_{-3.1}$; $a_5 = 205^{+70}_{-30}$

INDEX

D

E

F

G

H

Q

R

S

T

U

V

W