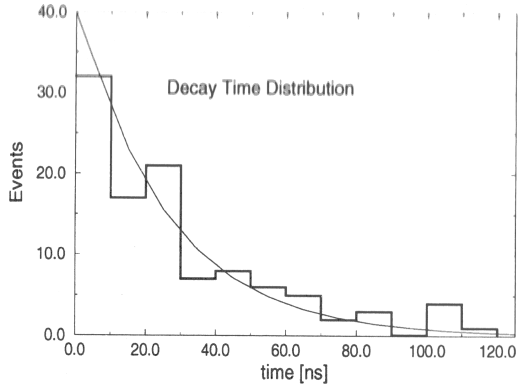

Fig. 5.3 A reproduction of the experimental measurement by [Stez 73] of the number of decays of a π meson as a function of time. Measurements are made during time intervals of 10-ns length.

5.10 PROBLEM: FITTING EXPONENTIAL DECAY

Fig. 5.3 presents experimental data of [Stez 73] on the number of decays ΔN of the π meson as a function of time. Notice that the time has been "binned" into $\Delta t = 10$ ns intervals and that the smooth curve gives the theoretical exponential decay law. The **problem** is to deduce the lifetime τ of the π meson from these data (the tabulated lifetime of the pion is 2.6×10^{-8} s).

5.11 MODEL: EXPONENTIAL DECAY

We start with a number N_0 of radioactive particles at time $t = 0$ that can decay to other particles.[2] If we wait a short time Δt, then a small number ΔN of the particles will decay spontaneously; that is, with no external influences. This decay is a stochastic process, which means that there is an element of chance involved and fluctuations are to be expected. The basic law of nature for spontaneous decay is that the number of decays ΔN in time interval Δt is proportional to the number of particles present at that time $N(t)$ and to the time interval

$$\Delta N(t) = -\frac{1}{\tau} N(t) \Delta t, \tag{5.19}$$

[2]Spontaneous decay is discussed further and simulated in §7.1.

where τ is the *lifetime* of the particle. This equation can be arranged into an equation for the average decay *rate*

$$\frac{\Delta N(t)}{\Delta t} = -\lambda N(t). \tag{5.20}$$

If the number of decays ΔN is very small compared to the number of particles N, and if we look at vanishingly small time intervals, then the difference equation (5.20) becomes the differential equation

$$\frac{dN(t)}{dt} \simeq -\lambda N(t). \tag{5.21}$$

This differential equation has an exponential solution for the number

$$N(t) = N_0 e^{-t/\tau}, \tag{5.22}$$

as well as an exponential solution for the decay rate

$$\frac{dN(t)}{dt} = -\frac{N_0}{\tau} e^{-t/\tau} = \frac{dN}{dt}(0) e^{-t/\tau}. \tag{5.23}$$

Equation (5.23) is the theoretical formula we wish to "fit" to the data in Fig. 5.3. The output of such a fit is a "best value" for the lifetime τ. Before we discuss how to carry out such a *least-squares fit*, we give some background information on probability theory. The reader familiar with probability theory may wish to skip ahead to §5.13.

5.12 THEORY: PROBABILITY THEORY

The field of statistics is an attempt to use mathematics to describe events, such as coin flips, in which there is an element of chance or randomness. A basic building block of statistics is the *binomial distribution* function

$$P_B(x) = \binom{N}{x} p^x (1-p)^{N-x} = \frac{N!}{(N-x)!x!} p^x (1-p)^{N-x}, \tag{5.24}$$

where $P_B(x)$ is the probability that the independent event (heads) will occur x times in the N trials. Here p is the probability of an individual event occurring; for example, the probability of "heads" in any one toss is $p = \frac{1}{2}$. The variable N is the number of *trials* or experiments in which that event can occur; for example, the number of times we flip the coin. For coin flipping, the probability of success p and the probability of failure $(1-p)$ are both $\frac{1}{2}$, but in the general case p can be any number between 0 and 1.

For calculational convenience, the factorials in the binomial distribution (5.24) are usually eliminated by considering the limit in which the number of

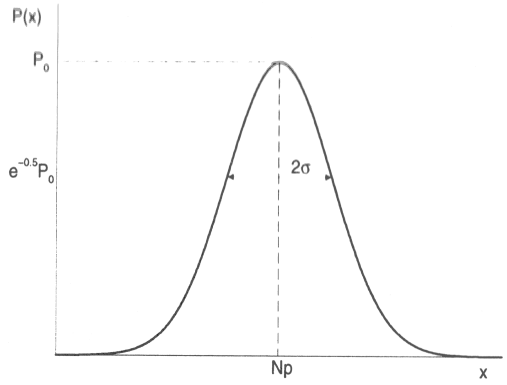

Fig. 5.4 A Gaussian distribution of m successes in N tries, each with probability p.

trials $N \to \infty$. In **Gaussian** or **normal** statistics, the probability p of an individual event (heads) remains finite as $N \to \infty$:

$$P_G(x) = \lim_{N\to\infty, p\neq 0} P_B(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{5.25}$$

This produces the function shown in Fig. 5.4, where $\mu \equiv \overline{x}$ is the mean and σ is the variance. These constants are related to the others by

$$\mu = Np\,, \quad \sigma = \sqrt{Np(1-p)}. \tag{5.26}$$

The Gaussian distribution is generally a very good approximation to the binomial distribution even for N as small as 10. To repeat, it describes an experiment in which N measurements of the variable x are made. The average of these measurements is μ and the "error" or uncertainty in μ is σ. As an example, in N =1000 coin flips, the probability of a head is $p = \frac{1}{2}$ and the average number of heads μ should be $Np = N/2 = 500$.

As shown in Fig. 5.4, the Gaussian distribution has a width

$$\sigma = \sqrt{Np(1-p)} \propto \sqrt{N}, \tag{5.27}$$

so that the distribution actually gets wider and wider as more measurements are made. Yet the *relative width*, whose inverse gives us an indication of the probability of obtaining the average μ, decreases with N:

$$\frac{\text{width}}{N} \propto \frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}} \to 0, \quad (N \to \infty). \tag{5.28}$$

Another limit of the binomial distribution is the **Poisson** distribution. In

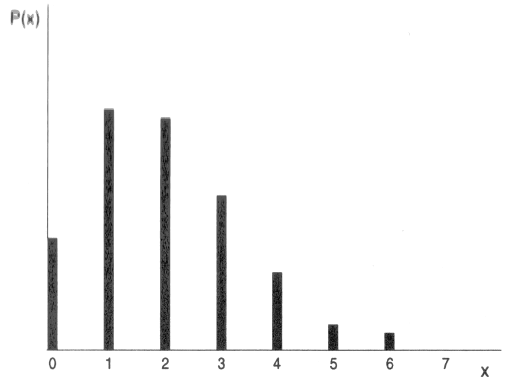

Fig. 5.5 A Poisson distribution for m successes with total success $a = 2$.

the Poisson distribution, the number of trials $N \to \infty$, yet the probability of an individual success $p \to 0$ in such a way that the product Np remains finite:

$$P_P(x) = \lim_{N\to\infty, p\to 0} P_B(x) = \frac{\mu^x e^{-\mu}}{x!}. \tag{5.29}$$

A **Poisson** distribution describes radioactive decay experiments or telephone interchanges where there may be a very large number of trials (each microsecond when the counter is on), but a low probability of an event (a decay or phone call) occurring in this one microsecond. As we see in Fig. 5.5, the Poisson distribution is quite asymmetric for small μ, and in this way quite different from a Gaussian distribution. For $\mu \gg 1$, the Poisson distribution approaches a Gaussian distribution.

5.13 METHOD: LEAST-SQUARES FITTING

Books have been written and careers have been spent discussing what is meant by a good fit to experimental data. We cannot do justice to the subject here and refer the reader to several other sources [B&R 92, Pres 94, M&W 65, Thom 92]. However, we will emphasize two points:

- If the data being fit contain errors, then the "best" fit in a statistical sense will not necessarily pass through the data points.
- Only for the simplest case of a linear, least-squares fit, can we write down a closed-form solution to evaluate and obtain the fit. More realistic problems are usually solved by *trial-and-error* search procedures using

library subroutines such as SLATEC.

Imagine that you have measured N_D values of the independent variable y as a function of the dependent variable x:

$$(x_i, y_i \pm \sigma_i), \quad i = 1, N_D, \tag{5.30}$$

where $\pm\sigma_i$ is the uncertainty in the ith value of y. (For simplicity we assume that all the errors σ_i occur in the dependent variable, although this is hardly ever true [Thom 92]). For our problem, y is the number of decays as a function of time.

Our goal is to determine how well a mathematical function $y = f(x)$ (also-called *theory*) can describe these data. Alternatively, if the theory contains some parameters or constants, our goal can be viewed as determining best values for these parameters. We assume that the model function $f(x)$ contains, in addition to the functional dependence on x, an additional dependence upon M_P parameters $\{a_1, a_2, \ldots, a_{M_P}\}$. For the exponential decay function (5.23), the parameter is the lifetime τ. We indicate this as

$$f(x) = f(x; \{a_1, a_2, \ldots, a_{M_P}\}) = f(x; \{a_m\}). \tag{5.31}$$

Notice that the parameters $\{a_m\}$ are not variables, in the sense of numbers read from a meter, but rather are parts of the theoretical model such as the size of a box, the mass of a particle, or the depth of a potential well.

We take the chi-squared (χ^2) measure as a gauge of how well a theoretical function f reproduces data

$$\boxed{\chi^2 \stackrel{\text{def}}{=} \sum_{i=1}^{N_D} \left(\frac{y_i - f(x_i; \{a_m\})}{\sigma_i} \right)^2}, \tag{5.32}$$

where the sum is over the N_D experimental points $(x_i, y_i \pm \sigma_i)$. The definition (5.32) is such that smaller values of χ^2 are better fits, with $\chi^2 = 0$ occurring if the theoretical curve went through the center of every data point. Notice also that the $1/\sigma_i^2$ weighting means that measurements with larger errors[3] contribute less to χ^2.

Least-squares fitting refers to adjusting the theory until a minimum in χ^2 is found; that is, finding a curve that produces the least value for the summed squares of the deviations of the data from the function $f(x)$. In general, this is the best fit possible or the best way to determine the parameters in a theory.

[3]If you are not given the errors, you can guess them on the basis of the apparent deviation of the data from a smooth curve, or you can weigh all points equally by setting $\sigma_i \equiv 1$ and continue with the fitting.

The M_P parameters $\{a_m, m = 1, M_P\}$ that make χ^2 an extremum are found by solving the M_P equations:

$$\frac{\partial \chi^2}{\partial a_m} = 0, \quad (m = 1, M_P), \tag{5.33}$$

$$\Rightarrow \sum_{i=1}^{N_D} \frac{y_i - f(x_i)}{\sigma_i^2} \frac{\partial f(x; \{a_m\})}{\partial a_m} = 0, \quad (m = 1, M_P). \tag{5.34}$$

More usually, the function $f(x; \{a_m\})$ has a sufficiently complicated dependence on the a_m values for (5.34) to produce M_P simultaneous, nonlinear equations in the a_m values. In these cases, solutions are found by a trial-and-error search through the M_P-dimensional parameter space. To be safe, when such a search is completed you need to check that the minimum χ^2 you found is *global* and not *local*. One way to do that is to repeat the search for a whole grid of starting values, and if different minima are found, to pick the one with the lowest χ^2.

5.14 THEORY: GOODNESS OF FIT

When the deviations from theory are due to random errors and when these errors are described by a Gaussian distribution, there are some useful rules of thumb to remember [B&R 92]. You know that your fit is good if the value of χ^2 calculated via the definition (5.32) is approximately equal to the number of degrees of freedom

$$\chi^2 \approx N_D - M_P, \tag{5.35}$$

where N_D is the number of data points and M_P the number of parameters in the theoretical function. If your χ^2 is much less than (5.35), it doesn't mean that you have a "great" theory or a really precise measurement; instead, you probably have too many parameters or have assigned errors (σ_i values) that are too large. In fact, too small a χ^2 may indicate that you are fitting the random scatter in the data rather than missing $\sim \frac{1}{3}$ of the error bars (as expected for Gaussian statistics). If your χ^2 is significantly greater than (5.35), the theory may not be good, you may have significantly underestimated your errors, or you may have errors which are not random.

If you think you obtained a good fit to the data, but cannot determine the χ^2 because you did not have values for the experimental errors (maybe you assumed $\sigma_i \equiv 1$), you can get an approximate σ^2 for use in calculating χ^2. First you fit the data using an arbitrary value for σ_i. Then you calculate the *variance* of your data,

$$\sigma_{\text{exp}}^2 \stackrel{\text{def}}{=} \frac{1}{N_D} \sum_{i=1}^{N_D} [y_i - f(x_i)]^2 . \tag{5.36}$$

Finally, you use $\sigma_{\exp}$ as an approximation to σ_i and apply (5.32) to obtain a meaningful χ^2.

5.15 IMPLEMENTATION: LEAST-SQUARES FITS, FIT.F (.C)

The M_P simultaneous equations (5.34) simplify considerably if the functions $f(x;\{a_m\})$ depend *linearly* on the a values. This happens, for example, when the theory function $f(x)$ is linear:

$$f(x;\{a_1,a_2\}) = a_1 + a_2 x. \tag{5.37}$$

In this case (also known as *linear regression*) there are $M_P = 2$ parameters, the slope a_2 and the y intercept a_1. Notice that while there are only two parameters to determine, there still may be an arbitrary number N_D of data points to fit. Remember, a unique solution is not possible unless the number of data points is equal to or greater than the number of parameters.

For the linear case, the χ^2 minimization equations (5.34) become two in number, and determine the parameters in terms of all the data points [Pres 94]:

$$a_1 = \frac{S_{xx}S_y - S_x S_{xy}}{\Delta}, \quad a_2 = \frac{SS_{xy} - S_x S_y}{\Delta}, \tag{5.38}$$

$$S = \sum_{i=1}^{N_D} \frac{1}{\sigma_i^2}, \quad S_x = \sum_{i=1}^{N_D} \frac{x_i}{\sigma_i^2}, \tag{5.39}$$

$$S_y = \sum_{i=1}^{N_D} \frac{y_i}{\sigma_i^2}, \quad S_{xx} = \sum_{i=1}^{N_D} \frac{x_i^2}{\sigma_i^2}, \tag{5.40}$$

$$S_{xy} = \sum_{i=1}^{N_D} \frac{x_i y_i}{\sigma_i^2}, \quad \Delta = SS_{xx} - S_x^2. \tag{5.41}$$

If you know the errors σ_i in your experimental measurements of the y_i, or have determined an approximate σ from the sample variance from your fitted function, the theory then gives you an expression for the *variance* or uncertainty in the deduced parameters:

$$\sigma_{a_1}^2 = \frac{S_{xx}}{\Delta}, \quad \sigma_{a_2}^2 = \frac{S}{\Delta}. \tag{5.42}$$

This is a measure of the uncertainties in the values of the fitted parameters arising from the uncertainties σ_i in the measured y_i values.

A measure of the dependence of the parameters on each other is given by

the *correlation coefficient*:

$$\rho(a_1,a_2) = \frac{\operatorname{cov}(a_1,a_2)}{\sigma_{a_1}\sigma_{a_2}}, \tag{5.43}$$

$$\operatorname{cov}(a_1,a_2) = \frac{-S_x}{\Delta}. \tag{5.44}$$

Here $\operatorname{cov}(a_1,a_2)$ is the *covariance* of a_1 and a_2 and vanishes if a_1 and a_2 are independent. The correlation coefficient $\rho(a_1,a_2)$ lies in the range $-1 \le \rho \le 1$. Positive ρ indicates that the errors in a_1 and a_2 are likely to have the same sign; negative ρ indicates opposite signs.

The preceding analytic solutions for the parameters are of the form found in statistics books, but are not optimal for numerical calculations because subtractive cancellation can make the answers unstable. As discussed in Chapter 3, *Errors and Uncertainties in Computations*, a rearrangement of the equations can decrease this type of error. For example, [Thom 92] gives improved expressions that measure the data relative to their averages:

$$a_1 = y_{\text{av}} - a_2 x_{\text{av}}, \quad a_2 = \frac{S_{xy}}{S_{xx}}, \tag{5.45}$$

$$S_{xy} = \sum_{i=1}^{N_d} \frac{(x_i - x_{\text{av}})(y_i - y_{\text{av}})}{\sigma_i^2}, \quad S_{xx} = \sum_{i=1}^{N_d} \frac{(x_i - x_{\text{av}})^2}{\sigma_i^2}, \tag{5.46}$$

$$x_{\text{av}} = \frac{1}{N}\sum_{i=1}^{N_d} x_i, \quad y_{\text{av}} = \frac{1}{N}\sum_{i=1}^{N_d} y_i. \tag{5.47}$$

5.16 ASSESSMENT: FITTING EXPONENTIAL DECAY

Fit the exponential decay law (5.23) to the data in Fig. 5.3. This means finding a value of τ that provides a best fit to the data, and then judging how good the fit is.

1. Construct a table $(dN/dt_i, t_i)$, for $i = 1, N_D$ from Fig. 5.3. Notice that because time was measured in bins, t_i should correspond to the middle of a bin.

2. Add an estimate of the error σ_i to obtain a table of the form $(dN/dt_i \pm \sigma_i, t_i)$. You can estimate the errors by eye, say, by estimating how much the histogram values appear to fluctuate about a smooth curve, or you can take $\sigma_i \simeq \sqrt{\text{Events}}$. (This last approximation is reasonable for large numbers, which this is not.)

3. In the limit of very large numbers, we would expect that a plot of $\ln|dN/dt|$ versus t is a straight line:

$$\ln\left|\frac{dN(t)}{dt}\right| \simeq \ln\left|\frac{dN_0}{dt}\right| - \frac{1}{\tau}t. \tag{5.48}$$

(While we do not have truly large numbers here, this result should be good on average.) This means that if we treat $\ln|\frac{dN(t)}{dt}|$ as the dependent variable and time t as the independent variable, we can use our linear fit results. Plot $\ln|dN/dt|$ versus t.

4. Make a least-squares fit of a straight line to your data and use it to determine the lifetime τ of the π meson. Compare your deduction to the tabulated lifetime of 2.6×10^{-8} s and comment on the difference.

5. Plot your best fit on the same graph as the data and comment on the agreement.

6. Use the formulas from statistics to deduce the goodness of fit of your straight line and the approximate error in your deduced lifetime. Do these agree with what your "eye" tells you?

5.17 ASSESSMENT: FITTING HEAT FLOW

Here is a table that gives the temperature T along a metal rod whose ends are kept at fixed constant temperatures. The temperature is a function of the distance x along the rod.

Position x_i (cm)	Temperature T_i (^{o}C)
1.0	14.6
2.0	18.5
3.0	36.6
4.0	30.8
5.0	59.2
6.0	60.1
7.0	62.2
8.0	79.4
9.0	99.9

1. Plot up the data to verify the appropriateness of a linear relation

$$T = a + bx. \tag{5.49}$$

2. Because you are not given the errors for each measurement, assume that the least-significant figure has been rounded off and so $\sigma \geq 0.05$. Use that to compute a least-square, straight-line fit to these data.

3. Plot your best $a + bx$ on the curve with the data.

4. After fitting the data, compute the variance and compare it to the deviation of your fit from the data. Verify that about one-third of the points miss the σ error band (that's what is expected for a normal distribution of errors).

5. Use your computed variance to determine the χ^2 of the fit. Comment on the value obtained.

6. Determine the variances σ_a and σ_b and check to see if they make sense as the errors in a and b.

7. What correlation is expected between a and b?

5.18 IMPLEMENTATION: LINEAR QUADRATIC FITS

As indicated earlier, as long as the function being fit depends *linearly* on the parameters a_i, the condition of minimum χ^2 leads to a set of simultaneous linear equations. These can be solved directly on the computer. For example, suppose we want to fit the experimental measurements $(x_i, y_i, i = 1, N_D)$ to the quadratic polynomial

$$y(x) = b_0 + b_1 x + b_2 x^2. \tag{5.50}$$

The χ^2 will be a minimum with respect to variation of these parameters (in other words, there will be maximum likelihood that these are the correct parameters describing the measurements), when we satisfy the three simultaneous linear equations:

$$N_D b_0 + S_x b_1 + S_{xx} b_2 = S_y, \tag{5.51}$$
$$S_x b_0 + S_{xx} b_1 + S_{xxx} b_2 = S_{xy}, \tag{5.52}$$
$$S_{xx} b_0 + S_{xxx} b_1 + S_{xxxx} b_2 = S_{xxy}. \tag{5.53}$$

Here the definitions of the S's are simple extensions of those used in (5.38). These equations can be written in matrix form:

$$\begin{bmatrix} N_D & S_x & S_{xx} \\ S_x & S_{xx} & S_{xxx} \\ S_{xx} & S_{xxx} & S_{xxxx} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_y \\ S_{xy} \\ S_{xxy} \end{bmatrix}. \tag{5.54}$$

The solution follows after finding the inverse of the S matrix:

$$[S]\mathbf{b} = \mathbf{s} \tag{5.55}$$

$$\mathbf{b} = [S]^{-1}\mathbf{s}. \tag{5.56}$$

The inversion can be accomplished with the techniques discussed in Chapter 15, *Matrix Computing and Subroutine Libraries.*

5.19 ASSESSMENT: QUADRATIC FIT

Fit a quadratic to the following data sets [given as $(x_1, y_1), (x_2, y_2), \ldots$]. In each case determine the solution to these equations, the number of degrees of freedom in the problem, *and* the value of χ^2.

1. $(0, 1)$
2. $(0, 1), (1, 3)$
3. $(0, 1), (1, 3), (2, 7)$
4. $(0, 1), (1, 3), (2, 7), (3, 15)$

5.20 METHOD: NONLINEAR LEAST-SQUARES FITTING

An example of a subroutine for conducting a nonlinear search is *snls1* from SLATEC.

5.21 ASSESSMENT: NONLINEAR FITTING

Return to Table 5.1, which gives the scattering cross section versus energy. Determine what values for the parameters E_r, σ_0, and γ in the Breit–Wigner formula (5.1) provide a best fit to the data in the table (that is, minimize χ^2).

6

Deterministic Randomness

6.1 PROBLEM: DETERMINISTIC RANDOMNESS

Some people are attracted to computing by its deterministic nature; it's nice to have something in life where nothing is left to chance. Barring random machine errors or undefined variables, you should get the same output every time you feed your program the same input. Nevertheless, many computer cycles are used for *Monte Carlo* calculations that at their very core strive to be random. These are calculations in which random numbers generated by the computer are used to *simulate* naturally random processes, such as thermal motion or radioactive decay, or to solve equations on the average. Indeed, much of the recognition of computational physics as a specialty has come about from the ability of computers to solve previously intractable thermodynamic and quantum mechanics problems using Monte Carlo techniques.

The **problem** in this chapter is explore how computers can generate random numbers and how well they can do it. To check whether it really works, you *simulate* some simple physical processes and evaluate some multidimensional integrals. Other applications, such as radioactive decay, magnetism, and lattice quantum mechanics, are considered in later chapters.

6.2 THEORY: RANDOM SEQUENCES

We define a sequence of numbers $r_1, r_2, \ldots$ as *random* if there are no correlations among the numbers in the sequence. Yet randomness does not necessar-