4. The Maximum-Likelihood Method

- *Basis*
- *Applications*
- *Relationship to the method of least-squares*

4. The Method of Maximum Likelihood

The method of maximum likelihood provides a basis for many of the techniques and methods discussed in this course. The reasons are:

- The method has a good intuitive foundation. The underlying concept is that the best estimate of a parameter is that giving the highest probability that the observed set of measurements will be obtained.
- The least-squares method and various approaches to combining errors or calculating weighted averages, etc., can be derived or justified in terms of the maximum likelihood approach.
- The method is of sufficient generality that most problems are amenable to a straightforward application of this method, even in cases where other techniques become difficult. Inelegant but conceptually simple approaches often provide useful results where there is no easy alternative.

a. Basis

The method of maximum likelihood provides one solution to the problem of *estimation.* Consider a set of observations $\{x_i\}$ from a population with the probability distribution function

$$\phi(x; a_1, a_2, \ldots a_n) = \phi(x, \{a\}). \tag{4.1}$$

The parameters $\{a\}$ influence the distribution function, but are generally unknown. The task of estimation is to determine functions of the observations $\{x\}$ to use as estimates of the parameters $\{a\}$.

An *estimator* of a_j can be any function $f_j(\{x\})$ used to estimate the true value of the parameter a_j. The sample mean and sample standard deviation are often used as estimators of the true population mean and standard deviation, for example. Desirable characteristics of estimators are:

- Absence of *bias* in the estimator, defined as $(\alpha_j - \langle f_j \rangle)$ where α_j is the true value of a_j. An example of a biased estimator is the use of the square root of the sample variance to estimate the population standard deviation. The factor $(n/(N-1))^{1/2}$ used to obtain (3.9) removes this bias.
- Low *variance,* defined as $V(f_j) = \langle (f_j - \langle f_j \rangle)^2 \rangle$. The mean squared error $\langle (f_j - \alpha_j)^2 \rangle$ of the estimator is the sum of the variance and the square of the bias.

The probability of obtaining a set of observations $\{x\}$ from a population with probability distribution function $\phi(x, \{a\})$ is the product of the probabilities of all the observations:

$$\mathcal{L}(a) = \prod_i \phi(x_i; \{a\}). \tag{4.2}$$

This joint probability function is called the *likelihood* and depends on the parameters $\{a\}$. If the likelihood function is plotted as a function of a for the case with a single parameter, the resulting distribution will have a shape somewhat like Fig. 4.1. The value a^*, for which the likelihood reaches its maximum value, is the maximum-likelihood estimate for the parameter a.

For numerical convenience, it is usually preferable to calculate instead the function W defined as

$$W = \ln \mathcal{L}(\{a\}). \tag{4.3}$$

Because W is a monotonic function of $\mathcal{L}$, the maximum in W will coincide with the maximum in $\mathcal{L}$. However, because the calculation of W involves a summation rather than a product, there are computational advantages to the use of W:

$$\begin{aligned} W &= \ln \left(\prod_i \phi(x_i; \{a\}) \right) \\ &= \sum_i \ln \left(\phi(x_i; \{a\}) \right) . \end{aligned} \tag{4.4}$$

The maximum-likelihood estimate of the parameters $\{a\}$ satisfies the simultaneous equations

$$\left. \frac{\partial W}{\partial a_j} \right|_{a_j = a_j^*} = 0. \tag{4.5}$$

The maximum-likelihood estimator has several desirable properties:

1. The estimator is *efficient* in the sense that there is no estimator with smaller variance.

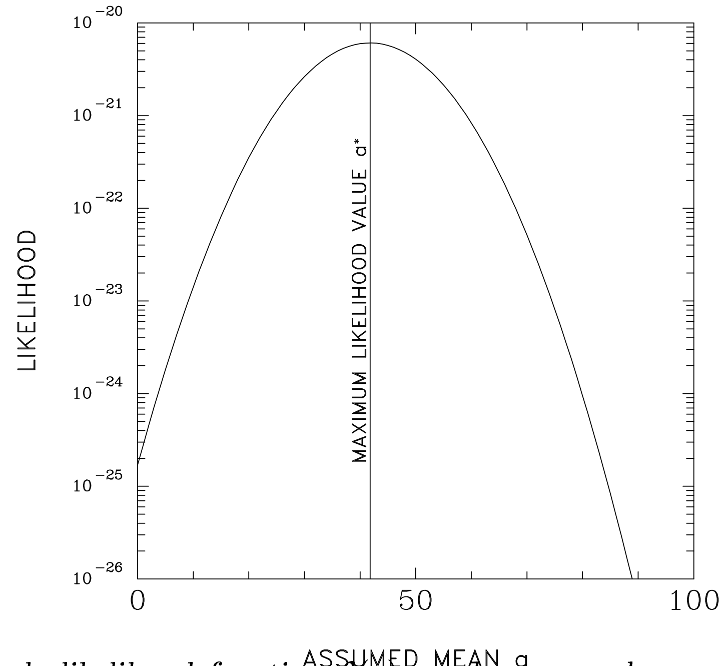

Fig. 4.1: Example likelihood function $\mathcal{L}(a)$ *vs the assumed parameter* a. *This example was generated using 10 randomly generated numbers in the interval (0–100): 67.6, 60.2, 41.5, 84.9, 40.1, 2.5, 9.6, 45.7, 50.2, 15.6. The actual mean is 41.8, and that is also the maximum in the likelihood function. The likelihood function was then calculated using (4.2) where* $\phi(x_i; a)$ *was taken to be a Gaussian probability function with mean* a *and standard deviation* $100/\sqrt{(12)}$.

2. The estimator approaches the true population parameter asymptotically as the number of observations increases.
3. The distribution of deviations of the estimator from the population parameter approaches a normal distribution for large numbers of observations.

The Gaussian behavior of the likelihood function for large sample sizes can be used to determine the uncertainty in the maximum-likelihood estimate of $\{a\}$. If the a_j are uncorrelated,

$$\mathcal{L}(\{a\}) = C_1 \ \exp\{-\frac{(a_1 - a_1^*)^2}{2\sigma_1^2}\} \exp\{-\frac{(a_2 - a_2^*)^2}{2\sigma_2^2}\} \cdots$$

and

$$W = C_2 - \sum_j \frac{(a_j - a_j^*)^2}{2\sigma_j^2} \tag{4.6}$$

where C_1 and C_2 are constants. Differentiating W twice isolates σ_j:

$$\frac{\partial^2 W}{\partial a_j^2} = -\frac{1}{\sigma_j^2}$$

$$\sigma_j = \left[-\frac{\partial^2 W}{\partial a_j^2}\Big|_{a_j^*}\right]^{-1/2}. \tag{4.7}$$

It is often simplest to use (4.6) directly, rather than evaluate the second derivative in (4.7), particularly when there is a single parameter to be determined. When a_j differs from a_j^* by σ_j, the term on the right side of (4.6) decreases by 1/2 from its maximum value, so for uncorrelated errors an estimate of the standard deviation in the result can be found by finding the deviation in a_j from a_j^* that causes W to reduce by 1/2.

Maximizing W is equivalent to minimizing the chisquare function, defined as

$$\chi^2 = \sum_j \frac{(a_j - a_j^*)^2}{\sigma_j^2} \ . \tag{4.8}$$

Because χ^2 increases by 1 when W decreases by 1/2, the standard deviation in the estimate of a_j^* can also be estimated from the deviation that causes unity increase in the chisquare.

In a case where the fit to the measurements is poor, perhaps because an inappropriate distribution function was used, the likelihood will have a value much smaller than expected. In that case, the estimates of uncertainty obtained from (4.6–4.8) should not be used. Instead, the proper conclusion is that the model used is inappropriate because it does not provide an adequate fit to the observations. Erroneously small estimates of uncertainty limits sometimes arise from using (4.6–4.8) when the fit is poor.

b. Applications

1). Weighted averages

If measurements $\{x\}$ are taken from a population with a Gaussian distribution, but are made with varying measurement uncertainties $\{\sigma\}$, what is the best estimate of the sample mean? The likelihood function for this case is

$$\mathcal{L}(a) = \prod_i \frac{1}{\sigma_i \sqrt{2\pi}} \exp\{-\frac{(x_i - a)^2}{2\sigma_i^2}\}$$

and

$$W(a) = -\sum_i \frac{(x_i - a)^2}{2\sigma_i^2} + \text{constant.}$$

The maximum occurs for

$$\frac{\partial W(a)}{\partial a}\Big|_{a^*} = \sum_i \frac{(x_i - a^*)}{\sigma_i^2} = \sum_i (\frac{x_i}{\sigma_i^2}) - \sum_i \frac{a^*}{\sigma_i^2} = 0$$

or

$$a^* = \frac{\sum_i (x_i/\sigma_i^2)}{\sum_i (1/\sigma_i^2)}. \tag{4.9}$$

This is a weighted average of the measurements, with weight factors inversely proportional to the square of the uncertainty.

Because

$$\frac{\partial^2 W}{\partial a^2}\Big|_{a^*} = -\sum_i \frac{1}{\sigma_i^2},$$

the estimated uncertainty in the weighted average, from (4.7), is

$$\sigma_{a^*} = [\sum_i \frac{1}{\sigma_i^2}]^{-1/2}. \tag{4.10}$$

2). Mean of the Poisson probability distribution function

EXERCISE 4.1: *Consider an experiment in which a number of events (e.g., ice crystals entering the sample volume of an airborne probe) are counted under conditions where the Poisson distribution function is expected to characterize the probability. Show that the maximum-likelihood estimate of the mean number of events is equal to the number of events counted; i.e., the estimate of the population mean μ is x. (The Poisson distribution is asymmetrical, so this result is not obvious.) Also show that the resulting estimator is unbiased.*

---

3). Mean of the binomial probability distribution function

The binomial distribution function describes the probability of observing n events in a given class out of N trials, when the population-average probability for the given class of event is p:

$$\Phi_B(n,p) = \binom{N}{n} p^n (1-p)^{N-n} \ . \tag{4.11}$$

If a trial is conducted and n^* events are observed, what is the best estimate for the parameter p? The logarithm of the likelihood function for p is

$$W(p) = n\ln p + (N-n)\ln(1-p) + \text{constant} \ . \tag{4.12}$$

The maximum likelihood occurs for

$$\frac{\partial W}{\partial p} = 0 = \frac{n}{p} - \frac{(N-n)}{(1-p)} = \frac{(n-pN)}{p(1-p)} \ . \tag{4.13}$$

Then

$$p^* = \frac{n^*}{N} \ . \tag{4.14}$$

is the resulting maximum-likelihood estimator for p.

The uncertainty in p can be found by use of

$$\frac{\partial^2 W}{\partial p^2} = -\frac{N}{p^*(1-p^*)} \tag{4.15}$$

in (4.7):

$$\sigma_{p^*} = \left[\frac{p^*(1-p^*)}{N}\right]^{1/2} \ . \tag{4.16}$$

Note that the standard deviation in n,

$$\sigma_n = N\sigma_p = (Np^*(1-p^*))^{1/2} \ , \tag{4.17}$$

is smaller than $\sqrt{N}$.

4). An example: Fitting to CCN measurements

EXAMPLE 4.1: *A cloud condensation nucleus (CCN) counter usually operates by exposing an air sample to a supersaturated environment and counting the number of droplets n that form in a specific sample volume V. As a function of the supersaturation S, the CCN spectrum is often reported in the form*

$$N(S) = C(\frac{S}{S_{ref}})^k \tag{4.18}$$

where $N(S)$ is the concentration of CCN active at supersaturations smaller than or equal to S, C is the concentration active at or below the reference supersaturation S_{ref}, and k characterizes the rate of increase of concentration with supersaturation. The estimated concentration at the supersaturation S_i is then $N_(S_i) = n_i/V$.

Often small numbers of CCN are detected at a given supersaturation, so statistical fluctuations have an important influence on the final estimates of the parameters C and k. Furthermore, when the number of droplets detected is small, the non-Gaussian nature of expected deviations also should be considered when fitted parameters are determined. The fit is complicated also by the correlations that are usually present between uncertainties in C and k.

The method of maximum likelihood can be used to estimate values of C and k using a set of observations $\{n_i\}$ corresponding to supersaturations $\{S_i\}$. As is always the case with maximum-likelihood approaches, the calculation of the likelihood starts with an initial assumption about the expected form of the probability function. In this case, we assume *that the spectrum has the form (4.18) and that the departures from that spectrum are caused by statistical fluctuations characterized by the Poisson distribution function. Either of these assumptions may be incorrect, for example because the real shape of the supersaturation spectrum may be more complicated than (4.18) or the changes measured in repeated samples may represent changing aerosol characteristics with time instead of the assumed supersaturation dependence. The maximum-likelihood result will always be dependent on the validity of these initial assumptions, and they should be tested when possible.*

If the spectrum obeys (4.18) and the measurements follow Poisson statistics, the likelihood can be calculated as follows:

- *For given values of C and k, the mean* number *of droplets expected at a given supersaturation S_i is $\mu_i = C\ (S_i/S_{ref})^k\ V$ where V is the sample volume.*
- *The (normalized) probability of observing n_i droplets at supersaturation S_i is given by the Poisson probability function (3.7) for mean value μ_i.*
- *The likelihood function is then*

$$\mathcal{L}(C,k) = \prod_i \Phi_P(n_i, \mu_i) \tag{4.19}$$

  where Φ_P is the Poisson probability function and the sum extends over all observations. This function is dependent on the assume values of C and k, so best values of these parameters correspond to the pair of values that maximize $\mathcal{L}$.

The resulting contour plot of the likelihood function might look like Fig. 4.2. The uncertainties in C and k are correlated, and that correlation poses special difficulties that will be discussed in the next section. The best value of the parameters can be found by *calculating enough values to plot the contours as in Fig. 4.2, or a numerical maximization procedure can be used to find the maximum value of the likelihood. A plot such as that shown is particularly informative because contours are chosen to correspond to deviation limits representing 1 and 2 standard deviations.*

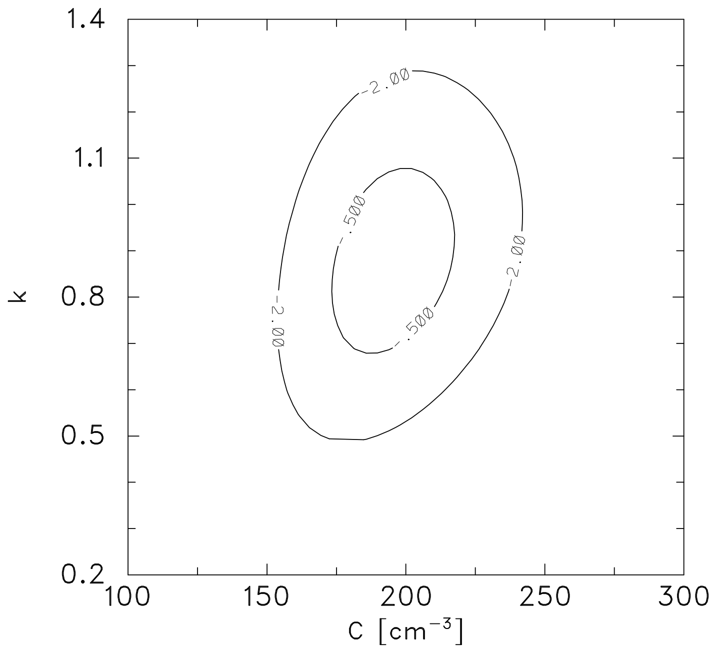

Fig. 4.2: Contours of an example likelihood function, in units relative to the maximum, for a fit to (4.18). Contours at -0.5 and -2.0 represent respectively the one- and two-standard-deviation limits for the fit. The maximum likelihood occurs for $C = 196$ cm^{-3} and $k = 0.87$. The data used to construct this plot were: for supersaturations of 0.2, 0.4, 0.6, 0.8, 1.0, and 1.5%, respective droplet counts of 2, 14, 11, 12, 26, and 25 representing a sample volume of 0.1 cm^{-3}. (These values were generated for the case where $C = 200$ cm^{-3} and $k = 0.8$, with random errors added to each assumed measurement.)

5). Maximum-likelihood estimation in cases with correlated errors

Consider the Taylor-series expansion of W about the maximum set of values for the parameters, $\{a^*\}$:

$$W(a) = W(a^*) + \sum_j \frac{\partial W}{\partial a_j}\Big|_{a^*}(a_j - a_j^*) + \frac{1}{2}\sum_j\sum_k \frac{\partial^2 W}{\partial a_j \partial a_k}\Big|_{a^*}(a_j - a_j^*)(a_k - a_k^*) + \cdots . \quad (4.20)$$

The second term on the right equals zero, because that is the condition for the validity of the maximum-likelihood solution. The reason for expanding W, rather than $\mathcal{L}$, is that the joint probability function can be approximated by a Gaussian distribution function and (4.20) gives a generalized Gaussian form for $\mathcal{L}$:

$$\mathcal{L}(\{a\}) = A \exp\{-\frac{1}{2}\sum_j \sum_k \delta a_j H_{jk} \delta a_k\} \tag{4.21}$$

where A=exp$\{W(a^*)\}$, $\delta a_j = (a_j - a_j^*)$, and

$$H_{jk} = -\frac{\partial^2 W}{\partial a_j \partial a_k}\Big|_{a^*} . \tag{4.22}$$

The matrix $\mathbf{H}$ is sometimes called the *information matrix,* and its inverse is the covariance or error matrix discussed in Chapter 2.

For uncorrelated fluctuations, the mixed derivatives of the likelihood function are zero because the variation of W with a_k is not a function of a_j if the two parameters are independent. In this case, the likelihood function reduces to the product of two independent Gaussian functions. However, if the parameters are not independent, there are terms in the resulting likelihood that depend on the mixed products of the fluctuations; e.g.,

$$\mathcal{L}(a_1, a_2) = \mathcal{L}_{max} \exp\{-\frac{H_{11}\delta a_1^2}{2} - \frac{H_{22}\delta a_2^2}{2} - \frac{(H_{12}+H_{21})\delta a_1 \delta a_2}{2}\} . \tag{4.23}$$

In all but degenerate cases, it is possible to transform the parameters $\{a\}$ into new parameters that are independent (or to diagonalize the error matrix). In the two-dimensional case, consider a rotation by an angle θ so that new parameters y and z are formed from the old parameters a_1 and a_2:

$$y = a_1 \cos\theta + a_2 \sin\theta$$

$$z = -a_1 \sin\theta + a_2 \cos\theta$$

Contours of constant probability, in terms of the new variables, are specified by the equation

$$\begin{aligned} \text{constant} = &H_{11}[\cos^2\theta\delta y^2 - 2\cos\theta\sin\theta\delta y\delta z + \sin^2\theta\delta z^2] \\ &+H_{22}[\sin^2\theta\delta y^2 + 2\cos\theta\sin\theta\delta y\delta z + \cos^2\theta\delta z^2] \\ &+(H_{12}+H_{21})[\cos\theta\sin\theta\delta y^2 + \cos^2\theta\delta y\delta z - \sin^2\theta\delta y\delta z - \cos\theta\sin\theta\delta z^2] . \end{aligned} \tag{4.24}$$

This has the form of an ellipse, and an appropriate choice of θ can give the standard form for an ellipse,

$$\frac{\delta y^2}{A^2} + \frac{\delta z^2}{B^2} = 1 , \tag{4.25}$$

which does not involve cross-terms in y and z. The required rotation can be found by setting the coefficient of the cross product in (4.24) to zero:

$$-2H_{11}\cos\theta\sin\theta + 2H_{22}\cos\theta\sin\theta + (H_{12}+H_{21})(\cos^2\theta - \sin^2\theta) = 0 \tag{4.26}$$

$$\sin 2\theta(H_{22}-H_{11}) + \cos 2\theta(H_{12}+H_{21}) = 0$$

$$\tan 2\theta = \frac{H_{12}+H_{21}}{H_{22}-H_{11}} . \tag{4.27}$$

In the transformed variables,

$$\mathcal{L} = \mathcal{L}_{max} \exp\{-\frac{H'_{11}\delta y^2}{2} - \frac{H'_{22}\delta z^2}{2}\} \tag{4.28}$$

and

$$\sigma_y^2 = \frac{1}{H'_{11}},$$
$$\sigma_z^2 = \frac{1}{H'_{22}}. \tag{4.29}$$

The variables y and z are uncorrelated, so the appropriate standard deviation in the initial parameters can be estimated from

$$\begin{aligned} \delta a_1 &= \delta y \cos\theta - \delta z \sin\theta \\ \overline{\delta a_1^2} &= \cos^2\theta \overline{\delta y^2} + \sin^2\theta \overline{\delta z^2} \\ \sigma_{a_1}^2 &= \cos^2\theta \sigma_y^2 + \sin^2\theta \sigma_z^2. \end{aligned} \tag{4.30}$$

The procedure can be generalized as follows. Write $\beta = (\mathbf{a} - \mathbf{a}^*)$, so that

$$\mathcal{L}(\mathbf{a}) = \mathcal{L}_{max} \exp\{-\frac{1}{2}\beta\mathbf{H}\beta^t\} \tag{4.31}$$

Transform according to the rotation matrix $\mathbf{U}$, to new parameters γ:

$$\gamma = \mathbf{U}\beta \tag{4.32}$$

Because the inverse of a rotation matrix is its transpose, (4.31) can be written

$$\mathcal{L}(\mathbf{a}) = \mathcal{L}_{max} \exp\{-\frac{1}{2}\beta\mathbf{U}^{-1}\mathbf{U}\mathbf{H}\mathbf{U}^{-1}\mathbf{U}\beta^t\} \tag{4.33}$$

$$= \mathcal{L}_{max} \exp\{-\frac{1}{2}\gamma\mathbf{h}\gamma\} \tag{4.34}$$

where

$$\mathbf{h} = \mathbf{U}\mathbf{H}\mathbf{U}^t. \tag{4.35}$$

For appropriate choice of $\mathbf{U}$, $\mathbf{h}$ can be made a diagonal matrix. In that case, $\mathcal{L}$ is the product of independent Gaussian distributions with variances specified by the components h_{jj}^{-1}. The variances in the initial variables β can then be found in terms of the (uncorrelated) variances in the variables γ:

$$\begin{aligned} \gamma &= \mathbf{U}\beta \\ \mathbf{U}^t\gamma &= \beta = \gamma\mathbf{U} \\ \beta^t\beta &= \mathbf{U}^t\gamma^t\gamma\mathbf{U} \\ \langle\beta^t\beta\rangle &= \mathbf{U}^t\langle\gamma^t\gamma\rangle U \\ &= \mathbf{U}^t\mathbf{h}^{-1}\mathbf{U} \\ &= (\mathbf{U}^t\mathbf{h}\mathbf{U})^{-1}. \end{aligned} \tag{4.36}$$

But, from (4.35),

$$\mathbf{H} = \mathbf{U}^t\mathbf{h}\mathbf{U} \tag{4.37}$$

so

$$\langle \beta^t \beta \rangle = \mathbf{H}^{-1}. \tag{4.38}$$

This again shows the connection between $\mathbf{H}$, the information matrix, and the variances and covariances of the observations.

One approach to determining the errors in the multivariate case with correlated errors is to determine the matrix $\mathbf{H}$ by finding the second derivatives of W, then invert $\mathbf{H}$ to find the error matrix. The elements of the error matrix then give the variances and covariances for the parameters.

6). Estimating exponential-decay time constants

Exponentially decaying functions arise in many experimental situations, either because of instrumental response characteristics or characteristics of the physical condition being measured. Some examples are these:

- Many thermometers respond to temperature change such that the difference between the measured and true temperature decays exponentially.
- Some observations of liquid water content in clouds appear to decrease exponentially with time.
- The time intervals between randomly occurring events obey an exponential probability distribution.

Because these situations are so common, there is often a need to estimate the time constant characterizing the exponential decay from measurements; i.e., to estimate τ in the formula

$$F(t) = F_0 e^{-t\lambda/\tau} \ . \tag{4.39}$$

The maximum-likelihood method provides a straightforward approach to this problem.[1] The approach will be illustrated by an example.

EXAMPLE 4.2: *The following are measurements of liquid water content averaged over 1 km in the center of a decaying cumulus cloud, at three minute intervals: 2.50, 2.15, 1.59, 1.10, 1.12, 0.93, 0.55, 0.65, 0.45 g m^{-3}. A plot of these shows them to be approximately consistent with an exponential decay. If the decay follows (4.39), what is the best estimate of the time constant τ, if the measurement uncertainty in each measurement of the liquid water content is the same and characterized by a Gaussian distribution about the true value?*

To calculate the likelihood function for this case, assume that (4.39) describes the mean value and measurements are distributed about this mean according to (3.2), so that the probability of observing liquid water content w_i at time t_i is

$$P(w_i, t_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \frac{-(w_i - A^{-t_i/\tau})^2}{2\sigma_i^2} \tag{4.40}$$

[1] Linear fits to the logarithm of F are sometimes used, but usually this approach distorts the effects of the measurement errors on the fit.

where A and τ are parameters to be determined from the fit. Then

$$W = -\sum_i \frac{(w_i - Ce^{-t_i/\tau})^2}{2\sigma_i^2} + C' \qquad (4.41)$$

where C' is constant with respect to variable parameters in the problem.

The two equations obtained from (4.5) by differentiating (4.41) with respect to the two parameters A and τ both give equations for A, so eliminating A by equating these expressions leads to the requirement that

$$f(\tau) = \frac{\sum_i \frac{w_i t_i e^{-t_i/\tau}}{\sigma_i^2}}{\sum_i \frac{t_i e^{-2t_i/\tau}}{sigma_i^2}} - \frac{\sum_i \frac{w_i e^{-t_i/\tau}}{\sigma_i^2}}{\sum_i \frac{e^{-2t_i/\tau}}{sigma_i^2}} = 0 \ . \qquad (4.42)$$

While not amenable to analytical solution, (4.42) is readily solved numerically or graphically, using methods that will be reviewed in Chapter 7. The solution can be obtained graphically by plotting $f(\tau)$ as a function of τ and finding the point at which $f(\tau)$=0. This occurs for τ=13.6 min for the data of this problem, and the result is the fit shown in Fig. 4.3.

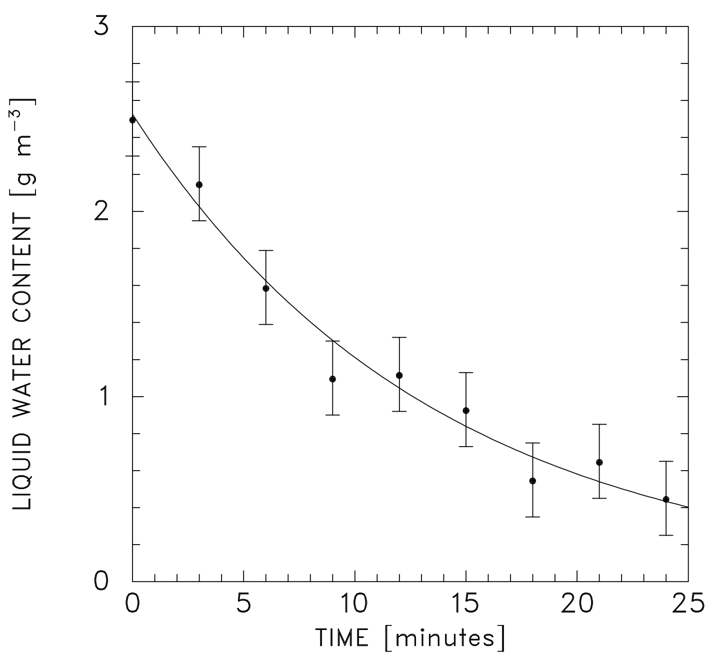

Fig. 4.3: Measurements of liquid water content as a function of time, with assumed constant uncertainty estimates, and the best-fit exponential obtained by maximum-likelihood analysis.

EXERCISE 4.2: *For the data of the preceding example, show that if the assumed accuracy in liquid water content is 0.2 g m^{-3} then the one-standard-deviation limit for τ is 1.6 minutes. Check also that the scatter of measurements about the exponential best-fit line is consistent with this estimate, and adjust it if necessary.*

---

c. *Application to experimental design*

A common problem of experimental design is to determine in advance if a proposed experimental configuration will be able to achieve a specified precision. In the preceding sections experimental results were used to estimate errors in parameters. For experimental design, it is useful to reformulate those expressions in ways suited to *a priori* estimation of experimental uncertainty.[2]

The expectation value for an element of the information matrix P**H** is

$$\langle H_{jk} \rangle = \left\langle \frac{\partial^2 W}{\partial a_j \partial a_k} \right\rangle \ . \tag{4.43}$$

If the observations are expected to occur according to some known probability distribution function ϕ, this form may be reduced to a function of ϕ. For N observations,

$$W(x, a) = \ln \mathcal{L}(x, a) = N \ln \phi(x, a) \ . \tag{4.44}$$

$$\langle H_{jk} \rangle = N \int \frac{\partial^2 \ln \phi}{\partial a_j \partial a_k} \phi dx \tag{4.45}$$

$$= N \int \frac{\partial}{\partial a_j} \left(\frac{1}{\phi} \frac{\partial \phi}{\partial a_k} \right) \phi dx \tag{4.46}$$

$$= N \left[- \int \frac{1}{\phi} \frac{\partial \phi}{\partial a_j} \frac{\partial \phi}{\partial a_k} dx + \int \frac{\partial^2 \phi}{a_j a_k} dx \right] \ . \tag{4.47}$$

If the integration is performed before differentiation in the last term, the integral is over the probability distribution which must give a constant (unity), so the last term in (4.47) vanishes. The expectation value for an element in the information matrix is then

$$\langle H_{jk} \rangle = -N \int \frac{1}{\phi} \frac{\partial \phi}{\partial a_j} \frac{\partial \phi}{\partial a_k} dx \ . \tag{4.48}$$

In particular, for the case without correlations,

$$V_{a_j a_j} = \frac{1}{N \int \frac{1}{\phi} (\frac{\partial \phi}{\partial a_j})^2 dx} = \frac{1}{N \left\langle (\frac{\partial \ln \phi}{\partial a_j})^2 \right\rangle} \tag{4.49}$$

---

[2] Cf. Orear, J., 1958: Notes on statistics for physicists, UCRL-8417, University of California Radiation Laboratory, Berkeley, CA.

EXAMPLE 4.3: *Suppose that a cloud droplet spectrum is expected to have a mean diameter μ of 15 μm and a standard deviation in diameter σ of 5 μm. How many droplets must be measured, if the measurement error is negligible, to permit determination of the standard deviation with 5% precision, if the droplet sizes are distributed approximately in a Gaussian distribution?*

From (3.2),

$$\ln \phi = -\ln \sigma - \frac{(d-\overline{d})^2}{2\sigma^2} + C$$

$$\frac{\partial \ln \phi}{\partial \sigma} = -\frac{1}{\sigma} + \frac{(d-\overline{d})^2}{\sigma^3}$$

$$\left(\frac{\partial \ln \phi}{\partial \sigma}\right)^2 = \frac{(d-\overline{d})^4 - 2\sigma^2(d-\overline{d})^2 + \sigma^4}{\sigma^6}$$

For a Gaussian distribution,

$$\langle (d-\overline{d})^2 \rangle = \sigma^2$$

and

$$\langle (d-\overline{d})^4 \rangle = 3\sigma^4$$

(obtained by integration of the probability distribution), so the expected uncertainty in measurement of σ, σ_σ, is given by

$$\sigma_\sigma^2 = \frac{1}{N\left(\frac{2\sigma^4}{\sigma^6}\right)} = \frac{\sigma^2}{2N}.$$

To obtain $\sigma_\sigma = 0.05(5) = 0.25$ μm, N must equal 200 droplets.

---

d. *Relationship to the method of least squares*

Consider a set of measurements $\{y\}$ at points $\{x\}$, with varying measurement precision $\{\sigma\}$. Find the parameters $\{a\}$ in the function $f(x; a)$ that are the best match to the observations. The situation might be as shown in Fig. 4.4.

Assume that the measurements $\{y\}$ are distributed according to a Gaussian probability distribution function (3.2). Then the likelihood function is

$$\mathcal{L} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(y_i - f(x_i,\{a\}))^2}{2\sigma_i^2}\right\} \tag{4.50}$$

and

$$W = \ln \mathcal{L} = -\frac{1}{2}\sum_i \frac{(y_i - f(x_i,\{a\}))^2}{\sigma_i^2} + C \tag{4.51}$$

where C is a constant not dependent on the parameters $\{a\}$. The maximum-likelihood solution is then equivalent to the solution that gives the minimum value for the sum

$$\chi^2 = \sum_i \frac{(y_i - f(x_i,\{a\}))^2}{\sigma_i^2}\ . \tag{4.52}$$

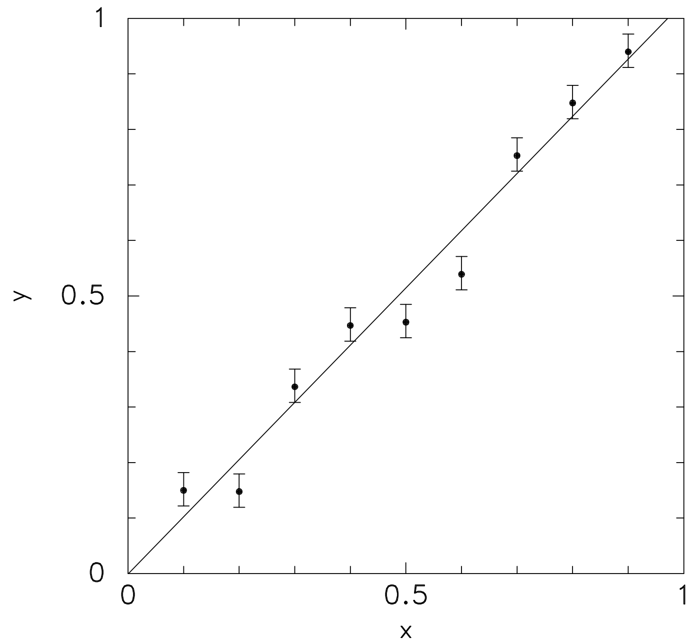

Fig. 4.4: Example of measurements of a variable y as a function of the variable x, for which each measurement is characterized by a measurement uncertainty designated by error bars above and below the data points representing deviations of one standard deviation. The line shown is an example of the desired representation of the measurements by a function, in this case a straight line, that provides a "best" fit to the observations.

This provides one justification for the *least-squares* method of the next section.

SOURCES AND FURTHER READING

Bevington, P. R., 1969: Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, New York, 336 pp.

Brownlee, K. A., 1965: Statistical Theory and Methodology in Science and Engineering. John Wiley and Sons, New York, 590 pp.

Orear, J., 1958: Notes on statistics for physicists, UCRL-8417, University of California Radiation Laboratory, Berkeley, CA.