7. The Method of Maximum Likelihood

7.1 Likelihood Ratio: Likelihood Function

In the last chapter we introduced the concept of parameter estimation. We have also described the desirable properties of estimators, though without specifying how such estimators can be constructed in a particular case. We have derived estimators only for the important quantities expectation value and variance. We now take on the general problem.

In order to specify explicitly the parameters

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_p) \quad ,$$

we now write the probability density of the random variables

$$\mathbf{x} = (\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n)$$

in the form

$$f = f(\mathbf{x}; \boldsymbol{\lambda}) \quad . \tag{7.1.1}$$

If we now carry out a certain number of experiments, say N, or we draw a sample of size N from a population, then we can give a number to each experiment j:

$$\mathrm{d}P^{(j)} = f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})\,\mathrm{d}\mathsf{x} \quad . \tag{7.1.2}$$

The number $\mathrm{d}P^{(j)}$ has the character of an *a posteriori probability*, i.e., given *after* the experiment, how probable it was to find the result $\mathbf{x}^{(j)}$ (within a small interval). The total probability to find exactly all of the events

$$\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(j)}, \ldots, \mathbf{x}^{(N)}$$

is then the product

$$\mathrm{d}P = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})\,\mathrm{d}\mathbf{x} \quad . \tag{7.1.3}$$

S. Brandt, *Data Analysis: Statistical and Computational Methods for Scientists and Engineers*,
DOI 10.1007/978-3-319-03762-2_7,

This probability still clearly depends on $\boldsymbol{\lambda}$. There are cases where the population is determined by only two possible sets of parameters, $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. Such cases occur, for example, in nuclear physics, where the parity of a state is necessarily "even" or "odd". One can construct the ratio

$$Q = \frac{\prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}_1)}{\prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}_2)} \tag{7.1.4}$$

and say that the values $\boldsymbol{\lambda}_1$ are "Q times more probable" than the values $\boldsymbol{\lambda}_2$. This factor is called the *likelihood ratio*.*

A product of the form

$$L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) \tag{7.1.5}$$

is called a *likelihood function*. One must clearly distinguish between a probability density and a likelihood function, which is a function of a sample and is hence a random variable. In particular, the a posteriori nature of the probability in (7.1.5) is of significance in many discussions.

Example 7.1: Likelihood ratio

Suppose one wishes to decide whether a coin belongs to type A or B by means of a number of tosses. The coins in question are asymmetric in such a way that A shows heads with a probability of $1/3$, and B shows heads with a probability of $2/3$.

	A	B
Heads	$1/3$	$2/3$
Tails	$2/3$	$1/3$

If an experiment yields heads once and tails four times, then one has $L_A = \frac{1}{3} \cdot (\frac{2}{3})^4$ and $L_B = \frac{2}{3} \cdot (\frac{1}{3})^4$,

$$Q = \frac{L_A}{L_B} = 8.$$

One would therefore tend towards the position that the coin is of type A. ■

*Although the likelihood ratio Q and the likelihood functions L and ℓ introduced below are random variables, since they are functions of a sample, we do not write them here with a special character type.

7.2 The Method of Maximum Likelihood

The generalization of the likelihood ratio is now clear. One gives the greatest confidence to that choice of the parameters $\boldsymbol{\lambda}$ for which the likelihood function (7.1.5) is a maximum. Figure 7.1 illustrates the situation for various forms of the likelihood function for the case of a single parameter λ.

The maximum can be located simply by setting the first derivative of the likelihood function with respect to the parameter λ_i equal to zero. The derivative of a product with many factors is, however, unpleasant to deal with. One first constructs therefore the logarithm of the likelihood function,

$$\ell = \ln L = \sum_{j=1}^{N} \ln f(\mathsf{x}^{(j)}; \boldsymbol{\lambda}). \tag{7.2.1}$$

The function ℓ is also often called the likelihood function. Sometimes one says explicitly "*log-likelihood function*". Clearly the maxima of (7.2.1) are identical with those of (7.1.5). For the case of a single parameter we now construct

$$\ell' = \mathrm{d}\ell/\mathrm{d}\lambda = 0. \tag{7.2.2}$$

The problem of estimating a parameter is now reduced to solving this *likelihood equation*. By application of (7.2.1) we can write

$$\ell' = \sum_{j=1}^{N} \frac{\mathrm{d}}{\mathrm{d}\lambda} \ln f(\mathsf{x}^{(j)}; \lambda) = \sum_{j=1}^{N} \frac{f'}{f} = \sum_{j=1}^{N} \varphi(\mathsf{x}^{(j)}; \lambda), \tag{7.2.3}$$

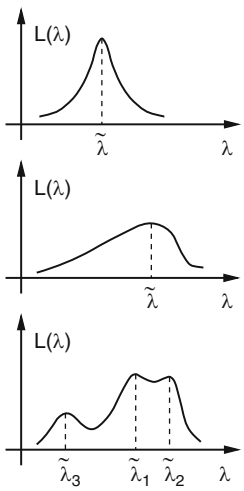

Fig. 7.1: Likelihood functions.

where

$$\varphi(\mathsf{x}^{(j)};\lambda) = \left(\frac{\mathrm{d}}{\mathrm{d}\lambda} f(\mathsf{x}^{(j)};\lambda)\right) \Big/ f(\mathsf{x}^{(j)};\lambda) \tag{7.2.4}$$

is the *logarithmic derivative* of the density f with respect to λ.

In the general case of p parameters the likelihood equation (7.2.2) is replaced by the system of p simultaneous equations,

$$\frac{\partial \ell}{\partial \lambda_i} = 0, \qquad i = 1, 2, \ldots, p. \tag{7.2.5}$$

Example 7.2: Repeated measurements of differing accuracy

If a quantity is measured with different instruments, then the measurement errors are in general different. The measurements $\mathsf{x}^{(j)}$ are spread about the true value λ. Suppose the errors are normally distributed, so that a measurement corresponds to obtaining a sample from a Gaussian distribution with mean λ and standard deviation σ_j. The a posteriori probability for a measured value is then

$$f(\mathsf{x}^{(j)};\lambda)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(\mathsf{x}^{(j)}-\lambda)^2}{2\sigma_j^2}\right)\mathrm{d}x.$$

From all N measurements one obtains the likelihood function

$$L = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(\mathsf{x}^{(j)}-\lambda)^2}{2\sigma_j^2}\right) \tag{7.2.6}$$

with the logarithm

$$\ell = -\frac{1}{2}\sum_{j=1}^{N} \frac{(\mathsf{x}^{(j)}-\lambda)^2}{\sigma_j^2} + \text{const.} \tag{7.2.7}$$

The likelihood equation thus becomes

$$\frac{\mathrm{d}\ell}{\mathrm{d}\lambda} = \sum_{j=1}^{N} \frac{\mathsf{x}^{(j)}-\lambda}{\sigma_j^2} = 0.$$

It has the solution

$$\tilde{\lambda} = \frac{\displaystyle\sum_{j=1}^{N} \frac{\mathsf{x}^{(j)}}{\sigma_j^2}}{\displaystyle\sum_{j=1}^{N} \frac{1}{\sigma_j^2}}. \tag{7.2.8}$$

Since $\mathrm{d}^2\ell/\mathrm{d}\lambda^2 = -\sum \sigma_j^{-2} < 0$, the solution is, in fact, a maximum. Thus we see that we obtain the maximum likelihood estimator as the mean of the N measurements weighted inversely by the variances of the individual measurements. ■

Example 7.3: Estimation of the parameter N of the hypergeometric distribution

As in the example with coins at the beginning of this chapter, sometimes parameters to be estimated can only take on discrete values. In Example 5.2 we indicated the possibility of estimating zoological population densities by means of tagging and recapture. According to (5.3.1), the probability to catch exactly n fish of which k are tagged out of a pond with an unknown total of N fish, out of which K are tagged, is given by

$$L(k; n, K, N) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

We must now find the value of N for which the function L is maximum. For this we use the ratio

$$\frac{L(k; n, k, N)}{L(k; n, k, N-1)} = \frac{(N-n)(N-k)}{(N-n-K+k)N} \begin{cases} > 1\,, & Nk < nK\,, \\ < 1\,, & Nk > nK\,. \end{cases}$$

The function L is thus maximum when N is the integer closest to nK/k. ■

7.3 Information Inequality. Minimum Variance Estimators. Sufficient Estimators

We now want to discuss once more the quality of an estimator. In Sect. 6.1 we called an estimator unbiased if for every sample the *bias* vanished,

$$B(\lambda) = E(\mathsf{S}) - \lambda = 0. \tag{7.3.1}$$

Lack of bias is, however, not the only characteristic required of a "good" estimator. More importantly one should require that the variance

$$\sigma^2(\mathsf{S})$$

is small. Here one must often find a compromise, since there is a connection between B and σ^2, described by the *information inequality*.[†]

One immediately sees that it is easy to achieve $\sigma^2(\mathsf{S}) = 0$ simply by using a constant for S. We consider an estimator $\mathsf{S}(\mathsf{x}^{(1)}, \mathsf{x}^{(2)}, \ldots, \mathsf{x}^{(N)})$ that is

[†]This inequality was independently found by H. Cramer, M. Fréchet, and C. R. Rao as well as by other authors. It is also called the Cramer–Rao or Fréchet inequality.

a function of the sample $\mathsf{x}^{(1)}, \mathsf{x}^{(2)}, \ldots, \mathsf{x}^{(N)}$. According to (6.1.3) and (6.1.4) the joint probability density of the elements of the sample is

$$f(x^{(1)}, x^{(2)}, \ldots, x^{(N)}; \lambda) = f(x^{(1)}; \lambda) f(x^{(2)}; \lambda) \cdots f(x^{(N)}; \lambda).$$

The expectation value of S is thus

$$E(\mathsf{S}) = \int \mathsf{S}(\mathsf{x}^{(1)}, \ldots, \mathsf{x}^{(N)}) f(x^{(1)}; \lambda) \cdots f(x^{(N)}; \lambda) \times \mathrm{d}x^{(1)} \, \mathrm{d}x^{(2)} \cdots \mathrm{d}x^{(N)}. \tag{7.3.2}$$

According to (7.3.1), however, one also has

$$E(\mathsf{S}) = B(\lambda) + \lambda.$$

We now assume that we can differentiate with respect to λ in the integral. We then obtain

$$1 + B'(\lambda) = \int \mathsf{S} \left(\sum_{j=1}^{N} \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)} \right) f(x^{(1)}; \lambda) \cdots f(x^{(N)}; \lambda) \, \mathrm{d}x^{(1)} \cdots \mathrm{d}x^{(N)},$$

which is equivalent to

$$1 + B'(\lambda) = E \left\{ \mathsf{S} \sum_{j=1}^{N} \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)} \right\} = E \left\{ \mathsf{S} \sum_{j=1}^{N} \varphi(x^{(j)}; \lambda) \right\}.$$

From (7.2.3) we have

$$\ell' = \sum_{j=1}^{N} \varphi(\mathsf{x}^{(j)}; \lambda)$$

and therefore

$$1 + B'(\lambda) = E\{\mathsf{S}\ell'\}. \tag{7.3.3}$$

One clearly has

$$\int f(x^{(1)}; \lambda) \cdots f(x^{(N)}; \lambda) \, \mathrm{d}x^{(1)} \cdots \mathrm{d}x^{(N)} = 1.$$

If we also compute the derivative with respect to λ, we obtain

$$\int \sum_{j=1}^{N} \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)} f(x^{(1)}; \lambda) \cdots f(x^{(N)}; \lambda) \, \mathrm{d}x^{(1)} \cdots \mathrm{d}x^{(N)} = E(\ell') = 0.$$

By multiplying this equation by $E(\mathsf{S})$ and subtracting the result of (7.3.3) one obtains

$$1 + B'(\lambda) = E\{\mathsf{S}\ell'\} - E(\mathsf{S})E(\ell') = E\{[\mathsf{S} - E(\mathsf{S})]\ell'\}. \tag{7.3.4}$$

In order to see the significance of this expression, we need to use the Cauchy–Schwarz inequality in the following form:

> If x and y are random variables and if x^2 and y^2 have finite expectation values, then
>
> $$\{E(\mathsf{xy})\}^2 \leq E(\mathsf{x}^2)E(\mathsf{y}^2). \tag{7.3.5}$$

To prove this inequality we consider the expression

$$E((a\mathsf{x}+\mathsf{y})^2) = a^2E(\mathsf{x}^2) + 2aE(\mathsf{xy}) + E(\mathsf{y}^2) \geq 0. \tag{7.3.6}$$

This is a non-negative number for all values of a. If we consider for the moment the case of equality, then this is a quadratic equation for a with the solutions

$$a_{1,2} = -\frac{E(\mathsf{xy})}{E(\mathsf{x}^2)} \pm \sqrt{\left(\frac{E(\mathsf{xy})}{E(\mathsf{x}^2)}\right)^2 - \frac{E(\mathsf{y}^2)}{E(\mathsf{x}^2)}}. \tag{7.3.7}$$

The inequality (7.3.6) is then valid for all a if the term under the square root is negative or zero. From this follows the assertion

$$\frac{\{E(\mathsf{xy})\}^2}{\{E(\mathsf{x}^2)\}^2} - \frac{E(\mathsf{y}^2)}{E(\mathsf{x}^2)} \leq 0.$$

If we now apply the inequality (7.3.5) to (7.3.4), one obtains

$$\{1 + B'(\lambda)\}^2 \leq E\{[\mathsf{S} - E(\mathsf{S})]^2\}E(\ell'^2). \tag{7.3.8}$$

We now use (7.2.3) in order to rewrite the expression for $E(\ell'^2)$,

$$\begin{aligned} E(\ell'^2) &= E\left\{\left(\sum_{j=1}^{N}\varphi(\mathsf{x}^{(j)};\lambda)\right)^2\right\} \\ &= E\left\{\sum_{j=1}^{N}(\varphi(\mathsf{x}^{(j)};\lambda))^2\right\} + E\left\{\sum_{i\neq j}\varphi(\mathsf{x}^{(i)};\lambda)\varphi(\mathsf{x}^{(j)};\lambda)\right\}. \end{aligned}$$

All terms on the right-hand side vanish, since for $i \neq j$

$$\begin{aligned} &E\{\varphi(\mathsf{x}^{(i)};\lambda)\varphi(\mathsf{x}^{(j)};\lambda)\} = E\{\varphi(\mathsf{x}^{(i)};\lambda)\}E\{\varphi(\mathsf{x}^{(j)};\lambda)\}, \\ &E\{\varphi(\mathsf{x};\lambda)\} = \int_{-\infty}^{\infty}\frac{f'(x;\lambda)}{f(x;\lambda)}f(x;\lambda)\,\mathrm{d}x = \int f'(x;\lambda)\,\mathrm{d}x, \end{aligned}$$

and

$$\int_{-\infty}^{\infty} f(x;\lambda)\,\mathrm{d}x = 1.$$

By differentiating the last line with respect to λ one obtains

$$\int_{-\infty}^{\infty} f'(x;\lambda)\,\mathrm{d}x = 0.$$

Thus one has simply

$$E(\ell'^2) = E\left\{\sum_{j=1}^{N}(\varphi(\mathbf{x}^{(j)};\lambda))^2\right\} = E\left\{\sum_{j=1}^{N}\left(\frac{f'(\mathbf{x}^{(j)};\lambda)}{f(\mathbf{x}^{(j)};\lambda)}\right)^2\right\}.$$

Since the individual terms of the sum are independent, the expectation value of the sum is simply the sum of the expectation values. The individual expectation values do not depend on the elements of the sample. Therefore one has

$$I(\lambda) = E(\ell'^2) = NE\left\{\left(\frac{f'(x;\lambda)}{f(x;\lambda)}\right)^2\right\}.$$

This expression is called the *information of the sample with respect to* λ. It is a non-negative number, which vanishes if the likelihood function does not depend on the parameter λ.

It is sometimes useful to write the information in a somewhat different form. To do this we differentiate the expression

$$E\left(\frac{f'(x;\lambda)}{f(x;\lambda)}\right) = \int_{-\infty}^{\infty}\frac{f'(x;\lambda)}{f(x;\lambda)}f(x;\lambda)\,\mathrm{d}x = 0$$

once more with respect to λ and obtain

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty}\left\{\frac{f'^2}{f} + f\left(\frac{f'}{f}\right)'\right\}\mathrm{d}x = \int_{-\infty}^{\infty}\left\{\left(\frac{f'}{f}\right)^2 + \left(\frac{f'}{f}\right)'\right\} f\,\mathrm{d}x \\ &= E\left\{\left(\frac{f'}{f}\right)^2\right\} + E\left\{\left(\frac{f'}{f}\right)'\right\}. \end{aligned}$$

The information can then be written as

$$I(\lambda) = NE\left\{\left(\frac{f'(x;\lambda)}{f(x;\lambda)}\right)^2\right\} = -NE\left\{\left(\frac{f'(x;\lambda)}{f(x;\lambda)}\right)'\right\}$$

or

$$I(\lambda) = E(\ell'^2) = -E(\ell''). \tag{7.3.9}$$

The inequality (7.3.8) can now be written in the following way:

$$\{1 + B'(\lambda)\}^2 \leq \sigma^2(\mathsf{S}) I(\lambda)$$

or

$$\sigma^2(\mathsf{S}) \geq \frac{\{1 + B'(\lambda)\}^2}{I(\lambda)}. \tag{7.3.10}$$

This is the *information inequality*. It gives the connection between the bias and the variance of an estimator and the information of a sample. It should be noted that in its derivation no assumption about the estimator was made. The right-hand side of the inequality (7.3.10) is therefore a lower bound for the variance of an estimator. It is called the *minimum variance bound* or *Cramer–Rao bound*. In cases where the bias does not depend on λ, i.e., particularly in cases of vanishing bias, the inequality (7.3.10) simplifies to

$$\sigma^2(\mathsf{S}) \geq 1/I(\lambda). \tag{7.3.11}$$

This relation justifies using the name information. As the information of a sample increases, the variance of an estimator can be made smaller.

We now ask under which circumstances the minimum variance bound is attained, or explicitly, when the equals sign in the relation (7.3.10) holds. In the inequality (7.3.6), one has equality if $(a\mathsf{x} + \mathsf{y})$ vanishes, since only then does one have $E\{(a\mathsf{x}+\mathsf{y})^2\} = 0$ for all values of a, x, and y. Applied to (7.3.8), this means that

$$\ell' + a(\mathsf{S} - E(\mathsf{S})) = 0$$

or

$$\ell' = A(\lambda)(\mathsf{S} - E(\mathsf{S})). \tag{7.3.12}$$

Here A means an arbitrary quantity that does not depend on the sample $\mathsf{x}^{(1)}, \mathsf{x}^{(2)}, \ldots, \mathsf{x}^{(N)}$, but may be, however, a function of λ. By integration we obtain

$$\ell = \int \ell' \, \mathrm{d}\lambda = B(\lambda)\mathsf{S} + C(\lambda) + D \tag{7.3.13}$$

and finally

$$L = d \exp\{B(\lambda)\mathsf{S} + C(\lambda)\}. \tag{7.3.14}$$

The quantities d and D do not depend on λ.

We thus see that estimators attain the minimum variance bound when the likelihood function is of the special form (7.3.14). They are therefore called *minimum variance estimators*.

For the case of an unbiased minimum variance estimator we obtain from (7.3.11)

$$\sigma^2(\mathsf{S}) = \frac{1}{I(\lambda)} = \frac{1}{E(\ell'^2)}. \tag{7.3.15}$$

By substituting (7.3.12) one obtains

$$\sigma^2(\mathsf{S}) = \frac{1}{(A(\lambda))^2 E\{(\mathsf{S} - E(\mathsf{S}))^2\}} = \frac{1}{(A(\lambda))^2 \sigma^2(\mathsf{S})}$$

or

$$\sigma^2(\mathsf{S}) = \frac{1}{|A(\lambda)|}. \tag{7.3.16}$$

If instead of (7.3.14) only the weaker requirement

$$L = g(\mathsf{S}, \lambda) c(\mathsf{x}^{(1)}, \mathsf{x}^{(2)}, \ldots, \mathsf{x}^{(N)}) \tag{7.3.17}$$

holds, then the estimator S is said to be *sufficient* for λ. One can show [see, e.g., Kendall and Stuart, Vol. 2 (1967)], that no other estimator can contribute information about λ that is not already contained in S, if the requirement (7.3.17) is fulfilled. Hence the name "sufficient estimator" (or statistic).

Example 7.4: Estimator for the parameter of the Poisson distribution

Consider the *Poisson distribution* (5.4.1)

$$f(k) = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}.$$

The likelihood function of a sample $\mathsf{k}^{(1)}, \mathsf{k}^{(2)}, \ldots, \mathsf{k}^{(N)}$ is

$$\ell = \sum_{j=1}^{N} \{\mathsf{k}^{(j)} \ln \lambda - \ln(\mathsf{k}^{(j)}!) - \lambda\}$$

and its derivative with respect to λ is

$$\frac{\mathrm{d}\ell}{\mathrm{d}\lambda} = \ell' = \sum_{j=1}^{N} \left\{ \frac{\mathsf{k}^{(j)}}{\lambda} - 1 \right\} = \frac{1}{\lambda} \sum_{j=1}^{N} \{\mathsf{k}^{(j)} - \lambda\},$$

$$\ell' = \frac{N}{\lambda} (\bar{\mathsf{k}} - \lambda). \tag{7.3.18}$$

Comparing with (7.3.12) and (7.3.16) shows that the arithmetic mean $\bar{k}$ is an unbiased minimum variance estimator with variance λ/N. ■

Example 7.5: Estimator for the parameter of the binomial distribution

The likelihood function of a sample from a *binomial distribution* with the parameters $p = \lambda$, $q = 1 - \lambda$ is given directly by (5.1.3),

$$L(\mathsf{k}, \lambda) = \binom{n}{\mathsf{k}} \lambda^{\mathsf{k}} (1-\lambda)^{n-\mathsf{k}}.$$

(The result of the sample can be summarized by the statement that in n experiments, the event A occurred k times; see Sect. 5.1.) One then has

$$\begin{aligned} \ell &= \ln L = \mathsf{k} \ln \lambda + (n - \mathsf{k}) \ln(1-\lambda) + \ln \binom{n}{\mathsf{k}}, \\ \ell' &= \frac{\mathsf{k}}{\lambda} - \frac{n-\mathsf{k}}{1-\lambda} = \frac{n}{\lambda(1-\lambda)} \left(\frac{\mathsf{k}}{n} - \lambda \right). \end{aligned}$$

By comparing with (7.3.12) and (7.3.16) one finds k/n to be a minimum variance estimator with variance $\lambda(1-\lambda)/n$. ■

Example 7.6: Law of error combination ("Quadratic averaging of individual errors")

We now return to the problem of Example 7.2 of repeated measurements of the same quantity with varying uncertainties, or expressed in another way, to the problem of drawing a sample from normal distributions with the same mean λ and different but known variances σ_j. From (7.2.7) we obtain

$$\frac{\mathrm{d}\ell}{\mathrm{d}\lambda} = \ell' = \sum_{j=1}^{N} \frac{\mathsf{x}^{(j)} - \lambda}{\sigma_j^2}.$$

We can rewrite this expression as

$$\begin{aligned} \ell' &= \sum \frac{\mathsf{x}^{(j)}}{\sigma_j^2} - \sum \frac{\lambda}{\sigma_j^2} \\ &= \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left\{ \frac{\sum \dfrac{\mathsf{x}^{(j)}}{\sigma_j^2}}{\sum \dfrac{1}{\sigma_j^2}} - \lambda \right\}. \end{aligned}$$

As in Example 7.2 we recognize

$$\mathsf{S} = \tilde{\lambda} = \frac{\sum \dfrac{\mathsf{x}^{(j)}}{\sigma_j^2}}{\sum \dfrac{1}{\sigma_j^2}} \tag{7.3.19}$$

as an unbiased estimator for λ. Comparing with (7.3.12) shows that it is also a minimum variance estimator. From (7.3.16) one determines that its variance is

$$\sigma^2(\tilde{\lambda}) = \left(\sum_{j=1}^{N} \frac{1}{\sigma_j^2} \right)^{-1} . \tag{7.3.20}$$

The relation (7.3.20) often goes by the name of the *law of error combination* or *quadratic averaging of individual errors*. It could have been obtained by application of the rule of error propagation (3.8.7) to (7.3.19). If we identify $\sigma(\tilde{\lambda})$ as the error of the estimator $\tilde{\lambda}$ and σ_j as the error of the jth measurement, then we can write it in its usual form

$$\Delta\tilde{\lambda} = \left(\frac{1}{(\Delta x_1)^2} + \frac{1}{(\Delta x_2)^2} + \cdots + \frac{1}{(\Delta x_n)^2} \right)^{-\frac{1}{2}} . \tag{7.3.21}$$

If all of the measurements have the same error $\sigma = \sigma_j$, Eqs. (7.3.19), (7.3.20) simplify to

$$\tilde{\lambda} = \bar{\mathsf{x}}, \qquad \sigma^2(\tilde{\lambda}) = \sigma^2/n,$$

which we have already found in Sect. 6.2. ■

7.4 Asymptotic Properties of the Likelihood Function and Maximum-Likelihood Estimators

We can now show heuristically several important properties of the likelihood function and maximum-likelihood estimators for very large data samples, that is, for the limit $N \to \infty$. The estimator $\mathsf{S} = \tilde{\lambda}$ was defined as the solution to the likelihood equation

$$\ell'(\lambda) = \sum_{j=1}^{N} \left(\frac{f'(\mathsf{x}^{(j)}; \lambda)}{f(\mathsf{x}^{(j)}; \lambda)} \right)_{\tilde{\lambda}} = 0. \tag{7.4.1}$$

Let us assume that $\ell'(\lambda)$ can be differentiated with respect to λ one more time, so that we can expand it in a series around the point $\lambda = \tilde{\lambda}$,

$$\ell'(\lambda) = \ell'(\tilde{\lambda}) + (\lambda - \tilde{\lambda})\ell''(\tilde{\lambda}) + \cdots . \tag{7.4.2}$$

The first term on the right side vanishes because of Eq. (7.4.1). In the second term one can write explicitly

$$\ell''(\tilde{\lambda}) = \sum_{j=1}^{N} \left(\frac{f'(\mathsf{x}^{(j)}; \lambda)}{f(\mathsf{x}^{(j)}; \lambda)} \right)'_{\tilde{\lambda}} .$$

This expression has the form of the mean value of a sample. For very large N it can be replaced by the expectation value of the population (Sect. 6.2),

$$\ell''(\tilde{\lambda}) = NE\left\{\left(\frac{f'(x;\lambda)}{f(x;\lambda)}\right)'_{\tilde{\lambda}}\right\}. \tag{7.4.3}$$

Using Eq. (7.3.9) we can now write

$$\ell''(\tilde{\lambda}) = E(\ell''(\tilde{\lambda})) = -E(\ell'^2(\tilde{\lambda})) = -I(\tilde{\lambda}) = -1/b^2. \tag{7.4.4}$$

In this way we can replace the expression for $\ell''(\tilde{\lambda})$, which is a function the sample $\mathsf{x}^{(1)}, \mathsf{x}^{(2)}, \ldots, \mathsf{x}^{(N)}$, by the quantity $-1/b^2$, which only depends on the probability density f and the estimator $\tilde{\lambda}$. If one neglects higher-order terms, Eq. (7.4.2) can be expressed as

$$\ell'(\lambda) = -\frac{1}{b^2}(\lambda - \tilde{\lambda}). \tag{7.4.5}$$

By integration one obtains

$$\ell(\lambda) = -\frac{1}{2b^2}(\lambda - \tilde{\lambda})^2 + c.$$

Inserting $\lambda = \tilde{\lambda}$ gives $c = \ell(\tilde{\lambda})$, or

$$\ell(\lambda) - \ell(\tilde{\lambda}) = -\frac{1}{2}\frac{(\lambda - \tilde{\lambda})^2}{b^2}. \tag{7.4.6}$$

By exponentiation one obtains

$$L(\lambda) = k \exp\{-(\lambda - \tilde{\lambda})^2/2b^2\}, \tag{7.4.7}$$

where k is a constant. The likelihood function $L(\lambda)$ has the form of a normal distribution with mean $\tilde{\lambda}$ and variance b^2. At the values $\lambda = \tilde{\lambda} \pm b$, where λ is one standard deviation from $\tilde{\lambda}$, one has

$$-(\ell(\lambda) - \ell(\tilde{\lambda})) = \frac{1}{2}. \tag{7.4.8}$$

We can now compare (7.4.7) with Eqs. (7.3.12) and (7.3.16). Since we are estimating the parameter λ, we must write $\mathsf{S} = \tilde{\lambda}$ and thus $E(\mathsf{S}) = \lambda$. The estimator $\tilde{\lambda}$ is therefore an unbiased minimum variance estimator with variance

$$\sigma^2(\tilde{\lambda}) = b^2 = \frac{1}{I(\tilde{\lambda})} = \frac{1}{E(\ell'^2(\tilde{\lambda}))} = -\frac{1}{E(\ell''(\tilde{\lambda}))}. \tag{7.4.9}$$

Since the estimator $\tilde{\lambda}$ only possesses this property for the limiting case $N \to \infty$, we call it *asymptotically unbiased*. This is equivalent to the statement

that the maximum likelihood estimator is consistent (Sect. 6.1). For the same reason the likelihood function is called *asymptotically normal*.

In Sect. 7.2 we interpreted the likelihood function $L(\lambda)$ as a measure of the probability that the true value λ_0 of a parameter is equal to λ. The result of an estimator is often represented in abbreviated form,

$$\lambda = \tilde{\lambda} \pm \sigma(\tilde{\lambda}) = \tilde{\lambda} \pm \Delta\tilde{\lambda}.$$

Since the likelihood function is asymptotically normal, at least in the case of large samples, i.e., many measurements, this can be interpreted by saying that the probability that the true value λ_0 lies in the interval

$$\tilde{\lambda} - \Delta\tilde{\lambda} < \lambda_0 < \tilde{\lambda} + \Delta\tilde{\lambda}$$

is 68.3 % (Sect. 5.8). In practice the relation above is used for large but finite samples. Unfortunately one cannot construct any general rule for determining when a sample is large enough for this procedure to be reliable. Clearly if N is finite, (7.4.3) is only an approximation, whose accuracy depends not only on N, but also on the particular probability density $f(x;\lambda)$.

Example 7.7: Determination of the mean lifetime from a small number of decays

The probability that a radioactive nucleus, which exists at time $t = 0$, decays in the time interval between t and $t + \mathrm{d}t$ is

$$f(t)\,\mathrm{d}t = \frac{1}{\tau}\exp(-t/\tau)\,\mathrm{d}t.$$

For observed decay times $t_1, t_2, \ldots, t_N$ the likelihood function is

$$L = \frac{1}{\tau^N}\exp\left\{-\frac{1}{\tau}\sum_{i=1}^{N} t_i\right\} = \frac{1}{\tau^N}\exp\left\{-\frac{N}{\tau}\bar{t}\right\}.$$

Its logarithm is

$$\ell = \ln L = -\frac{N}{\tau}\bar{t} - N\ln\tau$$

with the derivative

$$\ell' = \frac{N}{\tau}\left(\frac{\bar{t}}{\tau} - 1\right) = \frac{N}{\tau^2}(\bar{t} - \tau).$$

Comparing with (7.3.12) we see that $\tilde{\tau} = \bar{t}$ is the maximum likelihood solution, which has a variance of $\sigma^2(\tau) = \tau^2/N$. For $\tau = \tilde{\tau} = \bar{t}$ one obtains $\Delta\tilde{\tau} = \bar{t}/\sqrt{N}$.

For $\tilde{\tau} = \bar{t}$ one has

$$\ell(\tilde{\tau}) = \ell_{\max} = -N(1 + \ln \bar{t}).$$

We can write

$$-(\ell(\tau) - \ell(\tilde{\tau})) = N\left(\frac{\bar{t}}{\tau} + \ln\frac{\tau}{\bar{t}} - 1\right).$$

From this expression for the log-likelihood function one cannot so easily recognize the asymptotic form (7.4.6) for $N \to \infty$. For small values of N it clearly does not have this form. Corresponding to (7.4.8), we want to use the values $\tau_+ = \tilde{\tau} + \Delta_+$ and $\tau_- = \tilde{\tau} - \Delta_-$, where one has

$$-(\ell(\tau_\pm) - \ell(\tilde{\tau})) = \frac{1}{2}$$

for the *asymmetric errors* Δ_+, Δ_-. Clearly we expect for $N \to \infty$ that Δ_+, $\Delta_- \to \Delta\tilde{\tau} = \sigma(\tilde{\tau})$.

In Fig. 7.2 the N observed decay times t_i are marked as vertical tick marks on the abscissa for various small values of N. In addition the function $-(\ell - \ell_{\max}) = -(\ell(\tau) - \ell(\tilde{\tau}))$ is plotted. The points τ_+ and τ_- are found where a horizontal line intersects $-(\ell - \ell_{\max}) = 1/2$. The point $\tilde{\tau}$ is indicated by an additional mark on the horizontal line. One sees that with increasing N the function $-(\ell - \ell_{\max})$ approaches more and more the symmetric parabolic form and that the errors Δ_+, Δ_-, and $\Delta\tilde{\tau}$ become closer to each other. ■

7.5 Simultaneous Estimation of Several Parameters: Confidence Intervals

We have already given a system of equations (7.2.5) allowing the simultaneous determination of p parameters $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)$. It turns out that it is not the parameter determination but rather the estimation of their errors that becomes significantly more complicated in the case of several parameters. In particular we will need to consider correlations as well as errors of the parameters.

We extend our considerations from Sect. 7.4 on the properties of the likelihood function to the case of several parameters. The log-likelihood function

$$\ell(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)}; \boldsymbol{\lambda}) = \sum_{j=1}^{N} \ln f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) \tag{7.5.1}$$

can be expanded in a series about the point

$$\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_p) \tag{7.5.2}$$

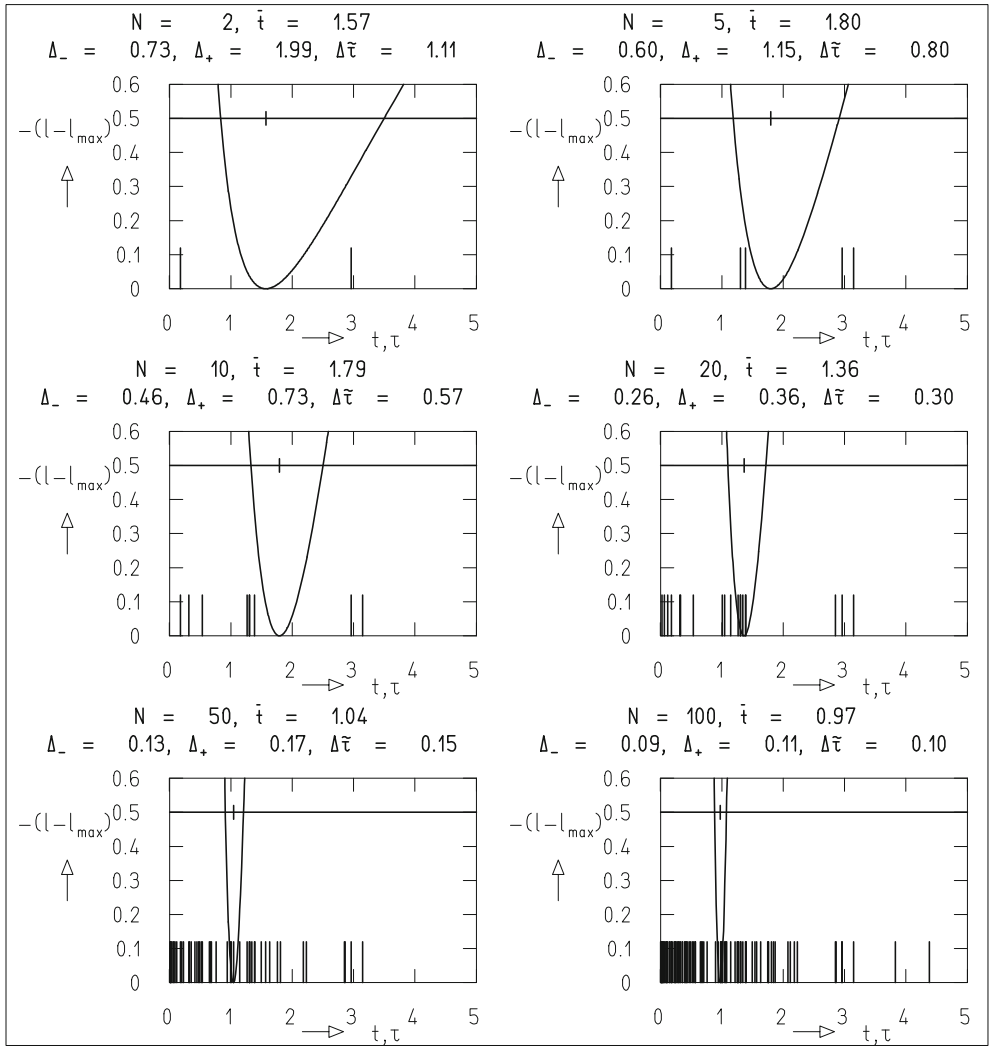

Fig. 7.2: Data and log-likelihood function for Example 7.7.

to give

$$\begin{aligned}\ell(\boldsymbol{\lambda}) &= \ell(\tilde{\boldsymbol{\lambda}}) + \sum_{k=1}^{p} \left(\frac{\partial \ell}{\partial \lambda_k}\right)_{\tilde{\boldsymbol{\lambda}}} (\lambda_k - \tilde{\lambda}_k) \\ &\quad + \frac{1}{2} \sum_{\ell=1}^{p} \sum_{m=1}^{p} \left(\frac{\partial^2 \ell}{\partial \lambda_\ell \partial \lambda_m}\right)_{\tilde{\boldsymbol{\lambda}}} (\lambda_\ell - \tilde{\lambda}_\ell)(\lambda_m - \tilde{\lambda}_m) + \cdots . \end{aligned} \tag{7.5.3}$$

Since by the definition of $\tilde{\boldsymbol{\lambda}}$ one has

$$\left(\frac{\partial \ell}{\partial \lambda_k}\right)_{\tilde{\boldsymbol{\lambda}}} = 0, \qquad k = 1, 2, \ldots, p, \tag{7.5.4}$$

which holds for all k, the series simplifies to

$$-\left(\ell(\boldsymbol{\lambda})-\ell(\tilde{\boldsymbol{\lambda}})\right)=\frac{1}{2}(\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}})^{\mathrm{T}}A(\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}})+\cdots \tag{7.5.5}$$

with

$$-A=\left(\begin{array}{cccc} \dfrac{\partial^2\ell}{\partial\lambda_1^2} & \dfrac{\partial^2\ell}{\partial\lambda_1\partial\lambda_2} & \cdots & \dfrac{\partial^2\ell}{\partial\lambda_1\partial\lambda_p} \\ \dfrac{\partial^2\ell}{\partial\lambda_1\partial\lambda_2} & \dfrac{\partial^2\ell}{\partial\lambda_2^2} & \cdots & \dfrac{\partial^2\ell}{\partial\lambda_2\partial\lambda_p} \\ \vdots & & & \\ \dfrac{\partial^2\ell}{\partial\lambda_1\partial\lambda_p} & \dfrac{\partial^2\ell}{\partial\lambda_2\partial\lambda_p} & \cdots & \dfrac{\partial^2\ell}{\partial\lambda_p^2} \end{array}\right)_{\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}}} . \tag{7.5.6}$$

In the limit $N\to\infty$ we can replace the elements of A, which still depend on the specific sample, by the corresponding expectation values,

$$B=E(A)= -\left(\begin{array}{cccc} E\left(\dfrac{\partial^2\ell}{\partial\lambda_1^2}\right) & E\left(\dfrac{\partial^2\ell}{\partial\lambda_1\partial\lambda_2}\right) & \cdots & E\left(\dfrac{\partial^2\ell}{\partial\lambda_1\partial\lambda_p}\right) \\ E\left(\dfrac{\partial^2\ell}{\partial\lambda_1\partial\lambda_2}\right) & E\left(\dfrac{\partial^2\ell}{\partial\lambda_2^2}\right) & \cdots & E\left(\dfrac{\partial^2\ell}{\partial\lambda_2\partial\lambda_p}\right) \\ \vdots & & & \\ E\left(\dfrac{\partial^2\ell}{\partial\lambda_1\partial\lambda_p}\right) & E\left(\dfrac{\partial^2\ell}{\partial\lambda_2\partial\lambda_p}\right) & \cdots & E\left(\dfrac{\partial^2\ell}{\partial\lambda_p^2}\right) \end{array}\right)_{\boldsymbol{\lambda}=\tilde{\boldsymbol{\lambda}}} . \tag{7.5.7}$$

If we neglect higher-order terms, we can give the likelihood function as

$$L=k\exp\{-\frac{1}{2}(\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}})^{\mathrm{T}}B(\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}})\}. \tag{7.5.8}$$

Comparing with (5.10.1) shows that this is a p-dimensional normal distribution with mean $\tilde{\boldsymbol{\lambda}}$ and covariance matrix

$$C=B^{-1}. \tag{7.5.9}$$

The variances of the maximum likelihood estimators $\tilde{\lambda}_1,\tilde{\lambda}_2,\ldots,\tilde{\lambda}_p$ are given by the diagonal elements of the matrix (7.5.9). The off-diagonal elements are the covariances between all possible pairs of estimators,

$$\sigma^2(\tilde{\lambda}_i) = c_{ii}, \tag{7.5.10}$$

$$\operatorname{cov}(\tilde{\lambda}_j, \tilde{\lambda}_k) = c_{jk}. \tag{7.5.11}$$

For the correlation coefficient between the estimators $\tilde{\lambda}_j, \tilde{\lambda}_k$ we can define

$$\varrho(\tilde{\lambda}_j, \tilde{\lambda}_k) = \frac{\operatorname{cov}(\tilde{\lambda}_j, \tilde{\lambda}_k)}{\sigma(\tilde{\lambda}_j)\sigma(\tilde{\lambda}_k)}. \tag{7.5.12}$$

As in the case of a single parameter, the square roots of the variances are given as the error or standard deviations of the estimators,

$$\Delta\tilde{\lambda}_i = \sigma(\tilde{\lambda}_i) = \sqrt{c_{ii}}. \tag{7.5.13}$$

In Sect. 7.4 we determined that by giving the maximum-likelihood estimator and its error one defines a region that contains the true value of the parameter with a probability of 68.3 %. Since the likelihood function in the several parameter case is asymptotically a Gaussian distribution of several variables, this region is not determined only by the errors, but rather by the entire covariance matrix. In the special case of two parameters this is the covariance ellipse, which we introduced in Sect. 5.10.

The expression (7.5.8) has (with the replacement $\mathbf{x} = \boldsymbol{\lambda}$) exactly the form of (5.10.1). We can therefore use it for all of the results of Sect. 5.10. For the exponent one has

$$-\frac{1}{2}(\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}})^{\mathrm{T}} B (\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}) = -\frac{1}{2} g(\boldsymbol{\lambda}) = -\left\{\ell(\boldsymbol{\lambda}) - \ell(\tilde{\boldsymbol{\lambda}})\right\}. \tag{7.5.14}$$

In the parameter space spanned by $\lambda_1, \ldots, \lambda_p$, the covariance ellipsoid of the distribution (7.5.8) is then determined by the condition

$$g(\boldsymbol{\lambda}) = 1 = 2\left\{\ell(\boldsymbol{\lambda}) - \ell(\tilde{\boldsymbol{\lambda}})\right\}. \tag{7.5.15}$$

For other values of $g(\boldsymbol{\lambda})$ one obtains the confidence ellipsoids introduced in Sect. 5.10. For smaller values of N, the series (7.5.3) cannot be truncated and the approximation (7.5.7) is not valid. Nevertheless, the solution (7.5.4) can clearly still be computed. For a given probability W one obtains instead of a confidence ellipsoid a *confidence region*, contained within the hypersurface

$$g(\boldsymbol{\lambda}) = 2\left\{\ell(\boldsymbol{\lambda}) - \ell(\tilde{\boldsymbol{\lambda}})\right\} = \text{const}. \tag{7.5.16}$$

The value of g is determined in the same way as for the confidence ellipsoid as in Sect. 5.10.

In Example 7.7 we computed the region $\tilde{\lambda} - \Delta_- < \lambda < \tilde{\lambda} + \Delta_+$ for the case of a single variable. This corresponds to a confidence region with the probability content 68.3 %.

Example 7.8: Estimation of the mean and variance of a normal distribution

We want to determine the mean λ_1 and standard deviation λ_2 of a normal distribution using a sample of size N. This problem occurs, for example, in the measurement of the range of α-particles in matter. Because of the statistical nature of the energy loss through a large number of independent individual collisions, the range of the individual particles is Gaussian distributed about some mean value. By measuring the range $\mathsf{x}^{(j)}$ of N different particles, the mean λ_1 and "straggling constant" $\lambda_2 = \sigma$ can be estimated. We obtain the likelihood function

$$L = \prod_{j=1}^{N} \frac{1}{\lambda_2\sqrt{2\pi}} \exp\left(-\frac{(\mathsf{x}^{(j)} - \lambda_1)^2}{2\lambda_2^2}\right)$$

and

$$\ell = -\frac{1}{2}\sum_{j=1}^{N} \frac{(\mathsf{x}^{(j)} - \lambda_1)^2}{\lambda_2^2} - N\ln\lambda_2 - \text{const.}$$

The system of likelihood equations is

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda_1} &= \sum_{j=1}^{N} \frac{\mathsf{x}^{(j)} - \lambda_1}{\lambda_2^2} = 0, \\ \frac{\partial \ell}{\partial \lambda_2} &= \frac{1}{\lambda_2^3}\sum_{j=1}^{N} (\mathsf{x}^{(j)} - \lambda_1)^2 - \frac{N}{\lambda_2} = 0. \end{aligned}$$

Its solution is

$$\begin{aligned} \tilde{\lambda}_1 &= \frac{1}{N}\sum_{j=1}^{N} \mathsf{x}^{(j)}, \\ \tilde{\lambda}_2 &= \sqrt{\frac{\sum_{j=1}^{N} (\mathsf{x}^{(j)} - \tilde{\lambda}_1)^2}{N}}. \end{aligned}$$

For the estimator of the mean, the maximum-likelihood method leads to the arithmetic mean of the individual measurements. For the variance it gives the quantity s'^2 (6.2.4), which has a small bias, and not s^2, the unbiased estimator (6.2.6).

Let us now determine the matrix B. The second derivatives are

$$\frac{\partial^2 \ell}{\partial \lambda_1^2} = -\frac{N}{\lambda_2^2},$$

$$\frac{\partial^2 \ell}{\partial \lambda_1 \partial \lambda_2} = -\frac{2\sum(\mathsf{x}^{(j)} - \lambda_1)}{\lambda_2^3},$$
$$\frac{\partial^2 \ell}{\partial \lambda_2^2} = -\frac{3\sum(\mathsf{x}^{(j)} - \lambda_1)^2}{\lambda_2^4} + \frac{N}{\lambda_2^2}.$$

We use the procedure of (7.5.7), substitute λ_1, λ_2 by $\tilde{\lambda}_1, \tilde{\lambda}_2$ and find

$$B = \begin{pmatrix} N/\tilde{\lambda}_2^2 & 0 \\ 0 & 2N/\tilde{\lambda}_2^2 \end{pmatrix}$$

or for the covariance matrix

$$C = B^{-1} = \begin{pmatrix} \tilde{\lambda}_2^2/N & 0 \\ 0 & \tilde{\lambda}_2^2/2N \end{pmatrix}.$$

We interpret the diagonal elements as the errors of the corresponding parameters, i.e.,

$$\Delta\tilde{\lambda}_1 = \tilde{\lambda}_2/\sqrt{N}, \qquad \Delta\tilde{\lambda}_2 = \tilde{\lambda}_2/\sqrt{2N}.$$

The estimators for λ_1 and λ_2 are not correlated. ■

Example 7.9: Estimators for the parameters of a two-dimensional normal distribution

To conclude we consider a population described by a two-dimensional normal distribution (Sect. 5.10)

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left[-\frac{1}{2(1-\varrho^2)} \times \left\{\frac{(x_1-a_1)^2}{\sigma_1^2} + \frac{(x_2-a_2)^2}{\sigma_2^2} - 2\varrho\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2}\right\}\right].$$

By constructing and solving a system of five simultaneous likelihood equations for the five parameters $a_1, a_2, \sigma_1^2, \sigma_2^2, \varrho$ we obtain for the maximum-likelihood estimators

$$\bar{\mathsf{x}}_1 = \frac{1}{N}\sum_{j=1}^{N} \mathsf{x}_1^{(j)}, \qquad \bar{\mathsf{x}}_2 = \frac{1}{N}\sum_{j=1}^{N} \mathsf{x}_2^{(j)},$$
$$\mathsf{s}_1'^2 = \frac{1}{N}\sum_{j=1}^{N} (\mathsf{x}_1^{(j)} - \bar{\mathsf{x}}_1)^2, \qquad \mathsf{s}_2'^2 = \frac{1}{N}\sum_{j=1}^{N} (\mathsf{x}_2^{(j)} - \bar{\mathsf{x}}_2)^2,$$

$$\mathsf{r} = \frac{\sum_{j=1}^{N} (\mathsf{x}_1^{(j)} - \bar{\mathsf{x}}_1)(\mathsf{x}_2^{(j)} - \bar{\mathsf{x}}_2)}{N \mathsf{s}_1' \mathsf{s}_2'}. \tag{7.5.17}$$

Exactly as in Example 7.8, the estimators for the variances $\mathsf{s}_1'^2$ and $\mathsf{s}_2'^2$ are biased. This also holds for the expression (7.5.17), the *sample correlation coefficient* r. Like all maximum likelihood estimators, r is consistent, i.e., it provides a good estimation of ϱ for very large samples. For $N \to \infty$ the probability density of the random variable r becomes a normal distribution with mean ϱ and variance

$$\sigma^2(\mathsf{r}) = (1 - \varrho^2)^2 / N. \tag{7.5.18}$$

For finite samples the distribution is asymmetric. It is therefore important to have a sufficiently large sample before applying Eq. (7.5.17). As a rule of thumb, $N \geq 500$ is usually recommended. ■

7.6 Example Programs

Example Program 7.1: The class E1MaxLife computes the mean lifetime and its asymmetric errors from a small number of radioactive decays

The program performs the computations and the graphical display for the problem described in Example 7.7. First by Monte Carlo method a total of N decay times t_i of radioactive nuclei with a mean lifetime of $\tau = 1$ are simulated. The number N of decays and also the seeds for the random number generator are entered interactively.

Example Program 7.2: The class E2MaxLife computes the maximum-likelihood estimates of the parameters of a bivariate normal distribution from a simulated sample

The program asks interactively for the number n_{exp} of experiments to simulate (i.e., of the samples to be treated consecutively), for the size n_{pt} of each sample and for the means a_1, a_2, the standard deviations σ_1, σ_2, and the correlation coefficient ρ of a bivariate Gaussian distribution.

The covariance matrix C of the normal distribution is calculated and the generator of random numbers from a multivariate Gaussian distribution is initialized. Each sample is generated and then analyzed, i.e., the quantities $\bar{x}_1$, $\bar{x}_2$, s_1', s_2', and r are computed, which are estimates of a_1, a_2, σ_1, σ_2, and ρ [cf. (7.5.17)]. The quantities are displayed for each sample.

Suggestions: Choose $n_{\text{exp}} = 20$, keep all other parameters fixed, and study the statistical fluctuations of r for $n_{\text{pt}} = 5, 50, 500$. Use the values $\rho = 0, 0.5, 0.95$.