The fundamental problem towards which the study of statistics is addressed, is that of inference. Some data are observed and we wish to make statements, inferences, about one or more unknown features of the physical system which gave rise to these data. (O'Hagan 1994, p. 1)

The variety of concepts relevant to the logical basis of the process of estimation of hypothetical quantities by the aid of observational material is considerable, ... (Fisher 1956, p. 5)

1 Likelihood based statistical inference

The prime difficulty lay in the uncertainty of such inferences, and it was a fortunate coincidence that the recognition of the concept of probability ... provided a possible means by which such uncertainty could be specified and made explicit. (Fisher 1956, p. 9)

1.1 Evidential meaning of statistical evidence

- **Statistical evidence**

  **Observed response**

  Let $y_{\text{obs}}$ be the observed response from some experiment $E$ containing uncontrolled haphazard variation. The first component of statistical evidence consists of $y_{\text{obs}}$.

  **Statistical model**

  The second necessary part of statistical evidence is a statistical model $\mathcal{M} = \{ p(y; \omega) : y \in \mathcal{Y}, \omega \in \Omega \}$, which is a set of possible probability distributions $p(y; \omega)$ for the observed data $y_{\text{obs}}$ with sample space $\mathcal{Y}$ and the parameter $\omega$ belonging to the parameter space $\Omega$. In some cases there may be other information, such as a priori distribution for the parameter. In the following we consider only the situation where statistical evidence consists of the pair $(y_{\text{obs}}, \mathcal{M})$. 

Evidential statements

... inference can most conveniently be thought of as concerned with statements about the unknown values of parameters. (O'Hagan 1994, p. 4)

Collection of evidential statements supported by statistical evidence

The result of a statistical inference procedure is a collection of statements concerning the unknown values of the parameter functions of interest. The statements are based on statistical evidence. The important problem of the theory of statistical inference is to characterize the form of statements that a given statistical evidence supports.

Uncertainties of statements

The Principle of Adequacy. A concept of statistical evidence is (very) inadequate if it does not distinguish evidence of (very) different strengths. (Pratt et al 1965)

Uncertainties or risks of the evidential statements

The uncertainties of the statements are also an essential part of the inference. The statements and their uncertainties are both based on statistical evidence and together form the evidential meaning of the statistical evidence. The theory of statistical inference has to characterize the way by which the uncertainties of the supported statements is determined. In the following we consider one approach to solve these problems, namely, the likelihood based approach to statistical inference or likelihood method.

1.2 Statistical principles

Principles

The problem of the foundation of statistics is to state a set of principles which entail the validity of all correct statistical inference, and which do not imply that any fallacious inference is valid. (Hacking 1965, p. 1)

Statistical principles

Statistical principles are certain descriptions of properties that the various statistical procedures derived by statistical methods like for example the likelihood method should have. They themselves have no other grounds than an intuitive appeal and more importantly they are judged by their results, that is, that they characterize as ‘good’ only those statistical procedures generally considered to be good. In the following we present the main principles considered generally to be the relevant principles with regard to informative statistical inference.

Likelihood

In the theory of estimation it has appeared that the whole of the information supplied by a sample is comprised in the likelihood, as a function known for all possible values of the parameter. (Fisher 1956, p. 70)
(Weak) Likelihood Principle

(Weak) Likelihood Principle: According to (weak) likelihood principle the evidential meaning of the statistical evidence \((y_{\text{obs}}, M)\) in a experiment \(E\) should depend on the observed data \(y_{\text{obs}}\) only through the observed likelihood function \(L_M(\omega; y_{\text{obs}})\) or equivalently through the observed log-likelihood function \(l_M(\omega; y_{\text{obs}})\).

Weak likelihood principle does not imply that evidential meanings in different experiments must be equivalent in case the observed data in those experiments have proportional likelihood functions.

Strong Likelihood Principle

Strong likelihood principle: According to strong likelihood principle the evidential meaning of the statistical evidence \((y_{\text{obs}}, M)\) in a experiment \(E\) should depend on the observed data \(y_{\text{obs}}\) and the model \(M\) only through the observed likelihood function \(L_M(\omega; y_{\text{obs}})\) or equivalently through the observed log-likelihood function \(l_M(\omega; y_{\text{obs}})\).

Strong likelihood principle does imply that evidential meanings in different experiments must be equivalent in case the observed data in those experiments have proportional likelihood functions.

■ Sufficiency

In this theory a case of peculiar simplicity arises when an estimate exists which, perhaps in conjunction with ancillary statistics, subsumes the whole of the information, relevant to the parameter, supplied by the observational record. (Fisher 1956, p. 49)

Minimal sufficient statistic

In a given statistical model \(M = \{p(y; \omega) : y \in Y, \omega \in \Omega\}\) a statistic \(S(y)\) is sufficient if the conditional distribution of \(y\) given the value of \(S(y) = s\) is independent of the model parameter \(\omega\). A sufficient statistic is minimal sufficient if it is a function of every other sufficient statistic. It can be shown that the likelihood function of the statistical model \(M\) depends on observed data \(y\) through the minimal sufficient statistic.

Sufficiency Principle

Sufficiency principle: The evidential meaning of the statistical evidence of an experiment \(E\) should depend on observed data only through minimal sufficient statistic.

Thus sufficiency principle implies that the evidential meaning does not depend on any such random aspects of the observed data, the distribution of which does not depend on the model parameter. Because there is one to one correspondence between the values of the minimal sufficient statistic and the possible likelihood functions, the weak likelihood principle and the sufficiency principle are equivalent.

■ Conditionality

In connection with methods of statistical inference, probability is used in two quite distinct ways. The first is,... to represent the haphazard behaviour of the empirical system under investigation, i.e. to define the stochastic model assumed to have generated the data. The second is to assess uncertainty in conclusions, via significance levels, confidence regions, posterior distributions and so forth. In essence, we consider methods of analysis to be analogous to measuring instruments and hence to be calibrated via their performance under repeated use. That is, we enquire, whether by analytical arguments or by simulation, as to how such a method would perform if, hypothetically, it were used repeatedly on data derived from the model under study. The probabilities used as a basis for inference are thus long-run frequencies under hypothetical repetition. The issue now arises, as before, how these long-run frequencies are to be made relevant to the data under study. The answer lies again conditioning the calculations so that the ‘long run’ matches the particular set of data in important respects.” (Barndorff-Nielsen and Cox 1994, p. 33)
Ancillary statistic

In a given statistical model $\mathcal{M} = \{p(y; \omega) : y \in \mathcal{Y}, \omega \in \Omega\}$ a function $A(y)$ of the minimal sufficient statistic $S(y)$ is ancillary if the marginal distribution of $A(y)$ is independent of the model parameter $\omega$.

Conditionality Principle

Conditionality principle: The evidential meaning of the statistical evidence $(y_{\text{obs}}, \mathcal{M})$ in an experiment $\mathcal{E}$ should depend on the model $\mathcal{M}$ only through the conditional model function $p_{\mathcal{M}}(y | a_{\text{obs}}; \omega)$ whenever $A(y)$ is an ancillary statistic.

Because the model function can be written in the form

$$p_{\mathcal{M}}(y; \omega) = p_{\mathcal{M}}(S(y); \omega) p_{\mathcal{M}}(y | S(y)) = p_{\mathcal{M}}(A(y)) p_{\mathcal{M}}(S(y)|A(y); \omega) p_{\mathcal{M}}(y | S(y)).$$

we see that the observed likelihood function $L_{\mathcal{M}}(\omega; y_{\text{obs}})$ is equivalent to the conditional observed likelihood function $L_{\mathcal{M}}(\omega; s_{\text{obs}} | a_{\text{obs}})$. This means that the original model function $p_{\mathcal{M}}(y; \omega)$ can be replaced by the conditional model function $p_{\mathcal{M}}(\omega; s | a_{\text{obs}})$ without losing information needed to satisfy all the principles mentioned above.

- Informative inference

Limited Strong Likelihood Principle

The strong likelihood principle demands that the whole evidential meaning, that is the collection of the statements and their uncertainties, depends on the observations and the model only through the observed likelihood function. On the other hand the weak likelihood principle (the sufficiency principle) and the conditionality principle allow that the some part of the evidential meaning may depend on other characteristics of the model. In the likelihood based approach only part of the evidential meaning obeys the strong likelihood principle, that is, the likelihood based approach to statistical inference satisfies the following version of the strong likelihood principle.

Limited strong likelihood principle: The collection of evidential statements of the statistical evidence $(y_{\text{obs}}, \mathcal{M})$ in an experiment $\mathcal{E}$ should depend on the observations $y_{\text{obs}}$ and the model $\mathcal{M}$ only through the observed likelihood function $L_{\mathcal{M}}(\omega; y_{\text{obs}})$.

Principle of Informative Inference

If it is allowed that the uncertainties of the evidential statements need not satisfy the strong likelihood principle but should satisfy the conditionality principle the following informative inference principle arises.

Principle of informative inference: The set of evidential statements of the statistical evidence $(y_{\text{obs}}, \mathcal{M})$ in a experiment $\mathcal{E}$ should depend on the observations $y_{\text{obs}}$ and the model $\mathcal{M}$ only through the observed likelihood function $L_{\mathcal{M}}(\omega; y_{\text{obs}})$ and the uncertainties of the evidential statements should depend on the observations $y_{\text{obs}}$ and the model $\mathcal{M}$ only through the conditional model function $p_{\mathcal{M}}(s | a_{\text{obs}})$ whenever $S(y)$ is a minimal sufficient statistic and $A(y)$ is an ancillary statistic.
1.3 Statistical laws

Laws

Statistical principles discussed above give only certain minimal requirements that a 'good' statistical procedure should satisfy, that is, they do not give unique answers and usually there is an infinite number of possible procedures satisfying those minimal requirements. The (strong) likelihood principle, for example, only tells that the conclusions should depend on the observed data (and the model) through likelihood function. It does not say what to do with the likelihood function.

In addition to statistical principles some other rules are needed to guide the search for a 'good' statistical procedure among those satisfying the relevant principles.

Law of direct inference

In some such cases, but not all, as will be seen, inferences can be drawn assigning a calculable mathematical probability to any assertion to the effect that the parameter lies between assigned limits. (Fisher 1956, p. 47)

Law of likelihood or indirect inference

The likelihood supplies a natural order of preference among the possibilities under consideration. (Fisher 1956, p. 70)

Law of informative inference

For all purposes, and more particularly for the communication of the relevant evidence supplied by a body of data, the values of the Mathematical Likelihood are better fitted to analyse, summarise, and communicate statistical evidence of types too weak to supply true probability statements; ... (Fisher 1956, p. 72)

Law of Informative Inference. The collection of evidential statements must satisfy the law of likelihood and the uncertainties or risks of the evidential statements must be calculated from the conditional model function $p(A | d_{obs})$ whenever $S(y)$ is a minimal sufficient statistic and $A(y)$ is an ancillary statistic.

This principle means that the evidential meaning, that is, both the collection of evidential statements and their uncertainties, must satisfy the conditionality principle and the collection of evidential statements but not their uncertainties also the strong likelihood principle.
1.4 Likelihood method

... neither effectively exact definitive statements, nor statements in terms of Mathematical Probability, are possible, yet in which some information is available, and we are not in a state of complete ignorance. (Fisher 1956, p. 65)

■ Pure likelihood approach

In the case under discussion a simple graph of the values of the Mathematical Likelihood expressed as a percentage of its maximum, against the possible values of the parameter p, shows clearly enough what values of the parameter have likelihoods comparable with the maximum, and outside what limits the likelihood falls to levels at which the corresponding values of the parameter become implausible. (Fisher, p. 72)

Relative likelihood function

Let \( y_{\text{obs}} \) be the observed response and \( \mathcal{M} = \{ p(\omega; y) : y \in \mathcal{Y}, \omega \in \Omega \} \) it's statistical model. According to law of likelihood the likelihood ratio \( \frac{L(\omega; y_{\text{obs}})}{L(\hat{\omega}; y_{\text{obs}})} \) measures, how much more or less the observed response \( y_{\text{obs}} \) supports the value \( \omega_1 \) of the parameter vector compared to the value \( \omega_2 \). Because the likelihood ratio does not change, if the likelihood function is multiplied by a number not depending on the parameter vector \( \omega \), it is convenient for the sake of various comparisons to multiply it by a appropriate number. A number generally used is the reciprocal of the maximum value of the likelihood function. The version \( R(\omega; y_{\text{obs}}) = \frac{L(\omega; y_{\text{obs}})}{L(\hat{\omega}; y_{\text{obs}})} \) of the likelihood function obtained in this way is called relative likelihood function, where \( \hat{\omega} \) the value of the parameter vector maximizing the likelihood function.

Logarithmic relative likelyhood function

The relative likelihood function takes values between 0 and 1 and it's maximum value is 1. Logarithmic relative likelihood function \( r(\omega; y_{\text{obs}}) = \ln \left( \frac{L(\omega; y_{\text{obs}})}{L(\hat{\omega}; y_{\text{obs}})} \right) \) is the logarithm of the relative likelihood function. The logarithmic relative likelihood function has it's values in the interval \( (-\infty, 0) \) and its maximum value is 0.

Pure likelihood approach

The pure likelihood approach gives statistical inference procedures, for which the collection of evidential statements is based wholly on the likelihood function. This means that the approach satisfies the strong likelihood principle. In a sense the likelihood function as a whole is often considered as the evidential statement and giving the evidential meaning of the statistical evidence. More generally, however, the likelihood function is summarized using some simpler evidential statements.

Likelihood Region

The most common way of summarizing the pure likelihood inference is to use the law of likelihood and based on that to give likelihood regions of appropriate or all percentage values. This method was enforced by Fisher (Fisher 1956) and consists of sets of the form

\[ \{ \omega : R(\omega; y_{\text{obs}}) \geq c \} = \left\{ \omega : \frac{L(\omega; y_{\text{obs}})}{L(\hat{\omega}; y_{\text{obs}})} \geq c \right\} = \{ \omega : r(\omega; y_{\text{obs}}) \geq \ln(c) \} \]

for some or all \( 0 < c < 1 \). The set is called the 100\(c\)% likelihood region. According to the law of likelihood the observed response supports every value of the parameter vector in the likelihood region more than any value outside the region. For each value \( \omega \) of the parameter vector in the likelihood region one can also say that there does not exist
any value of the parameter vector which would be supported by the statistical evidence more than \( \frac{1}{c} \)-fold compared to \( \omega \).

### Likelihood approach

It is customary to evaluate and compare statistical procedures by examining how they would behave in a series of hypothetical repetitions of the experiment. One imagines that the experiment which gave rise to the data is to be repeated over and over again under identical conditions. One then determines how frequently the statistical procedure would give correct results in this series of repetitions. (Kalbfleisch 1985, p. 96)

Note that the repetitions to which we keep referring are purely imaginary. Real experiments do not get repeated over and over again under identical conditions. The series of repetitions is invented by the statistician to give a theoretical framework within which statistical procedures may be investigated and compared. (Kalbfleisch 1985, p. 97)

It is suggested that the best way to construct confidence intervals is by calculating likelihood intervals of the appropriate size. Intervals constructed in this way will give valid information summaries in particular applications, as well as having the desired frequency properties in a series of imaginary repetitions. (Kalbfleisch 1985, p. 96)

**Likelihood based confidence region**

In the likelihood approach the collection of evidential statements is formed as in the pure likelihood approach. In addition the uncertainties of the evidential statements are calculated using the conditionality principle and so the approach satisfies the law of informative inference.

The given value \( \omega \) of the parameter vector does not belong to the 100\( c \)% likelihood region, if the response is such that

\[
r(\omega; y) < \log(c).
\]

The probability of this event calculated at the given value \( \omega \) of the parameter vector is used as a measure of uncertainty or risk of the likelihood region provided that the probability has the same value or at least approximately the same value for all possible values of the parameter vector.

The likelihood region for which the level \( c \) has been chosen so that the measure of the uncertainty of the region is equal to \( \alpha \) is called likelihood based \((1 - \alpha)\)-level confidence region.

**Likelihood Ratio Statistic**

Because asymptotically the likelihood ratio statistic

\[
-2 r(\omega; y) = 2 \left[ \ln(\hat{\omega}; y) - \ln(\omega; y) \right] = 2 \ln \left( \frac{L(\omega; y)}{L(\hat{\omega}; y)} \right)
\]

has \( \chi^2 \)-distribution with degrees of freedom \( q \) equal to the dimension of \( \omega \), the uncertainty of the likelihood region

\[
\{ \omega : -2 r(\omega; y_{\text{obs}}) \leq \chi^2_{\alpha}(q) \} = \{ \omega : \ln(\omega; y_{\text{obs}}) \geq \ln(\hat{\omega}; y_{\text{obs}}) - \chi^2_{1-\alpha}(q)/2 \}
\]

is the probability of the event

\[
r(\omega; y) < \chi^2_{1-\alpha}(q)/2,
\]

which is approximately equal to \( \alpha \). Thus the likelihood region is an approximate \((1-\alpha)\)-level confidence region for \( \omega \). The problem with this approach is that when sample size is small or more exactly when information in data is meager, the approximation may be pure. The well-known Bartlett-correction or other solutions given by Barndorff-Nielsen -Cox approach try to give regions with actual confidence levels closer the nominal ones, but this usually destroys the limited strong likelihood property of the regions.
It is to be noted that the collection of confidence sets must include sets of all confidence levels for the limited strong likelihood property to hold. This means that the sets essentially give the same information as the likelihood function as a whole, but they are calibrated using the measure of uncertainty given by the confidence level (Cox 1988). If for example the collection consists of only one confidence region, say the 95%-level region, the limited strong likelihood principle will not be satisfied, because generally the exact critical value and so the region depends on the model also otherwise than through the likelihood function.

## 1.5 Respiratory Disorders

The following table summarizes the information from a randomized clinical trial that compared two treatments (test, placebo) for a respiratory disorder (Stokes et al 1995, p. 4).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Favorable</th>
<th>Unfavorable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>16</td>
<td>48</td>
</tr>
<tr>
<td>Test</td>
<td>40</td>
<td>20</td>
</tr>
</tbody>
</table>

```
<< StatisticalInference`
```

Data

The following table summarizes the information from a randomized clinical trial that compared two treatments (test, placebo) for a respiratory disorder (Stokes et al 1995, p. 4).

```
dataFile = ToFileName[Directory[],"RespiratoryDisorders.dat"];
TableForm[dataOnRespiratoryDisorders = Import[dataFile],
  TableAlignments->Right,
  TableHeadings->{None, "Treatment", "Favorable", "Unfavorable"}]
```

```
treatment = dataOnRespiratoryDisorders[[All,1]]; favorable = dataOnRespiratoryDisorders[[All,2]]; unfavorable = dataOnRespiratoryDisorders[[All,3]]; patients = favorable + unfavorable;
```

```
proportions = unfavorable/patients;
proportions[[2]]-proportions[[1]] //N
-0.416667
```

```
odds = unfavorable/favorable;
Log[odds[[2]]/odds[[1]]] //N
-1.79176
```
Model

\[ M = \text{IndependenceModel}[[\text{BinomialModel}[64,p],\text{BinomialModel}[60,p]]] \]

--- IndependenceModel ---

Model fitting

\[ \mathcal{F} = \text{MLEFit}[M,\{16,40\}] \]

--- FittedModel ---  Convergence: True

\[[p_1, p_2] = \text{MLParameterEstimate}[\mathcal{F}]\] // N

\{0.25, 0.666667\}

Likelihood region

\[ L = \text{LikelihoodFunction}[M,\{16,40\},\{\omega_1,\omega_2\}]; \]

\[ R = \frac{L}{L/.\{\omega_1 \to p_1, \omega_2 \to p_2\}} \]

164473023104465444668166369181696 (1 - \omega_1)^{48} \omega_1^{16} (1 - \omega_2)^{20} \omega_2^{20} \]
ContourPlot[Evaluate[R], {ω1, 0.1, 0.4}, {ω2, 0.5, 0.8}, Contours -> {0.2}, ContourShading -> False, PlotLabel -> "20% likelihood region", FrameLabel -> {ω1, ω2}];

- Likelihood based confidence region

\[
l = \text{LogLikelihoodFunction}[M, \{16, 40\}, \{\omega_1, \omega_2\}];
\]

\[
r = l - (l /. (\omega_1 \to p1, \omega_2 \to p2))
= 48 \log\left[\frac{4}{3}\right] + 40 \log\left[\frac{3}{2}\right] + 20 \log[3] + 16 \log[4] + 48 \log[1 - \omega_1] + 16 \log[\omega_1] + 20 \log[1 - \omega_2] + 40 \log[\omega_2]
\]
Most practical problems of parametric inference, it seems to me, fall in the category of ‘inference in the presence of nuisance parameters,’ with the acknowledgement that various functions $g(\theta)$, which might be components of a parameter vector, are simultaneous interest. That is, in practice, we may wish to make inferences about each of many such functions $g(\theta)$ in turn, keeping in mind that joint inferences are being made, yet requiring separate statements for ease of contemplation. (Kass in Reid 1988)

### 2.1 Profile likelihood region

Most practical problems of parametric inference, it seems to me, fall in the category of ‘inference in the presence of nuisance parameters,’ with the acknowledgement that various functions $g(\theta)$, which might be components of a parameter vector, are simultaneous interest. That is, in practice, we may wish to make inferences about each of many such functions $g(\theta)$ in turn, keeping in mind that joint inferences are being made, yet requiring separate statements for ease of contemplation. (Kass in Reid 1988)

#### Parameter function of interest

Let $y_{\text{obs}}$ be the observed response and $M = \{ p(y; \omega) : y \in \mathcal{Y}, \omega \in \Omega \}$ its statistical model. In most practical problems only part of the parameter vector or more generally value of a given function of parameter vector is of interest. In the former case the rest of the parameter vector consists of *nuisance parameters*. 

```mathematica
ContourPlot[r, \{\omega_1, 0.1, 0.4\}, \{\omega_2, 0.5, 0.81\},
Contours -> {-QuantileAtParameter[ChiSquareModel[2, \mu], 0.95, 2] / 2},
ContourShading -> False,
PlotLabel -> "95% confidence region", FrameLabel -> \{\omega_1, \omega_2\}];
```
Let $g(\omega)$ be a given interest function with $q$-dimensional real vectors as values. Then the function

$$L_g(\psi; y_{\text{obs}}) = \max_{\omega \in \Omega : g(\omega) = \psi} L(\omega; y_{\text{obs}}) = L(\hat{\omega}_g; y_{\text{obs}})$$

is called the profile likelihood function of the interest function $g(\omega)$ based on the statistical evidence $y_{\text{obs}}, \mathcal{M}$.

The parameter vector $\hat{\omega}_g$ maximizes the likelihood function in the subset $\{\omega \in \Omega : g(\omega) = \psi\}$ of the parameter space.

Logarithmic Profile Likelihood Function

The function

$$l_g(\psi; y_{\text{obs}}) = \max_{\omega \in \Omega : g(\omega) = \psi} l(\omega; y_{\text{obs}}) = l(\hat{\omega}_g; y_{\text{obs}}) = \log(L_g(\psi; y_{\text{obs}}))$$

is called the logarithmic profile likelihood function of the interest function $g(\omega)$ based on the the statistical evidence $y_{\text{obs}}, \mathcal{M}$.

(Logarithmic) Relative Profile Likelihood Function

Furthermore the functions

$$R_g(\psi; y_{\text{obs}}) = \frac{L_g(\psi; y_{\text{obs}})}{L_g(\hat{\psi}; y_{\text{obs}})} = \frac{L(\hat{\omega}_g; y_{\text{obs}})}{L(\hat{\omega}; y_{\text{obs}})}$$

and

$$r_g(\psi; y_{\text{obs}}) = l_g(\psi; y_{\text{obs}}) - l_g(\hat{\psi}; y_{\text{obs}}) = l(\hat{\omega}_g; y_{\text{obs}}) - l(\hat{\omega}; y_{\text{obs}})$$

are called relative profile likelihood function and logarithmic relative profile likelihood function.

Profile likelihood region and it's uncertainty

Profile Likelihood Region

Based on the relative profile likelihood function one can construct so called profile likelihood regions. The following set of possible values of the interest function $g(\omega)$

$$\{\psi : R_g(\psi; y_{\text{obs}}) \geq p\} = \{\psi : r_g(\psi; y_{\text{obs}}) \geq \log(p)\}$$

is $100p\%$ profile likelihood region.

Profile Likelihood based Confidence Region

The value $\psi = g(\omega)$ of the interest function does not belong to the $100p\%$ profile likelihood region, if the response $y$ is such that

$$r_g(\psi; y) < \log(p).$$

The probability of this event calculated at a given value $\omega$ of the parameter vector is used as a measure of uncertainty or risk of the profile likelihood region provided that this probability takes the same or at least approximately the same value for all values of the parameter vector. If the uncertainty is equal to $\alpha$, then the profile likelihood region is called $(1 - \alpha)$-level confidence region.
Likelihood Ratio Statistic

Because the statistic

\[-2 r_g(\psi; y) = 2 \left[ l(\hat{\psi}; y) - l(\psi; y) \right] = 2 \left[ l(\hat{\omega}; y) - l(\tilde{\omega}_q; y) \right] \]

has asymptotically \( \chi^2 \)-distribution with degrees of freedom \( q \) equal to the dimension of \( \psi \), the profile likelihood region

\[ \{ \psi : r_g(\psi; y_{obs}) = -\chi^2_{1-a}(q)/2 \} = \{ \psi : l_g(\psi; y_{obs}) = l_g(\tilde{\psi}; y_{obs}) - \chi^2_{1-a}(q)/2 \} \]

is an approximate \( (1 - \alpha) \)-level confidence region for the interest function \( g(\omega) \).

### 2.2 Profile likelihood intervals

We regard the primary mode for summarizing evidence about a one-dimensional parameter of interest to be a nested series of upper and lower confidence limits based as directly as possible on the likelihood. (Barndorff-Nielsen & Cox 1994, p. 2)

In case of real valued interest function \( \psi = g(\omega) \) the approximate \((1 - \alpha)\)-level profile likelihood based confidence region has the following form

\[ \{ \psi : l_g(\psi; y_{obs}) \geq l(\hat{\omega}; y_{obs}) - \chi^2_{1-a}(1)/2 \} \]

The set need not be an interval of the real line, although it often is.

If the profile likelihood based confidence region is an interval, it’s end points \( \psi^L \) and \( \psi^U \) satisfy the relations \( \psi^L < \psi < \psi^U \) and

\[ l_g(\psi^L; y_{obs}) = l_g(\psi^U; y_{obs}) = l(\hat{\omega}; y_{obs}) - \chi^2_{1-a}(1)/2. \]

Thus \( \psi^L \) and \( \psi^U \) are roots of the equation \( l_g(\psi; y_{obs}) = l(\hat{\omega}; y_{obs}) - \chi^2_{1-a}(1)/2 \). Consequently most applications of the likelihood based intervals have determined \( \psi^L \) and \( \psi^U \) using some iterative root finding method. This approach involves the solution of an optimisation problem for all trial values of the interest function and depending on the method also the calculation the derivatives of the profile likelihood function at the same trial values.

### 2.3 New method for calculation of intervals

The approximate \((1 - \alpha)\)-level profile likelihood based confidence set for parameter function \( \psi = g(\omega) \) satisfies the relation

\[ \{ \psi : l_g(\psi; y_{obs}) > l(\hat{\omega}; y_{obs}) - \chi^2_{1-a}(1)/2 \} = \{ g(\omega) : \omega \in \mathcal{A}_{1-a}(y_{obs}) \}, \]

where

\[ \mathcal{A}_{1-a}(y_{obs}) = \{ \omega : l(\omega; y_{obs}) > l(\hat{\omega}; y_{obs}) - \chi^2_{1-a}(1)/2 \} \]
is a likelihood region for the whole parameter vector. This result is true, because the real number $\psi$ belongs to the left hand side of (*), if and only if there exist a parameter vector $\omega^*$ such that $g(\omega^*) = \psi$ and

$$\sup_{\omega \in \Omega : g(\omega) = \psi} l(\omega; y_{\text{obs}}) = l_\psi(\psi; y_{\text{obs}}) \geq l(\hat{\omega}; y_{\text{obs}}) > I(\hat{\omega}; y_{\text{obs}}) - \chi^2_{1-\alpha}[1]/2.$$  

That is if and only if there exists a parameter vector $\omega^*$ belonging to the likelihood region $\mathcal{A}_{1-\alpha}(y_{\text{obs}})$ such that it satisfies the equation $g(\omega^*) = \psi$. So the number $\psi$ belongs to the left hand side (*), if and only if it belongs to the right hand side of it.

Now if the confidence set (*) is an interval, the end points $\psi^L$ and $\psi^U$ satisfy the following relations

$$\psi^L = \inf_{\omega \in \mathcal{A}_{1-\alpha}(y_{\text{obs}})} g(\omega) \quad \text{and} \quad \psi^U = \sup_{\omega \in \mathcal{A}_{1-\alpha}(y_{\text{obs}})} g(\omega).$$

Let now the likelihood region be a bounded and connected subset of the parameter space. If the log-likelihood and the interest functions are continuous in the likelihood region, the latter with non-vanishing gradient, then the confidence set (*) is an interval $(\psi^L, \psi^U)$ with

$$\psi^L = g(\hat{\omega}_L^L) = \inf\{g(\omega) ; l(\omega; y_{\text{obs}}) = l(\hat{\omega}; y_{\text{obs}}) - \chi^2_{1-\alpha}[1]/2\}$$

and

$$\psi^U = g(\hat{\omega}_U^U) = \sup\{g(\omega) ; l(\omega; y_{\text{obs}}) = l(\hat{\omega}; y_{\text{obs}}) - \chi^2_{1-\alpha}[1]/2\}.$$  

This follows from the assumptions, because they imply, that the set $\mathcal{A}_{1-\alpha}(y_{\text{obs}})$ is open, connected, and bounded subset of the $p$-dimensional space. Thus the closure of $\mathcal{A}_{1-\alpha}(y_{\text{obs}})$ is a closed, connected, and bounded set. Form the assumptions concerning $g$ it follows, that it attains its infimum and supremum on the boundary of $\mathcal{A}_{1-\alpha}(y_{\text{obs}})$, and takes every value between infimum and supremum somewhere in $\mathcal{A}_{1-\alpha}(y_{\text{obs}})$.

Under the above assumptions the solutions of the following constrained minimisation (maximisation) problem

$$\min (\max) g(\omega) \quad \text{under} \quad l(\omega; y_{\text{obs}}) = l(\hat{\omega}; y_{\text{obs}}) - \chi^2_{1-\alpha}[1]/2.$$  

gives the left (right) end point of the profile likelihood interval. This problem is rarely explicitly solvable and requires use of some kind of iteration.

### 2.4 Poisson Regression

**Source**

Lindsey 1997, p. 50.
In a study on relationship between life stresses and illness one randomly chosen member of each randomly chosen household in a sample from Oakland, California, USA, was interviewed. In a list of 41 events, respondents were asked to note which had occurred within last 18 months. The results given in the following table are those recalling only one such stressful event (Haberman 1978, p. 3).

```plaintext

```

```plaintext

```
Generalised linear model

\[ M = \text{RegressionModel}[X, \text{Distribution} \to \text{PoissonModel}[\mu], \text{InverseLink} \to \text{Function}[z, \exp(z)]] \]

Log-likelihood function

\[ l = \text{LogLikelihoodFunction}[M, \text{count}, \{\beta_1, \beta_2\}] \]

Design

\[
\begin{align*}
x1 &= 1 & \text{Range}[\text{Length}[\text{dataStressfulEvents}]]; \\
x2 &= \text{month}; \\
X &= \{x1, x2\} & \text{Transpose}; \\
p &= \text{Dimensions}[X][[2]]; \\
df &= \text{Length}[\text{dataStressfulEvents}] - p;
\end{align*}
\]
ContourPlot[1, {β1, 0, 5}, {β2, -0.4, 0.2}];

**Model fitting**

\[
\mathcal{F} = \text{MLEFit}[M, \text{count}]
\]

--- FittedModel --- Convergence: True

\[
(b = \text{MLParameterEstimate}[\mathcal{F}])//N
\]

{2.80316, -0.0837689}

**Fitted values**

\[
fittedValues = \text{Table}[\text{Mean}[M, i, \{\beta1, \beta2\}], \{i, n\}] /.
\]

\[
\text{MapThread}[\#1 \rightarrow \#2 & \{\{\beta1, \beta2\}, b\}]
\]

plot1 = ListPlot[Transpose[{{month, count}}, PlotStyle -> PointSize[0.02], DisplayFunction -> Identity];
plot2 = ListPlot[Transpose[{{month, fittedValues}}, PlotJoined -> True, DisplayFunction -> Identity];
Show[plot1, plot2, DisplayFunction -> $DisplayFunction];

### Profile interval

\[ I = \text{ProfileInterval}[\mathcal{F}, \beta_2, \{\beta_1, \beta_2\}] \]

--- ProfileIntervalModel --- Convergence: True

ConfidenceLimits[I]

\{ -0.117203, -0.0512576 \}

### 2.5 Pharmacokinetic model

#### Source


#### Data

Data on the metabolism of sulfisoxazole were obtained. In this experiment, sulfisoxazole was administered to a subject intravenously, blood samples were taken at specified times, and the concentration of sulfisoxazole in plasma in micrograms per milliliter (\(\mu g/ml\)) was measured.
dataFile = ToFileName[Directory[],"Sulfisoxazole.dat"]; TableForm[dataOnSulfisoxazole = Import[dataFile], TableAlignments->Right, TableHeadings->{None,{"Time (min)","Concentration (µg/ml)"}}]

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Concentration (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>215.6</td>
</tr>
<tr>
<td>0.5</td>
<td>189.2</td>
</tr>
<tr>
<td>0.75</td>
<td>176.</td>
</tr>
<tr>
<td>1.0</td>
<td>162.8</td>
</tr>
<tr>
<td>1.5</td>
<td>138.6</td>
</tr>
<tr>
<td>2.0</td>
<td>121.</td>
</tr>
<tr>
<td>3.0</td>
<td>101.2</td>
</tr>
<tr>
<td>4.0</td>
<td>88.</td>
</tr>
<tr>
<td>6.0</td>
<td>61.6</td>
</tr>
<tr>
<td>12.0</td>
<td>22.</td>
</tr>
<tr>
<td>24.0</td>
<td>4.4</td>
</tr>
<tr>
<td>48.0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

time = dataOnSulfisoxazole[[All,1]]; concentration = dataOnSulfisoxazole[[All,2]];

ListPlot[dataOnSulfisoxazole];

■ Design matrix

\[ X = \text{Partition}[\text{time},1]; \]

■ Pharmacokinetic model

The pharmacokinetic model with two exponential terms is defined in the following way.
\[ M = \text{RegressionModel}(X, y_1 \text{ Exp}[t - \lambda_1 t] + y_2 \text{ Exp}[t - \lambda_2 t], \{t\}, \{\gamma_1, \lambda_1, \gamma_2, \lambda_2\}) \]
--- RegressionModel ---

### Log-likelihood function

\[
\text{LogLikelihoodFunction}[M, \text{concentration}, \{\gamma_1, \lambda_1, \gamma_2, \lambda_2, t\}] \\
(0.1 - e^{-(24.6 \lambda_1 y_1 - e^{-24.6 \lambda_2} \gamma_2)})^2 - \\
(4.4 - e^{-(24.6 \lambda_1 y_1 - e^{-24.6 \lambda_2} \gamma_2)})^2 - \\
(61.6 - e^{-(24.6 \lambda_1 y_1 - e^{-24.6 \lambda_2} \gamma_2)})^2 - \\
(101.2 - e^{-(24.6 \lambda_1 y_1 - e^{-24.6 \lambda_2} \gamma_2)})^2 - \\
(138.6 - e^{-(24.6 \lambda_1 y_1 - e^{-24.6 \lambda_2} \gamma_2)})^2 - \\
(176.2 - e^{-(24.6 \lambda_1 y_1 - e^{-24.6 \lambda_2} \gamma_2)})^2 - \\
(215.6 - e^{-(24.6 \lambda_1 y_1 - e^{-24.6 \lambda_2} \gamma_2)})^2 - \\
6 \log(2 \pi) - 12 \log(t)
\]

### Model fitting

The maximum likelihood estimates of parameters are

\[ \mathcal{F} = \text{MLEFit}[M, \text{concentration}, \{80, 1, 160, 0.15, 2.8\}, \text{ConvergenceTolerance} \rightarrow 0.001] \]
--- FittedModel ---  Convergence: True

### Fitted values

\[
m = \text{Table}[\text{Mean}[M, i, \{\gamma_1, \lambda_1, \gamma_2, \lambda_2, t\}], \\
\{i, \text{Length}[\text{dataOnSulfisoxazole}]\}] / . \\
\text{MapThread}[\{\#1 \rightarrow \#2\} \&, \{\{\gamma_1, \lambda_1, \gamma_2, \lambda_2, t\}, \text{MLParameterEstimate}[\mathcal{F}]\}]
\]
\{213.231, 191.97, 174.854, 160.869, 139.475, 123.77,
101.426, 85.2173, 61.4931, 23.5842, 3.47645, 0.075538\}
plot1 = ListPlot[Transpose[{time, Flatten[concentration]}],
   PlotStyle → PointSize[0.02],
   DisplayFunction → Identity];
plot2 = ListPlot[Transpose[{time, m}], PlotJoined → True,
   DisplayFunction → Identity];
Show[plot1, plot2, DisplayFunction → $DisplayFunction];

Profile interval

The profile likelihood based confidence interval for $\lambda_2$ is

\[
I = \text{ProfileInterval}[\mathcal{F}, \lambda_2, \{\gamma_1, \lambda_1, \gamma_2, \lambda_2, \tau\}]
\]

ConfidenceLimits[I]

--- ProfileIntervalModel --- Convergence: True

\{0.14609, 0.177081\}

The profile likelihood based confidence interval for the mean $\gamma_1 e^{-\lambda_1 10} + \gamma_2 e^{-\lambda_2 10}$ at time 10 is

\[
I = \text{ProfileInterval}[\mathcal{F}, \gamma_1 \text{Exp}[-\lambda_1 10] + \gamma_2 \text{Exp}[-\lambda_2 10], \{\gamma_1, \lambda_1, \gamma_2, \lambda_2, \tau\}]
\]

ConfidenceLimits[I]

--- ProfileIntervalModel --- Convergence: True

\{29.3645, 35.1813\}

The profile likelihood based confidence interval for AUC $\frac{\gamma_1}{\lambda_1} + \frac{\gamma_2}{\lambda_2}$ is
\[ I = \text{ProfileInterval}[\mathcal{F}, \frac{\gamma_1}{\lambda_1} + \frac{\gamma_2}{\lambda_2}, \{\gamma_1, \lambda_1, \gamma_2, \lambda_2, \tau\}] \]

ConfidenceLimits[I]

--- ProfileIntervalModel --- Convergence: True

\{1020.8, 1118.98\}

References


