USER 4 MANUAL

USER-SPECIFIED ESTIMATION ROUTINE

COLUMBIA BASIN RESEARCH
SCHOOL OF AQUATIC AND FISHERY SCIENCES
UNIVERSITY of WASHINGTON

USER 4: User-Specified Estimation Routine

James M. Lady, John R. Skalski

Columbia Basin Research
School of Aquatic and Fishery Sciences
University of Washington
1325 4th Avenue, Suite 1820
Seattle, WA 98101-2509

Prepared for
U.S. Department of Energy
Bonneville Power Administration
Division of Fish and Wildlife
P.O. Box 3621
Portland, OR 98208-3621
Project No. 1989-107-00
Contract No. 39987

May 2009

Contents

List of Figures

1.0 Overview

Program **USER 4** is a tool that allows investigators to estimate parameters of a study with the following characteristics:

- All possible outcomes of the study can be characterized in terms of a finite number of discrete categories.
- The categories are mutually exclusive and exhaustive; i.e., an individual in the study can be classified into one and only one category.
- The data from the study consist of the number of individuals in the study that fall into each category.

A model within the USER framework consists of a likelihood with zero or more auxiliary likelihoods. Each likelihood consists of two or more categories, and each category consists of:

- A unique **label** for identifying the category.
- The **probability** for the category, defined as a function of the model parameters.
- A **count** indicating the number of observations for the category.

Since the categories are mutually exclusive and exhaustive, the probabilities must sum to 1.0.

Program USER includes an integrated, context-sensitive help system that guides the user on how to interact with the program. Therefore, this document concentrates on describing some applications, setting them up in USER, and estimating the parameters of interest.

The outline of this manual is as follows.

1. A brief description of the USER interface (Chapter 2).
2. A description of the integrated help system (Chapter 3).
3. The analysis of four hypothetical studies
    (a) A study with one likelihood (Section 4.1).
    (b) A study requiring an auxiliary likelihood (Section 4.2).

(c) A study for estimating abundance, requiring the use of an unobserved category (Section 4.3).

(d) An example of performing hypothesis testing (Section 4.4).

This document describes version 4.4 of Program USER.

2.0 Interface Description

The user interface for Program USER upon startup is illustrated in Figure 2.1. The left-hand side consists of the navigation panel, and the right-hand side consists of the content page. The active content page at startup is the "Parameters" page.

Along the top of the USER dialog is a toolbar with buttons that provide shortcuts to corresponding actions in the "File" menu.

2.1 Navigation Panel

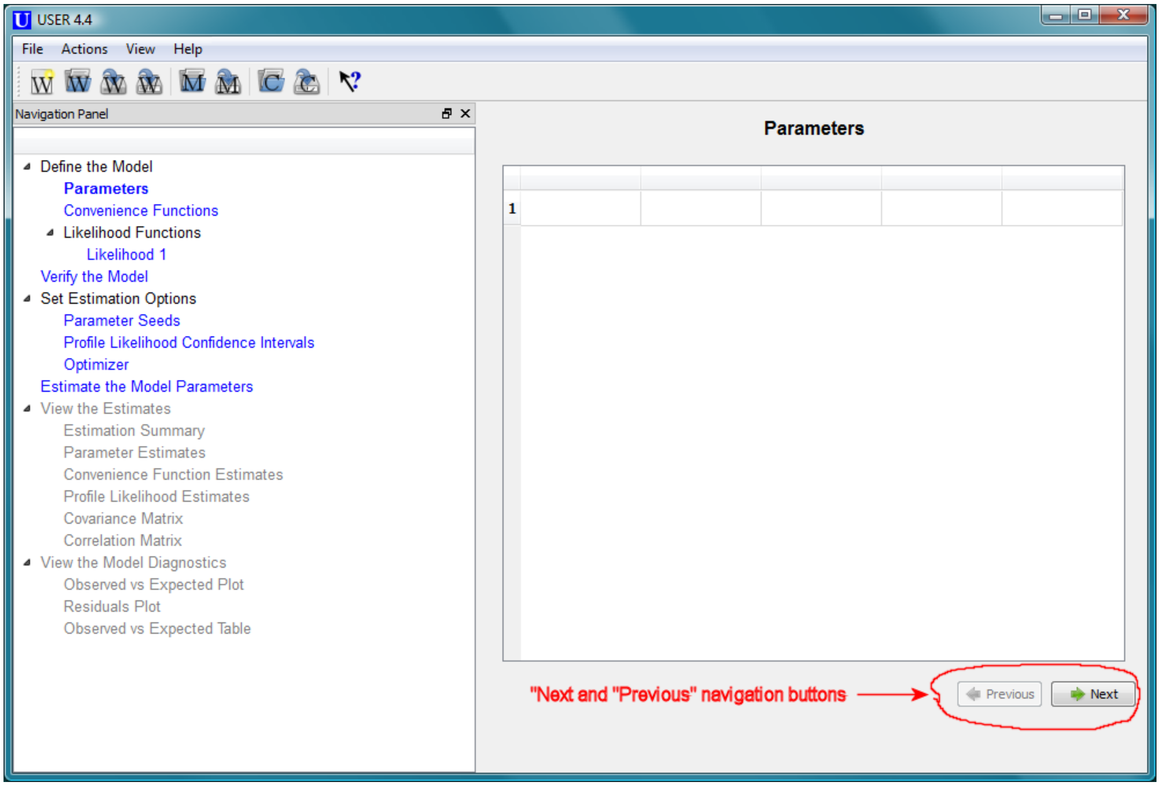
By default, the navigation panel is always displayed on the left-hand side of the USER dialog. Its visibility can be toggled on and off via the "View" menu.

The navigation panel is organized to show the usual progression of steps one would normally take in defining a model, estimating the parameters, and viewing the results. Each content page has a "Next" and "Previous" button (see Figure 2.1) in the lower right side of the screen to allow the user to progress through these steps. However, the user can also double-click on the title of any content page to go directly to that page (unless it is grayed out and thus unavailable).

Headers used strictly for organizational purposes are distinguished from content pages by the use of blue text for actual content pages. The title of the currently active page is displayed in bold text.

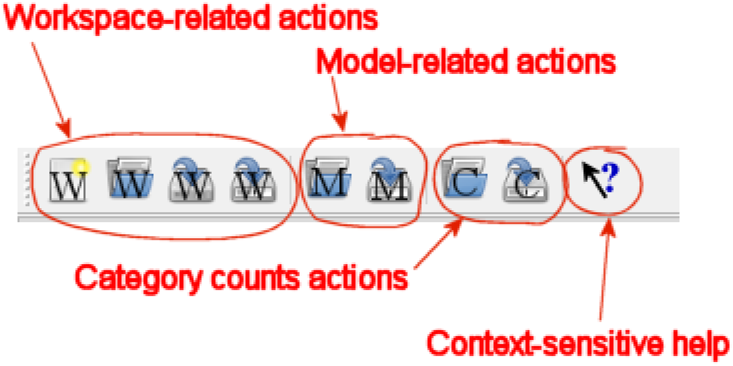
2.2 Toolbar

The USER toolbar (Figure 2.2) provides shortcut buttons for commands in the "File" menu. The buttons with "W" on them pertain to a workspace file, those with "M" pertain to the model definition file, and those with "C" pertain to the category counts file. The toolbar also contains a context-sensitive help button. The visibility of the toolbar can be toggled off and on similarly to the Navigation Panel.



"Next and "Previous" navigation buttons

Figure 2.1: The USER 4 dialog as it appears at startup.



Workspace-related actions

Model-related actions

Category counts actions

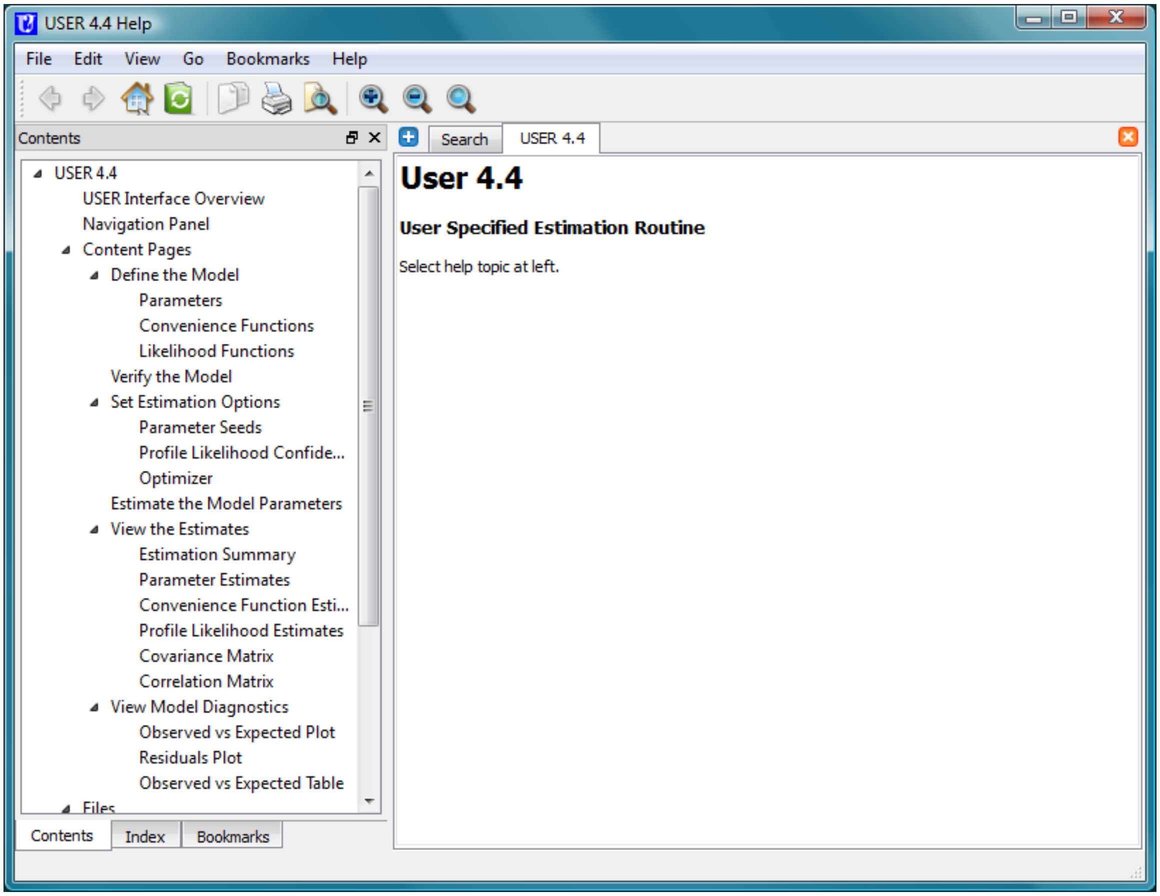
Context-sensitive help

Figure 2.2: The USER 4 toolbar provides shortcut buttons for commands in the "File" menu. "W" pertains to the workspace file, "M" to the model definition file, and "C" to the category counts file.

3.0 Integrated Help System

Program USER 4 includes an integrated, context-sensitive help system. The “Contents” action under the “About” menu will bring up a separate help viewer as shown in Figure 3.1. The navigation panel on the left side shows the list of available topics. Under the “Content Pages,” the structure parallels the structure of the content pages of Program USER 4.

To obtain context-sensitive help on a specific item, the user clicks on the help button on the toolbar (see Figure 2.2), and then clicks on the item of interest; if the user clicks on the currently active content page, the help for that particular content page will be displayed on the help viewer.



User 4.4

User Specified Estimation Routine

Select help topic at left.

Figure 3.1: The integrated help viewer for USER 4.4.

4.0 Examples

4.1 One Likelihood

A simple method for estimating the number of animals in a closed population using a single mark-release of individuals is the Petersen Method. A sample of n_1 individuals is taken from the population and marked for future identification and returned to the population. After some interval, a second sample of n_2 individuals is taken and it is found that m of them are marked. We are interested in estimating the parameter N, the number of animals in the population.

For this example we will use the following expository data:

n_1:	220
n_2:	235
m:	40

If we let P represent the proportion marked after initial tagging, then $P = \frac{n_1}{N}$.

Assuming that the proportion of marked individuals in the second sample is a reasonable estimate of the proportion marked in the population, then the probability that an animal in n_2 is marked is P.

The categories and their corresponding probabilities of occurrence are as follows:

Category	**Probability**
m	P
$n_2 - m$	$1 - P$

The resulting likelihood is

$$L = \binom{n_2}{m} \left(\frac{n_1}{N}\right)^m \left(1 - \frac{n_1}{N}\right)^{n_2 - m} \tag{4.1}$$

4.1.1 Model Definition

Define the Parameters

On startup, the "Parameters" content page is active on the USER dialog. In this example, we have one parameter to estimate: N, so we place the cursor in the first cell of the parameter definitions table and type "N" and press "Tab" to move to next cell (Figure 4.1).

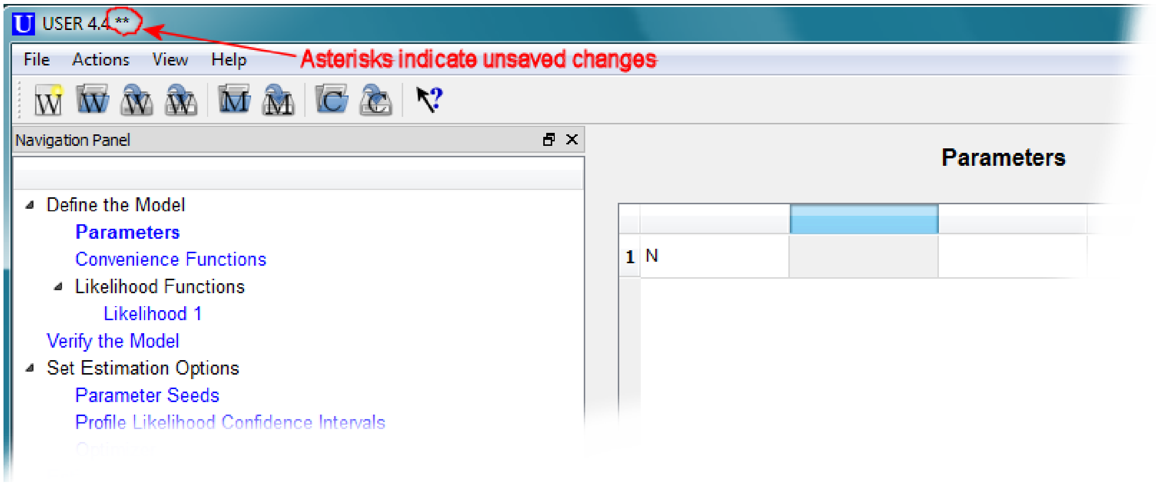


Figure 4.1: The "Parameters" page for the one likelihood example

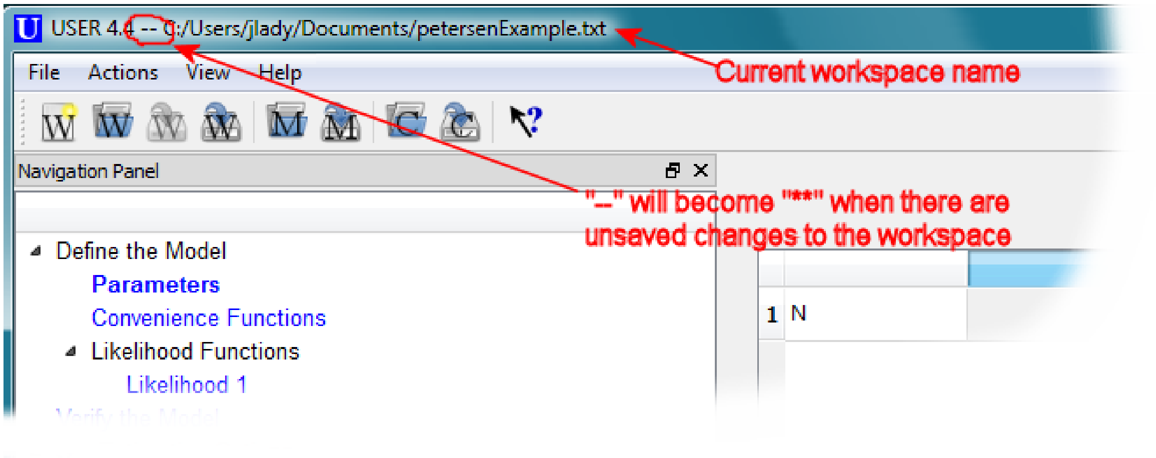
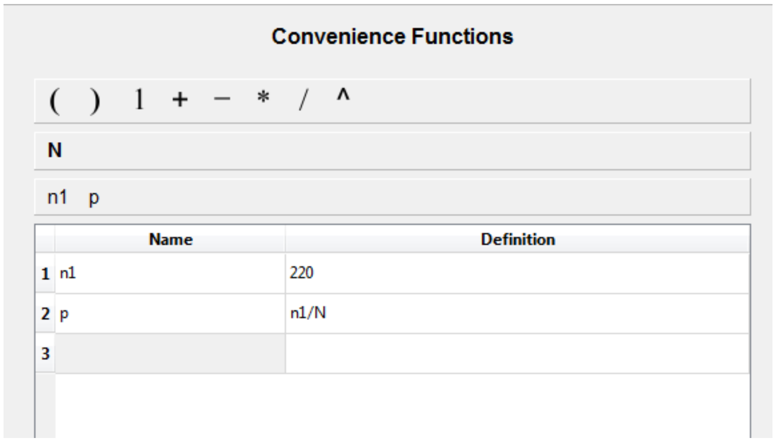


Figure 4.2: Using a workspace file

Notice in Figure 4.1 that after entering the parameter N, two asterisks appear in the window title after the version number. This indicates that there are unsaved changes to the workspace. Use the "Save Workspace" or "Save Workspace as" commands to save your work to a workspace file. If you then exit USER and start it up again, you can use the "Load Workspace" command to pick up where you left off. (see Figure 4.2).

Define the Convenience Functions

From the "Parameters" content page, press the "Next" button (or double-click on "Convenience Functions" on the navigation panel). This will make the "Convenience Functions" content page active.



	Name	Definition
1	n1	220
2	p	n1/N
3		

Figure 4.3: Convenience functions for the one likelihood example

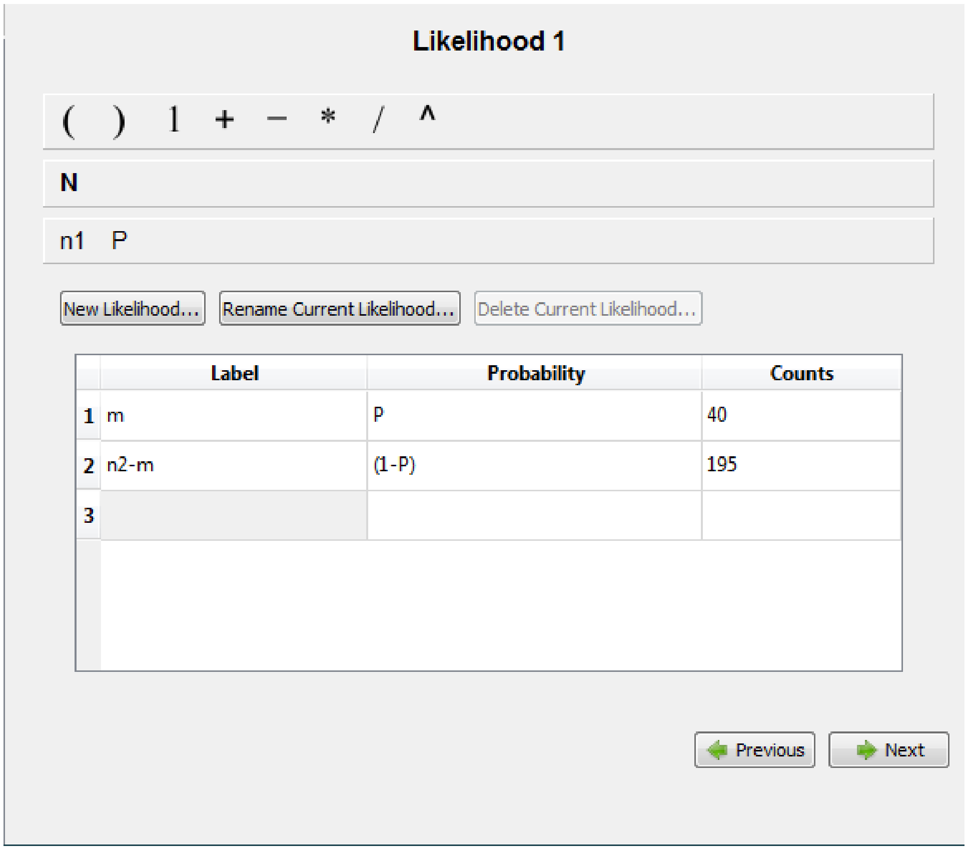
Convenience functions allow a user to define a symbol as a substitute for a function of the parameters and/or other convenience functions; or, as in this case, a substitute for a constant value. Convenience functions are optional, but they can be useful in making it easier to define the likelihood and make the likelihood more readable and flexible.

The likelihood for this example uses the value $n1$, which in this case is 220. We could simply use the number "220" directly in the likelihood, but instead we will define it as a convenience function. In the first row under name, enter "n1". Tab to the "Definition" column and enter "220".

Next, we define P as $\frac{n_1}{N}$. In the second row enter "P" for the name. In the Definition column, enter the function $\frac{n_1}{N}$. You could simply type "n1/N", or you could use the shortcut buttons on the toolbars on the top of the "Convenience Functions" page as shown in Figure 4.3. The topmost toolbar provides some commonly used mathematical operators. The second toolbar provides the parameter names, and the third toolbar provides the names of convenience functions defined so far. Thus, an alternative way to enter the definition for P is as follows:

- With the cursor in the definition cell for P, click on the "n1" on the third toolbar.
- Click on the divide symbol ("/") on the top toolbar.
- Click on "N" on the second toolbar.

The right side of the USER dialog (the content page) should look like Figure 4.3 when completed.



	Label	Probability	Counts
1	m	P	40
2	n2-m	(1-P)	195
3			

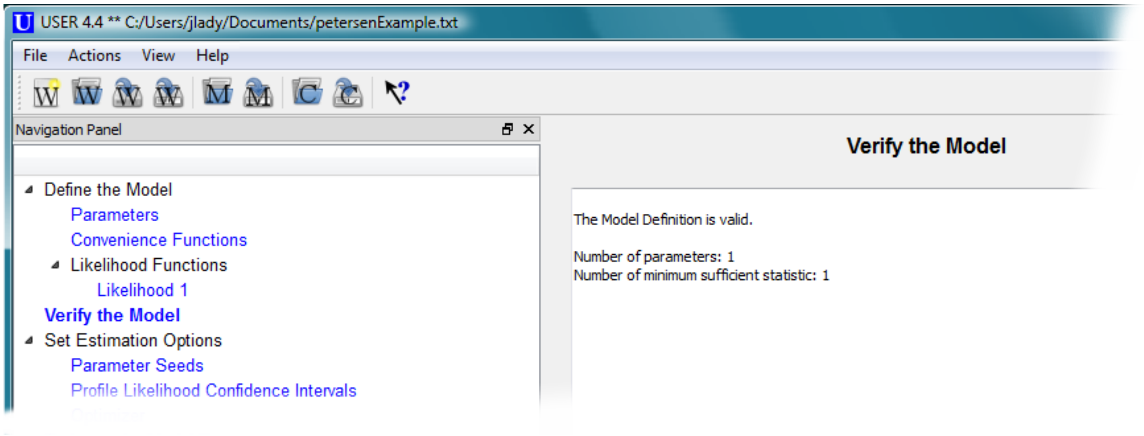
Figure 4.4: Likelihood definition for the one likelihood example

Define the Likelihood

Press "Next" to go to the Likelihood Definition Page. Each likelihood consists of two or more mutually exclusive and exhaustive categories. From Equation 4.1 and the mark-recapture data on page 7, we can define the categories as follows:

Label	Probability	Count
m	P	40
$n2 - m$	$1 - P$	195

Consequently, the Likelihood Definition page is completed in as shown in Figure 4.4. Notice the "(1-P)" probability for the "n2-m" category. This can be filled in by holding the Control key on the keyboard and pressing the "P" on the toolbar. 1 minus any symbol (parameter or convenience function) can be entered by holding the Control key and pressing the symbol name on the appropriate toolbar.



The Model Definition is valid.

Number of parameters: 1
Number of minimum sufficient statistic: 1

Figure 4.5: Verify the Model for the one likelihood example

Verify the Model

Pressing the "Next" key from the Likelihood Definition page makes the Verify the Model page active, verifying that

1. There are no undefined symbols used in the likelihood definition, and
2. The probabilities of occurrence sum to 1.0.

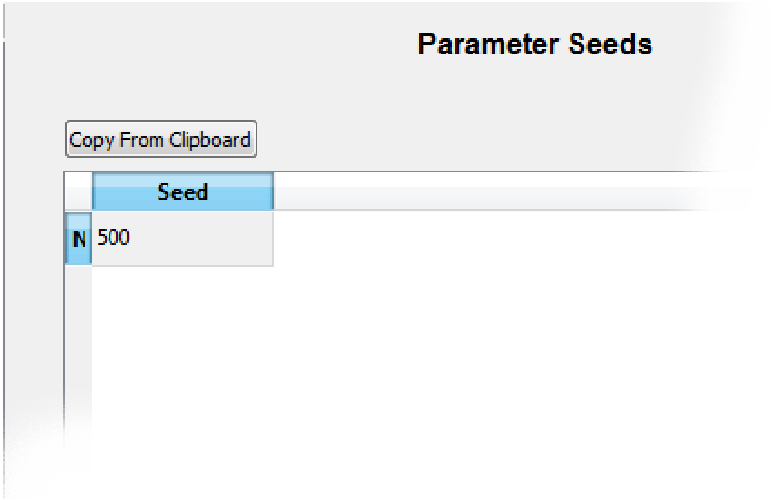
The Verify the Model page also dispays the number of parameters and the dimension of the Minimum Sufficient Statistic. If the dimension is less than the number of parameters for a model with one likelihood, the parameters may not be estimable. If everything has been done correctly, the USER dialog should now appear as on Figure 4.5. If there are any undefined symbols, or if the probabilities do not sum to 1.0, you will not be able to estimate the model parameters.

We have now finished the model definition phase. The parameters and likelihoods are defined, and we move to the estimation phase.

4.1.2 Estimation

Specify the Parameter Seeds

The "Next" button takes the investigator to the "Parameter Seeds" content page. Parameter estimation in USER is performed by maximizing the likelihood function using numerical optimization. All numerical optimization procedures require initial seeds, or starting points for the parameters. USER uses a default value of 0.5 for all parameter seeds. This is usually a good seed for estimating



	Seed
N	500

Figure 4.6: Parameter seed for the one likelihood example

probabilities that are between 0.0 and 1.0. In this case, however, the parameter to be estimated is animal abundance, so we need to change the seed to a more reasonable number, say, 500 (Figure 4.6).

The next content page is the “Profile Likelihood Confidence Intervals” page. No profile likelihood confidence intervals will be requested for this example, so proceed to the “Optimizer” content page.

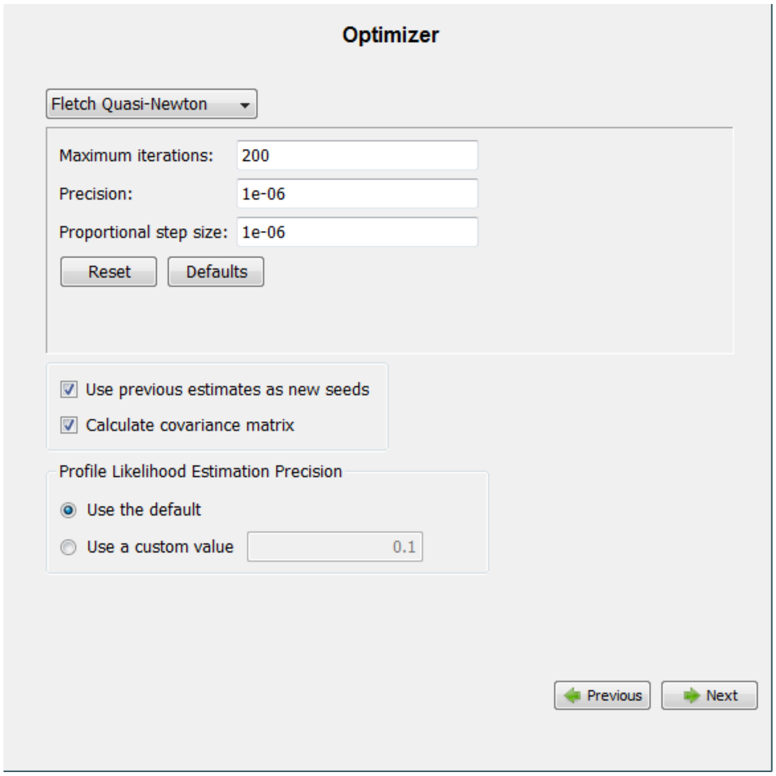
Optimizer

The next content page is titled “Optimizer,” and allows the user to specify the numerical optimization settings. The default optimizer is called “Fletch” which uses the “quasi Newton-Rhapson method” to numerically solve for the parameter estimates (Figure 4.7).

Refer to the context-sensitive help in the integrated help system (Chapter 3) for more information on the settings on the “Optimizer” content page. In most cases, it is sufficient to use the default settings.

Estimate the Model Parameters

The next content page is the “Estimate the Model Parameters” page. This is where, after defining the model, specifying the counts, and setting up the optimizer, the parameters are actually estimated. Press the “Estimate” button on the upper left, and the output of the optimizer will appear in the text area of this content page (Figure 4.8).



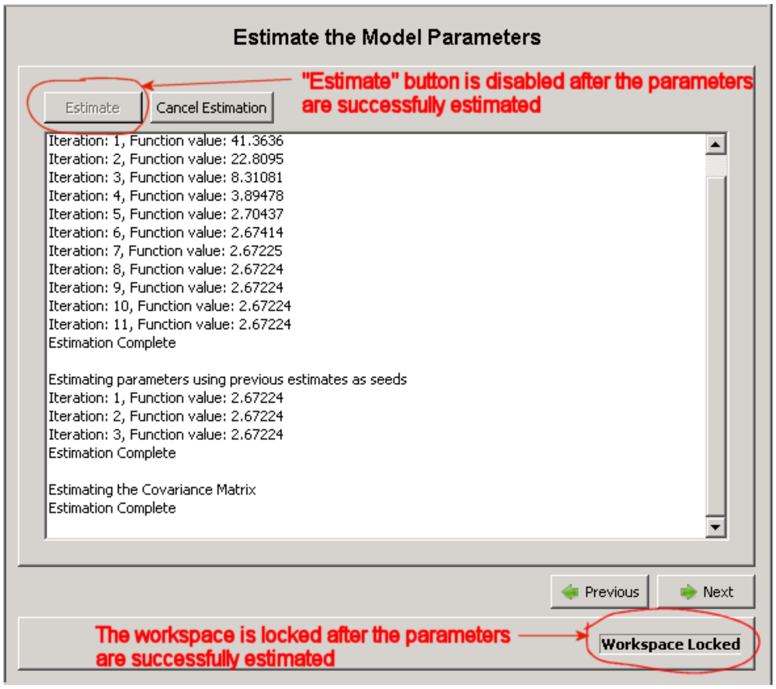
Fletch Quasi-Newton

Maximum iterations: 200
Precision: 1e-06
Proportional step size: 1e-06

☑ Use previous estimates as new seeds
☑ Calculate covariance matrix

Profile Likelihood Estimation Precision
● Use the default
○ Use a custom value 0.1

Figure 4.7: Optimizer settings for the one likelihood example



```
Iteration: 1, Function value: 41.3636
Iteration: 2, Function value: 22.8095
Iteration: 3, Function value: 8.31081
Iteration: 4, Function value: 3.89478
Iteration: 5, Function value: 2.70437
Iteration: 6, Function value: 2.67414
Iteration: 7, Function value: 2.67225
Iteration: 8, Function value: 2.67224
Iteration: 9, Function value: 2.67224
Iteration: 10, Function value: 2.67224
Iteration: 11, Function value: 2.67224
Estimation Complete

Estimating parameters using previous estimates as seeds
Iteration: 1, Function value: 2.67224
Iteration: 2, Function value: 2.67224
Iteration: 3, Function value: 2.67224
Estimation Complete

Estimating the Covariance Matrix
Estimation Complete
```

"Estimate" button is disabled after the parameters are successfully estimated

The workspace is locked after the parameters are successfully estimated

Figure 4.8: "Estimate the Model Parameters" page after the successful estimation of the model parameters for the one likelihood example

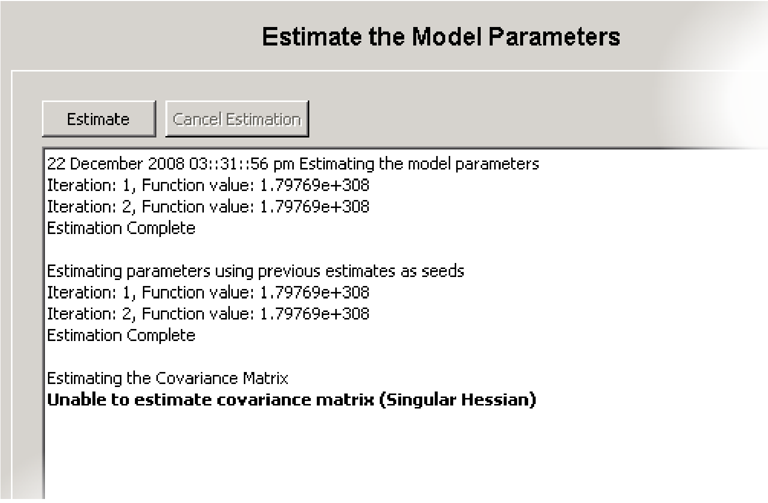
If the model parameters are successfully estimated, the "Workspace Locked" indicator will appear in the lower right (circled in red in Figure 4.8). This means that you can no longer make any changes to the workspace (parameters, convenience functions, likelihood definitions, counts) without first explicitly unlocking the workspace. This ensures that the model results and diagnostics are consistent with what appears in the "Model Definition" section (Section 4.1.1). If you wish to make changes, you must use the "Clear Current Estimates" command under the "Action" menu. This will erase all current estimation results.

If you left the parameter seed at the default value of 0.5 , pressing the "Estimate" button might produce something like what we see in Figure 4.9. If the seed is bad, the optimizer may fail to find parameter estimates. If this happens, go back to the "Parameter Seeds" content page and change the seed. For complex models, it may take some trial and error to find appropriate parameter seeds.

4.1.3 Estimates

The next content page under the "Estimates" header is titled "Estimation Summary". It summarizes the results and provides:

- The categories with their corresponding counts.

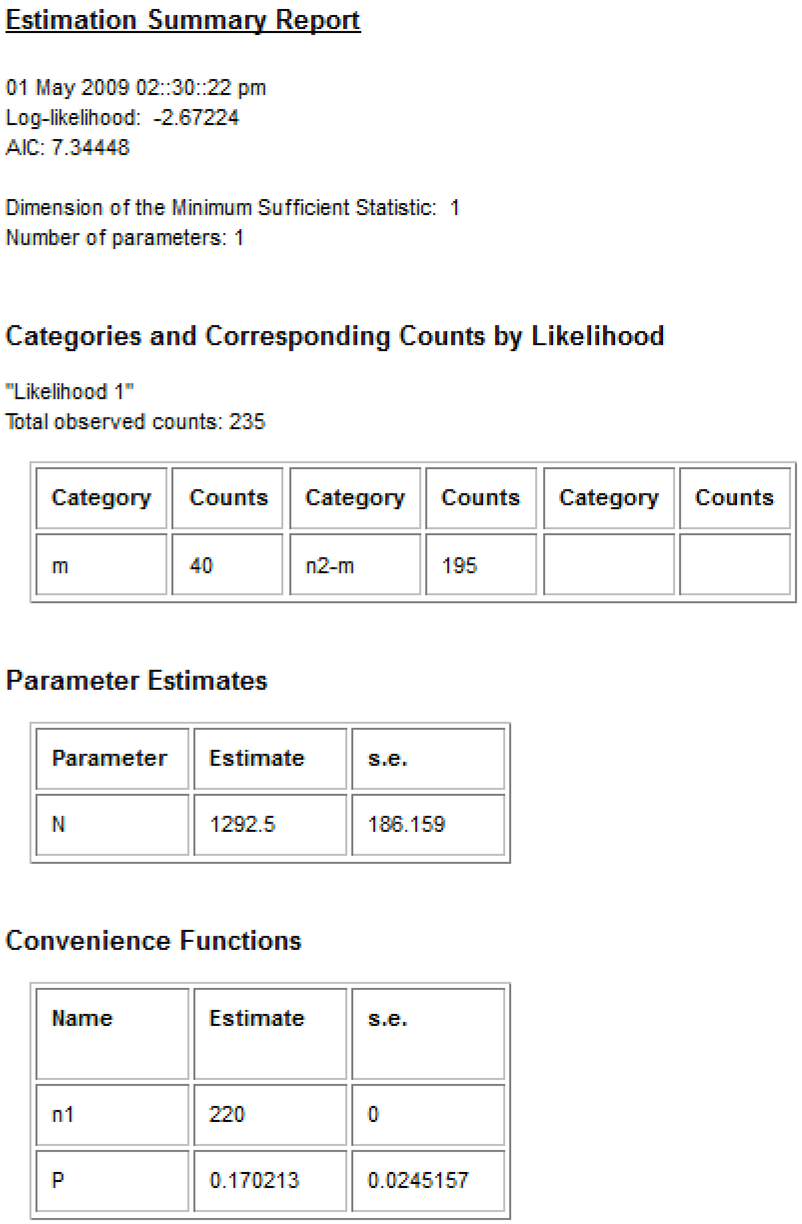


```
22 December 2008 03::31::56 pm Estimating the model parameters
Iteration: 1, Function value: 1.79769e+308
Iteration: 2, Function value: 1.79769e+308
Estimation Complete

Estimating parameters using previous estimates as seeds
Iteration: 1, Function value: 1.79769e+308
Iteration: 2, Function value: 1.79769e+308
Estimation Complete

Estimating the Covariance Matrix
Unable to estimate covariance matrix (Singular Hessian)
```

Figure 4.9: "Estimate the Model Parameters" page after estimation failure for the one likelihood example

- The parameter estimates along with their standard errors.
- The convenience function estimates along with their standard errors, calculated using the delta-method.
- The requested profile likelihood confidence intervals.

The "Estimation Summary" page for this example is shown in Figure 4.10.



Estimation Summary Report

01 May 2009 02::30::22 pm
Log-likelihood: -2.67224
AIC: 7.34448

Dimension of the Minimum Sufficient Statistic: 1
Number of parameters: 1

Categories and Corresponding Counts by Likelihood

"Likelihood 1"
Total observed counts: 235

Category	Counts	Category	Counts	Category	Counts
m	40	n2-m	195		

Parameter Estimates

Parameter	Estimate	s.e.
N	1292.5	186.159

Convenience Functions

Name	Estimate	s.e.
n1	220	0
P	0.170213	0.0245157

Figure 4.10: "Estimation Summary" page for the one likelihood example

4.2 Joint Likelihood Example

We now look at an example in which not all the parameters are estimable in a single likelihood; an auxiliary likelihood is required to estimate the model parameters.

In a simplified, hypothetical study to estimate the survival of downstream migrating juvenile salmon in a given river reach, acoustic-tagged salmon are released and detected at a downstream detection site as show in Figure 4.11. The parameter of interest is the survival probability S. Not all fish are detected at the downstream detection site, so the model must include a detection probability P.

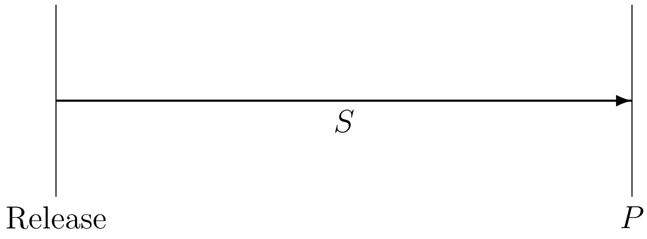


Figure 4.11: Joint likelihood example: Diagram of the study design

The categories and corresponding probabilities of occurrence are as follows:

Category	**Description**	**Probability**
n_1	Detected	SP
n_2	Not detected	$1 - SP$

The resulting likelihood is as follows:

$$L_1 \propto (SP)^{n1}(1 - SP)^{n2} \tag{4.2}$$

Notice that the parameters S and P always occur together in the likelihood model, and therefore are not separately estimable; there is no way of distinguishing between mortality and non-detection. A solution is to modify the study so that there are two independent detection sites, as show in Figure 4.12.

This modified design makes is possible to estimate the detection probability separately from the survival probability by using an auxiliary likelihood. Instead of one detection probability, we now have two detection probabilities, P_1 and P_2, associated with each of the two detection arrays.

We now define the categories for the detection process. All probabilities in the auxiliary likelihood are conditional on detection somewhere at the downstream

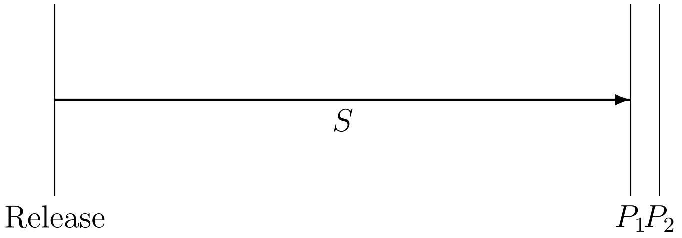


Figure 4.12: Joint likelihood example: Diagram of the modified study design with two independent detection sites

detection site. The overall detection probability at the two arrays is now defined as $P = 1 - (1 - P_1)(1 - P_2)$.

Category	**Description**	**Probability**
n_a	Detected at site #1 only	$P_1(1 - P_2)/P$
n_b	Detected at site #2 only	$(1 - P_1)P_2/P$
n_{ab}	Detected at both site #1 and site #2	P_1P_2/P

The auxiliary likelihood is defined as:

$$L_2 \propto (P_1(1 - P_2)/P)^{n_a} \, ((1 - P_1)P_2/P)^{n_b} \, (P_1P_2/P)^{n_{ab}}$$

The joint likelihood is defined as

$$L = L_1 \times \text{Ł}_2$$

We can now define the joint model in USER 4.2.

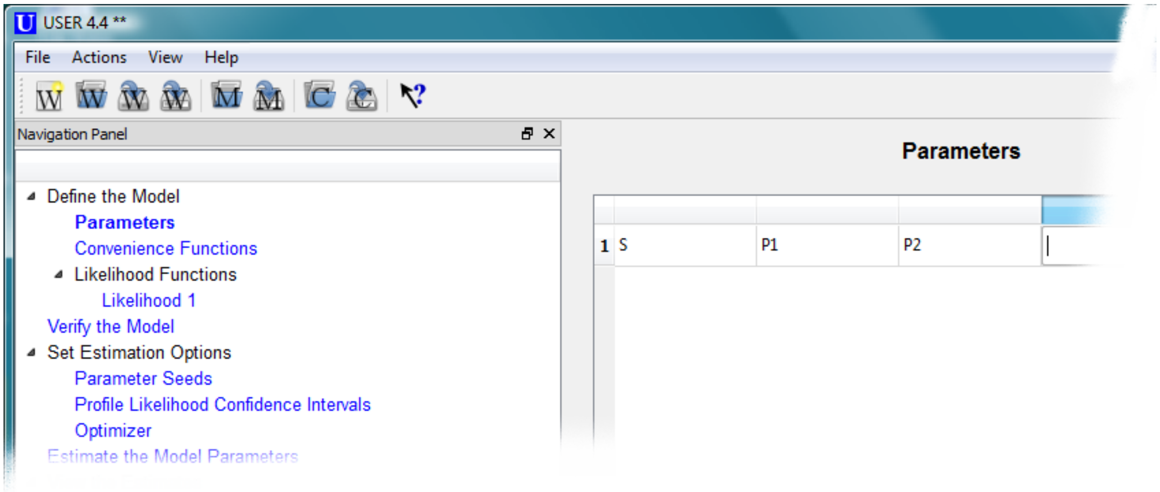
4.2.1 Define the Model

Parameters

The first step is to define the parameters S, P_1, and P_2 as shown in Figure 4.13.

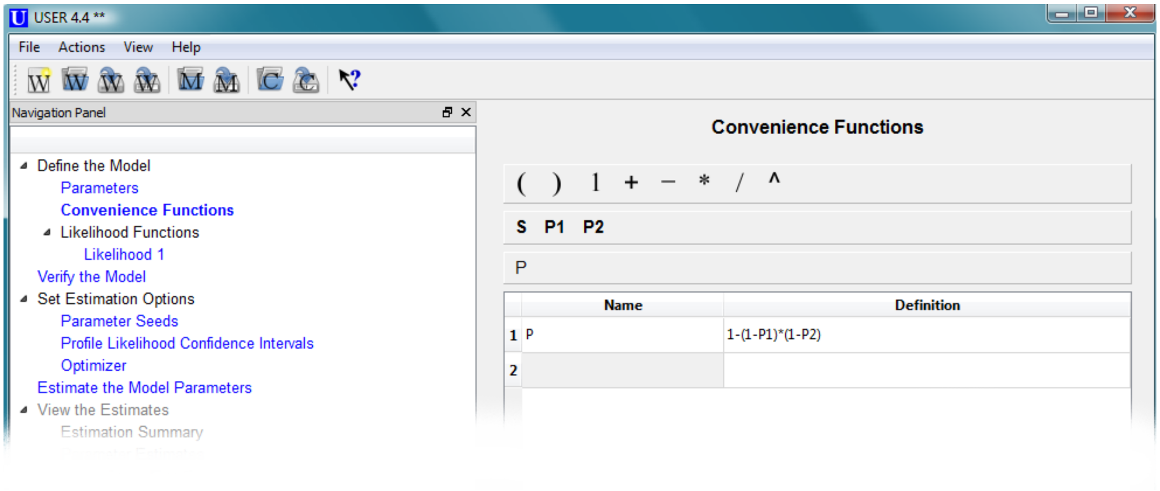
Convenience Functions

Move to the next content page in the "Define the Model" section, labeled "Convenience Functions." The two likelihoods make use of the overall detection



1	S	P1	P2	

Figure 4.13: Parameters for the joint likelihood example



	Name	Definition
1	P	1-(1-P1)*(1-P2)
2		

Figure 4.14: Convenience Function for the joint likelihood example

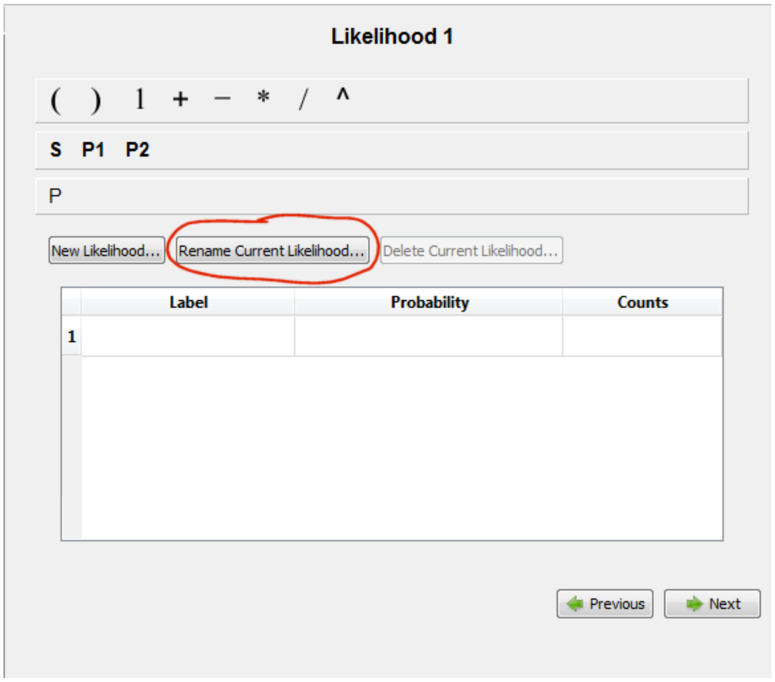


Figure 4.15: Button for renaming the current likelihood

probability $P = 1 - (1 - P_1)(1 - P_2)$, a function of the two model parameters P_1 and P_2. Define the convenience function as shown in Figure 4.14.

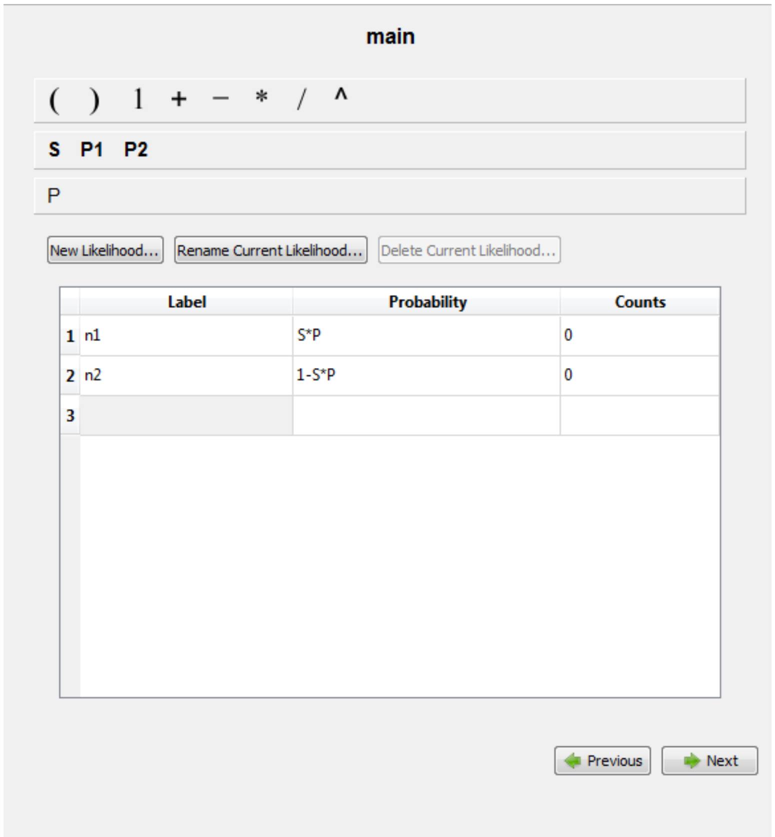
Now may be good time to make use of the "Save Workspace" function to save your work to a workspace file, say "Example2.txt".

Likelihood Functions

The next content page is labeled "Likelihood 1". In this example, there are two likelihoods: The main likelihood (L_1) and the auxiliary likelihood (L_2). Press the "Rename Current Likelihood" button (Figure 4.15) and rename the current likelihood "main". Notice that the title on the current content page and the navigation panel are updated to reflect the change. Enter the likelihood definition for the main likelihood as shown in Figure 4.16.

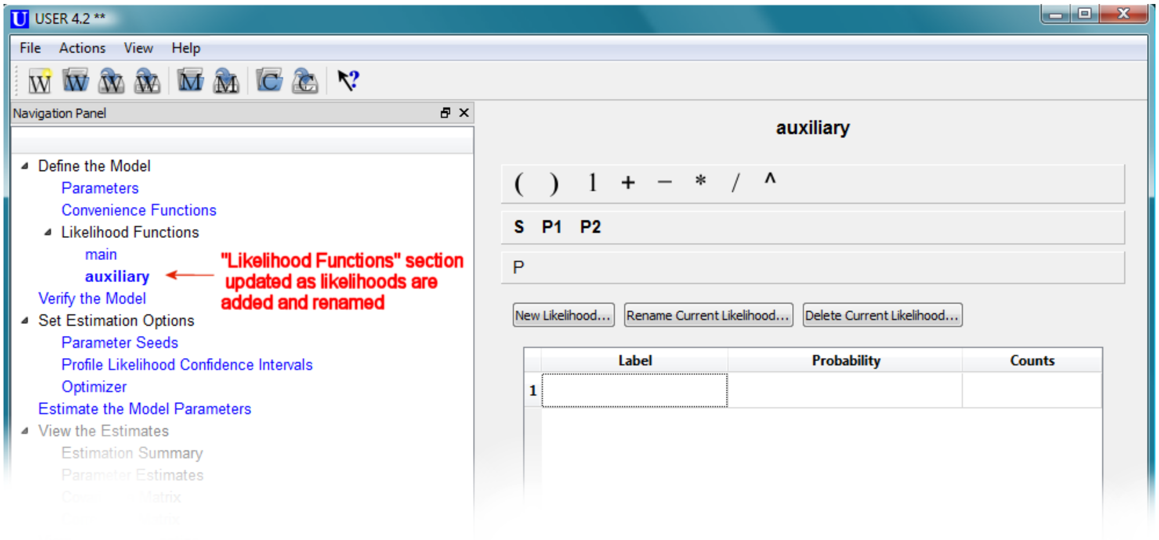
In order to define the auxiliary likelihood, press the "New Likelihood" button (to the left of the "Rename Current Likelihood" button), and give the new likelihood the name "auxiliary". This will add a new content page under "Likelihood Functions" labeled "auxiliary" (Figure 4.17).

Fill in the auxiliary likelihood definition as shown in Figure 4.18.



	Label	Probability	Counts
1	n1	S*P	0
2	n2	1-S*P	0
3			

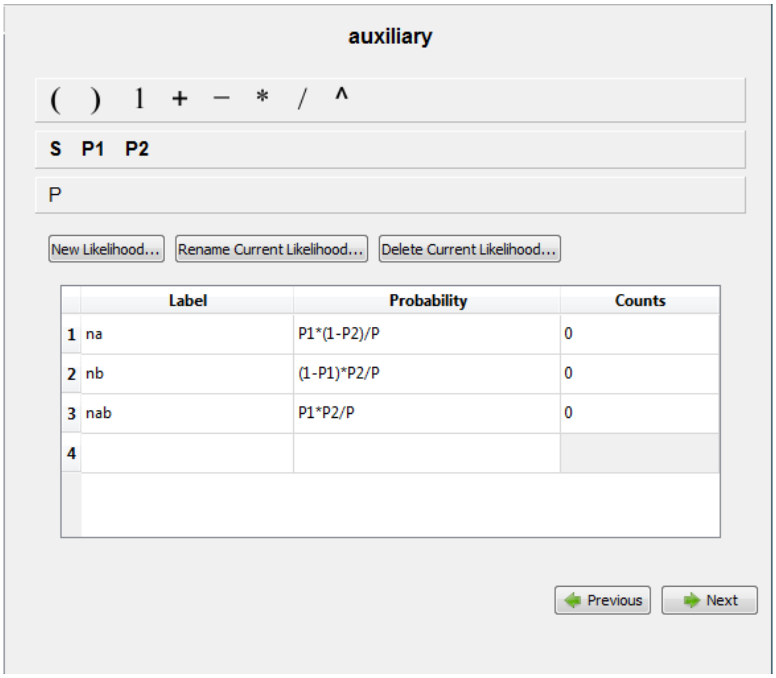
Figure 4.16: “main” likelihood in the joint likelihood example



"Likelihood Functions" section updated as likelihoods are added and renamed

	Label	Probability	Counts
1			

Figure 4.17: Begin the auxiliary likelihood definition for the joint likelihood example



	Label	Probability	Counts
1	na	P1*(1-P2)/P	0
2	nb	(1-P1)*P2/P	0
3	nab	P1*P2/P	0
4			

Figure 4.18: Completed auxiliary likelihood definition for the joint likelihood example

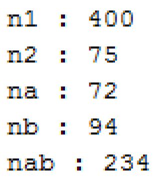
4.2.2 Verify the Model

Once the auxiliary likelihood is complete, press the “Next” button and, if everything was entered correctly, you should see a message indicating that the model definition is valid. If not, use the navigation panel to return to the pages where corrections need to be made, and then return to the “Verify Model” page.

4.2.3 Model Definition File and Category Counts File

You may have noticed that we have not entered any category counts for the likelihoods, but have used the default value of zero. USER allows you to keep the model definition separate from the actual data so that more than one dataset may be used with the same model, or, conversely, a single dataset may be fitted with multiple models.

From the “File” menu, select “Save Model Definition”, and enter a filename. Start a text editor, and enter the category counts as show in Figure 4.19. Each line includes a category label and a colon (“:”), separated by white space, followed by the corresponding counts. The category labels must match what was entered in the likelihood definitions; otherwise you will get an error when loading the counts into



```
n1 : 400
n2 : 75
na : 72
nb : 94
nab : 234
```

Figure 4.19: Category counts file for the joint likelihood example

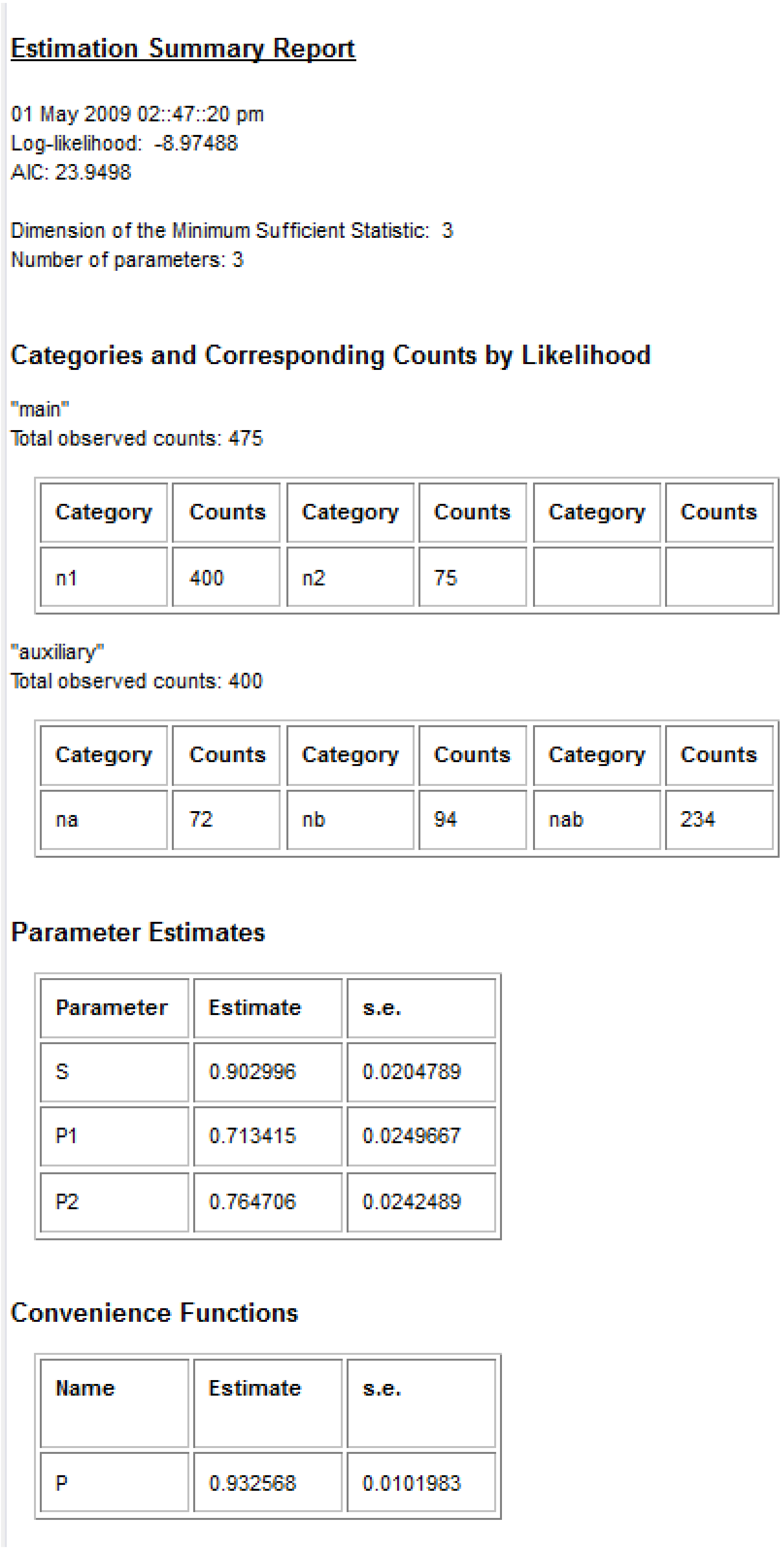
USER. If the colon and the counts are omitted, the counts are assumed to be one; if a category label is listed more than once, the counts are added together.

Save the category counts file and return to the USER program. If you exited USER previously, Use the “Load Model Definition” action under the “File” menu, and load the model definition file you created earlier. Now use the “Load Category Counts” action, also under the “File” menu to load the category counts text file you created.

After the counts have been loaded successfully, you may go back to the likelihood definition pages and see that the “Counts” column has been updated.

4.2.4 Estimate

Next, go to the “Estimate the Model Parameters” page and press the “Estimate” button. If all was done correctly and the estimation is successful, you may now proceed to the pages under the headings “View the Estimates” and “View Model Diagnostics” to examine the estimates and the suitability of the model for the data. Figure 4.20 shows the Estimation Summary report.



Estimation Summary Report

01 May 2009 02::47::20 pm
Log-likelihood: -8.97488
AIC: 23.9498

Dimension of the Minimum Sufficient Statistic: 3
Number of parameters: 3

Categories and Corresponding Counts by Likelihood

"main"
Total observed counts: 475

Category	Counts	Category	Counts	Category	Counts
n1	400	n2	75		

"auxiliary"
Total observed counts: 400

Category	Counts	Category	Counts	Category	Counts
na	72	nb	94	nab	234

Parameter Estimates

Parameter	Estimate	s.e.
S	0.902996	0.0204789
P1	0.713415	0.0249667
P2	0.764706	0.0242489

Convenience Functions

Name	Estimate	s.e.
P	0.932568	0.0101983

Figure 4.20: Estimation Summary report for the joint likelihood example

4.3 Estimating Abundance

In this example, we look at a situation where we are interested in estimating the total abundance of a population, and we must take into account individuals never observed.

This example uses a constant effort removal technique, where the population of interest is a pest species, and it is advantageous to decimate the population. A constant effort P is used at each sampling occasion. Let N represent the total number of individuals at the beginning of the study.

For the initial sampling event, the number of individuals removed will be NP, the number of individuals times the sampling effort. For the second event, the number of individuals still remaining is $N(1-P)$, so the expected number of animals removed will be $N(1-P)P$. For a study with four removal events, we can define the categories and probabilities of occurrence as follows.

Category	**Probability**
n1	P
n2	$(1-P)P$
n3	$(1-P)^2P$
n4	$(1-P)^3P$

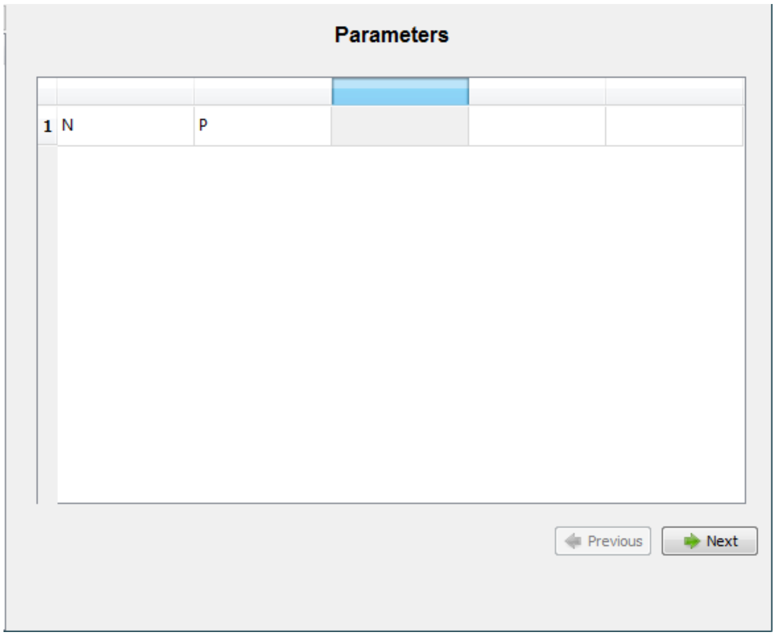
But this table is incomplete. Remember that the categories for a likelihood must be mutually exclusive and exhaustive, and we haven't taken into account the animals never recaptured. The complete table is as follows:

Category	**Probability**
n1	P
n2	$(1-P)P$
n3	$(1-P)^2P$
n4	$(1-P)^3P$
n0	$(1-P)^4$

4.3.1 Define the Model

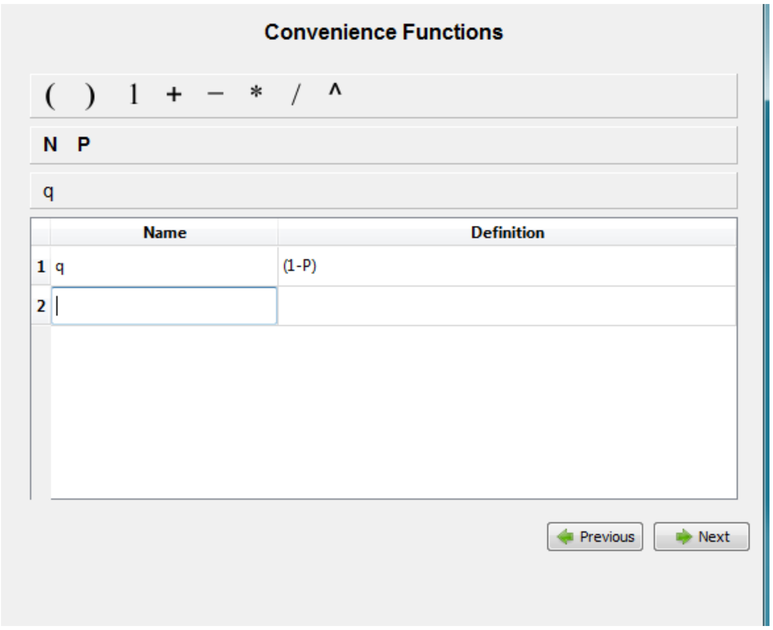
In USER, define the parameters as shown in Figure 4.21. In order to make the likelihood definition cleaner, we will define the convenience function $q = 1 - P$ as show in Figure 4.22.

For the likelihood definition, enter the first four categories as shown in Figure 4.23. The final category represents individuals never observed - an "unobserved category." We have no counts to enter for this category. In order to indicate that this is an unobserved category, in the Label column enter the



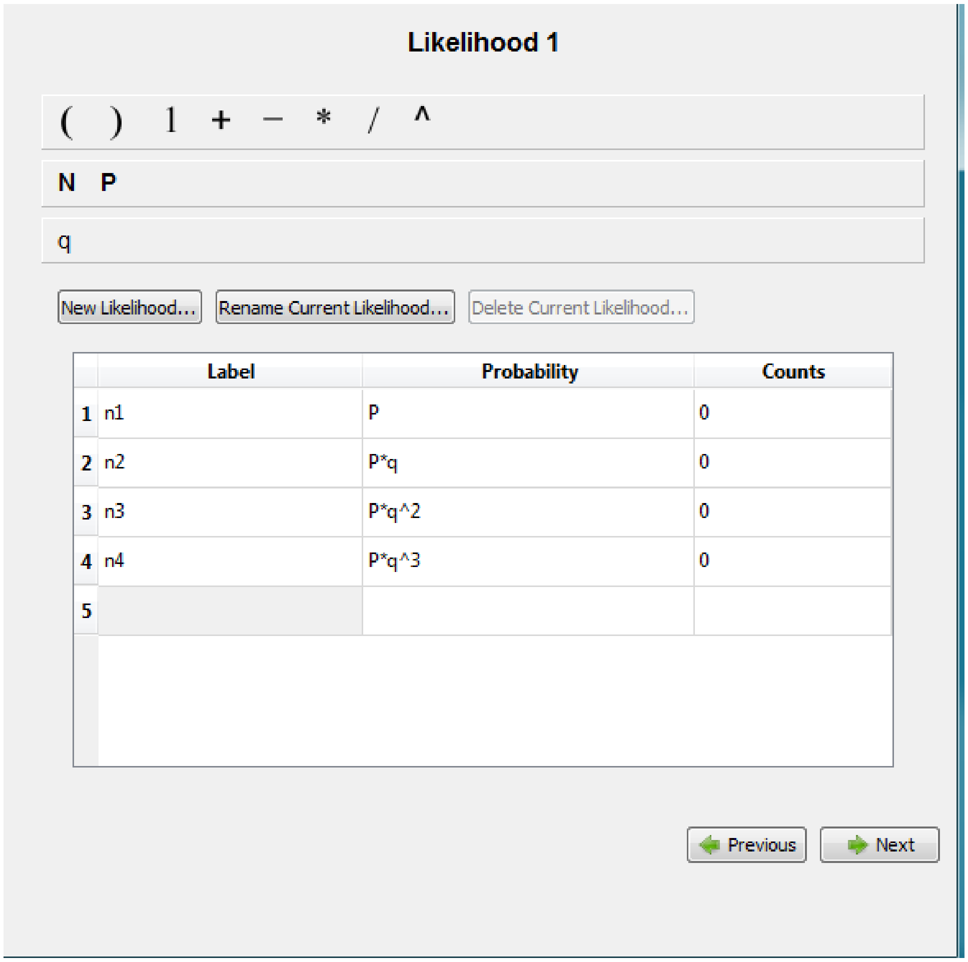
1	N	P			

Figure 4.21: Parameters for the abundance estimation example



	Name	Definition
1	q	(1-P)
2		

Figure 4.22: Convenience function for the abundance estimation example



	Label	Probability	Counts
1	n1	P	0
2	n2	P*q	0
3	n3	P*q^2	0
4	n4	P*q^3	0
5			

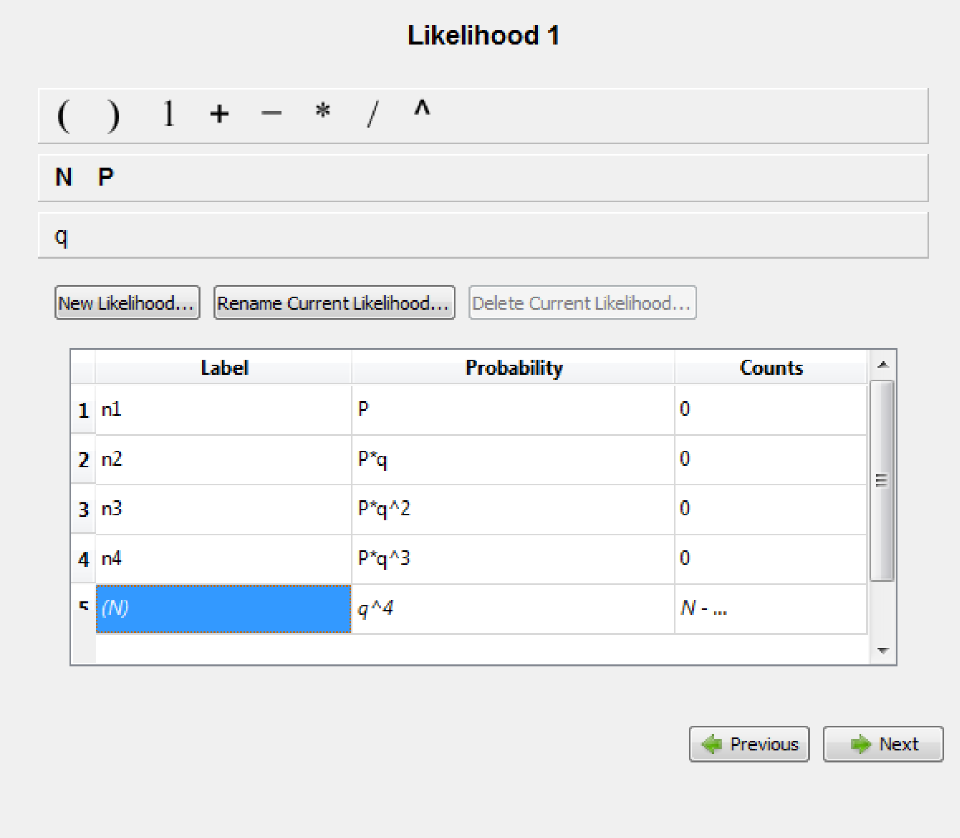
Figure 4.23: Partial likelihood definition for the abundance estimation example

abundance parameter in parentheses - "(N)", and press <Tab> to move to the Probability column. USER will put the text "N - ..." in the Counts column to indicate that the counts are assumed to be N minus the sum of all the other categories in this likelihood. USER also italicizes the text to distinguish it from the observed categories. Enter the probability as you would for any other category (Figure 4.24).

4.3.2 Estimate the Parameters

Load the Category Counts

Create category counts file as shown in Figure 4.25, and load them using the "Load Category Counts" action under the "File" menu.



	Label	Probability	Counts
1	n1	P	0
2	n2	P*q	0
3	n3	P*q^2	0
4	n4	P*q^3	0
5	(N)	q^4	N - ...

Figure 4.24: Likelihood definition for the abundance estimation example

```
n1 : 10000
n2 : 2500
n3 : 500
n4 : 75
```

Figure 4.25: Category counts file for the abundance estimation example

Entering the Parameter Seed for an Abundance Parameter

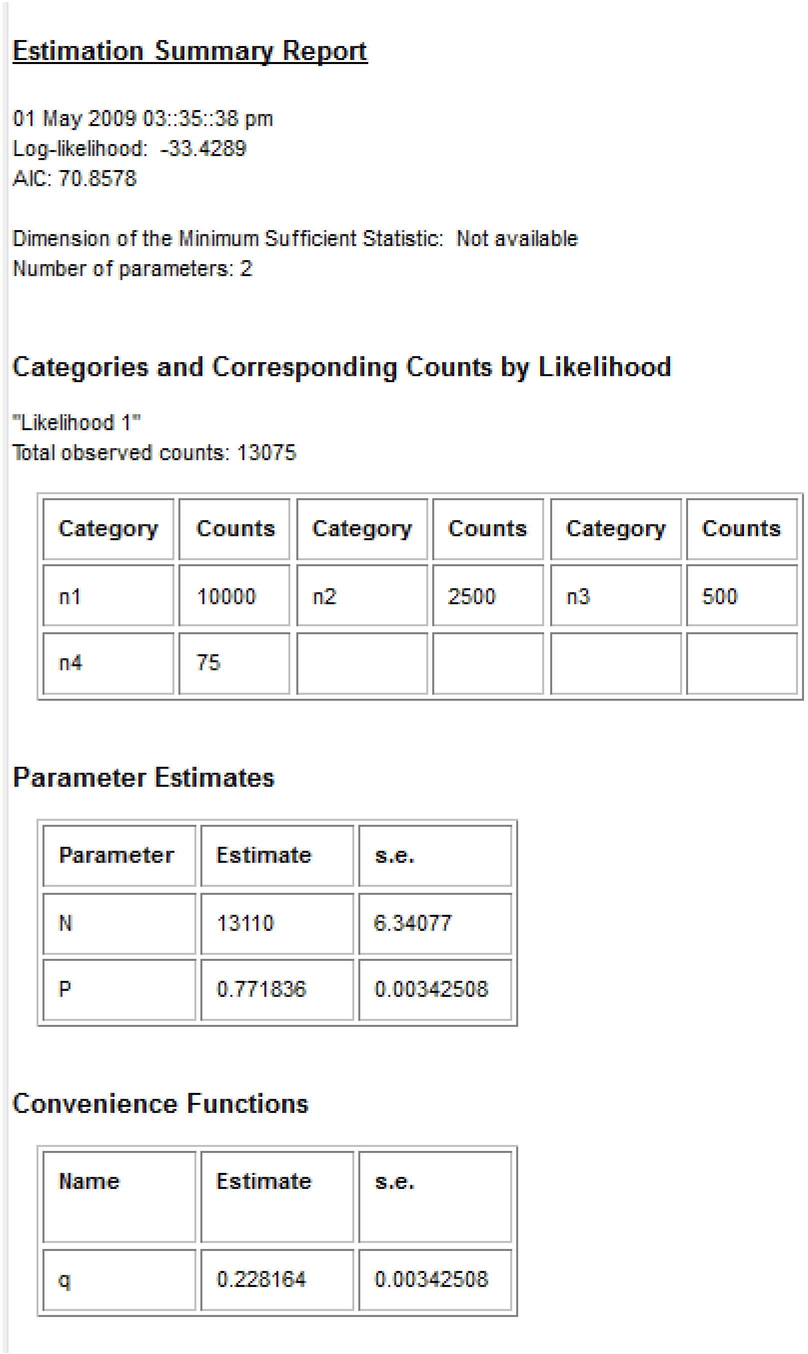
At the “Parameter Seeds” content page, it is clear that the seed for the parameter N needs to be changed to be at least as large as the counts for “n1” - the number captured at the initial event. For this type of model, it may take several attempts to find a seed which will allow the optimizer to converge to find the parameter estimates. It may be helpful to go to the “Optimizer” content page and use the “Downhill Simplex” method instead of the default optimizer. Some sources suggest that this method may be more robust but less precise. Once an estimate is found using the Downhill Simplex method, the resulting estimates could be used as seeds for the default “Fletch Quasi-Newton” method.

For this example, enter 20000 as the seed for N. This should lead to successful estimation of the model parameters. Figure 4.26 shows the Estimation Summary report.

Model Diagnostics

We can now look at the model diagnostics to see how well the model fits the data. Proceed to the “Observed vs Expected Plot” under “View Model Diagnostics.” Note that this plot, as with all plots in USER, can be made larger simply by enlarging the USER dialog window, or by using the zoom slider at the top (Figure 4.27).

If you go to the “Residuals Plot” content page (Figure 4.28), you will observe that two of the four observations fall well outside the ± 1.96 range. If the data fit the model well, the residuals will follow a standard normal distribution, and 95% of the observations should fall within the ± 1.96 range. Note that by holding the cursor over a data point, USER will display the category label for that point, the observed counts, and the expected counts. Perhaps, in this case, the assumption of constant effort removal was violated.



Estimation Summary Report

01 May 2009 03::35::38 pm
Log-likelihood: -33.4289
AIC: 70.8578

Dimension of the Minimum Sufficient Statistic: Not available
Number of parameters: 2

Categories and Corresponding Counts by Likelihood

"Likelihood 1"
Total observed counts: 13075

Category	Counts	Category	Counts	Category	Counts
n1	10000	n2	2500	n3	500
n4	75				

Parameter Estimates

Parameter	Estimate	s.e.
N	13110	6.34077
P	0.771836	0.00342508

Convenience Functions

Name	Estimate	s.e.
q	0.228164	0.00342508

Figure 4.26: Estimation Summary for the abundance estimation example

Figure 4.27: Observed vs Expected Plot for the abundance estimation example

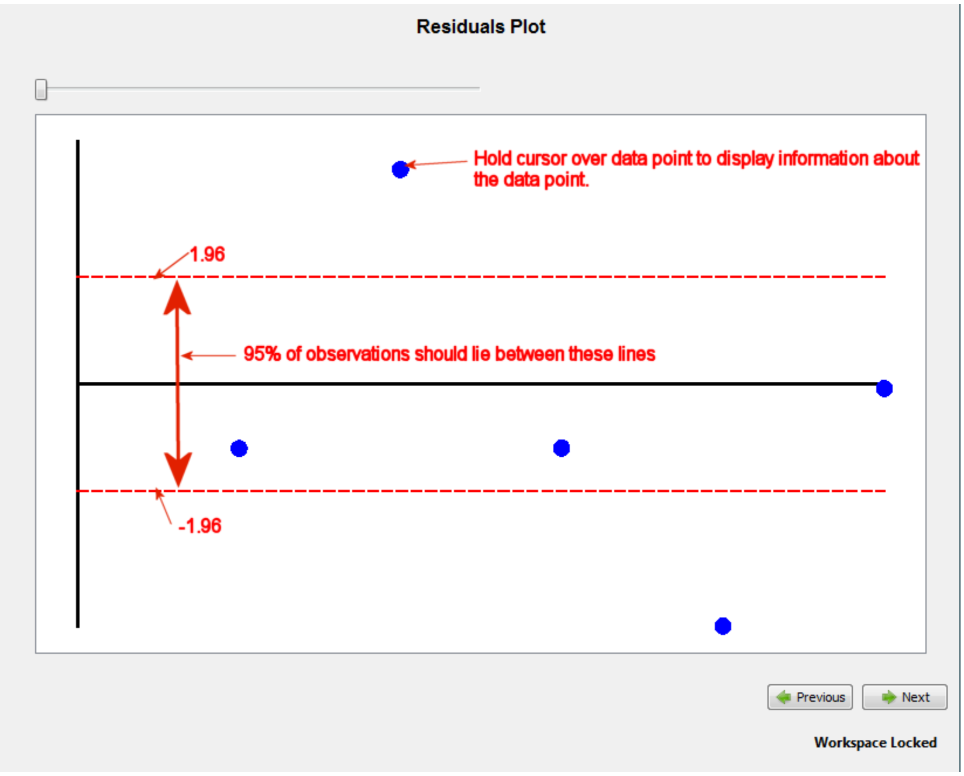


Figure 4.28: Residuals Plot for the abundance estimation example

4.4 Hypothesis Testing Example

This example demonstrates using USER to perform hypothesis testing and model selection.

Animals are captured and radio-tagged at an initial tagging event. The study area is then surveyed at two subsequent sampling events, evenly spaced in time, and the number of marked individuals still alive are recorded. This is a known-fate study, meaning that the detection probability is 1.0. If a marked individual is not detected, it is assumed to have died since the previous event.

The model parameters are as follows:

S_1	The probability of survival from the initial tagging occasion to the first survey event
S_2	The conditional probability of survival from the first survey event to the second event, given survival to the first event.

In addition to the survival estimates, we are also interested in knowing:

1. If the two survival probabilities are equal.
2. The overall survival probability for the study ($S_{total} = S_1 S_2$), along with a confidence interval for the overall survival.

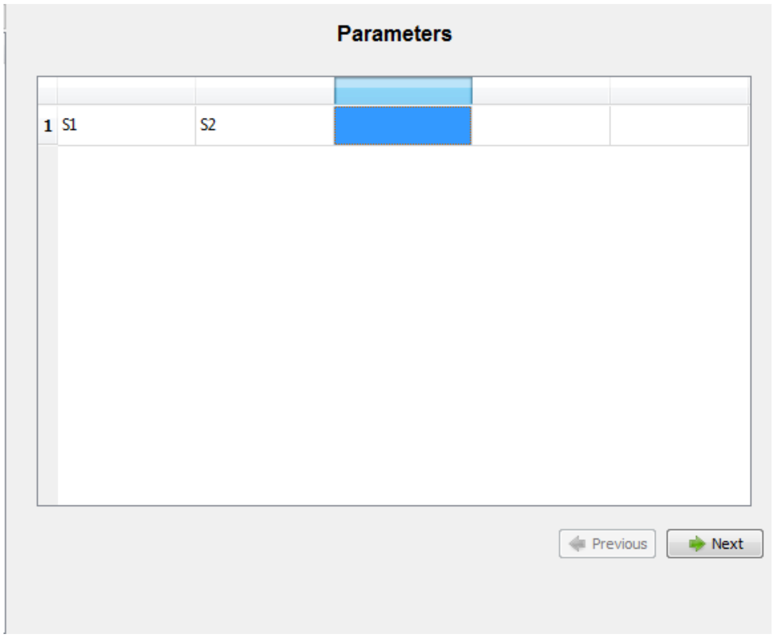
The likelihood categories are as follows:

Category	**Description**	**Probability**
1 0 0	The individual dies after initial tagging and before the first survey event	$1 - S_1$
1 1 0	The individual dies after the first survey event and before the second survey event	$S_1(1 - S_2)$
1 1 1	The individual is still alive at the end of the study	$S_1 S_2$

The data from the study are as follows.

Category	**Counts**
1 0 0	40
1 1 0	75
1 1 1	223

The full model has two separate period-specific survival probabilities S_1 and S_2. A reduced model would assume a common survival probability $S_1 = S_2$. We can test if the full model better describes the data than the reduced model.



1	S1	S2			

Figure 4.29: Parameter definitions for the hypothesis testing example

4.4.1 Full Model Definition

We define the parameters as shown in Figure 4.29. In addition, we are interested in the overall survival $S_{total} = S_1 S_2$. Therefore we need to define S_{total} as a convenience function as shown in Figure 4.30. The likelihood definition is shown in Figure 4.31.

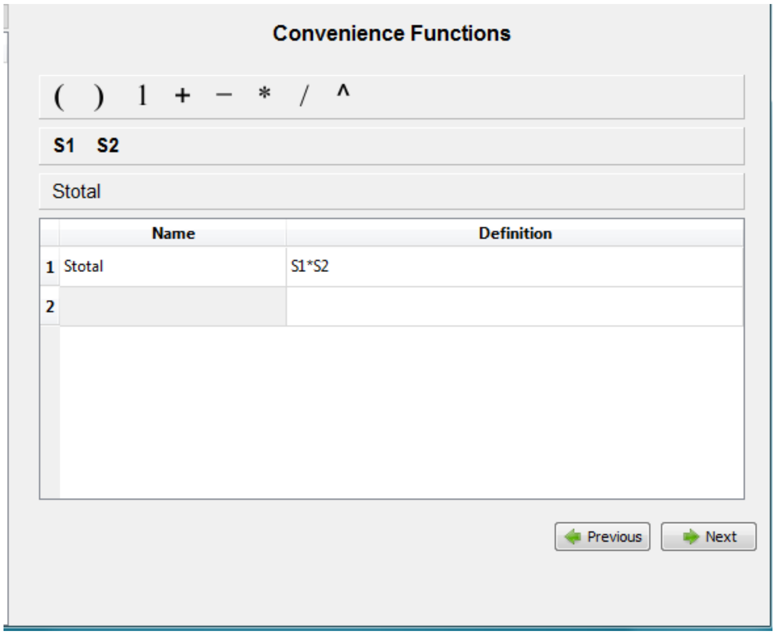
After defining the likelihood, go to the “Verify the Model” page. Once everything has been entered correctly, go to the “Parameter Seeds” page. Since we are estimating probabilities, the default seeds of 0.5 should suffice.

Profile Likelihood Confidence Intervals

We defined the convenience function $S_{total} = S_1 S_2$ because we were interested in a confidence interval for the overall survival. Figure 4.32 shows that we are requesting profile likelihood confidence intervals (alpha-level = 0.05) for the individual survival probabilities as well as the overall survival.

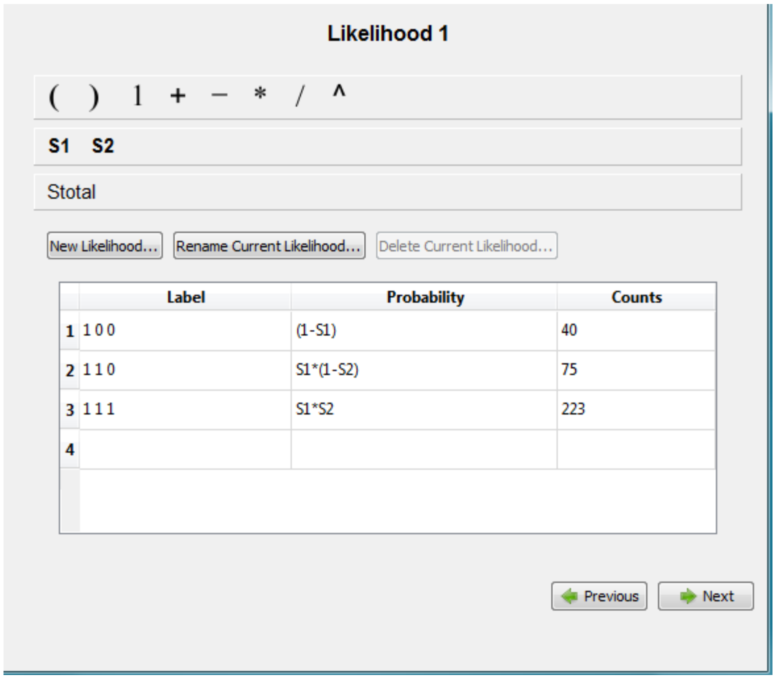
Estimate the Parameters

Advance to the “Estimate the Model Parameters” page, and press the “Estimate” button. Figure 4.33 shows the estimation summary report for the full



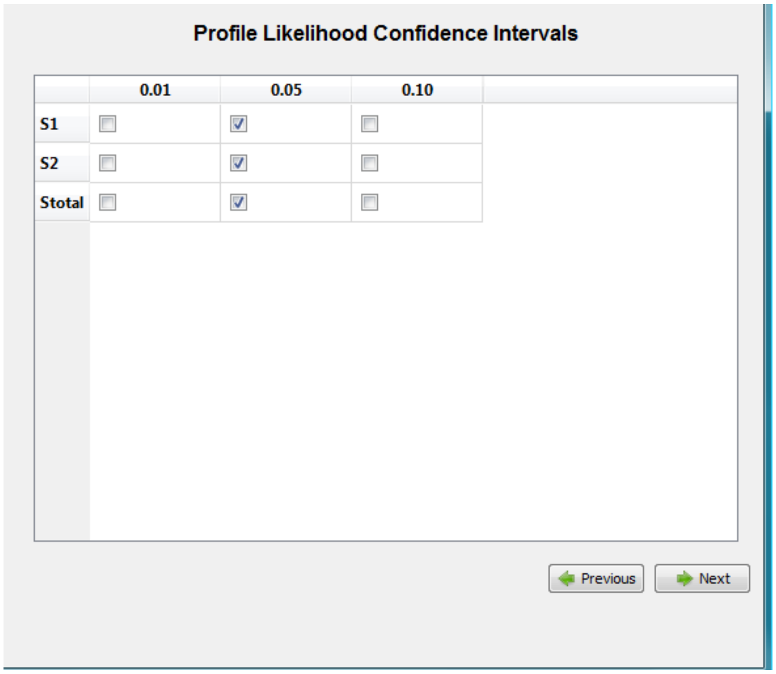
	Name	Definition
1	Stotal	S1*S2
2		

Figure 4.30: Convenience function definition for the hypothesis testing example



	Label	Probability	Counts
1	1 0 0	(1-S1)	40
2	1 1 0	S1*(1-S2)	75
3	1 1 1	S1*S2	223
4			

Figure 4.31: Likelihood definition for the hypothesis testing example



Profile Likelihood Confidence Intervals

	0.01	0.05	0.10
S1	☐	☑	☐
S2	☐	☑	☐
Stotal	☐	☑	☐

Figure 4.32: Profile likelihood confidence interval requests for the hypothesis testing example

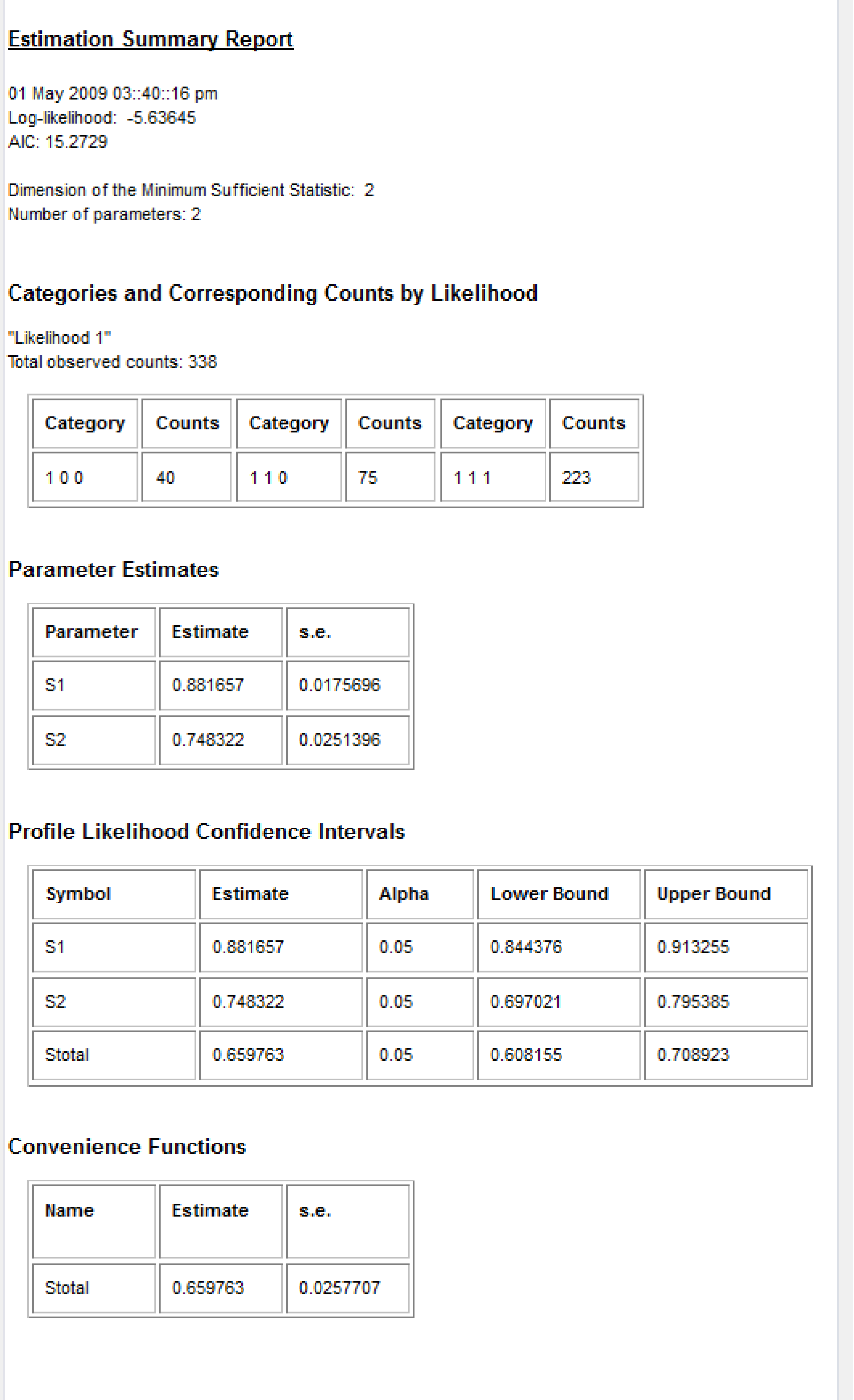
model. In order to perform the the the hypothesis test, we need to record the log-likelihood (-5.63645).

4.4.2 Reduced Model Definition

We are interested in comparing the full and reduced models. In the context of hypothesis testing, the null hypothesis is: $H_0 : S_1 = S_2$ (reduced model), vs $H_A : S_1 \neq S_2$ (full model). To define the reduced model, we can use the full model definition from the previous section as starting point, as follows:

1. Unlock the workspace so that the model may be changed. Go to the "Actions" menu and select "Clear Current Estimates." Press "OK" when asked for confirmation. This will clear all estimates from the null model.

2. Double-click on "Parameters" on the navigation panel to return to the "Parameters" page. Right-click on the "S2" cell and select "Delete" as shown in Figure 4.34. We now have only one parameter: S_1.



Estimation Summary Report

01 May 2009 03::40::16 pm
Log-likelihood: -5.63645
AIC: 15.2729

Dimension of the Minimum Sufficient Statistic: 2
Number of parameters: 2

Categories and Corresponding Counts by Likelihood

"Likelihood 1"
Total observed counts: 338

Category	Counts	Category	Counts	Category	Counts
1 0 0	40	1 1 0	75	1 1 1	223

Parameter Estimates

Parameter	Estimate	s.e.
S1	0.881657	0.0175696
S2	0.748322	0.0251396

Profile Likelihood Confidence Intervals

Symbol	Estimate	Alpha	Lower Bound	Upper Bound
S1	0.881657	0.05	0.844376	0.913255
S2	0.748322	0.05	0.697021	0.795385
Stotal	0.659763	0.05	0.608155	0.708923

Convenience Functions

Name	Estimate	s.e.
Stotal	0.659763	0.0257707

Figure 4.33: Estimation Summary Report for the hypothesis testing example, full model

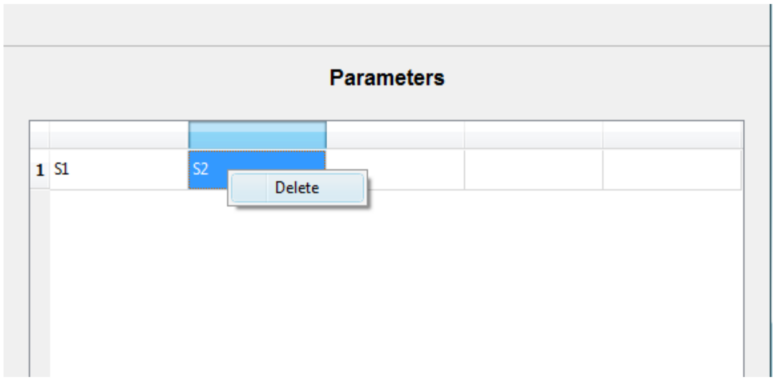


Figure 4.34: Delete the parameter S2 for the hypothesis testing example

3. Go to the “Convenience Functions” page, and add the convenience function “S2” as shown in Figure 4.35. Note that the actual likelihood definition does not need to be changed. The difference is that S_2 is no longer a model parameter to be estimated, but is simply another name for the one model parameter S_1.
4. Go to the “Estimate the Model Parameters” page and press “Estimate”. Figure 4.36 shows the resulting estimation summary report.

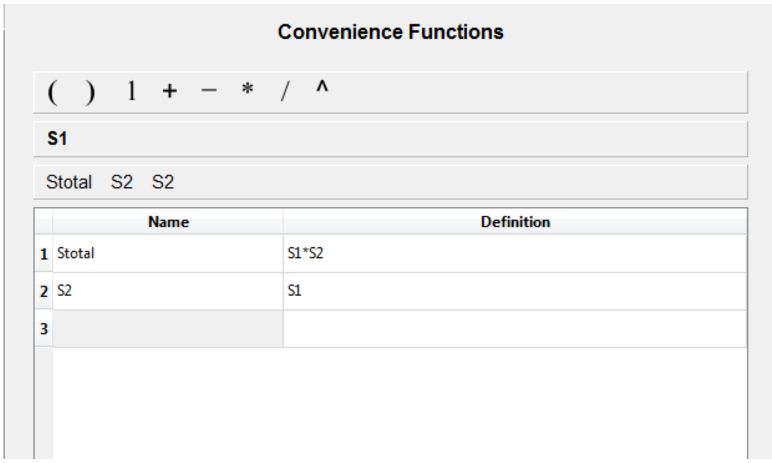
As with the full model, we must note the log-likelihood (-15.2063).

4.4.3 Test the Hypothesis

We can now test the hypothesis using a Likelihood Ratio Test:

- Null hypothesis: $S_1 = S_2$
- Alternative hypothesis: $S_1 \neq S_2$

The likelihood ratio test statistic is $\chi^2 = 2(-5.63645 - (-15.2063)) = 19.1397$. χ^2 has an asymptotic chi-square distribution with one degree of freedom, giving a p-value of 0.00001, rejecting the null hypothesis in favor of the alternative hypothesis $S_1 \neq S_2$.

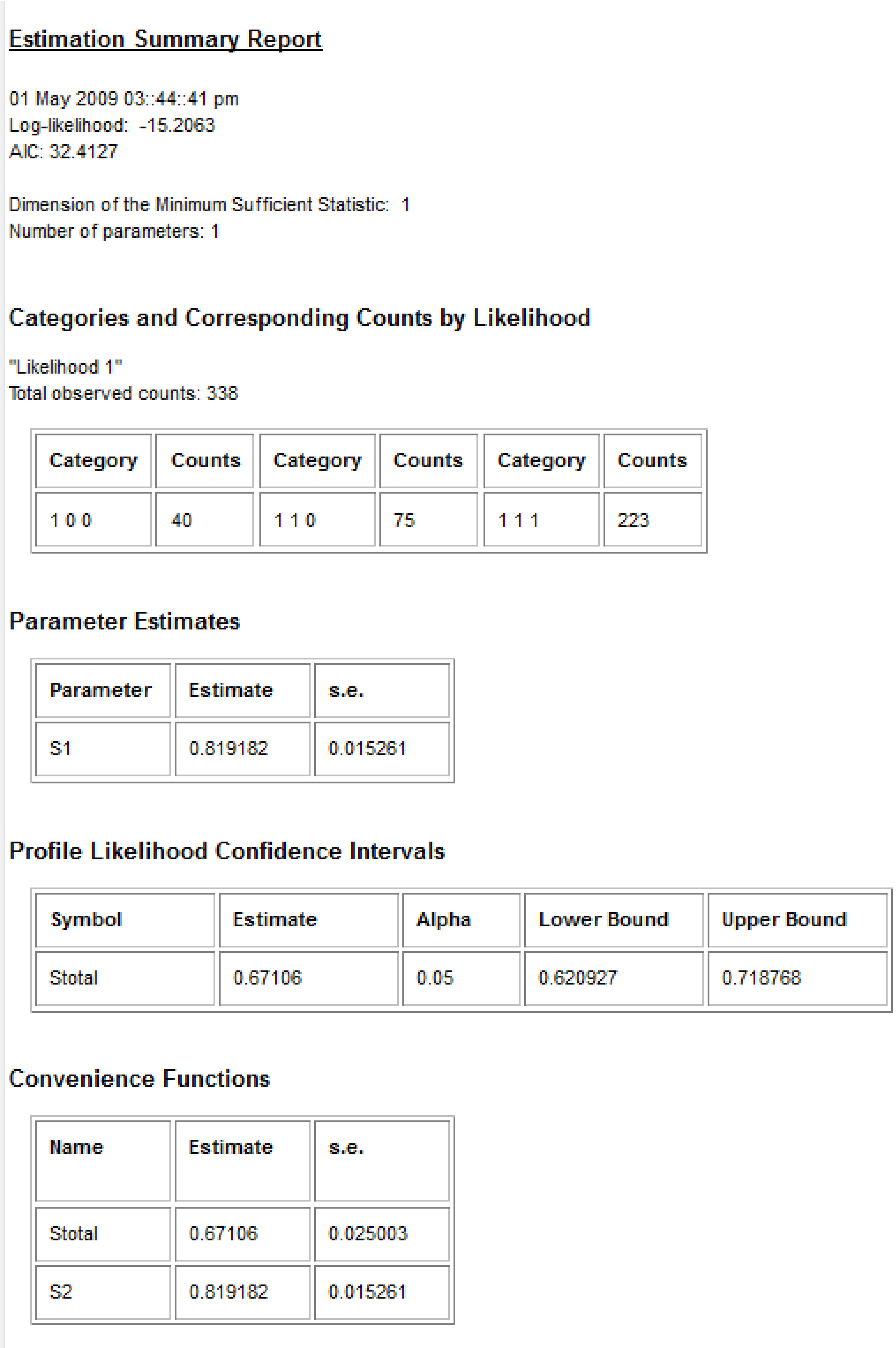


	Name	Definition
1	Stotal	S1*S2
2	S2	S1
3		

Figure 4.35: Define the convenience function S2 for the hypothesis testing example

4.4.4 Model Selection Using an Information-Theoretic Approach

An alternative approach to hypothesis testing is model selection based on an information-theoretic approach(*Model Selection and Multinomial Inference. A Practical Information-Theoretic Approach.* 2nd Ed, Burnham, Kenneth P. and David R. Anderson. 2002). With this approach, model selection is based on Akaike's Information Criterion (AIC). The model with the smaller AIC is the preferred model. In this example, the full model has an AIC of 15.2729, and the reduced model has an AIC of 32.4127. Based on the AIC, the full model is the preferred model.



Estimation Summary Report

01 May 2009 03::44::41 pm
Log-likelihood: -15.2063
AIC: 32.4127

Dimension of the Minimum Sufficient Statistic: 1
Number of parameters: 1

Categories and Corresponding Counts by Likelihood

"Likelihood 1"
Total observed counts: 338

Category	Counts	Category	Counts	Category	Counts
1 0 0	40	1 1 0	75	1 1 1	223

Parameter Estimates

Parameter	Estimate	s.e.
S1	0.819182	0.015261

Profile Likelihood Confidence Intervals

Symbol	Estimate	Alpha	Lower Bound	Upper Bound
Stotal	0.67106	0.05	0.620927	0.718768

Convenience Functions

Name	Estimate	s.e.
Stotal	0.67106	0.025003
S2	0.819182	0.015261

Figure 4.36: Estimation Summary Report for the hypothesis testing example, reduced model