USER 4 MANUAL

USER-SPECIFIED ESTIMATION ROUTINE

COLUMBIA BASIN RESEARCH

SCHOOL OF AQUATIC AND FISHERY SCIENCES

W 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

Contents

List of Figures

1.0 Overview

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

1. All possible outcomes of the study can be characterized in terms of a finite number of discrete categories.

2. The categories are mutually exclusive and exhaustive; i.e., an individual in the study can be classified into one and only one category.

3. 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, and

- a **count** indicating the number of observations for the category.

Because 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. 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:

- Chapter 2: a brief description of the USER interface and integrated help system;

- Chapter 3: the analysis of four hypothetical studies, including:
  - Section 3.1: a study with a single likelihood,
  - Section 3.2: a study requiring an auxiliary likelihood,
  - Section 3.3: a study for estimating abundance, requiring the use of an unobserved category, and
  - Section 3.4: an example of performing hypothesis testing.

This document describes version 4.7 of Program USER, which can be found on the [Program USER page on the Columbia Basin Research website](#).

2.0 Interface Description

The user interface for Program USER upon startup is illustrated in Figure 2.1. The left side consists of the navigation panel and the right side consists of the content page. The active content page at startup is the **Parameters** page.
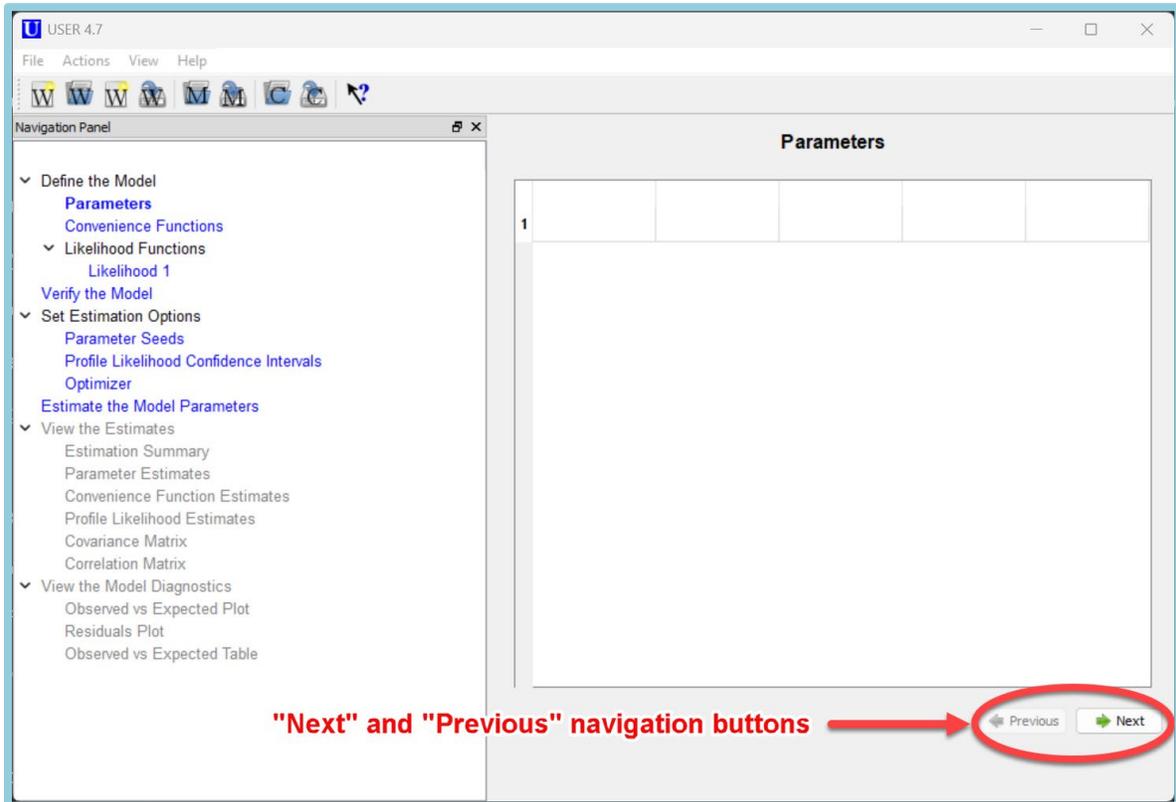
Figure 2.1: The USER dialog as it appears at startup.

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 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 that are used strictly for organizational purposes are distinguished from content pages by the blue text denoting the 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 via the **View** menu.

Figure 2.2: The USER 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.

2.3   Integrated Help System

Help for users is designed to be provided in two ways within the program: navigating to **Help** on the **File** menu and selecting **Contents**, or clicking on the **Context-sensitive help icon** mentioned in Figure 2.2. Please note that the Help system within USER is undergoing revisions and is not currently functional. These features will be available in future versions of this program. If the user requires assistance, further help can be found on the Columbia Basin Research USER page or by contacting emailing Columbia Basin Research directly.

3.0 Examples

3.1 Single 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:

$$
\begin{aligned}
n_1 &= 220 \\
n_2 &= 235 \\
m &= 40
\end{aligned}
$$

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

$$
L = \binom{n_2}{m} \left(\frac{n_1}{N}\right)^m \left(1 - \frac{n_1}{N}\right)^{n_2 - m} \tag{3.1}
$$

3.1.1 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, type **N**, and press **Tab** to move to next cell (Figure 3.1).
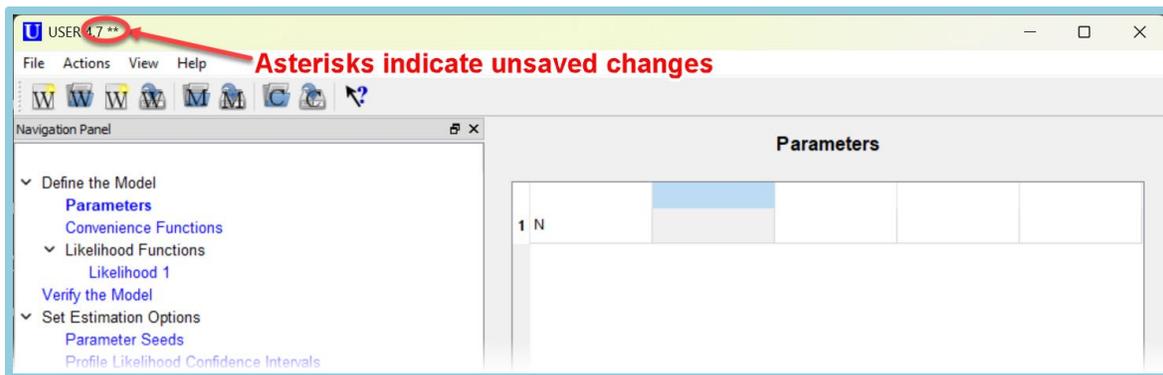
Figure 3.1: The **Parameters** content page for the single likelihood example.

Notice in Figure 3.1 that after entering the parameter N, two asterisks (**) appear in the window title to the right of 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 3.2).
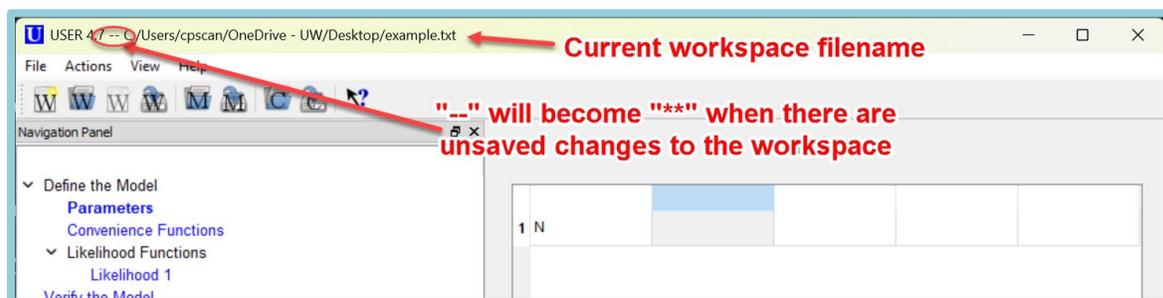
Figure 3.2: Using a workspace file.

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.

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 the likelihood easier to define, more readable, and more flexible.

The likelihood for this example uses the value n_1, 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 3.3. 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 3.3 when completed.

Convenience Functions

( ) 1 + − * / ^

N

n1 P

Name	Definition
1 n1	220
2 P	n1/N
3	
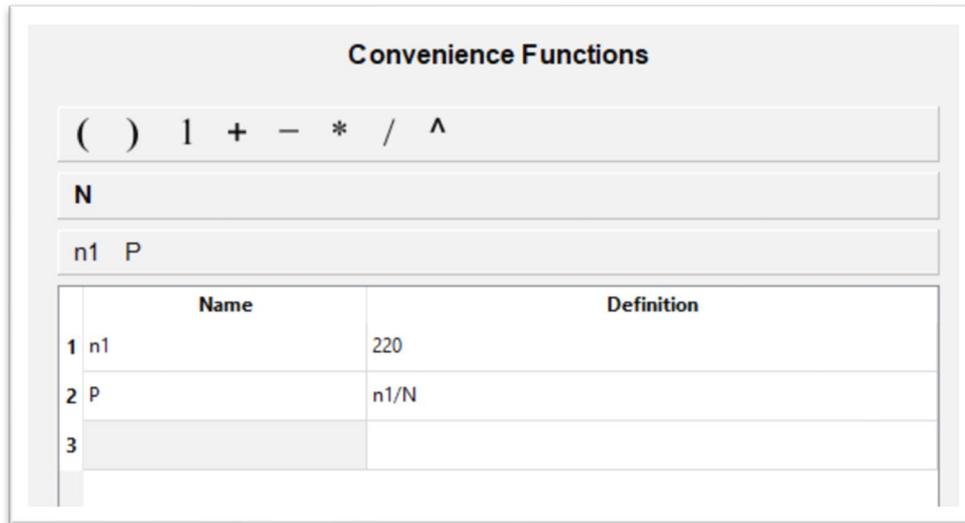
Figure 3.3: The **Convenience Functions** content page for the single likelihood example.

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 3.1 and the mark-recapture data on page 7, we can define the categories as follows:

Label	Probability	Count
m	P	40
$n_2 - m$	$1 - P$	195

Consequently, the **Likelihood Definition** page is completed as in as shown in Figure 3.3. 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. One (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.
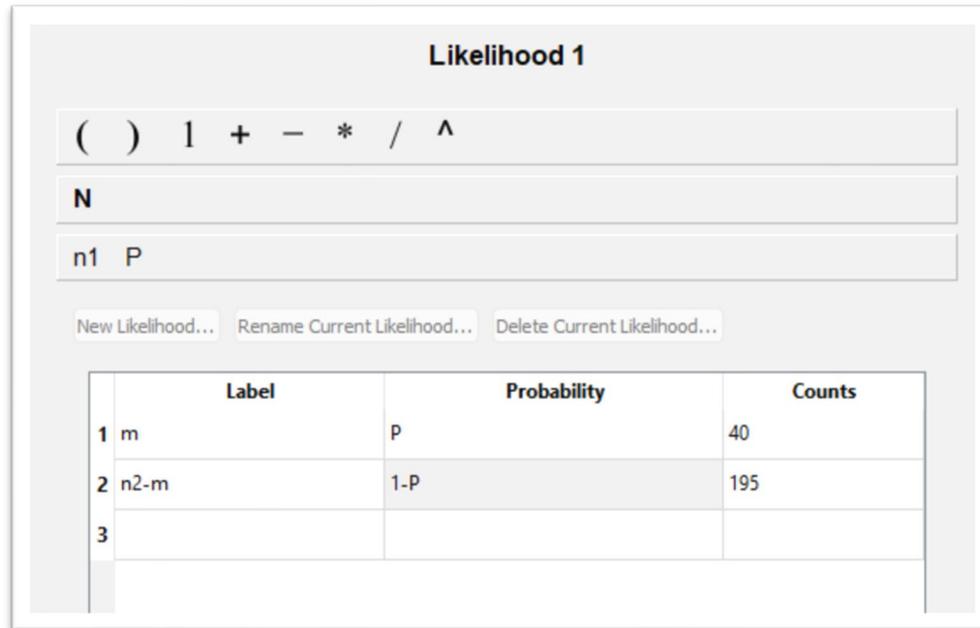
Figure 3.4: The **Likelihood** definition content page for the single likelihood example.

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 displays 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 3.5. 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.
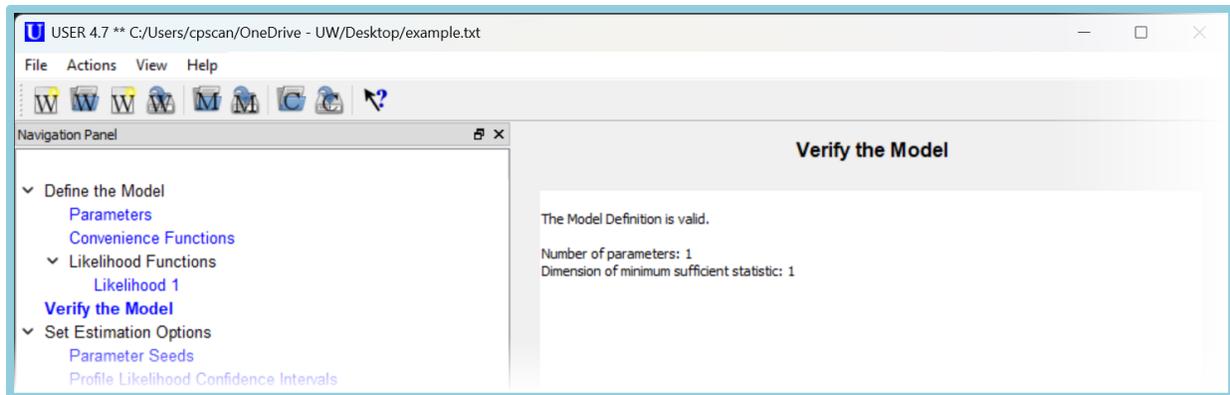
Figure 3.5: The **Verify the Model** content page for the single likelihood example.

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

3.1.2 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 probabilities that are between 0.0 and 1.0. For this example, however, the parameter to be estimated is animal abundance, so we need to change the seed to a more reasonable number, say, 500 (Figure 3.6).
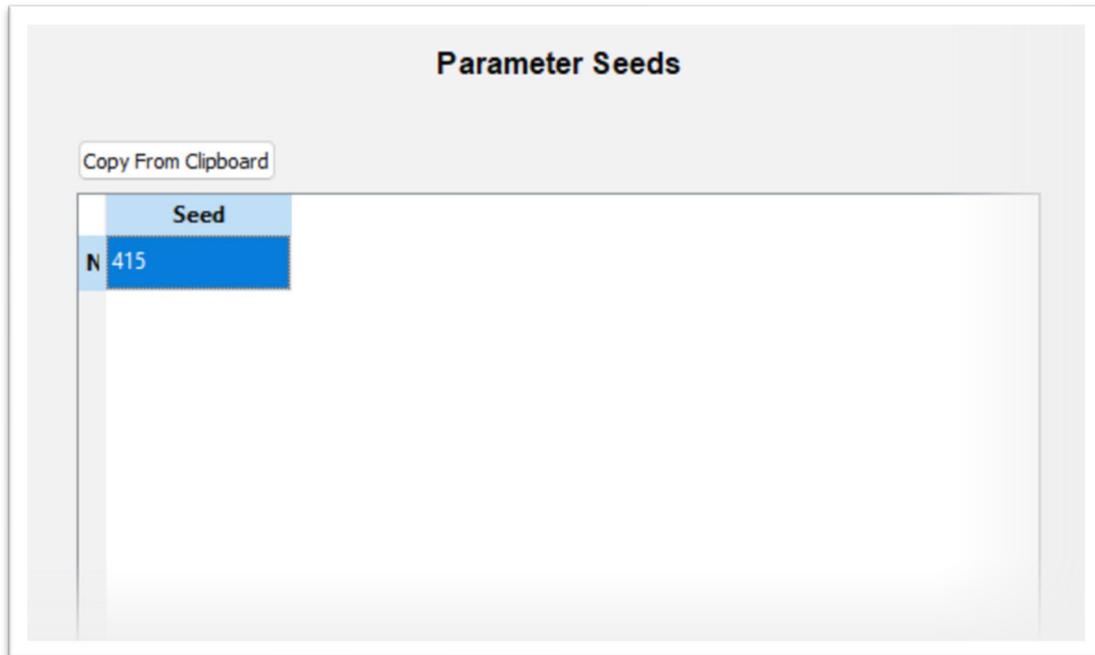
Figure 3.6: The **Parameter Seeds** content page for the single likelihood example.

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 by again clicking **Next**.

Optimizer

The next content page is titled **Optimizer** and allows the user to specify the numerical optimization settings (Figure 3.7). The default optimizer is **Powell's Direction Set**, also known as Powell's method. This method finds the local minimum of a numerical function by minimizing in a single search direction at a time and then updating the search set based on the direction of the most progress. It has the benefit of not requiring the function to be differentiable and thus is useable for a wide range of functions. Powell's method is more efficient than **Downhill Simplex** in terms of the required number of iterations but may be less efficient than quasi Newton-Raphson methods if the likelihood function is relatively smooth.

The next numerical optimizer offered is **Fletch Quasi-Newton** (Figure 3.7), which uses the "quasi Newton-Raphson method" to numerically solve for the

parameter estimates. Quasi Newton-Raphson methods use an approximation of the function's Hessian matrix to efficiently converge to the function's local maximum. While efficient and often robust, this method performs best for functions that are relatively smooth.

The last numerical optimizer offered is the **Nelder-Mead Downhill Simplex** (or simply **Downhill Simplex**) method. Like **Powell's Direction Set**, this method is a direct search method that does not require derivatives. Also known as the amoeba method, this method uses a simplex (e.g., a triangle generalized to one, two, or more dimensions) to search the parameter space, adaptively reflecting, expanding, shrinking, and contracting the shape based on evaluations of the objective function. While this method does not require derivatives and is more robust than other methods, it performs best on smooth functions, can be imprecise, and can also be inefficient in terms of the number of function evaluations.

Both **Fletch Quasi-Newton** and **Downhill Simplex** require configuration settings such as the maximum number of iterations, precision, and proportional step size (Figure 3.7). The default configuration settings are expected to perform well in most cases. 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 optimizer. In other cases, it may be advisable to use multiple optimizers to find the best convergence (i.e., to minimize the function value). For example, use the **Downhill Simplex** method to find estimates in the neighborhood of the maximum likelihood estimates, and then use those resulting estimates as seeds for the default **Powell's Direction Set** method or **Fletch Quasi-Newton method**.

Figure 3.7: The **Optimizer** content page settings for the single likelihood example.

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 seeds and 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 3.8).
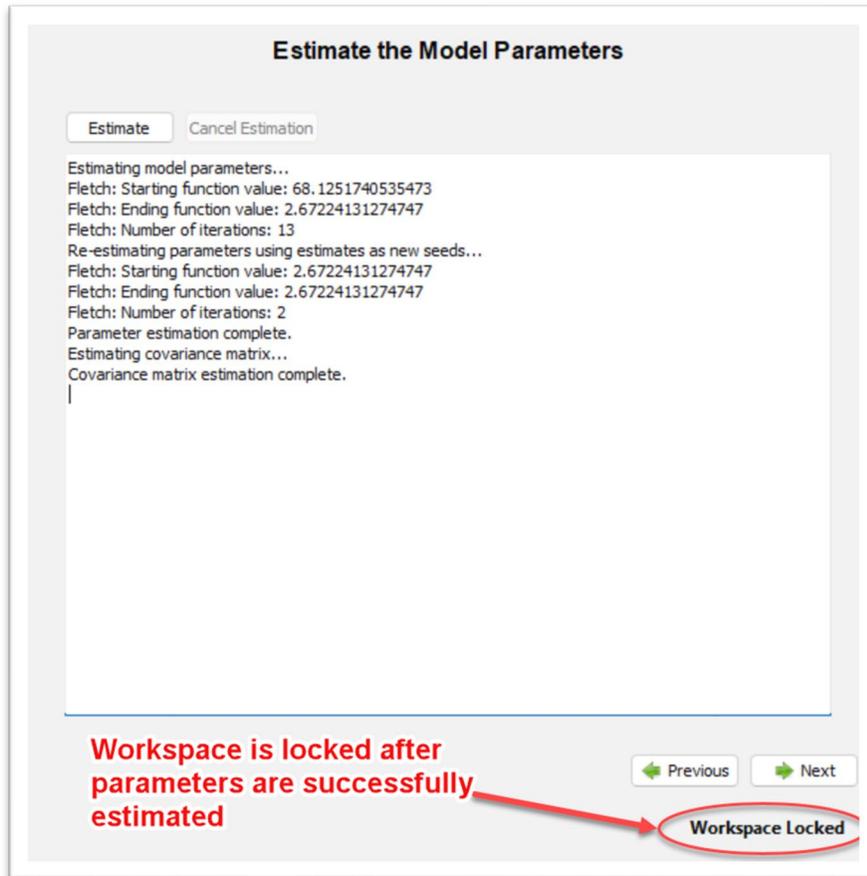
Figure 3.8: The **Estimate the Model Parameters** page after the successful estimation of the model parameters for the single likelihood example.

If the model parameters are successfully estimated, the **Workspace Locked** indicator will appear in the lower right (circled in red in Figure 3.8). 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 3.1.1). If you wish to make changes, you must use the **Clear Current Estimates** command under the **Actions** 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 3.9. In this case, the covariance matrix could not be estimated because the Hessian

matrix was singular (non-invertible). This happened because the starting function value was invalid: note that the starting function value was reported as 1.79e+308 – this is the machine integer limit (i.e., the highest non-infinite value available) and it indicates a bad seed. If the seed is bad, the optimizer may fail to find parameter estimates. If this happens, clear the current estimates if necessary, 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.

Estimate the Model Parameters

| Estimate | Cancel Estimation |

Estimating model parameters...
Fletch: Starting function value: 1.79769313486232e+308
Fletch: Ending function value: 1.79769313486232e+308
Fletch: Number of iterations: 2
Parameter estimation complete.
Estimating covariance matrix...
Singular Hessian in calculating covariance matrix
Covariance matrix estimation failed.

Figure 3.9: The **Estimate the Model Parameters** page after estimation failure for the single likelihood example.

3.1.3 Estimates

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

- the categories with their corresponding counts,
- the parameter estimates along with their standard errors,

- the convenience function estimates and their standard errors calculated using the delta-method, and

- the requested profile likelihood confidence intervals (if any).

The **Estimation Summary** page for this example is shown in Figure 3.10.

The various components of the **Estimation Summary** are also reported on separate pages following the **Estimation Summary** page: **Parameter Estimates**, **Convenience Function Estimates**, and **Profile Likelihood Estimates** (if requested). Additionally, the estimated **Covariance Matrix** and **Correlation Matrix** associated with the parameter estimates are also provided on their own pages. Each of the pages under **View the Estimates** provides the ability to save the report to a file and either print (for **Estimation Summary**) or copy to the clipboard (all other pages under **View the Estimates**).

Estimation Summary Report

29 January 2026 09::45::22 am
AIC: 7.34448
Log-likelihood: -2.67224

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	186.158

Convenience Functions

Name	Estimate	s.e.
n1	220	0
P	0.170213	0.0245156

Figure 3.10: The **Estimation Summary Report** output for the single likelihood example.

3.2   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 3.11. 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.
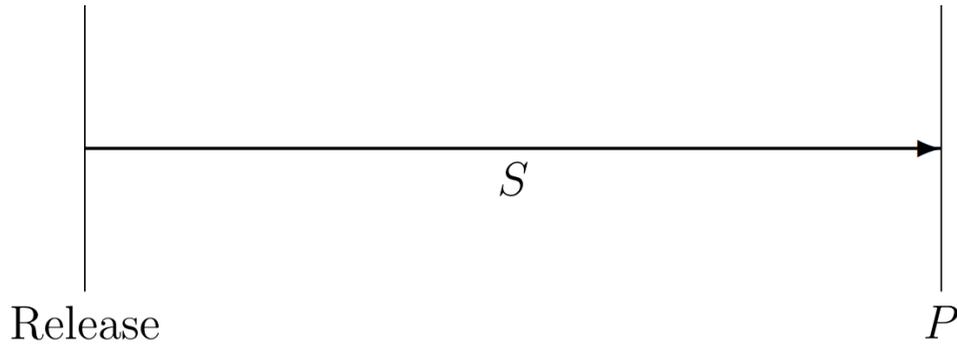
Figure 3.11: Joint likelihood example: Diagram of the study design.

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:

$$L_1 \propto (SP)^{n_1}(1 - SP)^{n_2} \tag{3.2}$$

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 shown in Figure 3.12.
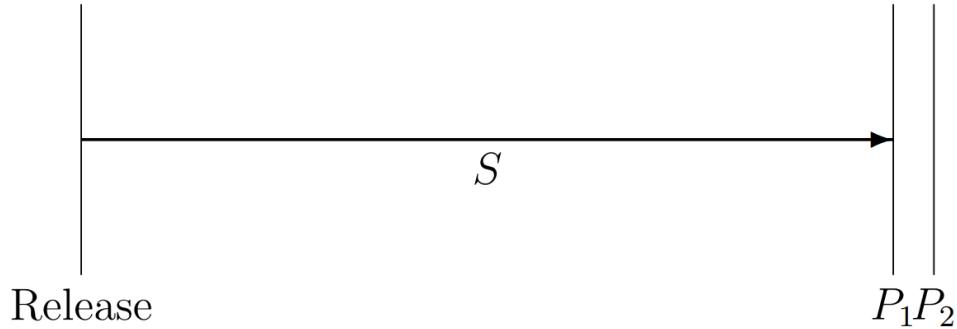
Figure 3.12: Joint likelihood example: diagram of the modified study design with two independent detection sites.

This modified design makes it 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 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	Detection at site #1 only	$P_1(1 - P_2)/P$
n_b	Detection at site #2 only	$(1 - P_1)P_2/P$
n_{ab}	Detection at both site #1 and site #2	$P_1 P_2/P$

The auxiliary likelihood is defined as:

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

The joint likelihood is defined as

23

$$L = L_1 \times L_2.$$

We can now define the joint model in USER 3.2.

3.2.1 Define the model

Parameters

The first step is to define the parameters S, P_1, and P_2 as shown in Figure 3.13.
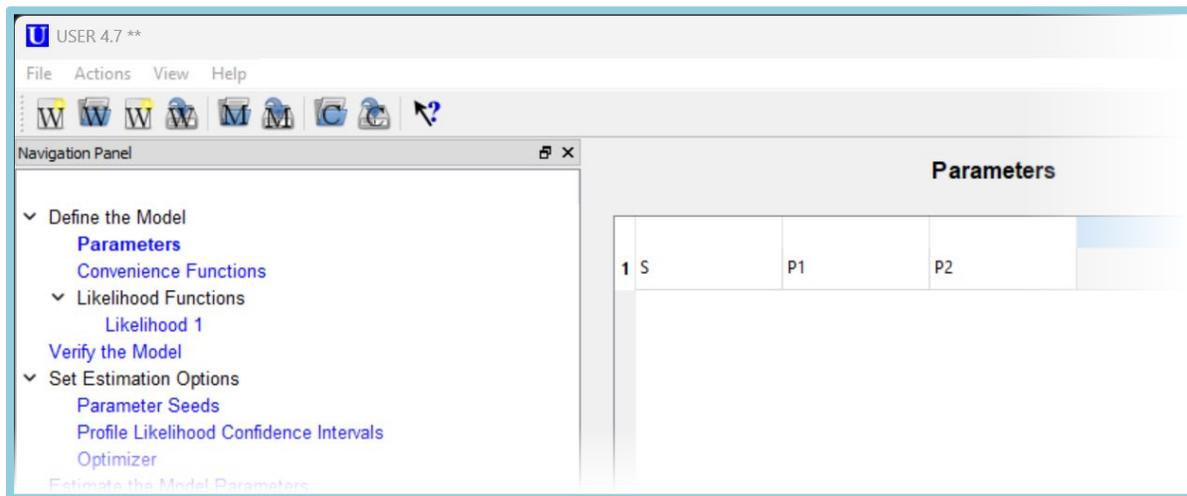
Figure 3.13: The **Parameters** content page for the joint likelihood example.

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 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 3.13.
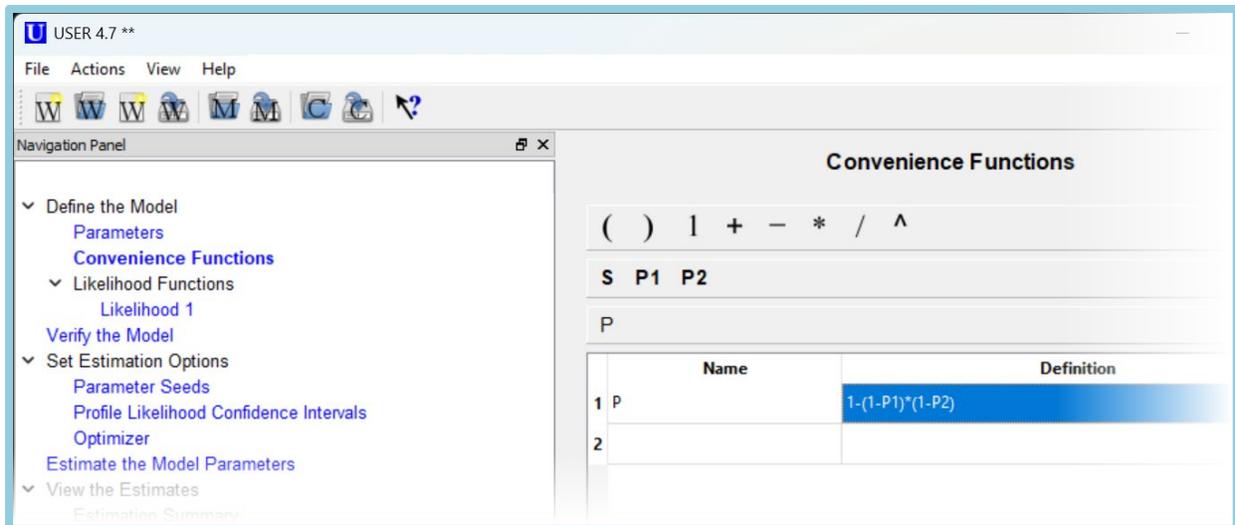
Figure 3.14: The **Convenience Functions** page for the joint likelihood example.

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 3.15) 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 3.16.
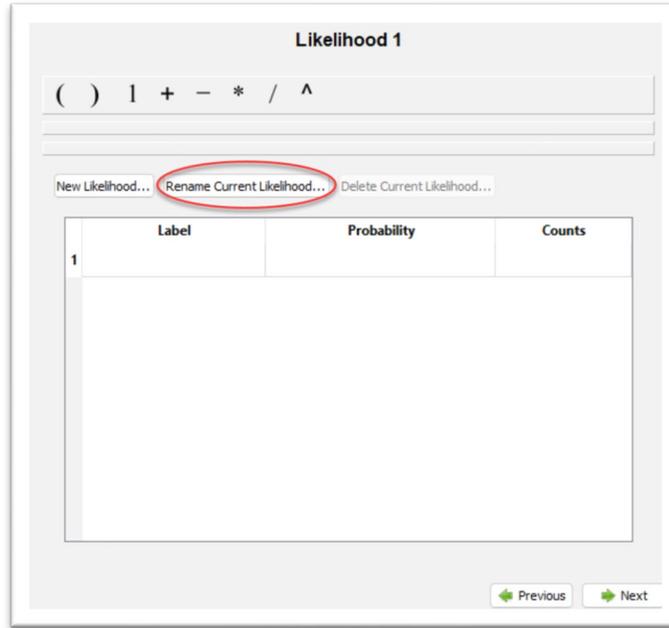
Figure 3.15: Button **Rename Current Likelihood…** for renaming the current likelihood probability.
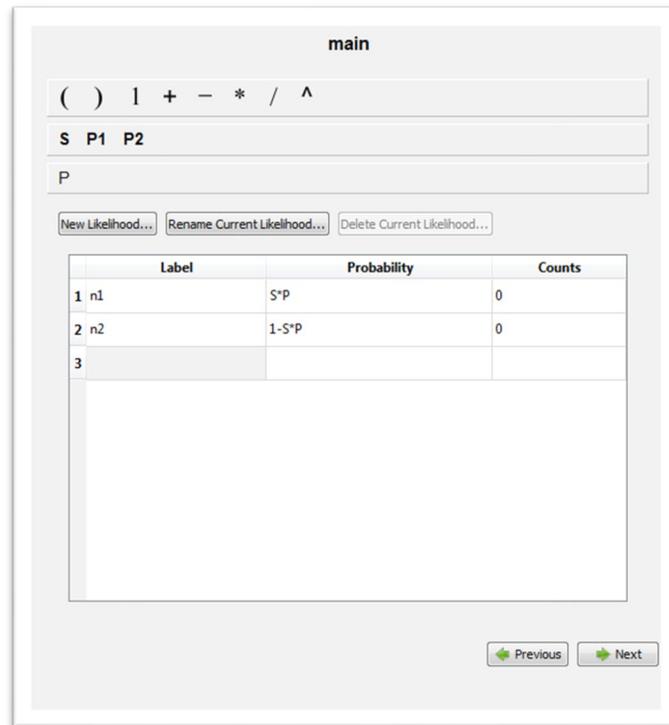
Figure 3.16: The "main" likelihood in the joint likelihood example.

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 3.17).
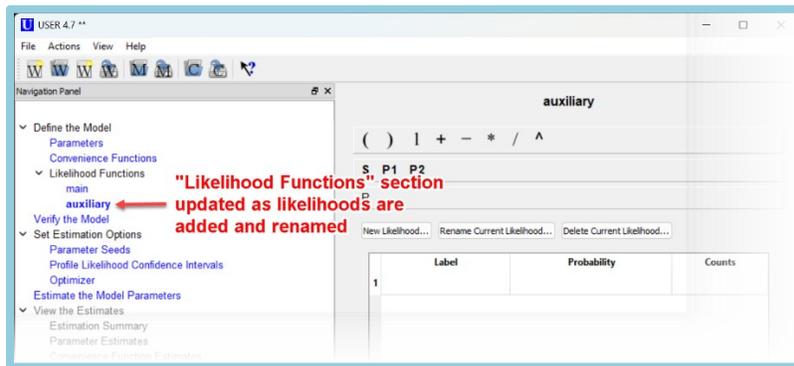
Figure 3.17: Creating the auxiliary likelihood definition for the joint likelihood example.

Fill in the auxiliary likelihood definition as shown in Figure 3.18.
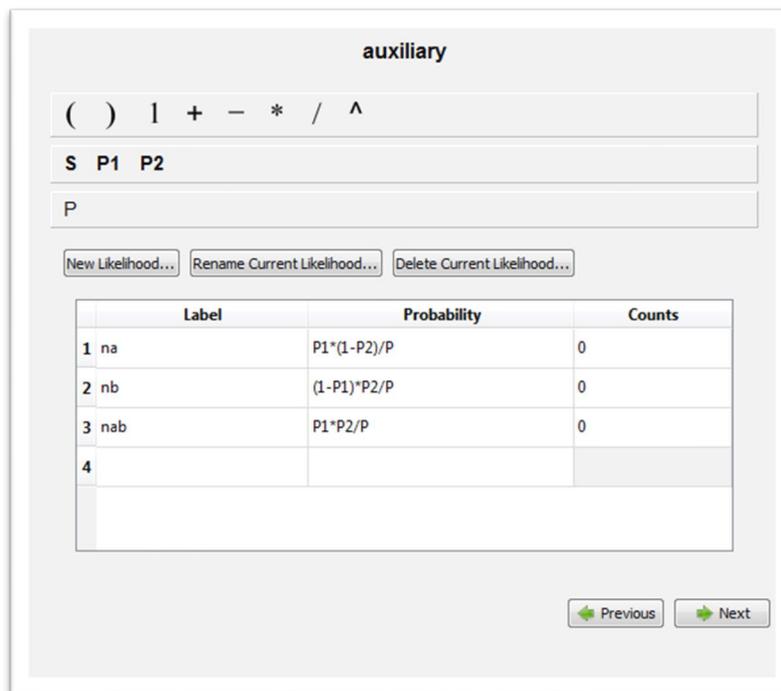
Figure 3.18: Completed auxiliary likelihood definition for the joint likelihood example.

3.2.2 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.

3.2.3 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 3.19. 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 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.

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

Figure 3.19: Category counts file for the joint likelihood example.

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 to 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.

3.2.4 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 3.20 shows the **Estimation Summary Report**.

Estimation Summary Report

29 January 2026 10::06::18 am
AIC: 23.9498
Log-likelihood: −8.97488
  main: −2.9929
  auxiliary: −5.98198

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 3.20: **Estimation Summary Report** for the joint likelihood example.

3.3 Estimating Abundance

In this example, we are interested in estimating the total abundance of a population, and we must account for 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 expected number of individuals removed will be $NP,$ the number of individuals multiplied by the sampling effort. For the second event, the expected 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
n_1	P
n_2	$(1 - P)P$
n_3	$(1 - P)^2 P$
n_4	$(1 - P)^3 P$

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

Category	Probability
n_1	P
n_2	$(1 - P)P$
n_3	$(1 - P)^2 P$
n_4	$(1 - P)^3 P$
n_0	$(1 - P)^4$

3.3.1 Define the model

In USER, define the parameters as shown in Figure 3.21. In order to make the likelihood definition cleaner, we will define the convenience function $q = 1 - P$ as show in Figure 3.22.

Figure 3.21: **Parameters** for the abundance estimation example.

Figure 3.22: **Convenience Functions** for the abundance estimation example.

For the likelihood definition, enter the first four categories as shown in Figure 3.23. 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 abundance parameter in parentheses as **(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 3.24).

Likelihood 1

( ) 1 + − * / ^

N P

q

| New Likelihood... | Rename Current Likelihood... | Delete Current Likelihood... |

	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			

Previous   Next

Figure 3.23: Partial likelihood definition for the abundance estimation example.
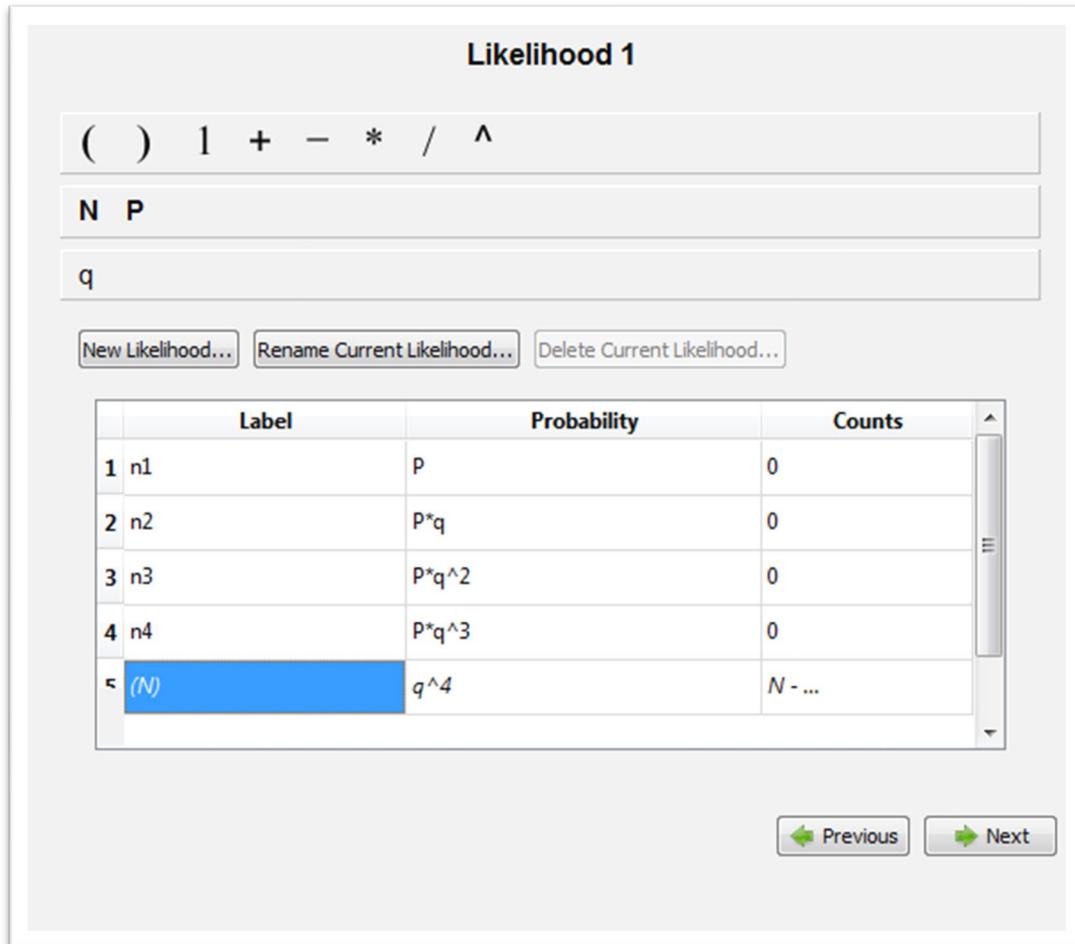
Figure 3.24: Likelihood definition for the abundance estimation example.

3.3.2 Estimate the parameters

Load the category counts

Create category counts file as shown in Figure 3.25, and load them using the **Load Category Counts** action under the **File** menu.

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

Figure 3.25: 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 also be helpful to use more than one optimizer, starting with the default **Powell's Direction Set** method and then using the resulting estimates as seeds for the **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 3.26 shows the Estimation Summary report.

Estimation Summary Report

29 January 2026 11::15::19 am
AIC: 70.8578
Log-likelihood: -33.4289

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 3.26: Estimation Summary for the abundance estimation example.

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 the 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 3.27).
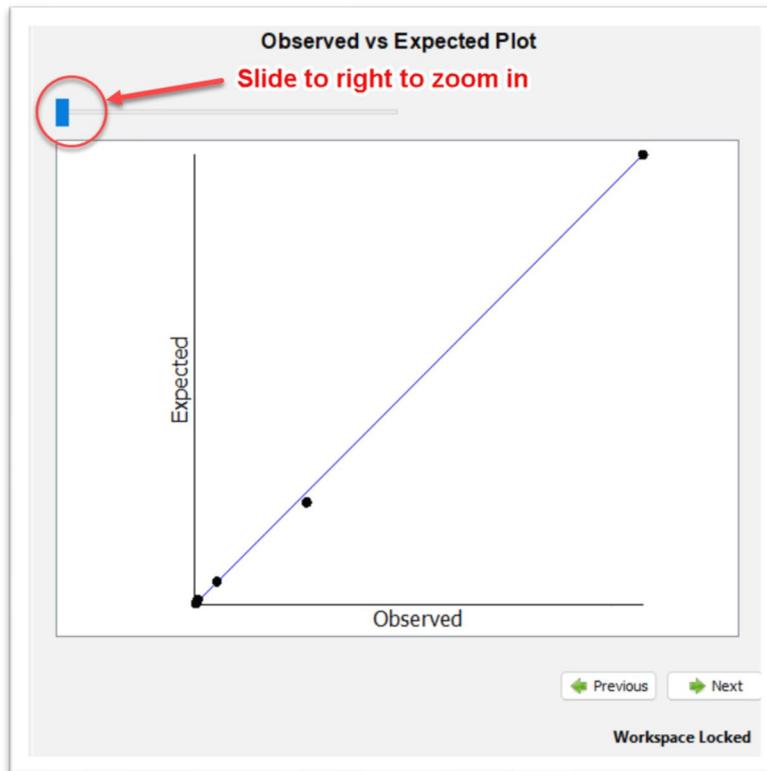
Figure 3.27: **Observed vs Expected Plot** for the abundance estimation example.

The **Residuals Plot** content page (Figure 3.28) shows a plot of the standardized Anscombe residuals, while the **Observed vs Expected Table** content page (Figure 3.29) shows the same residuals in a tabular format that can be easily saved to a file or copied to a clipboard. If the model assumptions are met, Anscombe residuals will be approximately normally distributed even for nonnormal response variables.

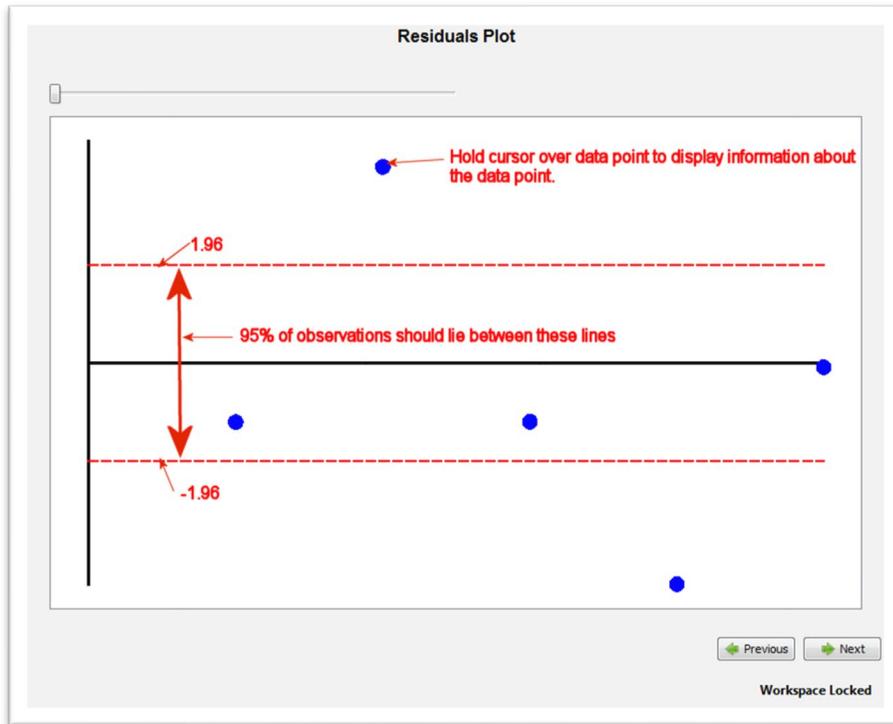Figure 3.28: **Residuals Plot** for the abundance estimation example.

Figure 3.29: **Observed vs Expected Table** for the abundance estimation example.

Looking at the **Residuals Plot** content page for this example (Figure 3.28), you will observe that two of the five residuals 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.

3.4 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.

3.3.1 Full model definition

We define the parameters as shown in Figure 3.30. 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 3.31. The likelihood definition is shown in Figure 3.32.

Figure 3.30: **Parameters** definition for the hypothesis testing example.

Figure 3.31: **Convenience Functions** definition for the hypothesis testing example.

Likelihood 1

( ) 1 + − * / ^

S1 S2

Stotal

[New Likelihood...] [Rename Current Likelihood...] [Delete Current Likelihood...]

	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			

[← Previous] [→ Next]

Figure 3.32: **Likelihood** definition for the hypothesis testing example.

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 3.33 shows that we are requesting profile likelihood confidence intervals (alpha-level = 0.05) for the individual survival probabilities as well as the overall survival.

Figure 3.33: **Profile Likelihood Confidence Intervals** request for the hypothesis testing example.

Estimate the parameters

Advance to the **Estimate the Model Parameters** page and press the **Estimate** button. Figure 3.34 shows the **Estimation Summary Report** for the full model. In order to perform the hypothesis test, we need to record the log-likelihood (-5.63645).

Estimation Summary Report

29 January 2026 11::34::13 am
AIC: 15.2729
Log-likelihood: –5.63645

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	0.84438	0.91318
S2	0.748322	0.05	0.69706	0.79534
Stotal	0.659763	0.05	0.63187	0.68335

Convenience Functions

Name	Estimate	s.e.
Stotal	0.659763	0.0257707

Figure 3.34: **Estimation Summary Report** for the hypothesis testing example, full model.

3.3.2 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 3.35. We now have only one parameter: **S1**.

3. Go to the **Convenience Functions** page and add the convenience function **S2** as shown in Figure 3.36. 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 **Profile Likelihood Confidence Intervals** page and re-select alpha-level 0.05 for both S_2 and S_{total}.

5. Go to the **Estimate the Model Parameters** page and press **Estimate**. Figure 3.37 shows the resulting estimation summary report.

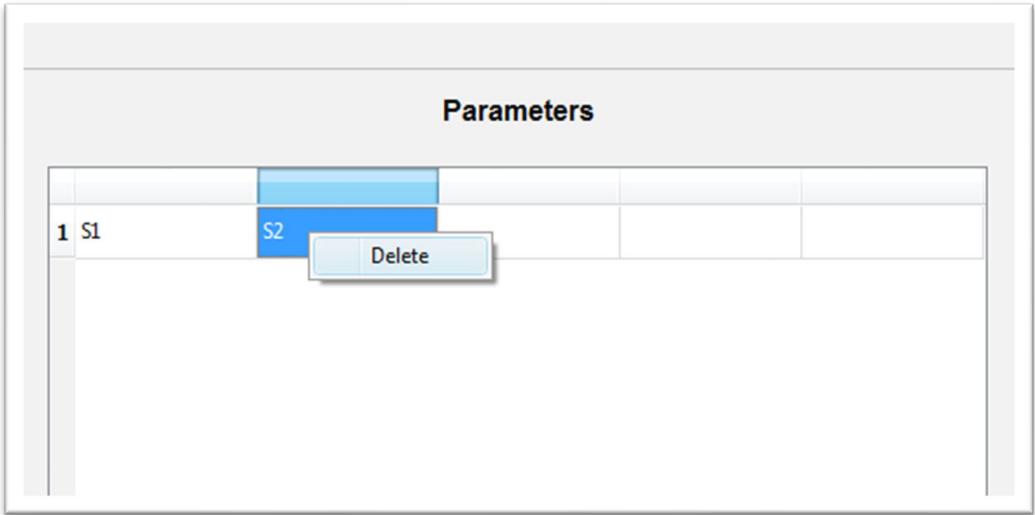As with the full model, we must note the log-likelihood (-15.2063).

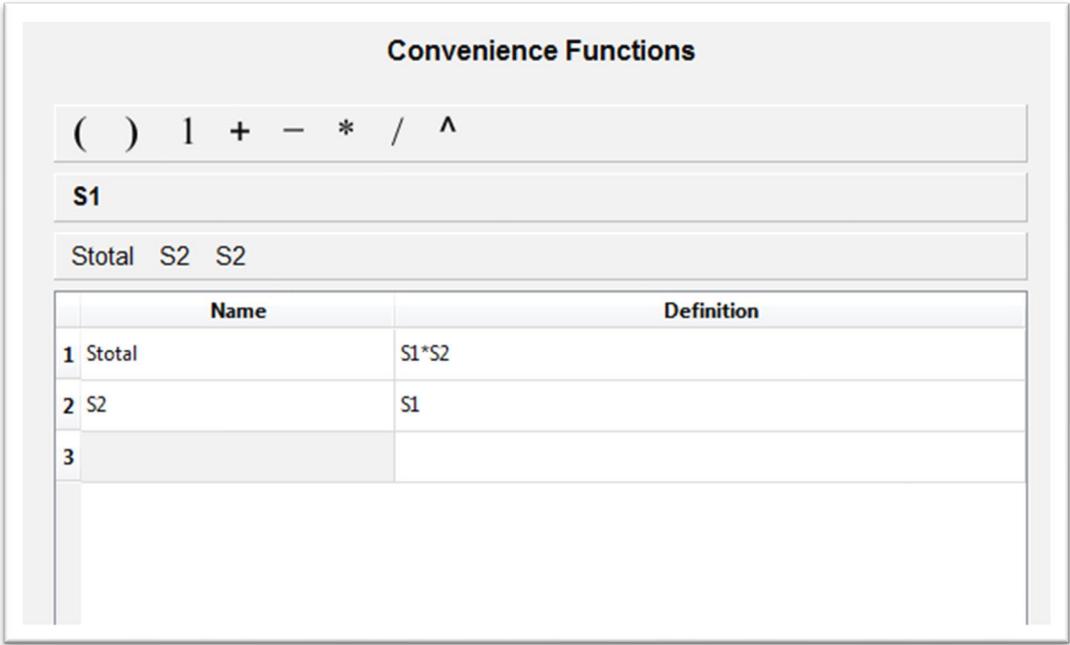Figure 3.35: Delete parameter **S2** for the hypothesis testing example.

Figure 3.36: Define the convenience function **S2** for the hypothesis testing example.

Estimation Summary Report

29 January 2026 11::56::04 am
AIC: 32.4127
Log-likelihood: -15.2063

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	0.62097	0.71874

Convenience Functions

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

Figure 3.37: **Estimation Summary Report** for the hypothesis testing example, reduced model.

3.3.3 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 $S_1 \neq S_2$.

3.3.4 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.