

USER 2.1

User Specified Estimation Routine

Technical Manual
2003



This Document should be cited as follows:

Lady, James, Peter Westhagen, John Skalski, "USER 2.1", Project No. 1989-10700, 51 electronic pages, (BPA Report DOE/BP-00012494-2)

Bonneville Power Administration
P.O. Box 3621
Portland, Oregon 97208

This report was funded by the Bonneville Power Administration (BPA), U.S. Department of Energy, as part of BPA's program to protect, mitigate, and enhance fish and wildlife affected by the development and operation of hydroelectric facilities on the Columbia River and its tributaries. The views in this report are the author's and do not necessarily represent the views of BPA.

USER 2.1

User Specified Estimation Routine

Developed by
James M. Lady, Peter Westhagen, and John R. Skalski
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 97208-3621
Project No. 198910700

March 2003

Contents

Preface	iv
1 Overview	1
2 Interface Description	2
2.1 Defining symbols	2
2.2 Modifying symbol definitions	2
2.3 Defining likelihoods	4
2.4 Entering function definitions	5
2.5 Using the Entry Mode Toggle	6
2.6 Setting the initial parameter seeds	7
2.7 Verifying the likelihoods	7
2.8 Requesting profile likelihood confidence intervals	8
2.9 Estimating the model parameters	9
2.10 Saving/Restoring model definitions and model data	11
2.11 The covariance matrix	12
2.12 Other Functions	13
3 File Formats	14
3.1 Model definition file	14
3.2 Model data file	16
4 Examples	17
4.1 Petersen method	17
4.1.1 Model definition	17
4.1.2 Model definition in USER	18
4.2 Joint likelihood	26
4.2.1 Model definition	26

4.2.2	Model definition in USER	27
4.3	Hypothesis testing	35
4.3.1	Model definition	35
4.3.2	Model Definition in USER	36
4.3.3	Estimate the Parameters	39
4.3.4	Create the Null Hypothesis Model	40
4.3.5	Test the Hypothesis	42

List of Figures

2.1	Main USER window	3
2.2	USER results output window	10
3.1	Sample model definition file	15

Preface

This document is primarily a description of the user interface for USER2.1; it is not a description of the statistical theory and calculations behind USER.

This project is funded by the Bonneville Power Administration, U.S. Department of Energy, under Contract No. 004126, Project No. 198910700 as part of the BPA's program to protect, mitigate, and enhance fish and wildlife affected by the development and operation of hydroelectric facilities on the Columbia River and its tributaries.

The analysis of fish and wildlife data requires investigators to have the ability to develop statistical models tailored to their study requirements. Hence, a flexible platform to develop statistical likelihood models to estimate demographic parameters is necessary. To this end, Program USER (User Specified Estimation Routine) was developed to provide a convenient platform for investigators to develop statistical models and analyze tagging and count data.

The program is capable of developing models and analyzing any count data that can be described by multinomial or product multinomial distributions. Such data include release-recapture studies using PIT-tags, radio-tags, balloon-tags, and acoustic-tags to estimate survival, movement, and demographic data on the age and/or sex structure of wild populations. The user of the program can specify the parameters and model structure at will to tailor the analyses to the specific requirements of the field sampling program, the data, and populations under investigation. All of this is available without the need for the user to know advanced programming languages or numerical analysis techniques, and without involving cumbersome software developed for extraneous purposes. Program USER represents a powerful statistical modeling routine that can be readily used by investigators with a wide range of interests and quantitative skills.

The USER software, along with an electronic version of this document, is available for download at <http://www.cbr.washington.edu/paramEst/USER/>, or contact:

Jim Lady
Columbia Basin Research
1325 4th Ave. Suite #1820
Seattle, WA 98101
206-221-3252 jlady@u.washington.edu

1.0 Overview

USER is a tool that allows a user 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.

USER makes a distinction between a model definition and the model data. Both are needed in order to estimate the parameters.

1. A **model definition** consists of one or more likelihood definitions. Each likelihood definition consists of:
 - (a) **Category labels** for each of the possible outcomes,
 - (b) A **probability** corresponding to each category defined in terms of the model parameters.

Within a likelihood, the probabilities for all categories must sum to 1.0.

2. The **model data** consist of the number of observations for each category. **USER** stores model definitions and model data separately, thereby allowing a given set of data to be used with more than one model definition, or, conversely, allowing more than one set of data to be used with a given model definition.

2.0 Interface Description

Figure 2.1 shows the USER dialog window with a simple trinomial model likelihood. The names of the components of the dialog window mentioned in the following text are italicized.

2.1 Defining symbols

The USER program allows the user to create two types of symbols:

1. **Parameters** of the model to be estimated,
2. **Variables** defined as a function of the parameters and/or other variables.

Symbol names must begin with an alphabetic character, and can be followed by any number of alphanumeric characters.

Both parameters and variables are defined via the *symbol name dialog* shown in Figure 2.1. To define parameters, a user enters a name for the parameter in the *symbol name dialog* and leaves the *symbol definition dialog* blank; to define symbols, a user enters the name in the *symbol name dialog* and its definition in the *symbol definition dialog* (see Section 2.4 on page 5). If using the mouse along with the *math entry buttons* and the symbol definitions to enter a symbol definition, make sure the *entry mode toggle* is not depressed.

A symbol definition is added to the *symbol definitions* by clicking on the “Add” button to the right of the *symbol definition dialog*, or by pressing the “Return” key.

2.2 Modifying symbol definitions

A symbol definition may be deleted by right-clicking on the symbol name in the *symbols definition area* and selecting “Delete” from the context-sensitive menu.

To modify the name or definition of a symbol, select “Edit” from the context-sensitive menu. The symbol and its definition will then appear in the *symbol name dialog* and the *symbol definition dialog* for the user to edit.

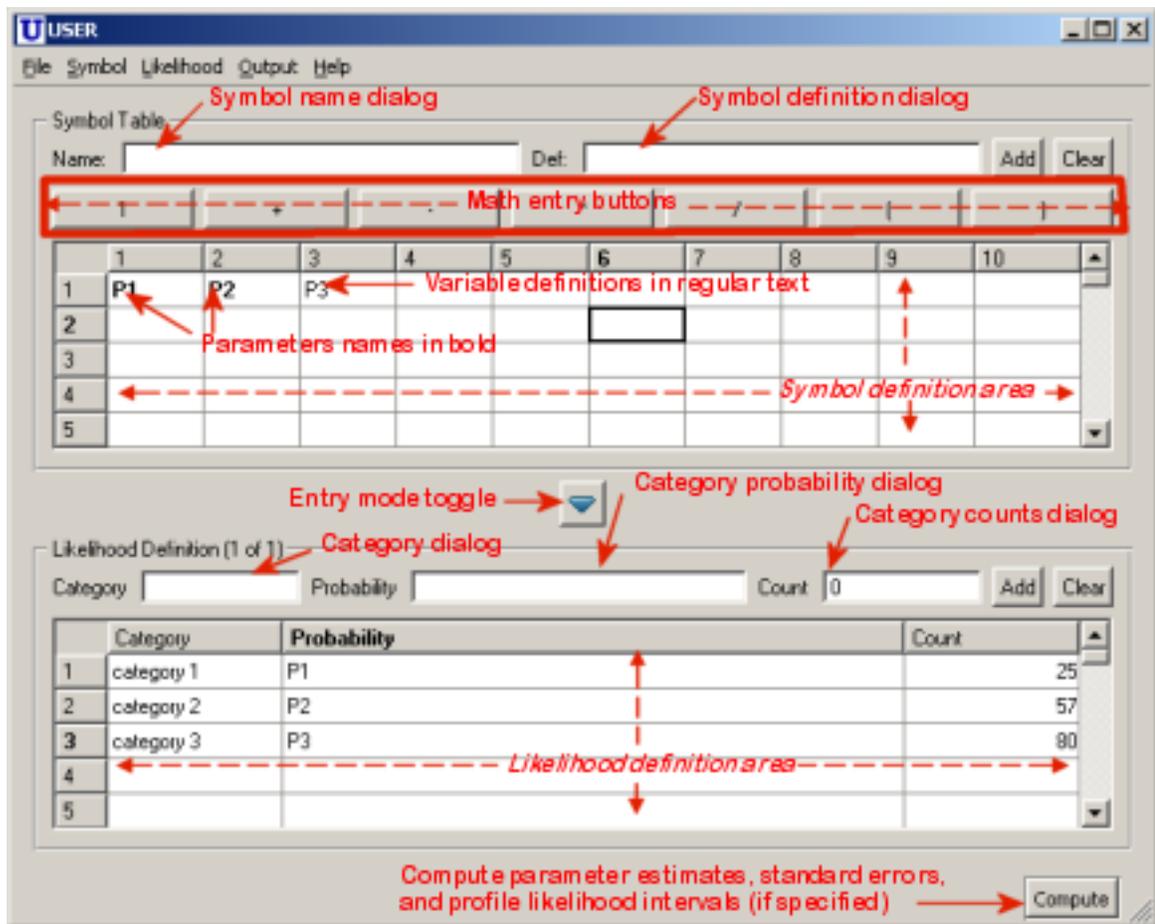
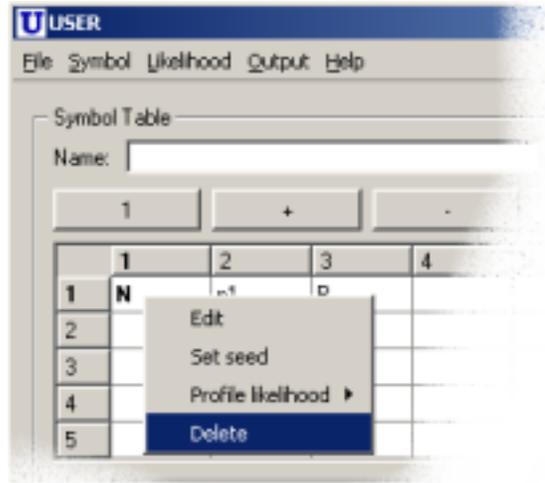


Figure 2.1: Main USER window



2.3 Defining likelihoods

A likelihood definition is comprised of a list of category definitions with their corresponding probabilities and observed counts for the category. The category counts default to zero, giving the user the option of reading the category counts from a separate input file. The categories, their probabilities, and their counts are entered in the *category dialog*, *category probability dialog* and the *category counts dialog*, respectively.

If using the mouse, the *entry mode toggle* must be depressed to ensure that clicking on the *math entry buttons* and the symbol definitions affect the *category probability dialog* rather than the *symbol definition dialog*.

A category definition within a likelihood can be modified by clicking on the appropriate line in the likelihood definition, causing the definition to appear in the *category dialog*, *category probability dialog*, and the *category count dialog* for editing.

Category labels may consist of any string of characters, including spaces, with one exception: They may not contain a single colon (":") surrounded by white space since it is used by USER to indicate the end of a category definition (see Figure 3.1 on page 15).

For models with multiple likelihoods, a new likelihood may be defined by selecting the "New Likelihood" item from the "Likelihood" menu.

variable definition, they must already have been defined, or else they will be flagged as undefined when the model is evaluated. The mathematical operations available are:

<u>Operation</u>	<u>Symbol</u>
Addition	+
Subtraction	-
Multiplication	*
Division	/

Operations are evaluated in the usual order of precedence from left to right. Parentheses may be used to specify the order of evaluation.

In addition to above, the following functions are available:

- $invLogit(u)$, defined as $exp(u)/(1 + exp(u))$,
- $logit(u)$, defined as $log_e(u/1 - u)$.

where u is any valid USER expression.

A user may enter a function definition directly with the keyboard, or by using the mouse in conjunction with the *math entry buttons* and the already defined parameters and variables as follows.

- Insert a mathematical operation, a parenthesis, or the number one (“1”) by clicking on the appropriate *math entry button*.
- Insert an already defined parameter or variable by clicking on the desired symbol in the *symbol definitions area*.
- Insert 1.0 minus (“1.0 -”) a symbol by holding down the control key on the keyboard while clicking on the symbol in the *symbol definitions area*.

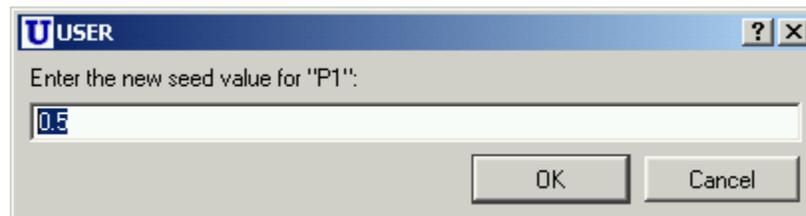
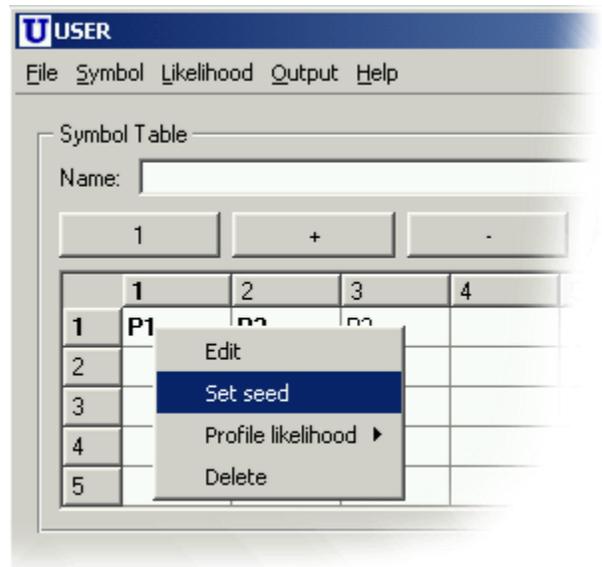
2.5 Using the Entry Mode Toggle

The *entry mode toggle* has two modes:

1. If depressed, clicking on one of the *math entry buttons* or a symbol definition modifies the *category probability dialog*.
2. If not depressed, clicking on one of the *math entry buttons* or a symbol definition modifies the *symbol definition dialog*.

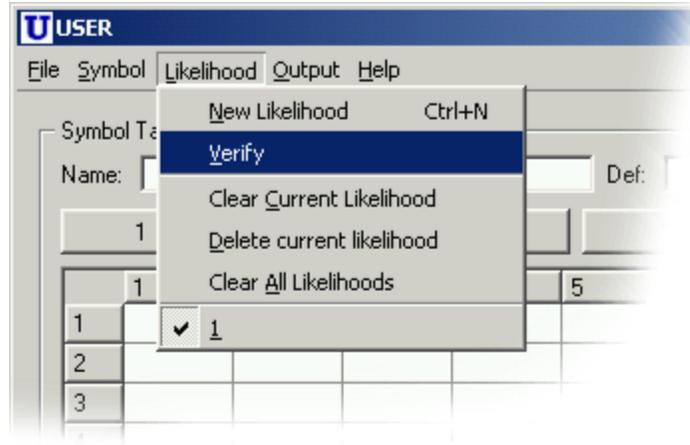
2.6 Setting the initial parameter seeds

For the parameters to be estimated, an initial “guess”, or seed, must be set for each parameter. The default initial seed is 0.5 for all parameters. This may not be adequate for some models, so USER allows the initial seed to be set by the user by right-clicking on the parameter and selecting “Set seed”.



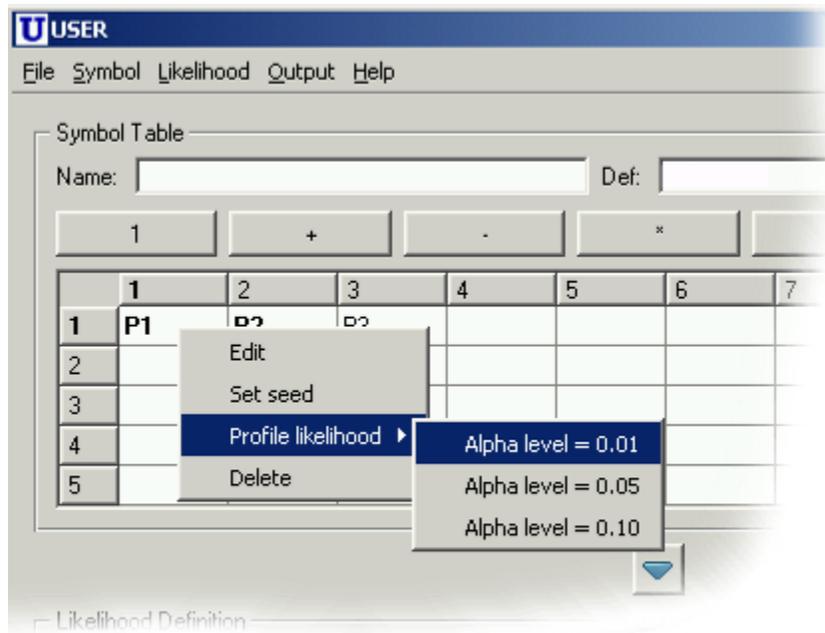
2.7 Verifying the likelihoods

USER provides the capability of checking if all the probabilities for each likelihood defined in the model sum to one. Once the “Verify” option is selected, USER will display a message indicating whether or not the cell probabilities sum to one. The model parameters cannot be estimated until all cell probabilities sum to one.



2.8 Requesting profile likelihood confidence intervals

A user may request the profile likelihood confidence intervals for any of the model parameters by right-clicking on the parameter or function of parameters in the *symbol definitions area* and selecting the “Profile likelihood” from the context-sensitive menu. The user may then select from three alpha-levels: 0.01, 0.05, or 0.10. Profile likelihood confidence intervals may be selected for multiple parameters and functions.



2.9 Estimating the model parameters

Once the model has been fully defined and the user has selected all the desired profile likelihood confidence intervals, the “Compute” button is clicked whereupon USER computes and displays the parameter estimates, their standard errors, and the profile likelihood confidence intervals as shown in Figure 2.2.

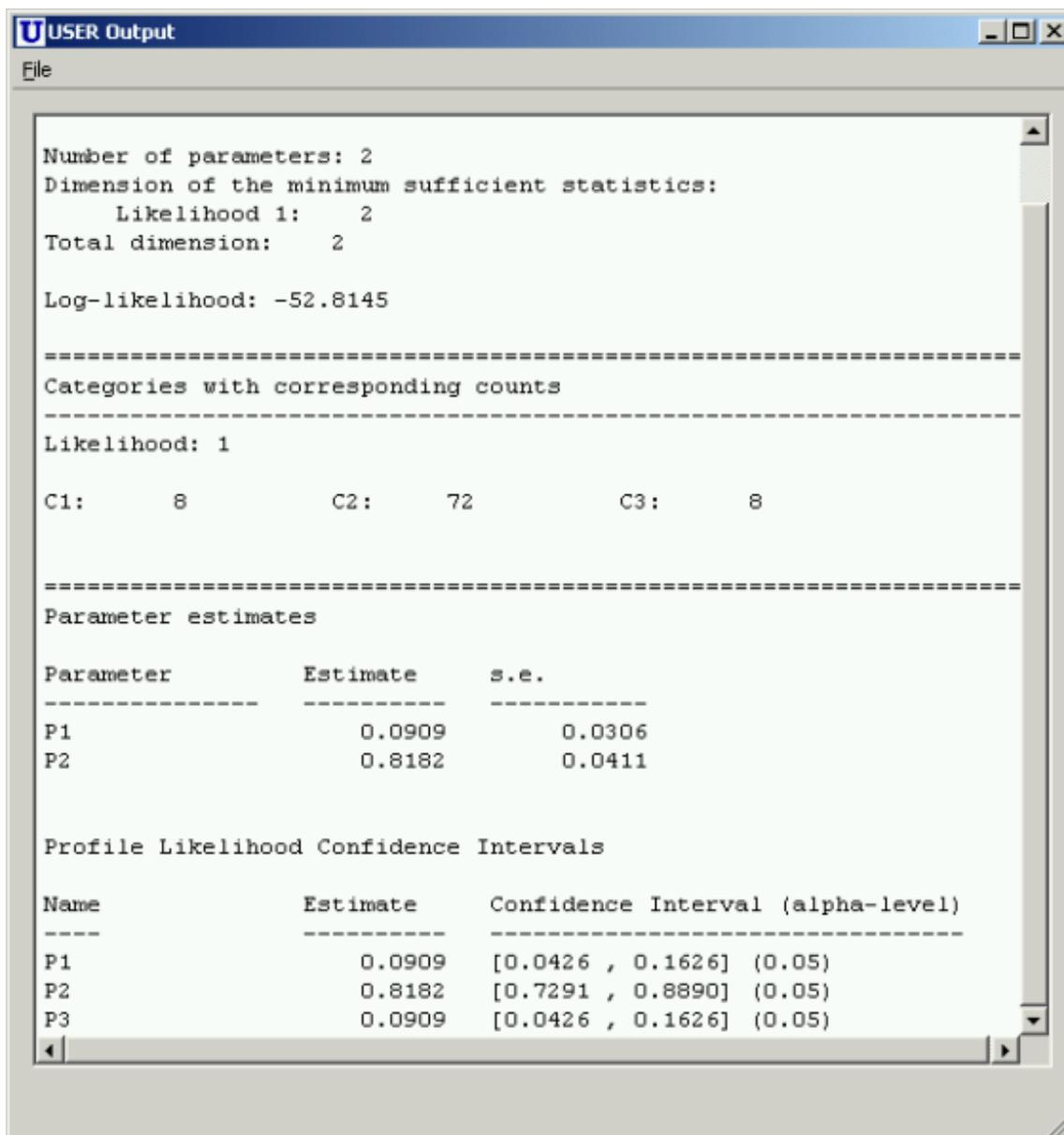
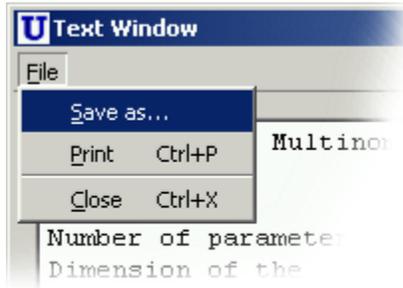
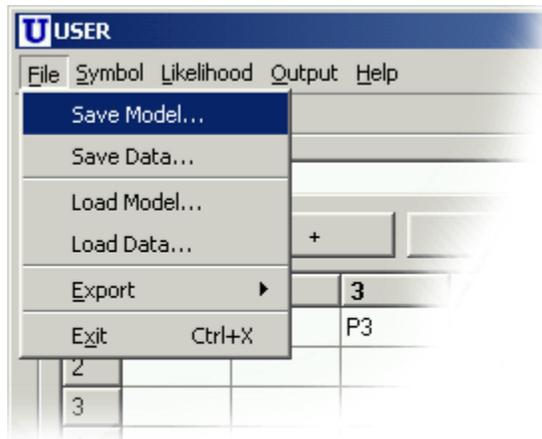


Figure 2.2: USER results output window

The “File” menu of the output window allows the user to save the results to a text file or to print the results.



2.10 Saving/Restoring model definitions and model data



The USER program allows the model definition and the model data to be stored in separate files for use at a later time. A user may select “Save Model...” to save the model definition (but not the category counts!) to a file, and “Load Model...” to load a previously defined model.

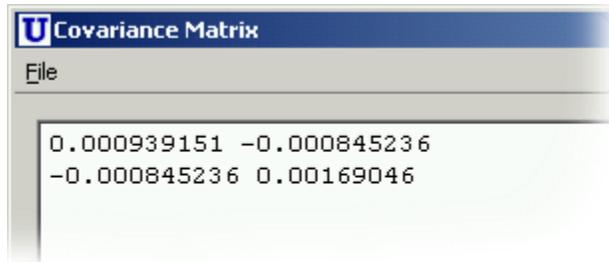
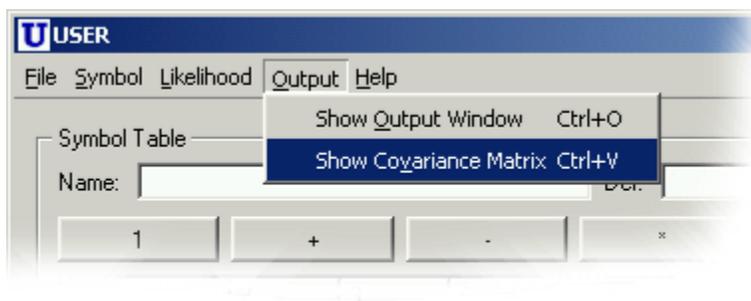
Similarly, the model data (i.e., the counts per category) may be saved and restored using the “Save Data...” and “Load Data...” options in the file menu shown above.

Both the model definition file and the model data file are text files that may be edited directly by the user if so desired; a user may use any text editor to create

a model definition file to be loaded into USER at a later time. Similarly, the data may be entered in a text file outside of the USER program (see Chapter 3).

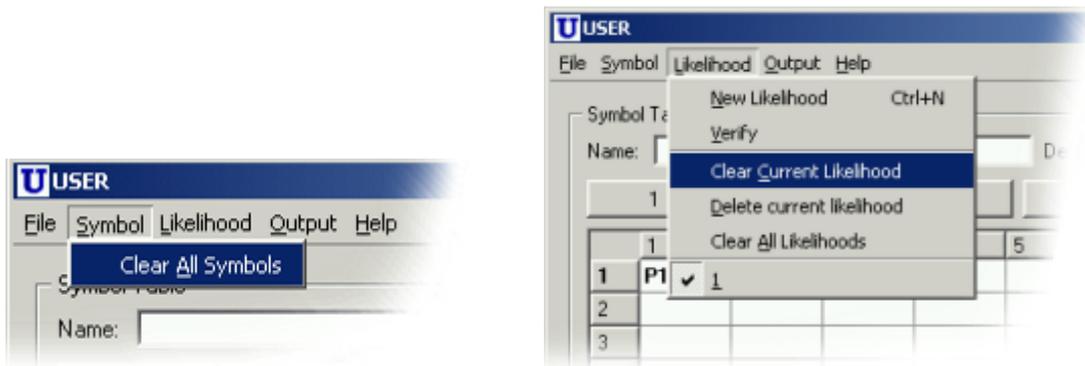
2.11 The covariance matrix

If a user wants to calculate the variance of a function of the model parameters, the covariances of the parameters are needed. USER provides the facility to display the covariance matrix as shown below. The second image below is the covariance matrix for a model with two parameters.



The covariance matrix is formatted with spaces between the parameters, allowing it to be easily saved to a file and imported into a statistical software package or spreadsheet.

2.12 Other Functions



Other functions available in USER include the ability to delete (clear) all currently defined symbols, clear the current likelihood or all likelihoods, or delete the current likelihood.

3.0 File Formats

3.1 Model definition file

Figure 3.1 is an example of the model definition file from a somewhat complicated model definition. USER saves a model definition in this format when a user saves a model definition to a file (Section 2.10).

Some important things to note:

- Keywords such as “parameters” and “likelihood” are not case sensitive; parameter and variable names, however, ARE case sensitive.
- All keywords and tokens must be surrounded by whitespace (tab, space, newline). This applies to the braces (“{” and “}”) as well.
- The model definition file is free format in that new lines can be inserted wherever needed with the exception of within a category label and within a symbol name.

```

parameters { Spool E G Psp1 Psp2 Pph1 Pph2 Psl1 Psl2 Ssl Sph Ssp lambda C }
parametersInit { 0.2 0.3 0.1 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.7 0.7 0.5 0.1 }
variableDefs
{ Psp = 1 - (1-Psp1)*(1-Psp2) }
{ Psl = 1 - (1-Psl1)*(1-Psl2) }
{ Pph = 1 - (1-Pph1)*(1-Pph2) }
{ Eff = F*G }
profileIntervals
{ S : 0.05 }
{ Eff : 0.05 }
likelihood
{ 1 0 0 : 1 - Spool + Spool*(1-C)*((1-E)*G*(1-Psl)*(1-Ssl+Ssl*(1-lambda)) +
(1-E)*(1-G)*(1-Pph)*(1-Sph+Sph*(1-lambda)) +
E*(1-Psp)*(1-Ssp+Ssp*(1-lambda))) }
{ 1 0 1 : Spool*(1-C)*((1-E)*G*(1-Psl)*Ssl + (1-E)*(1-G)*(1-Pph)*Sph +
E*(1-Psp)*Ssp)*lambda }
{ 1 2 0 : Spool*(1-C)*(1-E)*G*Psl*(1-Ssl+Ssl*(1-lambda)) }
{ 1 2 1 : Spool*(1-C)*(1-E)*G*Psl*Ssl*lambda }
{ 1 3 0 : Spool*(1-C)*(1-E)*(1-G)*Pph*(1-Sph+Sph*(1-lambda)) }
{ 1 3 1 : Spool*(1-C)*(1-E)*(1-G)*Pph*Sph*lambda }
{ 1 4 0 : Spool*(1-C)*E*Psp*(1-Ssp+Ssp*(1-lambda)) }
{ 1 4 1 : Spool*(1-C)*E*Psp*Ssp*lambda }
{ 1 5 0 : Spool*C }

likelihood
{ 1 0 : 1-lambda }
{ 1 1 : lambda }

likelihood
{ 2a : Psl1*(1-Psl2)/Psl }
{ 2b : (1-Psl1)*Psl2/Psl }
{ 2ab : Psl1*Psl2/Psl }

likelihood
{ 3a : Pph1*(1-Pph2)/Pph }
{ 3b : (1-Pph1)*Pph2/Pph }
{ 3ab : Pph1*Pph2/Pph }

likelihood
{ 4a : Psp1*(1-Psp2)/Psp }
{ 4b : (1-Psp1)*Psp2/Psp }
{ 4ab : Psp1*Psp2/Psp }

```

Required parameter definitions - must be the first section of the model definition file.
 Optional initial "seed" values for the parameters. All default to 0.5 if this section is omitted.
 The number of values must match the number of parameters.
 Optional variable definitions
 "Eff" is not used in the likelihood definitions, but rather it represents a function of the parameters for which we want a profile confidence interval calculated (next section)
 Request profile likelihood intervals at alpha = 0.05 level for one parameter ("S") and one function of the parameters ("Eff")
 The keyword "likelihood" must begin each likelihood in the model definition.

Figure 3.1: Sample model definition file

3.2 Model data file

Each line of the model data file consists of a category label (the labels must correspond to the category labels in the model definition), optionally followed by a colon (":") and a count. If a count is not specified, it is assumed to be 1. A category may be repeated, in which case the counts are simply added to the previous counts of the same category. The following examples are all equivalent.

<u>Example 1</u>	<u>Example 2</u>	<u>Example 3</u>
C1 : 3	C1	C1 : 2
C2 : 4	C2	C2 : 2
C3 : 5	C1	C3 : 3
	C2	C1 : 1
	C2	C2 : 2
	C1	C3 : 2
	C3	

4.0 Examples

4.1 Petersen Method for estimating animal abundance

4.1.1 Model definition

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 made-up 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 .

We can define the categories and their expected values as follows:

<u>Category</u>	<u>Probability</u>
m	P
$n_2 - m$	$1 - P$

The resulting likelihood is

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

4.1.2 Model definition in USER

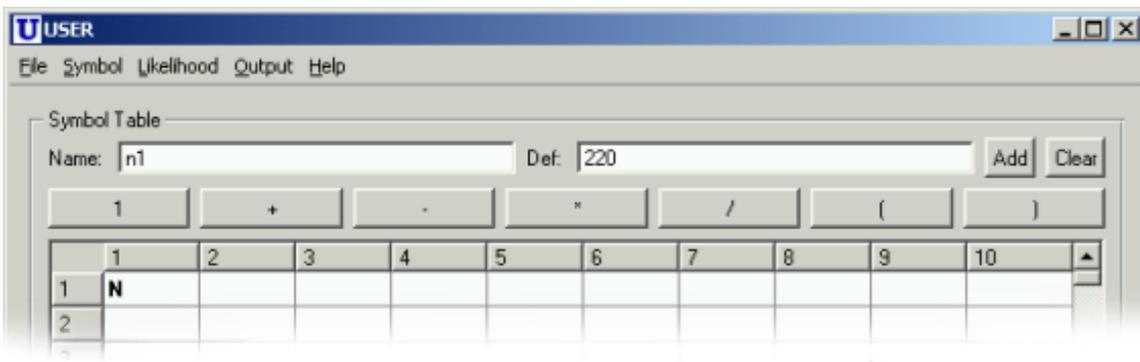
Define the parameters

We can now define the model in the USER program. First, the model parameter N is defined by typing it in the *symbol name dialog* and typing “Return” or clicking on the “Add” button; since the *symbol definition dialog* was left blank, USER recognizes N as a model parameter.

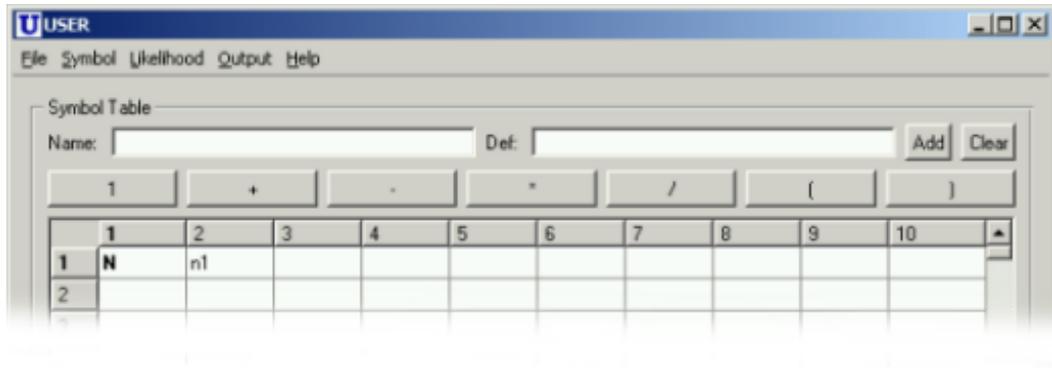
Define the variables

n_1 is a constant for the model and the category probabilities are written as a function of n_1 . We can define it as a variable as follows:

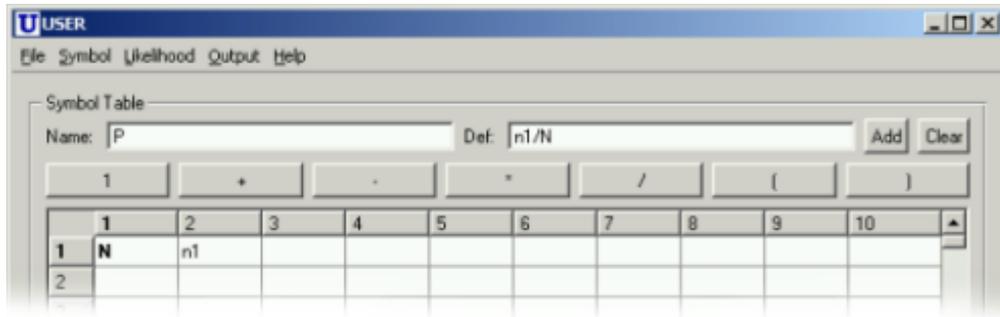
1. Type “n1” in the symbol name dialog.
2. Press “Tab” to advance to the symbol definition dialog (or click inside it) and type in the value “220”.



3. Press “Return” or click “Add” to add the symbol n_1 to the symbol definitions.
4. Define the variable $P = \frac{n_1}{N}$ by typing “P” in the symbol name dialog, and press “Tab” to proceed to the symbol definition dialog. The USER dialog window now appears as shown below.



5. The equation “ $n1/N$ ” can be entered in the symbol definition dialog by typing the expression, or as follows:
 - (a) Making sure that the entry mode toggle is not depressed, click on the “ $n1$ ” in the symbol definitions.
 - (b) Click on the “/” button.
 - (c) Click on the “ N ” in the symbol definitions.

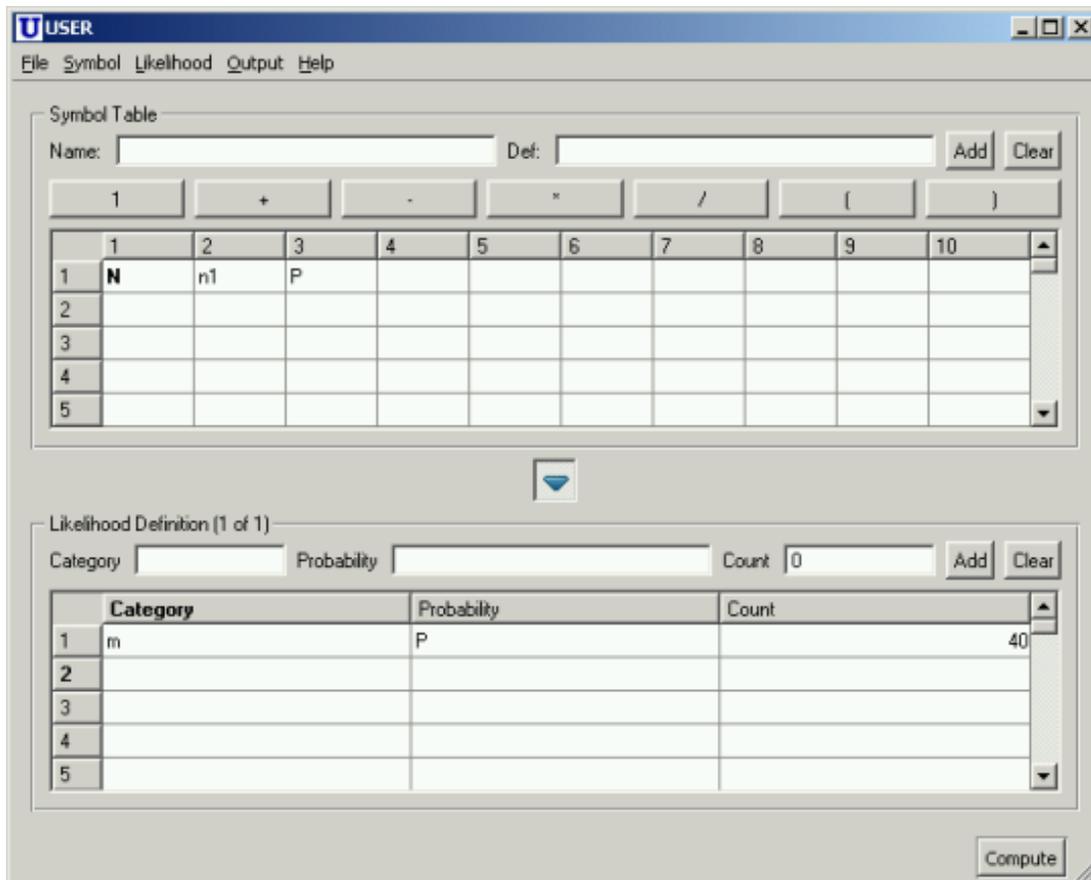


Define the likelihood and enter the category counts

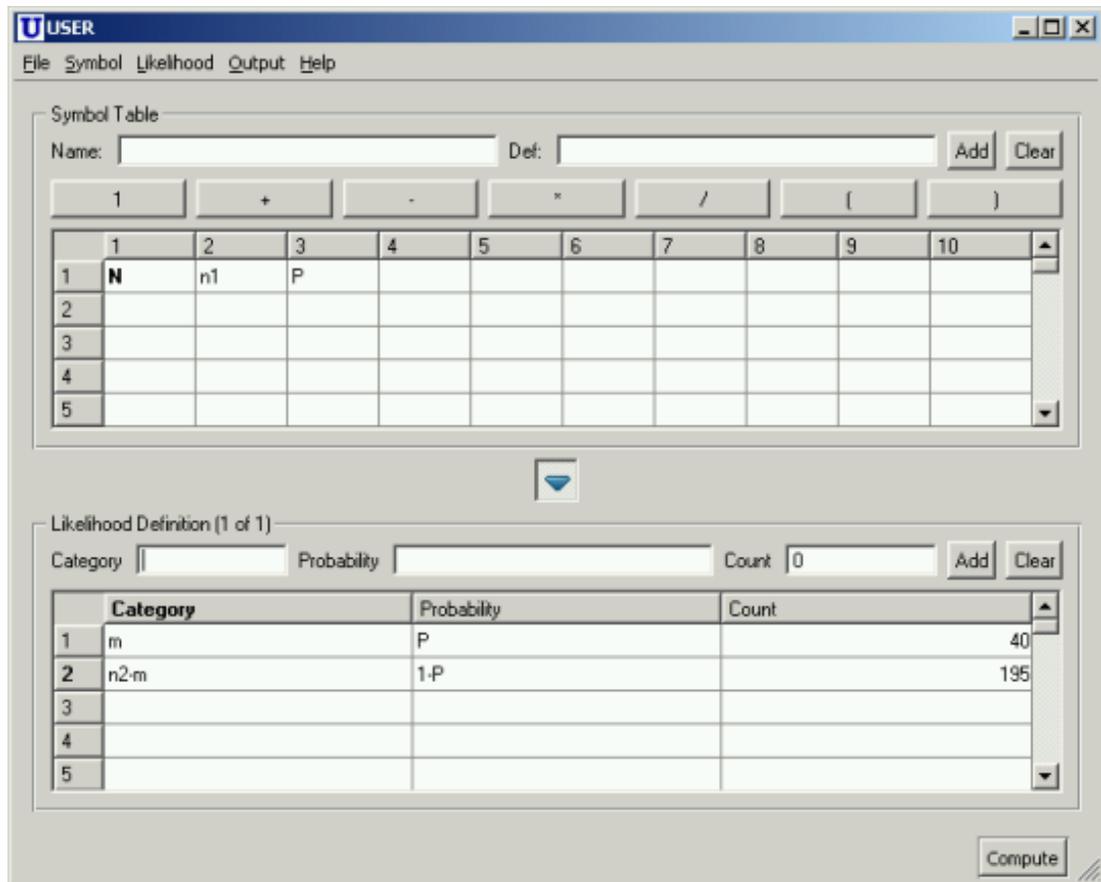
We now have all the necessary symbols defined, and we can proceed to defining the likelihood in terms of the categories and their probabilities.

1. Click on the *entry mode toggle* to depress it and have all changes affect the category definitions rather than the symbol definitions.
2. Enter “ m ” in the *category dialog*.
3. Press “Tab” or click inside the *category probability dialog* and enter “ P ” for the probability by either clicking on it in the *symbol definitions area* or typing it from the keyboard.

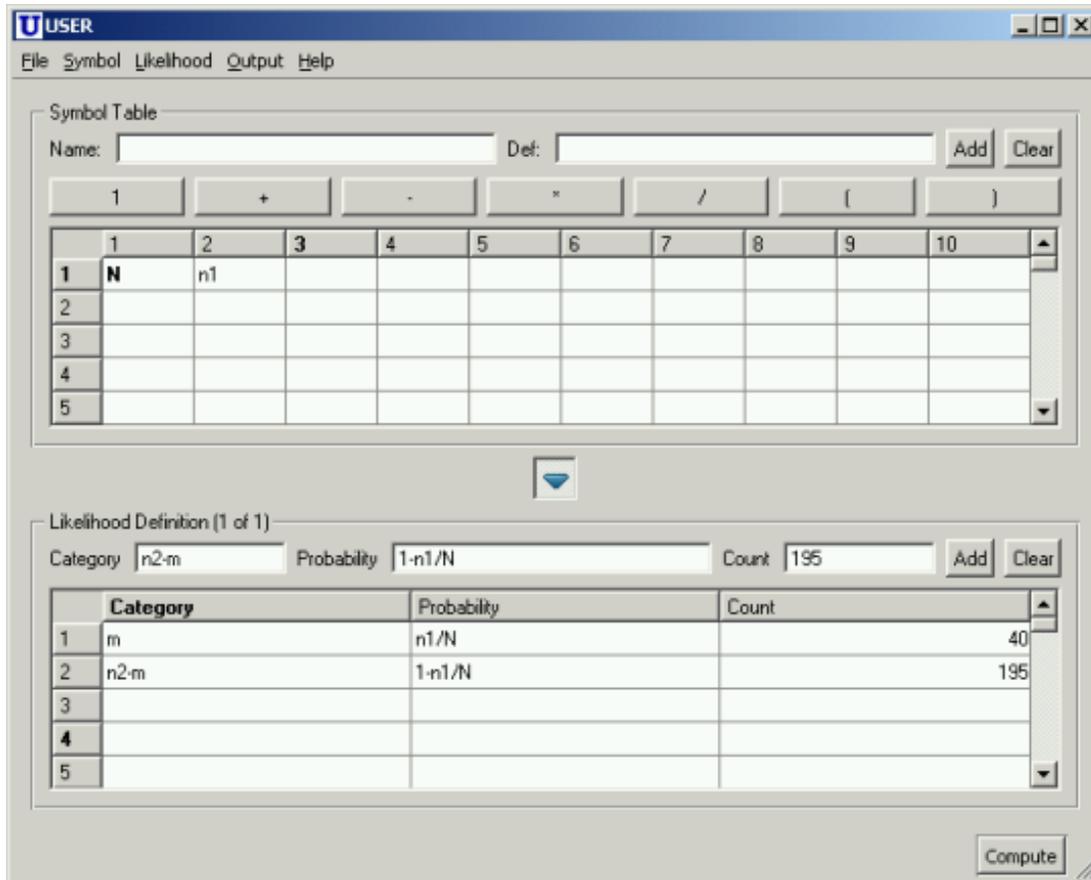
4. Advance to the *category counts dialog* and enter “40”
5. Press “Return” or click “Add”. The USER dialog should now look as shown below.



6. For the second and final category, enter “n2-m” in the *category dialog*, “1-P” in the *category probability dialog*, and “195” (235 minus 40) in the *category counts dialog*.
7. Press “Return” or click “Add” to finish the model definition. The USER dialog should now look as follows:

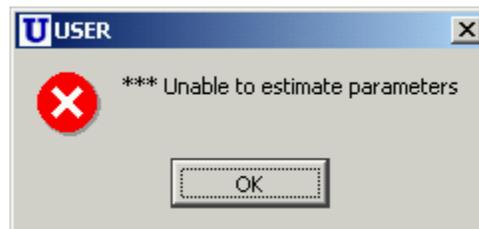


Note that defining P as a variable was not absolutely necessary. We could have defined the model as shown:



Estimate the parameters

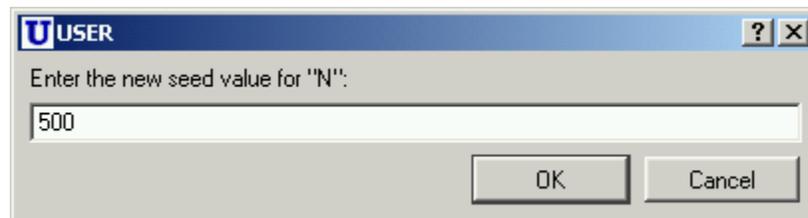
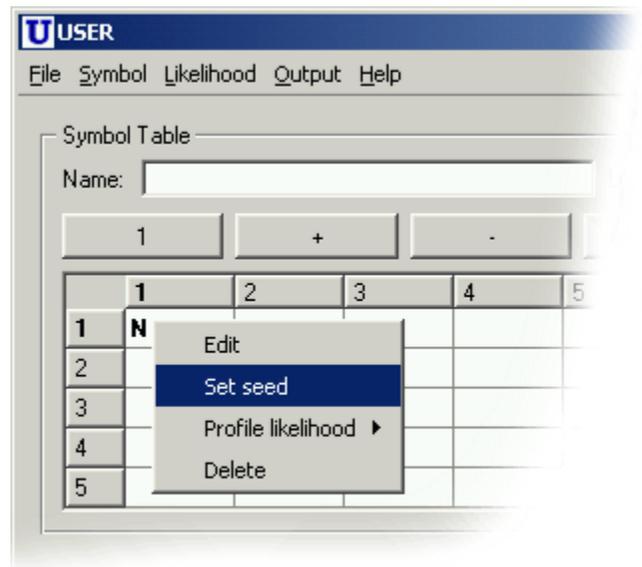
At this point, the model and data have been defined in USER. If the “Compute” button is clicked, the following pop-up appears.



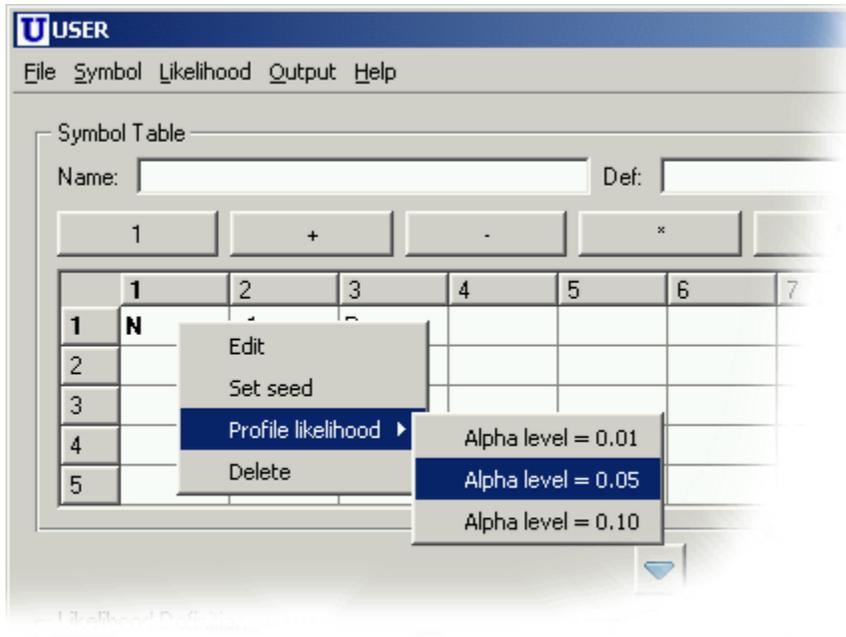
One common cause of this error is the use of inappropriate seeds for the parameters. Since the parameter we are estimating, N , is an estimate of abundance and not a probability, the default seed of 0.5 is probably not appropriate. A seed that is well above the value of n_2 would be more appropriate.

To change the seed, right-click on “N” in the symbol definitions, and set the

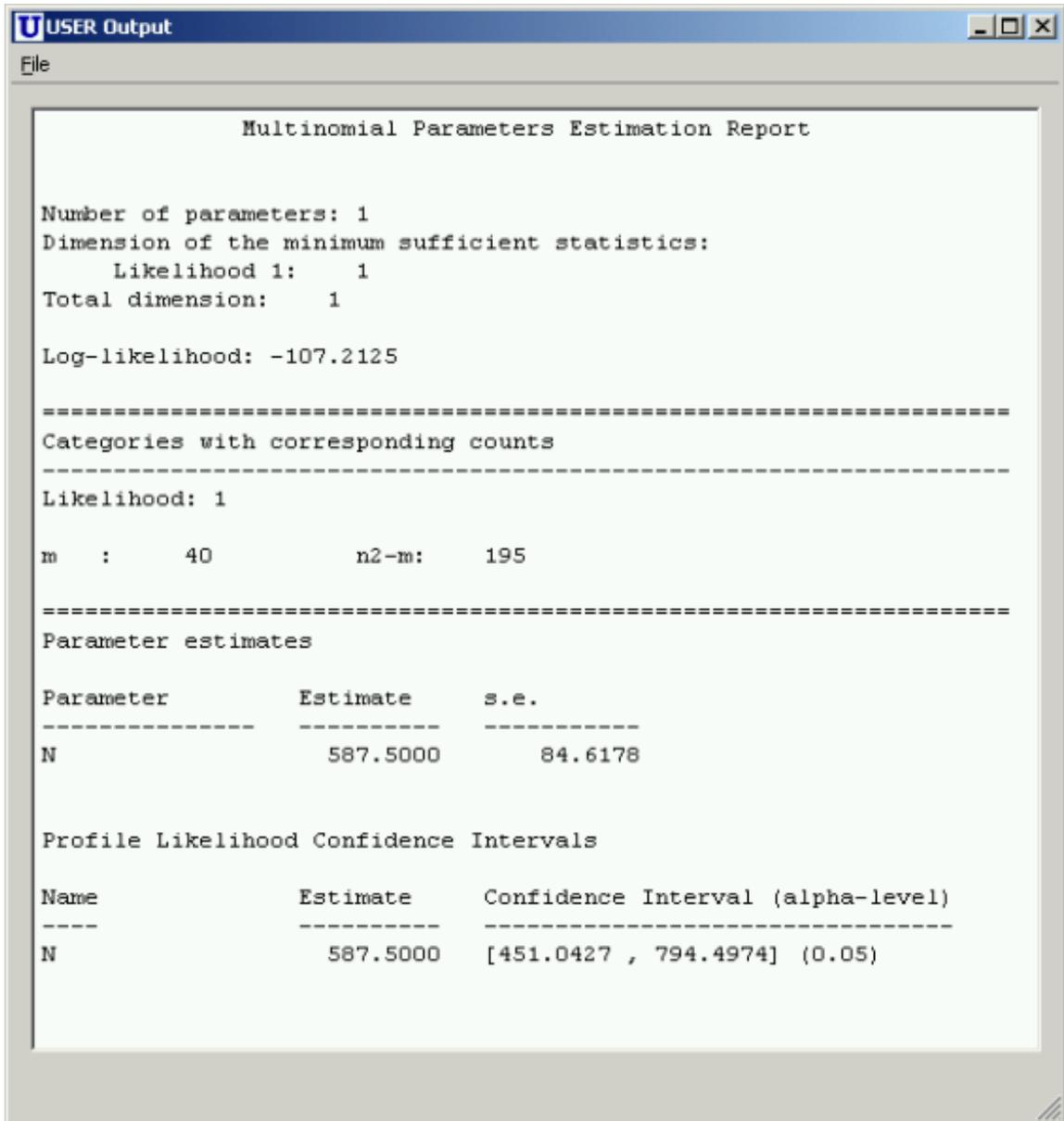
seed to, say, 500.



In addition, we can request a profile likelihood confidence interval for N .



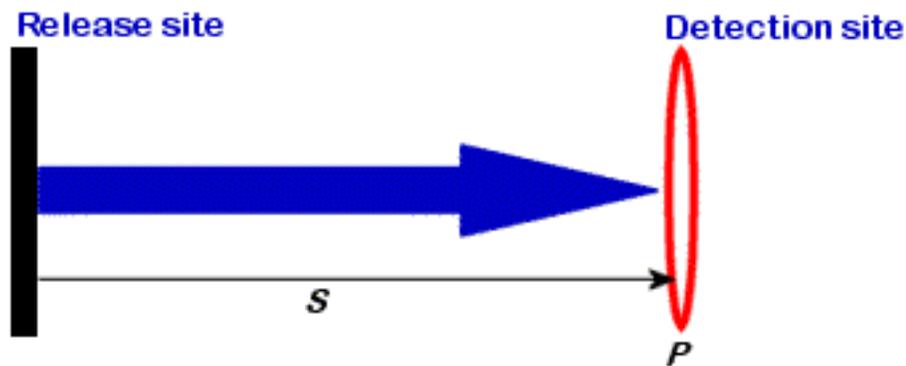
Now if the “Compute” button is pressed, the following output report appears.



4.2 Joint likelihood

4.2.1 Model definition

In a simplified, hypothetical study to estimate the survival of downstream migrating juvenile salmon in a given river reach, radio-tagged salmon are released at a release site and detected at a detection site at the downstream end of the river reach. The parameter of interest is the survival probability S . Not all of the fish are detected at the detection site, so we must also estimate the detection efficiency P .



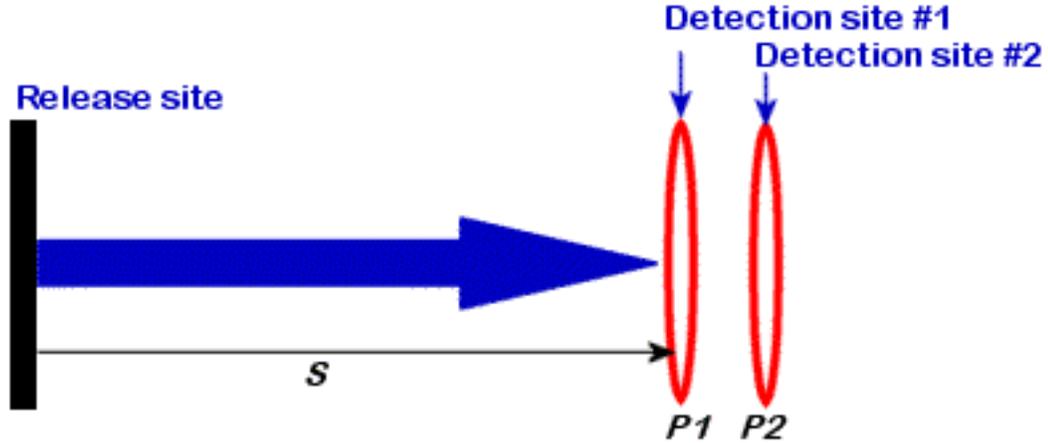
The model can be defined as follows:

<u>Category counts</u>	<u>Description</u>	<u>Probability</u>
N_1	Detected	SP
N_0	Not detected	$1 - SP$

The likelihood is defined as:

$$L_1 \propto (SP)^{N_1}(1 - SP)^{N_0} \quad (4.1)$$

The parameters S and P always occur together and are not separately estimable; if a fish is not detected, there is no way of knowing if it did not survive or simply avoided detection. One solution is to have two independent detection sites downstream, as shown below.



This gives us the ability to estimate the detection probability separately from the survival probability. Instead of one parameter for detection efficiency, we have two: P_1 is the detection efficiency for site #1, and P_2 for site #2. The overall detection efficiency is now defined as: $P = 1 - (1 - P_1)(1 - P_2)$ We can now define an auxiliary likelihood conditional on being detected somewhere at the detection sites.

<u>Category Counts</u>	<u>Description</u>	<u>Probability</u>
N_a	Detected at #1 only	$P_1(1 - P_2)/P$
N_b	Detected at #2 only	$(1 - P_1)P_2/P$
N_{ab}	Detected at both #1 and #2	P_1P_2/P

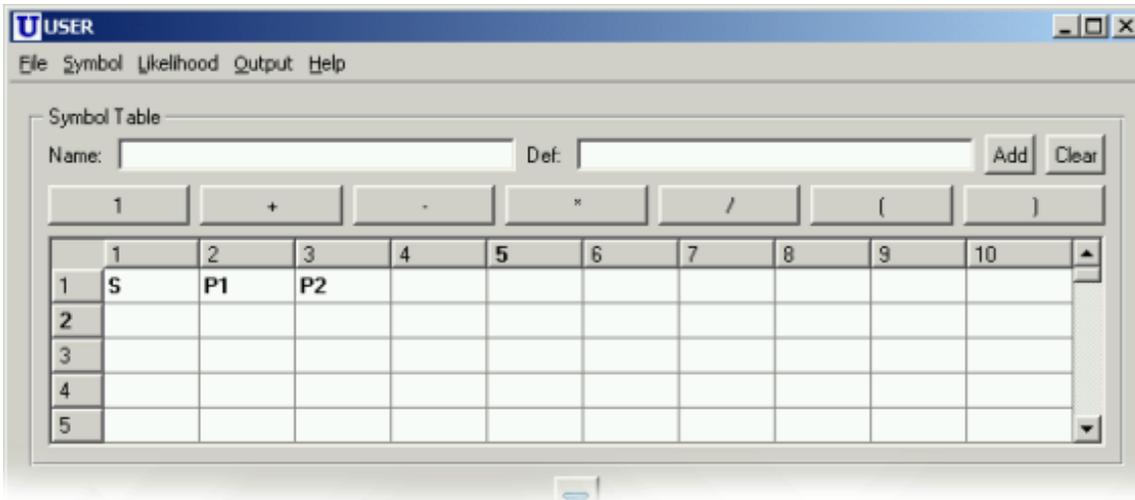
$$L_2 \propto [P_1(1 - P_2)]^{N_a} [(1 - P_1)P_2]^{N_b} (P_1P_2)^{N_{ab}}$$

The full likelihood is $L \propto L_1L_2$

4.2.2 Model definition in USER

Define the parameters

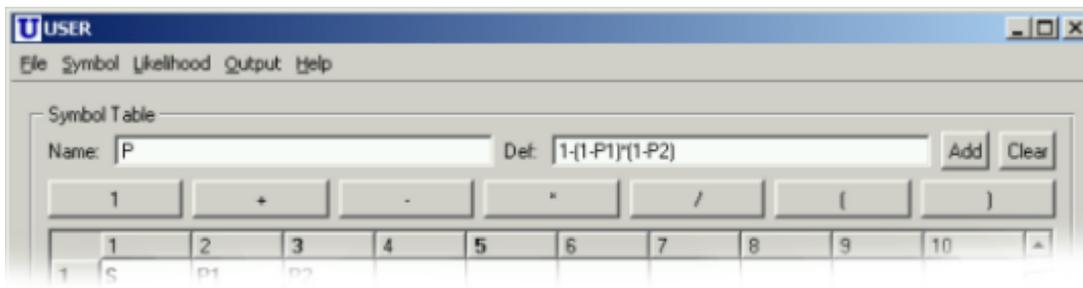
The first step in defining the model in USER is to define the model parameters. The names of the three parameters are entered in the *symbol name dialog* and “Return” key is pressed or the “Add” button is clicked, leaving the *symbol definition dialog* blank. The *symbol definitions area* should now appear as shown below.



Define the variables

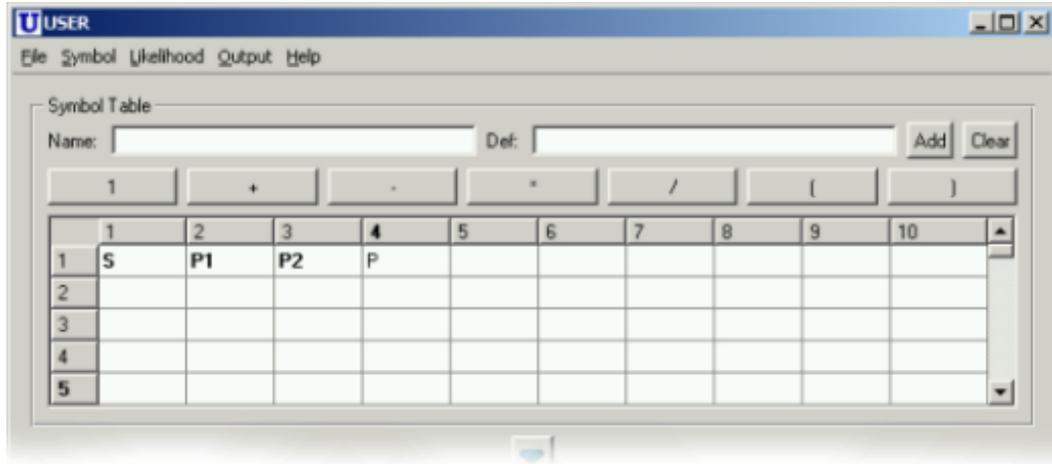
The model definition defines P as a function of the two parameters P_1 and P_2 as the overall detection efficiency. It makes sense, therefore, to define this as a variable. The name “P” is entered in the *symbol name dialog*, and the expression “ $1-(1-P_1)(1-P_2)$ ” is entered in the *symbol definition dialog* as follows:

1. Ensuring that the *entry mode toggle* is NOT depressed, click on the *math entry buttons* labeled “1”.
2. Click on the “-” button.
3. Holding the “Ctrl” key on the keyboard, click on the “P1” parameter name in order to enter “(1-P1)”. Click on the “*” button.
4. Hold the “Ctrl” key and click on the “P2” key.



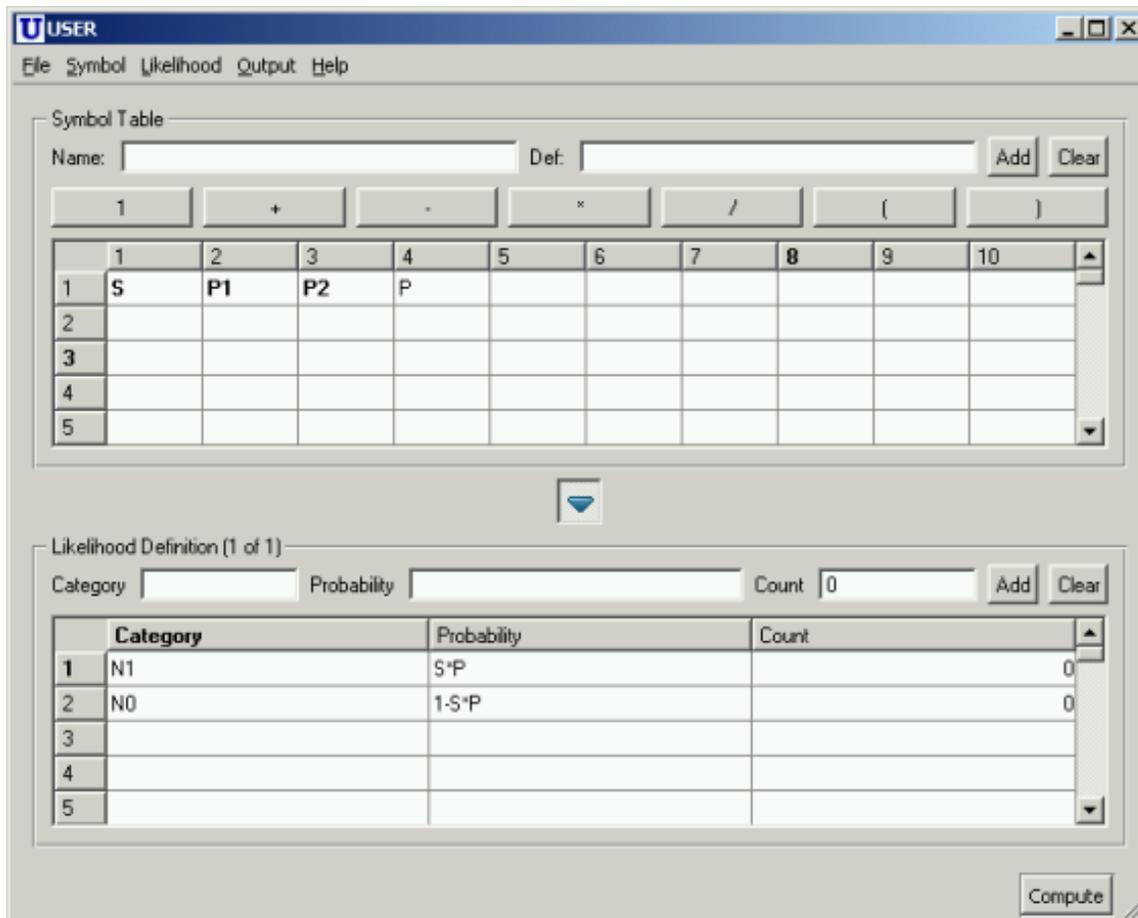
The expression could also simply be typed from the keyboard, or by using any combination of the keyboard and the *math entry buttons*.

The user clicks “Add” or presses “Return” to finish the definition.



Define the first likelihood

The first likelihood (Equation 4.1 above) is defined by entering the category labels in the *category dialog* and the corresponding expression for the probability in the *category probability dialog*. The category probabilities are entered in the same manner as the definition for the symbol P above. If the *math entry buttons* are to be used, the *entry mode toggle* must be depressed so that they affect the *category probability dialog* rather than the *symbol definition dialog*. The complete likelihood appears below. The category counts could be entered at the same time, but in this example we will enter the data later.

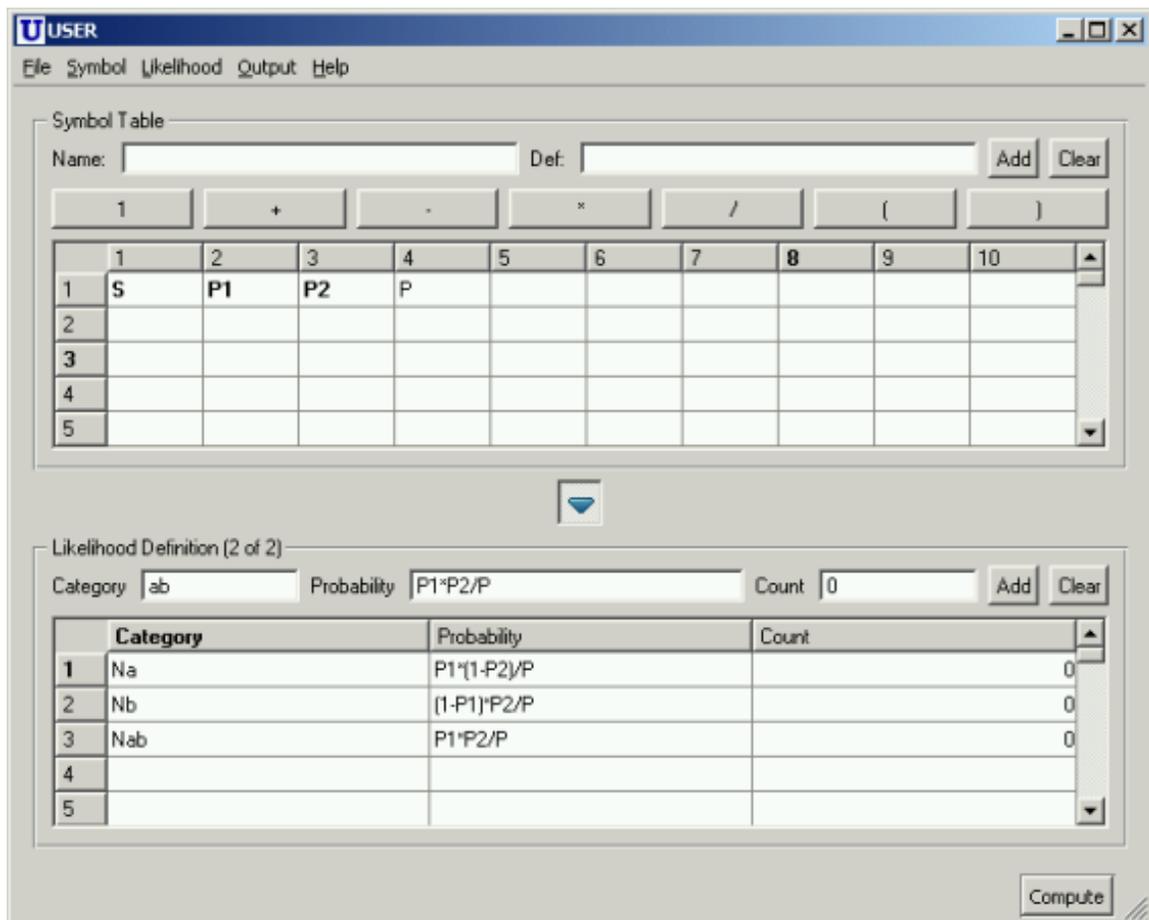


Define the auxiliary likelihood

In order to define the auxiliary likelihood, select “New Likelihood” from the “Likelihood” menu,



and enter the likelihood definition in the same way as before.



Enter the category counts

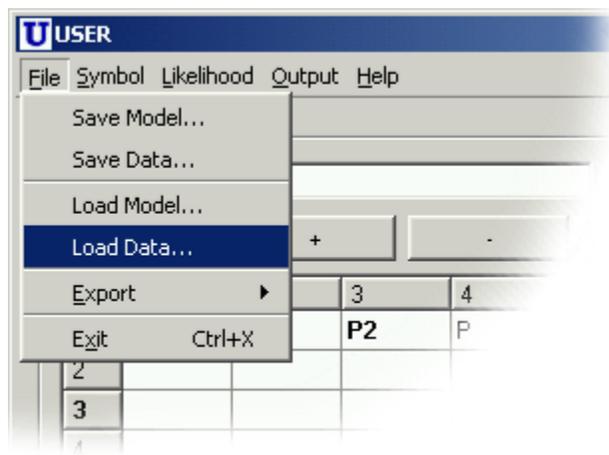
To enter the counts for each category, we could go back and select each category for each likelihood and enter the counts within USER. In this example, however, we shall create a text file containing the category counts and load the data from the text file into USER.

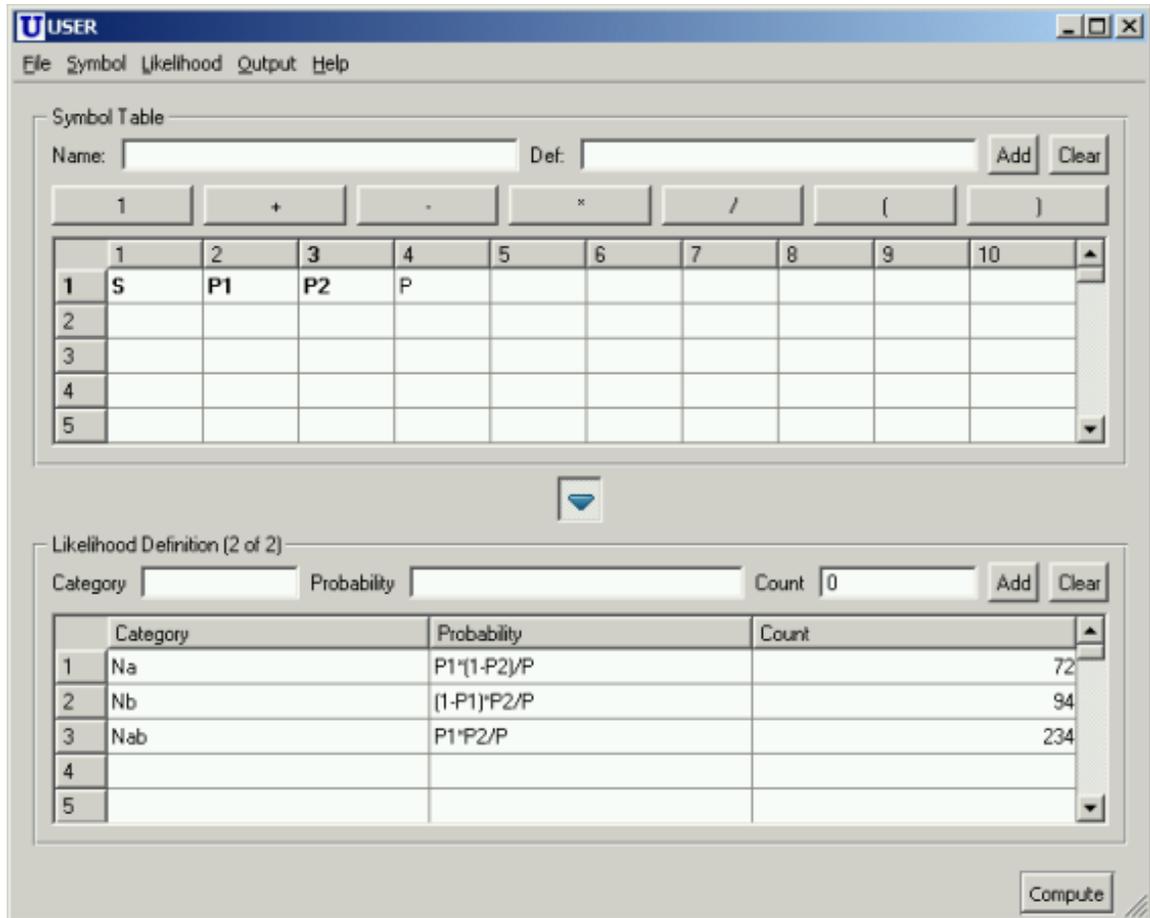
Using made-up data, the text file contains the following:

```
N1 : 400
N0 : 75
Na : 72
Nb : 94
Nab : 234
```

Load the category counts

To load the counts, select “Load Data” from the “File” menu, and select the data file.





Estimate the parameters

To estimate the parameters, simply click on the “Compute” button, and the following window appears.

U Text Window

File

Multinomial Parameters Estimation Report

Number of parameters: 3
 Dimension of the minimum sufficient statistics:
 Likelihood 1: 1
 Likelihood 2: 2
 Total dimension: 3

Log-likelihood: -592.2281

=====
 Categories with corresponding counts
 =====

Likelihood: 1

N1: 400 NO: 75

=====
 Likelihood: 2

Na : 72 Nb : 94 Nab: 234

=====
 Parameter estimates

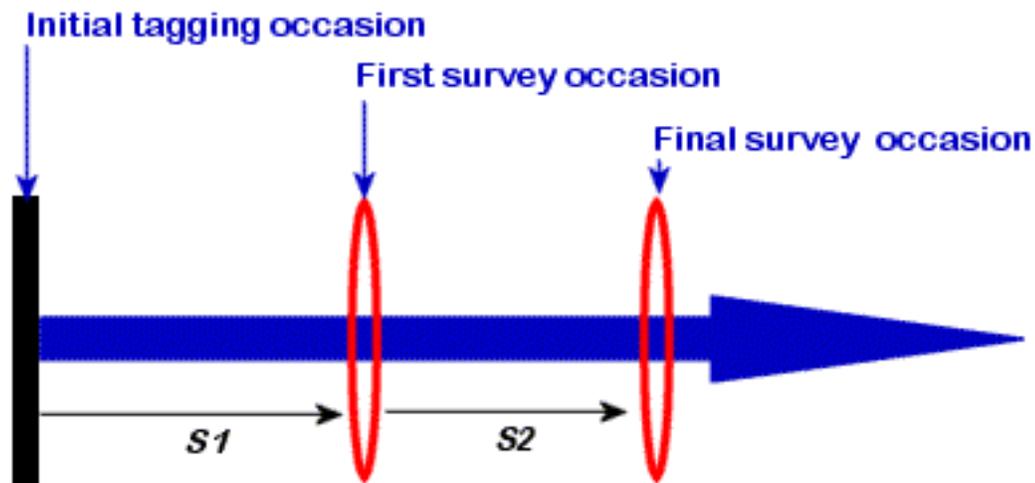
Parameter	Estimate	s.e.
S	0.9030	0.0205
P1	0.7134	0.0250
P2	0.7647	0.0242

4.3 Hypothesis testing

4.3.1 Model definition

USER provides the capability of performing hypothesis testing of an alternative model relative to a null model, as the following example demonstrates.

In a known-fate study, individuals are tagged at the beginning of the study, and the study area is surveyed periodically over regular time intervals. The assumption in a known-fate study is that all live tagged individuals in the study area are detected, so if a previously marked individual is not detected during a survey, it is assumed that the individual has died since the last survey. A simple known-fate study with an initial tagging followed by two subsequent surveys is diagrammed below.



The model parameters are:

<u>Parameter</u>	<u>Description</u>
S_1	The probability of survival from the initial tagging occasion to the first survey occasion
S_2	The conditional probability of survival from the first survey occasion to the second and final survey occasion, given survival up to the first survey occasion

In addition to the estimates and the precision of the estimates of the model parameters, we may also be interested in:

- Whether or not the two survival estimates S_1 and S_2 are significantly different,

- The overall survival probability for the study ($S = S_1S_2$) along with a confidence interval for the overall survival.

The study can be modeled as follows:

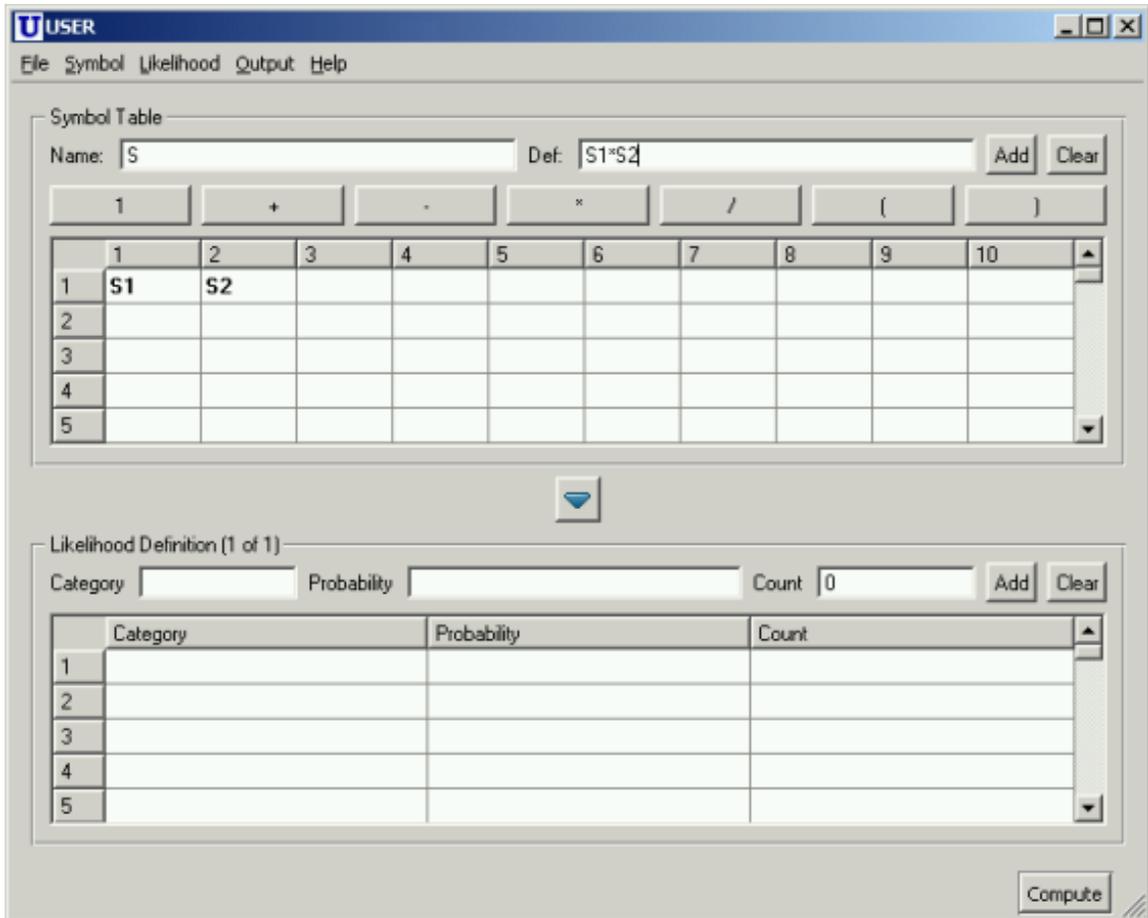
<u>Category</u>	<u>Description</u>	<u>Probability</u>	<u>Count</u>
1 0 0	The individual dies before the first survey	$1 - S_1$	40
1 1 0	The individual is alive at the first survey, but dies before the second	$S_1(1 - S_2)$	75
1 1 1	The individual is alive at the final survey	S_1S_2	223

We can now define the model in the USER program.

4.3.2 Model Definition in USER

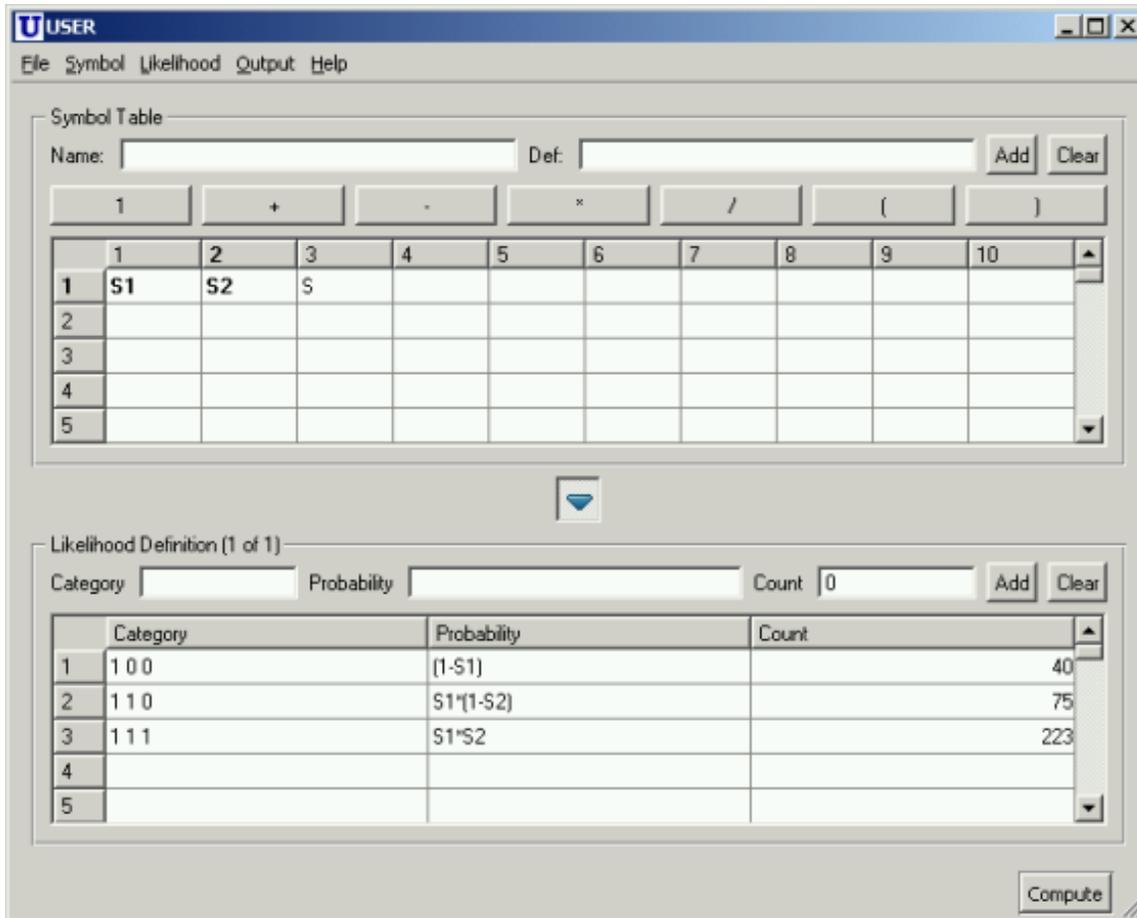
Define the Parameters and Variables

The parameters are defined in the same manner as in the previous examples. We also want to define a variable $S = S_1S_2$ in order to compute a confidence interval around the overall survival.

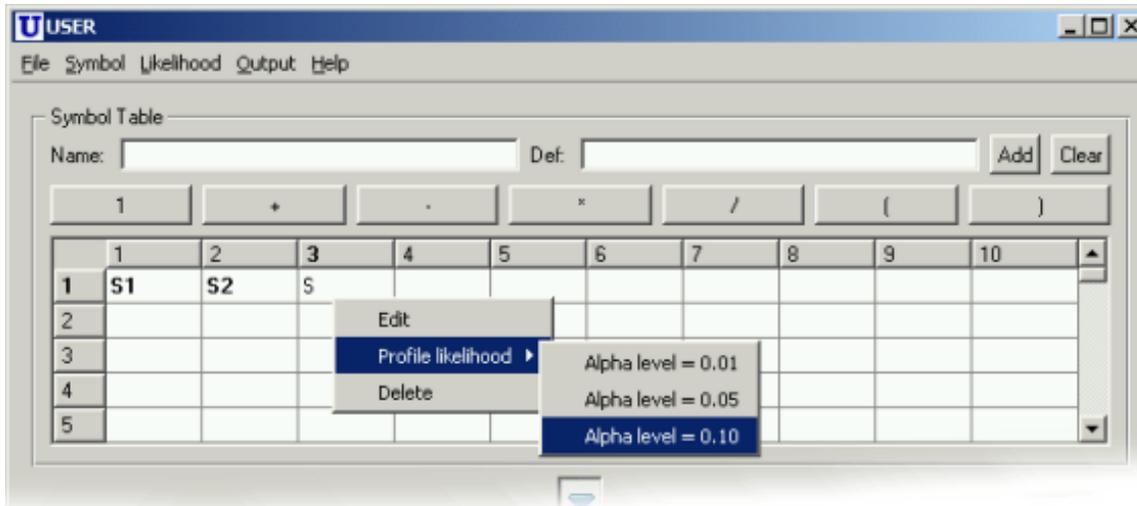


Define the likelihood and enter the category counts

The categories, their probabilities, and the corresponding counts are defined as shown below.

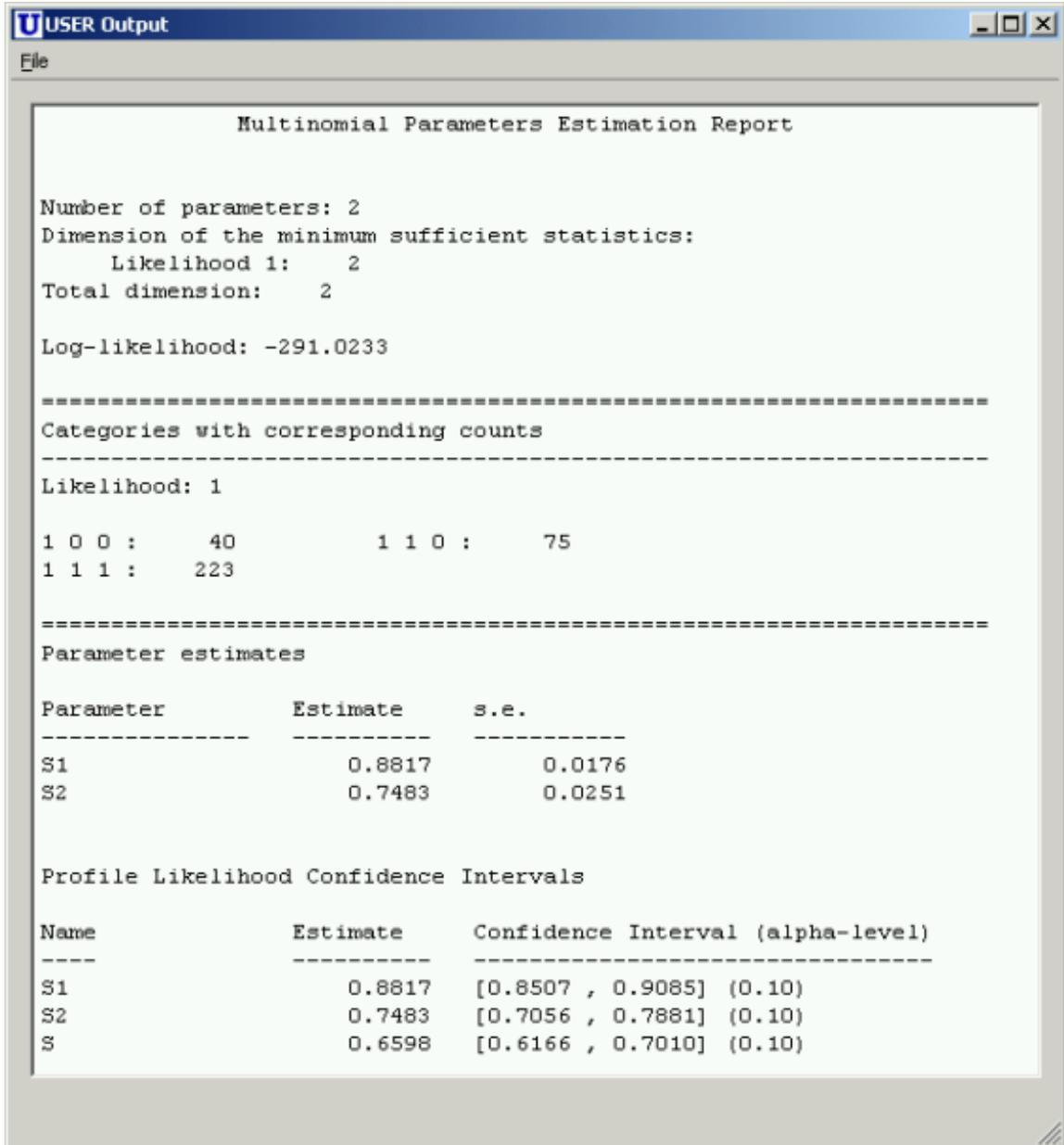


We can now request profile confidence intervals for the individual survival estimates as well as the overall survival probability. We will use an alpha-level of 0.10.



4.3.3 Estimate the Parameters

We now press the “Compute” button to produce the parameter estimates and the profile likelihood confidence intervals.



```
USER Output
File

Multinomial Parameters Estimation Report

Number of parameters: 2
Dimension of the minimum sufficient statistics:
  Likelihood 1: 2
Total dimension: 2

Log-likelihood: -291.0233

=====
Categories with corresponding counts
=====
Likelihood: 1

1 0 0 : 40          1 1 0 : 75
1 1 1 : 223

=====
Parameter estimates

Parameter      Estimate      s.e.
-----
S1             0.8817       0.0176
S2             0.7483       0.0251

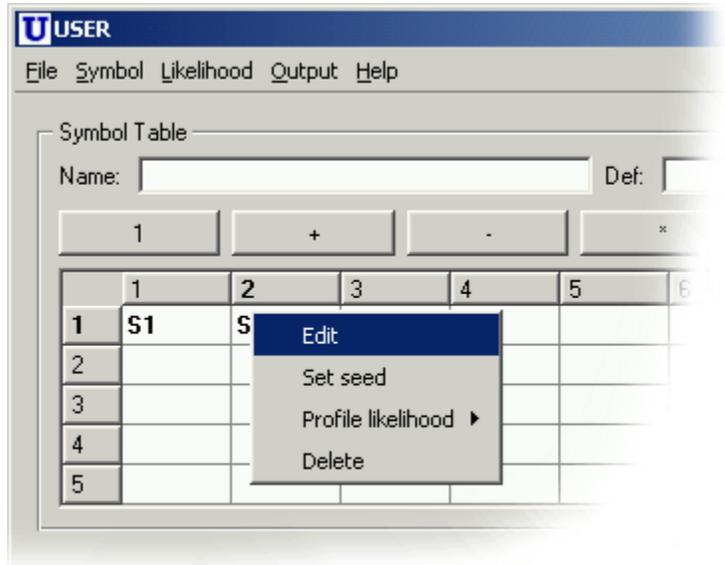
Profile Likelihood Confidence Intervals

Name           Estimate      Confidence Interval (alpha-level)
-----
S1             0.8817       [0.8507 , 0.9085] (0.10)
S2             0.7483       [0.7056 , 0.7881] (0.10)
S              0.6598       [0.6166 , 0.7010] (0.10)
```

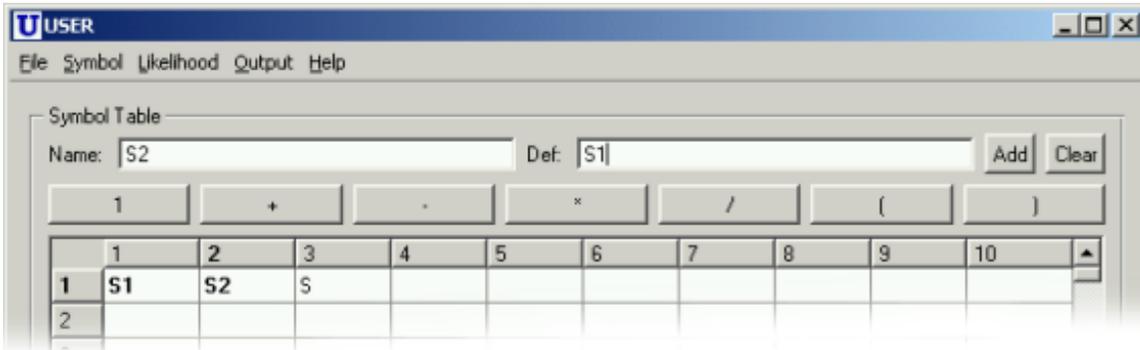
4.3.4 Create the Null Hypothesis Model

In order to test the hypothesis of equal survival probabilities, a model must be created that assumes equal survival probabilities. The easiest way of doing this is to change S_2 from a parameter to a variable, equating it to S_1 . In this way, the actual likelihood definition does not need to be changed.

Right click on “S2” in the symbol definitions, and select “Edit”.

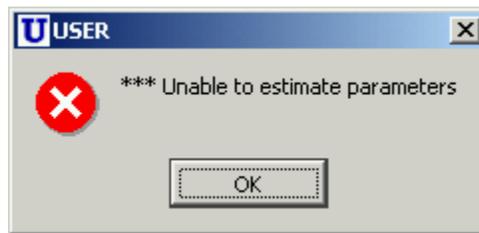


In the symbol definition dialog, enter “S1” via the keyboard or by clicking on it in the *symbol definitions area* and press “Return” or click on the “Add” button.

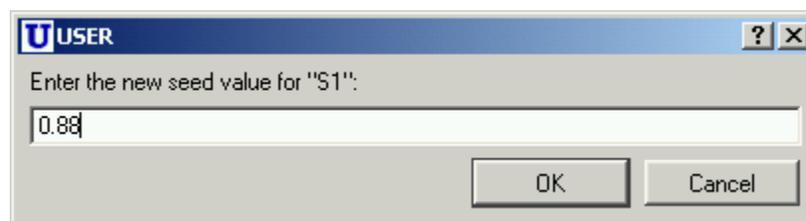
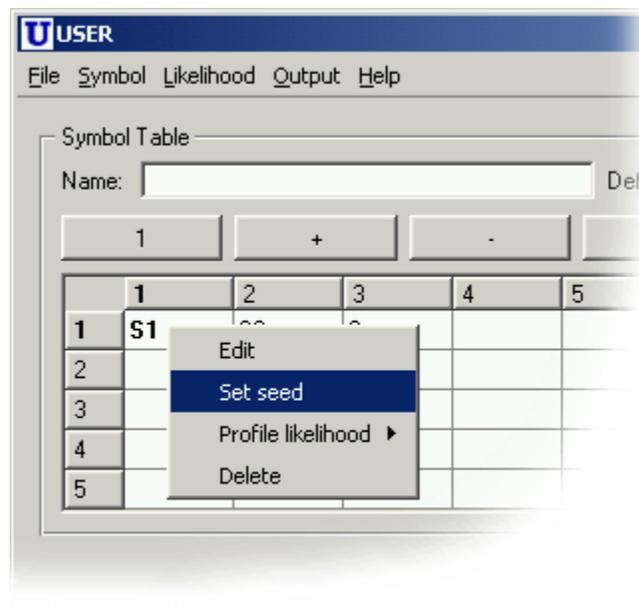


Verify that the profile likelihood confidence intervals are still requested for S_1 and S ; we no longer need a confidence interval for S_2 since it is equated to S_1 . Now

press the “Compute” button. The following error window will be displayed



In the original model, S_1 was estimated at about 0.88, so set the initial seed to that value and try again. Right click on the parameter name “S1”, select “Set Seed”, and type in 0.88.



Try again by pressing the “Compute” button. This time the output window appears. Note that now there is only one parameter.

```

USER Output
File

Multinomial Parameters Estimation Report

Number of parameters: 1
Dimension of the minimum sufficient statistics:
    Likelihood 1:    2
Total dimension:    2

Log-likelihood: -300.5932

=====
Categories with corresponding counts
-----
Likelihood: 1

1 0 0 :    40          1 1 0 :    75
1 1 1 :   223

=====
Parameter estimates

Parameter      Estimate      s.e.
-----
S1              0.8192       0.0153

Profile Likelihood Confidence Intervals

Name           Estimate      Confidence Interval (alpha-level)
-----
S1              0.8192      [0.7932 , 0.8434] (0.10)
S                0.6711      [0.6292 , 0.7113] (0.10)

```

4.3.5 Test the Hypothesis

Note that the output windows show the log-likelihood value for each model, allowing us to use a likelihood ratio test (LRT) to test the hypothesis of equal survival probabilities.

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

The likelihood ratio test statistic is $\Lambda = 2[-291.0232 - (-300.5932)] = 19.1400$. Λ 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.