Lower Confidence Bounds for Seized Material Sampling App Documentation

Jack Prothero and Will Guthrie

1 Introduction

This app provides a set of tools for creating and testing sampling plans for qualitative analysis of seized material. Similar to many other resources for this task, this app uses the *hypergeometric distribution* as the model for probabilities of certain outcomes of analyses of samples from seized material populations. The app provides fine-grain inputs for controlling sample size selection, statistical significance, misidentification control, random number generation for sampling, and diagnostic testing.

The layout of the documentation is as follows. Section 2 provides an overview of the app's available features for sampling for identification. Section 3 provides an overview of the app's available features for estimating unit counts for large seizures.

2 Identification Features

2.1 Input Sidebar

All input fields for the app appear in the left-hand sidebar. Figure 1 shows all of the available inputs when the app is first opened. Changing the selections of many dropdown menus and checkboxes will expose and/or hide certain inputs such that only relevant inputs for the current settings are available to edit. Detailed explanations of the meaning, use, and acceptable values for each input are provided in Section 2.4. Each subsection covers one portion of the input panel. The first chunk, covered in Section 2.4.1, concerns overall information about the seizure and about desired statistical properties of the output of the app. The second chunk, covered in Section 2.4.2, concerns choosing the size of the sample to take from the population. The third chunk, covered in Section 2.4.3, concerns accounting for potential misidentification errors in the analysis. The fourth chunk, covered in Sections 2.4.4 and 2.4.5, concerns the diagnostic plot as well as implementation tools for using the calculations from the app in sampling protocols.

2.2 Lower Confidence Bound Table

Figure 2 shows an example of table of lower confidence bounds generated by the app. The output in this particular example corresponds to the input settings shown in Figure 1.

At the top of the table, above the column labels, the currently selected or recommended sample size is provided based on app inputs. This sample size can either be input directly or calculated in order to guarantee a certain amount of information about the population if enough units in the sample are positive. In the example table in Figure 2, this sample size was chosen as 9 units out of the 150 in the population. This calculated sample size depends on all of the inputs in the first two chunks of the input panel in Figure 1. See Sections 2.4.1 and 2.4.2 for details on each of these input fields.

The first column of the table, labeled "Observed Positives," will always range from the recommended/chosen sample size down to 0. Each row of the table corresponds to the outcome of observing that many positives in the sample.

The second column of the table, labeled "Lower Confidence Bounds," shows the minimum number of units that likely contain illicit material at the input statistical significance level for each sample analysis outcome. The values in this column are the statistical results that would be reported after observing the corresponding first-column value in the analysis. These values are calculated using an implementation of Equation (16) from [Wang, 2015]. Validation calculations have shown this implementation to be equivalent to that in other commonly used tools.

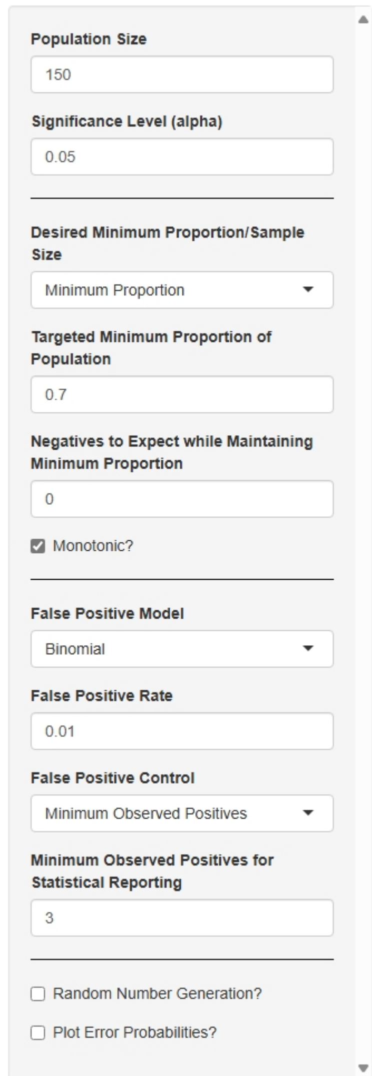

Figure 1: Input panel for identification sampling mode.

In the rows for 2 and 1 observed positives in the table in Figure 2, the lower confidence bounds have been marked with an asterisk (*). This asterisk is an indication that the lower confidence bound is being reduced by the false positive control protocol in an attempt to maintain the input statistical significance level despite the chance for misidentifications in the qualitative analysis. If any lower confidence bounds are altered by false positive control, the note below the table in Figure 2 appears as an extra indication. For more information on the false positive modeling options available in the app, see Section 2.4.3.

The third column of the table, labeled "Inferred Proportion," shows the lower confidence bounds as a proportion of the population size. These values contain the same statistical information as the values in the second column, just reported in terms of proportions rather than absolute numbers of units. Since the population size for this table is 150 (see Figure 1), the entries in the third column are the entries in the second column divided by 150.

Recommended Sample Size: 9.

Observed.Positives	Lower.Confidence.Bounds	Inferred.Proportion
9	109	0.727
8	87	0.58
7	69	0.46
6	53	0.353
5	39	0.26
4	27	0.18
3	16	0.107
2	0*	0
1	0*	0
0	0	0

Lower Confidence Bounds marked with an asterisk (*) have been adjusted to 0 in accordance with the selected false positive control scheme. While the observed number of positives should still be reported, no statistical statement should be included.

Figure 2: Example of table of lower confidence bounds. Output matches input settings from Figure 1.

2.3 Achieved Confidence Plot

The achieved confidence plot is a tool to examine how a particular sampling plan might perform in practice. The user inputs a statistical significance level to strive for, but between the discreteness of the hypergeometric distribution and the potential for misidentification the achieved significance may be higher or lower than the nominal significance. A third factor that influences performance in practice is the kinds of seizures typically analyzed by the lab. If many seizures with little or no illicit material are analyzed in casework, that may suggest a different sort of approach than if most analyzed seizures are all or mostly illicit material. Therefore, evaluating the achieved confidence in practice also involves

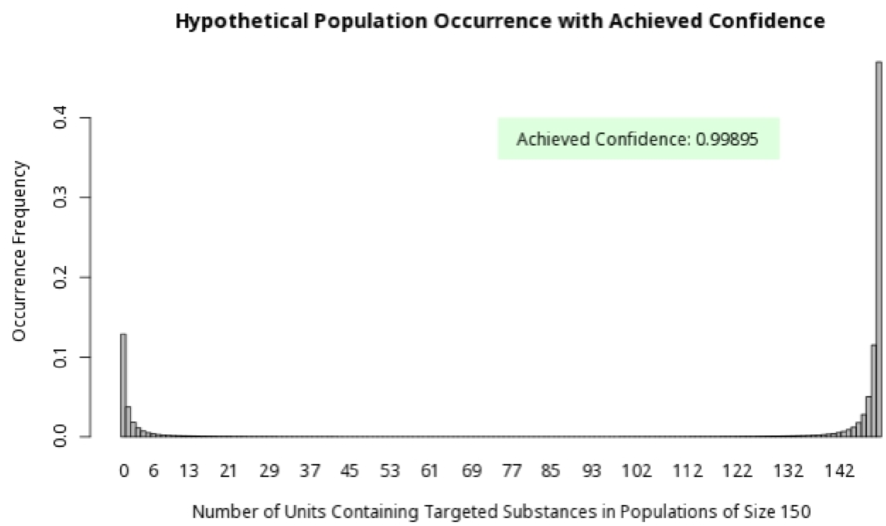

Figure 3: Example of achieved confidence plot. This plot can be reproduced by checking the "Plot Error Probabilities" box with other settings preserved from Figure 1, setting the First Shape Parameter slider to 0.35, and the Second Shape Parameter to 0.32.

inputting an estimate of the kinds of seizures typically seen in the lab. Detailed explanations of how to input that information for the purposes of this plot are provided in Section 2.4.4.

Figure 3 shows an example of this plot using settings from the input panel in Figure 1 and checking the "Plot Error Probabilities" box. The horizontal axis represents the number of positive units in a hypothetical population, and the vertical axis represents the relative frequency of each kind of population. The height of each bar corresponds to the relative frequency of that bar's associated population type based on the input expectations about the kinds of seizures typically seen in the lab. In this example, those expectations are set at high frequencies all-illicit or mostly-illicit seizures with some small percentage of no-illicit or barely-illicit seizures. Once the expected seizure types are set, the plot calculates the achieved confidence level and displays it in a colored box on the plot. If the box is green, the achieved confidence is at least as high as the input statistical significance level would suggest. If the box is red, the achieved confidence level is below the input statistical significance level. As a remark, if the false positive rate is 0, the achieved confidence will always be at least as high as the input statistical significance level.

2.4 Inputs

2.4.1 Population Parameters

- **Population Size:** This field requires a positive integer. It represents the number of superficially-indistinguishable units in the portion of the seizure under current study. For this parameter, any input other than a positive integer (0, negative values, fractional/decimal values, letters, punctuation) will generate an error in the app. Any positive integer value is acceptable in this field, but inputs larger than 10000 may result in noticeably longer computation times, especially if the "Minimum Proportion" sample size determination is used and especially if the "Monotonic?" box is checked in that mode.

- **Significance Level** (α)**:** This field requires a decimal value strictly between 0 and 1. It represents the level of statistical significance associated with the calculated lower bounds. More formally, the significance level, or α, is the assumed probability of overstating the contents of a seizure using this procedure. Smaller significance levels will result in larger recommended sample sizes in "Minimum Proportion" mode and smaller lower confidence bounds in "Sample Size" mode. The value $1 - \alpha$ (times 100%) is referred to as the *confidence level* and is also often of interest. A significance level of $\alpha = 0.05$ therefore implies a confidence level of 0.95 (95%). For this parameter, any numerical input ≤ 0 or ≥ 1, as well as any non-numeric input, will generate an error in the app.

2.4.2 Sampling Threshold Parameters

The tables of lower confidence bounds can be calculated under two paradigms. The first ("Minimum Proportion" option in the Desired Minimum Proportion/Sample Size dropdown) follows the procedure and philosophy described in [Schiavone et al., 2011] which focuses on choosing sample sizes in order to attempt to positively identify a certain proportion of the seizure as illicit. The second ("Sample Size" option in the Desired Minimum Proportion/Sample Size dropdown) allows the user to directly input their own sample size.

Minimum Proportion

- **Targeted Minimum Proportion of Population:** This field requires a decimal value between 0 and 1 (inclusive). It represents the proportion of the population the analyst wishes to be able to claim contains drugs with statistical significance. The app will choose the smallest sample size that achieves this goal under the constraints of the **Negatives to Expect while Maintaining Minimum Proportion** and **Monotonic?** inputs. Larger inputs for this parameter will lead to larger recommended sample sizes. The rows in the resulting table of lower confidence bounds will range from 0 to the calculated sample size. For this parameter, any numerical input less than 0 or greater than 1, as well as any non-numeric input, will generate an error in the app

- **Negatives to Expect while Maintaining Minimum Proportion:** This field requires a non-negative integer value between 0 and the population size. It represents the number of negatives in the sample under which the **Targeted Minimum Proportion of Population** should still be achieved. For example, if this input is set to 1, the first two rows of the resulting table of lower confidence bounds will describe outcomes where the minimum proportion is achieved, since those two rows correspond to samples that are either all positive or one negative and the rest positive. Larger inputs for this parameter will lead to larger recommended sample sizes. For this parameter, any integer less than 0 or greater than the population size, any fractional/decimal value, and any non-numeric input will generate an error in the app. Inputs to this parameter that exceed $(1-MinimumProportion)*PopulationSize$ will not produce a table since in these cases the desired minimum proportion cannot be guaranteed even if the entire population is analyzed. As a remark, the analyst/laboratory must determine which analytical results constitute "positive" and "negative" for a given seizure according to their own standard operating procedures and/or professional judgment.

- **Monotonic?:** This field is a yes-or-no option for whether recommended sample sizes should be adjusted to ensure that they always either increase or stay the same as population sizes increase. Due to the discrete nature of the hypergeometric distribution, for some population sizes the effective statistical significance level for the lower confidence bounds is slightly more stringent than the input significance level α. This leads to situations where the recommended sample size for the larger of two consecutive population sizes might actually be smaller than for the smaller of the two. For example, at $\alpha = 0.05$, $MinimumProportion = 0.5$, and $NegativesToExpect = 0$, the recommended sample size for a population of 49 is 5, while for a population of 50 it is 4. If this box is checked, this phenomenon will be smoothed over and the recommended sample size will always be at minimum the recommended sample size of the previous consecutive population size. Recommended sample sizes will be larger overall if this box is checked, but may also be more easily justifiable in reports/procedures/testimonies.

Sample Size

- **Sample Size:** This field requires a non-negative integer between 0 and the population size. It represents the sample size chosen by the analyst for this seizure. The rows in the resulting table of lower confidence bounds will range from 0 to the input sample size value. For this parameter, any integer less than 0 or greater than the population size, any fractional/decimal value, and any non-numeric input will generate an error in the app.

2.4.3 False Positive Modeling

Many documents and recommendations for seized drug analysis do not incorporate considerations for a substance, either controlled or uncontrolled, being misidentified as the substance of interest. If so-called "false positives" occur at a high enough rate in the analysis, the stated level of statistical significance for the lower confidence bounds calculated in the app may not be achieved in practice. This collection of parameters enables modeling of this possibility.

- **False Positive Rate:** This field requires a decimal value between 0 and 1 (inclusive). It represents the probability of any individual analyzed unit without the substance of interest being misidentified as containing the substance of interest. Without employing any correction protocols (see below) to the lower confidence bounds, higher false positive rates will generally result in lower achieved confidence levels. For this parameter, any numerical input less than 0 or greater than 1, as well as any non-numeric input, will generate an error in the app.

- **False Positive Model:** This dropdown provides the user with a selection of probability distributions used to model the incidence of false positives in the analysis. All of them use the false positive rate and chosen/recommended sample size as inputs. Any additional required inputs for a given distribution are revealed when that distribution is chosen.

  - **Binomial:** This distribution assumes that each true negative unit's chance of being misidentified as a positive is independent of the other units. It requires no additional inputs.

  - **Correlated Binomial:** This distribution assumes that each true negative unit's chance of being misidentified as a positive is not independent of the other units. This option may be more realistic than the typical Binomial distribution as if a cross-contamination has occurred in one unit, it's likely other units in the sample and/or population have also been affected. This distribution requires one additional input of **Correlation** specifying how interrelated the chances of misidentification are for each unit in the sample. This field requires a decimal value strictly between -1 and 1. The range of acceptable values is further restricted depending on the currently-specified sample size and false positive rate. The error message associated with this field specifies the exact range for this input depending on the state of other app inputs. Choosing 0 correlation duplicates the functionality of the typical Binomial distribution. Positive values for this input inflate the chances of very low and very high numbers of false positives relative to typical Binomial. Negative values for this input inflate the chances for medium numbers of false positives relative to the typical Binomial distribution. Without employing any correction protocols (see below), any values other than 0 for this input will generally result in lower achieved confidence levels. The app uses the implementation of the correlated binomial model found in the `fitODBOD` R package [Mahendran and Wijekoon, 2023].

- **False Positive Control:** This dropdown provides the user with a selection of protocols through which the lower confidence bounds should be altered in an effort to preserve the reported level of statistical significance in the face of potential false positives. Values in the lower confidence bounds table that are altered by a false positive control protocol are marked with an asterisk (*).
  - **None:** There is no alteration to any lower confidence bounds in this protocol.
  - **Minimum Observed Positives:** Statistical lower confidence bounds are reported only in cases where enough positives are observed in the sample in this protocol. It uses one input, **Minimum Observed Positives for Statistical Reporting**, which requires a non-negative integer between 0 and the sample size. Lower confidence bounds are reduced from the statistical value to 0 for rows in the table associated with fewer positive units in the sample than this input. For example, if the sample size is 5 and **Minimum Observed Positives for Statistical Reporting** is set to 3, the lower confidence bounds for 1 and 2 observed positives are set to 0, while the lower confidence bounds for 3, 4, and 5 observed positives are left alone. Larger inputs for this parameter will generally result in lower error rates, as requiring the observation of more positives in order to report statistical lower confidence bounds is more conservative. As a remark, this protocol does not depend on any other input parameters, and will in fact still alter the lower confidence bound table even if the false positive rate is set at 0.
  - **Adaptive:** Statistical lower confidence bounds are reported only in cases where the probability that the observed positives are false positives is sufficiently small in this protocol. It uses one input, **Maximum Error Probability for a Null Population**, which requires a decimal value between 0 and 1 (inclusive). Lower confidence bounds are reduced to 0 for rows in the table that are deemed too likely to be observed via misidentification. The tolerable probability of observing a positive via misidentification is what is controlled with the **Maximum Error Probability for a Null Population** input. Notably the the alterations to the table induced by this input will also depend on the sample size, the false positive rate, and false positive model (and its associated parameters, if any). As an example, with a sample size of 5, a false positive rate of 0.1, a Binomial false positive model, and a **Maximum Error Probability for a Null Population** of 0.05, the lower confidence bounds for 1 and 2 observed positives are set to 0, while the lower confidence bounds for 3, 4, and 5 observed positives are left alone. Smaller inputs for this parameter

will generally result in higher achieved confidence levels, as a smaller maximum probability lowers the threshold for an analysis outcome being considered as potentially all-false-positives.

2.4.4 Error Rate Diagnostic Graphics Parameters

The plot that evaluates achieved confidence for the current lower confidence bound table can be shown by checking the **Plot Error Probabilities?** box. As discussed above, the achieved confidence depends on both the table itself and the prior expectation of what kinds of seized populations typically enter the lab for analysis. The app provides two methods to express this prior expectation, which can be chosen using the **Population Occurrence Selection Mode** dropdown.

- **Smooth Curve:** This method controls the overall shape of the expected seized population distribution with two sliders. Roughly, the first slider controls the frequency of populations of little or no controlled material, and the second slider controls the frequency of populations with mostly or all controlled material. Setting the sliders to the left will increase the frequency of their corresponding populations, while setting them to the right will decrease the frequency. Specifically, the two sliders correspond to the two shape parameters of a Beta distribution and have a range from 0.2 to 2. This Beta distribution is then discretized and transformed to cover the integers from 0 to the population size. Choosing 1 for both parameters results in equal frequency for all types of populations. Choosing 0.2 for both parameters results in equal high frequencies for no-positive-unit and all-positive-unit populations and vanishingly low frequencies for all others. Slightly increasing one of the parameters above 0.2 results in favoring either no-positive-unit or all-positive-unit populations.

- **Table Specification:** This method allows for direct input of frequencies for individual population types. Clicking the "Add Row" button produces two new inputs: the left-hand box requires a non-negative integer as input to represent the number of positives in a population type of interest, and the right-hand box requires a decimal between 0 and 1 (inclusive) to represent the probability of the population type occurring in casework. If the total probability of all input population types is less than 1, the remaining probability is distributed evenly across all other population types not specified in the inputs. The total probability of all specified population types cannot exceed 1.

2.4.5 Sampling Plan Implementation

Random Number Generation Once a sample size is selected, the units for analysis must then be appropriately randomly sampled from the population. To execute a simple random sample where each unit is equally likely to be chosen for analysis, it is often best to label each unit and then use an external tool to choose which labels to sample. Direct choice of units by the analyst risks introducing unintentional bias in the kinds of units that are analyzed. For this purpose, the app can generate a list of random integers with length equal to the sample size by clicking the **Random Number Generation** check box. The **Number of Populations** parameter requires a positive integer and controls the number of lists of sample indices generated. Integers may be repeated across sub-population lists but will never be repeated within a list.

3 Unit Count Features

3.1 Input Sidebar

All input fields for the app appear in the left-hand sidebar. Figure 4 shows all of the available inputs when the "Counting" tab is selected. Changing the selections of checkboxes will expose and/or hide certain inputs such that only relevant inputs for the current settings are available to edit. Detailed explanations of the meaning, use, and acceptable values for each input are provided in Section 3.4.

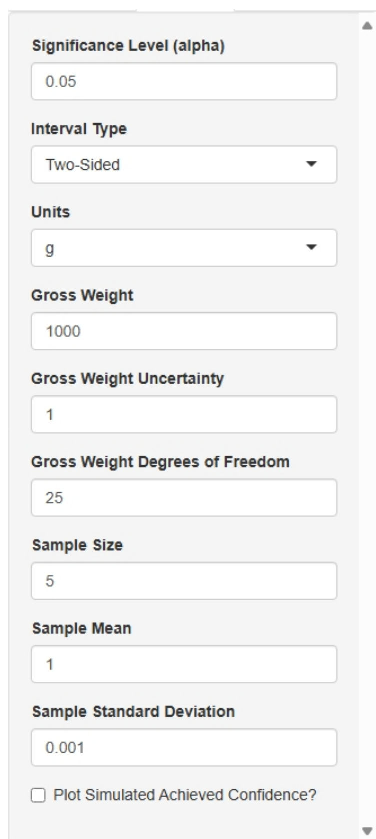

Figure 4: Input panel for unit count mode.

3.2 Results Readout

Figure 5 shows an example readout of estimated unit count and associated expanded uncertainty intervals. The output in this particular example corresponds to the input settings shown in Figure 4. The first line provides the unrounded best estimate of the number of units based on the weights provided as input. The second line provides a confidence interval at the desired significance level with endpoints rounded conservatively (i.e. in a way that results in a larger interval) to the nearest whole number. The third line provides a confidence interval with unrounded endpoints and the corresponding margin of error. The fourth line provides additional information about the propagation of uncertainty combining the variability observed in the sample and the uncertainty of measurement associated with the weighings.

Estimated unit count: 1000
95% Rounded Interval: (997, 1003)
95% Interval: (997.758887169227, 1002.24111283077) , Margin of Error: 2.24111283077277
Standard Uncertainty: 1.09544511501033 ; Degrees of Freedom: 28.8

Figure 5: Example of estimated unit count readout. Output matches input settings from Figure 4.

3.3 Achieved Confidence Plot

The achieved confidence plot is a tool to examine how the uncertainty propagation performed with current app inputs. You can reveal the achieved confidence plot for unit count estimation by checking the **Plot Simulated Achieved Confidence?** checkbox. The statistics are relatively straightforward in this situation so the chosen confidence level will likely be achieved comfortably except in extreme cases where very few samples are taken and the degrees of freedom associated with uncertainty of measurement are also quite low. The app takes the provided inputs and "resimulates" the weighing experiment 1000 times as if the provided inputs were the true values for the seizure. In each resimulation the confidence interval either succeeds or fails to include the "true" value, and we expect the percentage of simulations that fail to include the true value to match the input significance level. The plot displayed (an example of which can be found in Figure 6) shows the first 50 of these resimulations with intervals that include the expected value colored in blue and intervals that fail to include the expected value colored in orange. The proportion of successful simulations is included in the plot's title. While the plot is revealed, a **Resimulate** button is available that when clicked will redo the whole resimulation experiment with 1000 new replications and update the plot accordingly. A helpful diagnostic in this situation is to check that the achieved confidence in simulation seems to "hover" around the desired confidence across several runs of the resimulation. The achieved confidence in simulation values should seem equally likely to fall either above or below the desired confidence.

3.4 Inputs

- **Significance Level** (α)**:** This field requires a decimal value strictly between 0 and 1. It represents the level of statistical significance associated with the calculated lower bounds. The value $1-\alpha$ (times 100%) is referred to as the *confidence level* and is also often of interest. A significance level of $\alpha = 0.05$ therefore implies a confidence level of 0.95 (95%). For this parameter, any numerical input ≤ 0 or ≥ 1, as well as any non-numeric input, will generate an error in the app.

- **Interval Type:** This dropdown provides the user with a selection of the types of confidence intervals generated in the readout. **Two-Sided** produces an interval with both a lower bound and an upper bound. **One-**

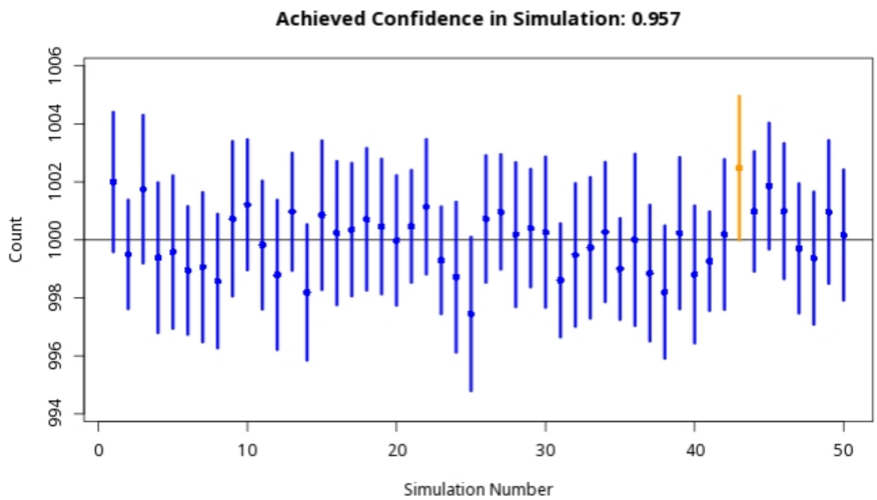

Figure 6: Example of achieved confidence plot. Output matches input settings from Figure 4

Sided produces an interval with only a lower bound. The lower bound for the one-sided interval will be larger than the lower bound for the two-sided interval. Both options are equally valid statistically, and a lab should choose which interval is more fit for purpose for their casework. If as specific estimate of the of the unit count is of primary interest for the exhibit, a two-sided interval may be appropriate as a representation of the range of plausible values for the total unit count given the weights measured. If the unit count is estimated primarily for the purpose of informing statistical sampling, a one-sided lower bound may be more efficient for that use case.

- **Units:** This dropdown selects the units each measurement were taken in with options for kg, g, and mg. This input does not affect any outputs. It is provided as an avenue for documentation and as a reminder that the gross weight, sample mean weight, and all associated uncertainties are expected to be provided in the same units.
- **Gross Weight:** This field requires a positive real number. It represents the total measured weight of all of the units, likely on a balance meant for large weights. This weight may include the weight of capsules or small bags containing powder, but should exclude the weight of any seizure-size containers like large bags or storage bins holding the tablets/capsules/etc.
- **Gross Weight Uncertainty:** This field requires a positive real number. It represents the uncertainty associated with the weighing on the balance used for the gross weight measurement. This uncertainty likely comes from an intra-laboratory study or protocol.

- **Gross Weight Degrees of Freedom:** This field requires a positive real number. It represents the degrees of freedom associated with the uncertainty of measurement provided in the previous field. This value likely comes from an intra-laboratory study or protocol regarding the uncertainty of measurement for the balance used to measure the gross weight. One possible way to choose this value is to count how many observations are used in the calculation of the gross weight uncertainty in the laboratory's existing protocol, then subtract 1 from that value. A value derived from the number of measurements of the gross weight of the units in an exhibit should not be used for this input in most cases.

- **Sample Size:** This field requires a positive integer. It represents the number of units sampled from the seizure for the purpose of estimating unit count.

- **Sample Mean:** This field requires a positive real number. It represents the mean of the measured weights of each unit sampled for the purpose of estimating unit count. The units weighed to determine this value should be in the same state as the population of units were for the gross weight determination. For example, if a seizure of capsules is being studied and the capsule casings were included in the gross weight, the capsule casings should also be included for the sample weights.

- **Sample Standard Deviation:** This field requires a positive real number. It represents the standard deviation of the measured weights of each unit sampled for the purpose of estimating unit count. The units weighed to determine this value should be in the same state as the population of units were for the gross weight determination. For example, if a seizure of capsules is being studied and the capsule casings were included in the gross weight, the capsule casings should also be included for the sample weights.

References

[Mahendran and Wijekoon, 2023] Mahendran, A. and Wijekoon, P. (2023). *fitODBOD: Modeling Over Dispersed Binomial Outcome Data Using BMD and ABD*. R package version 1.5.0.

[Schiavone et al., 2011] Schiavone, S., Perrin, M., Coyle, H., Huizer, H., Bolck, A., and Cardinetti, B. (2011). Guidelines on representative drug sampling. Technical Report ST/NAR/38, United Nations Office on Drugs and Crime.

[Wang, 2015] Wang, W. (2015). Exact optimal confidence intervals for hypergeometric parameters. *Journal of the American Statistical Association*, 110(512):1491–1499.