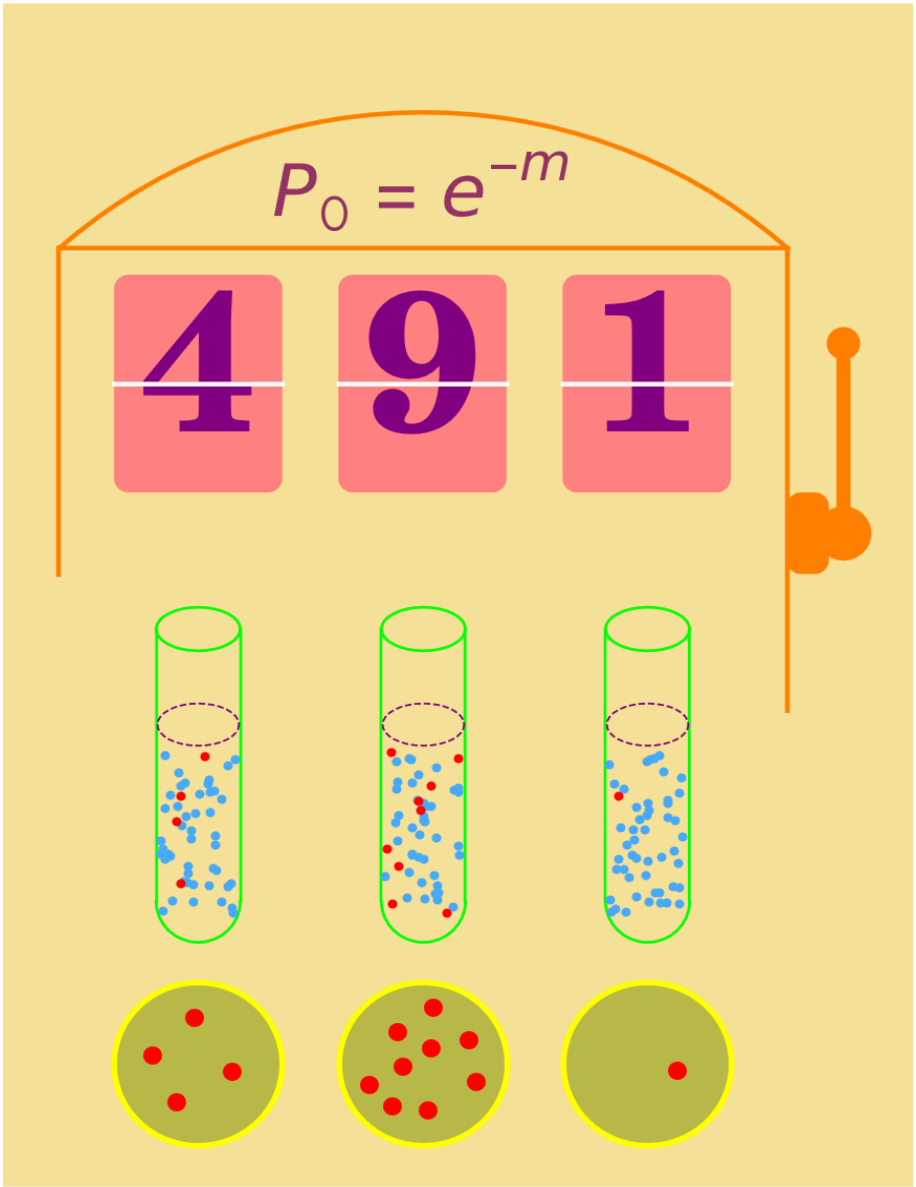
Getting Started with webSalvador

Qi Zheng
Texas A&M University School of Public Health
College Station, Texas 77843
qzheng@tamu.edu



1. Introduction

The fluctuation experiment, proposed by Luria and Delbrück [6], remains an important laboratory protocol for determining microbial mutation rates. Data generated by a fluctuation experiment are mutant counts, which must be analyzed by statistical methods to obtain meaningful information about the mutation rate under investigation. Methods developed before 2000 have been thoroughly reviewed and critiqued by Foster [4]. Many of the analytic methods developed after 2000 are available in the R package called rSalvador [15]. webSalvador, a web interface to rSalvador, allows the user to access most of the methods in rSalvador without having to learn the R language.

webSalvador's capabilities fall into two groups: mutation rate estimation and mutation rate comparison. Methods in the first group analyze mutant count data generated by a single fluctuation experiment, whereas methods in the second group analyze mutant count data coming from two separate fluctuation experiments. In addition to mutant count data, all methods require an important parameter called N_t for each experiment. The symbol N_t in fluctuation assay data analysis refers to the average final number of viable cells per culture. If only a portion of a culture is plated (called partial plating), N_t refers to the average number of cells in the whole culture, not just the plated portion. Each method available in webSalvador assumes a particular mutant distribution, under which a fluctuation experiment generates its mutant count data. The most popular mutant distribution is due to Lea and Coulson, but is widely called the Luria-Delbrück distribution.

2. The classic method

This is an estimation method assuming the Luria-Delbrück distribution. An overwhelming majority of published fluctuation experiments were analyzed using this method. The experi-

ment conducted by Demerec [2] is among the earliest experiments analyzed using this method [10]. Here N_t was determined to be about 1.9×10^8. Note that webSalvador abandons the familiar two-column data input format, because measuring N_t for each culture will impose partial plating on the investigator, rendering the classic method inappropriate. Figure 1 illustrates the data input process. In the "Mutant counts" box, mutant count data can be separated by one or multiple spaces, and input data can be divided into multiple lines. In the "Nt" box, the final number of cells can be given either as a positive integer or in its scientific notation. That is, both "190000000" and "1.9e8" are acceptable. After the data input process, the user clicks the green button to obtain analysis results. As shown in Figure 2, The output (results) array gives two kinds of estimates. The row labeled m is about the expected number of mutations per culture, and the row labeled μ is about the mutation rate. While m is a fundamental parameter mathematically, investigators are mainly interested in the mutation rate μ. In addition to a so-called point estimate, webSalvador also gives a corresponding 95% confidence interval (CI). Note that webSalvador always gives likelihood ratio CIs, as this kind of CI is more appropriate when an experiment comprises only a small number of cultures. As Pawitan [8, p. 48] puts it, "the applicability of the likelihood-based CI is much wider and, consequently, it is much safer to use than the Wald interval." In this example, the estimated mutation rate is 5.71×10^{-8}, while a 95% CI for the mutation rate is $[4.55 \times 10^{-8}, 6.94 \times 10^{-8}]$. Note that almost all algorithms in fluctuation data assay are algorithmically iterative. As a result, these algorithms need appropriate initial values to work properly. Fortunately, webSalvador can choose good initial values automatically most of the time. Hence, the user need not to put any values in the three initial value input boxes (a value of zero indicates that the initial value would be determined by webSalvador automatically).

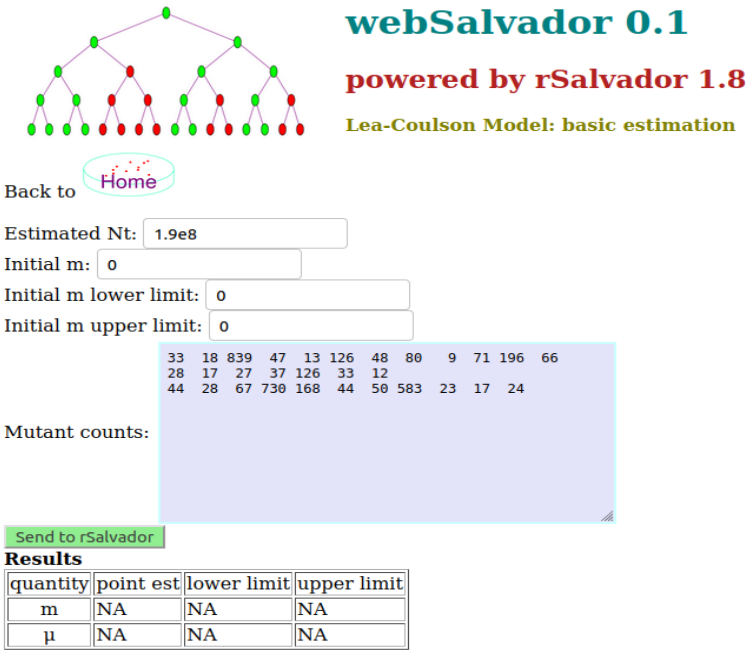


Figure 1: Data input for classic analysis of the Demerec data.

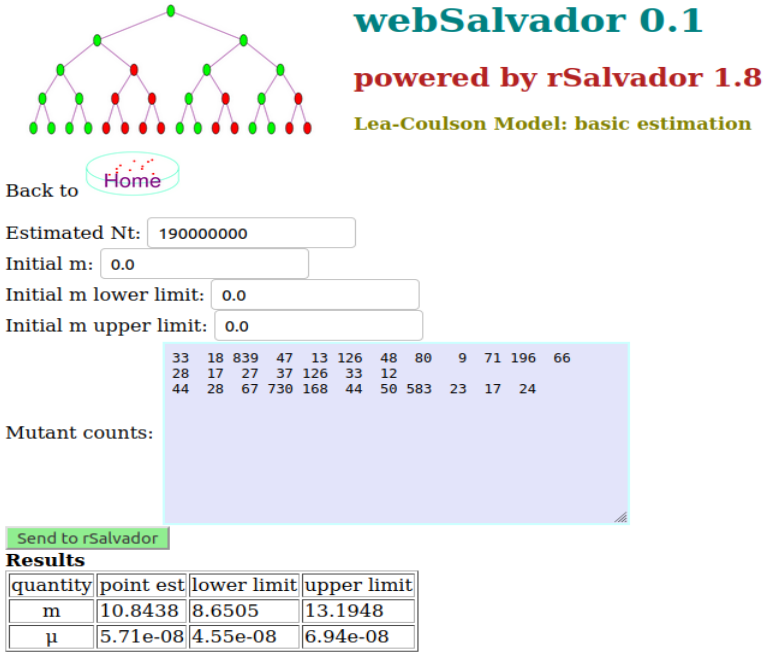


Figure 2: Results for the analysis of the Demerec data.

3. A Luria-Delbrück model accommodating partial plating

This kind of analysis makes the same assumptions as does the classic analysis discussed in the previous section, except that it modifies the mutant distribution to account for partial

plating. When only a portion of a culture (say 10 percent) is plated, the experimenter induces an additional kind of uncertainty in data. Some investigators new to the Luria-Delbrück protocol may attempt to inflate the mutant count data before subjecting the data to the same kind of analysis as presented in the previous section. For example, an investigator may change a mutant count of 3 to 30 when 10 percent of the culture is plated. This intuitive approach is incorrect.

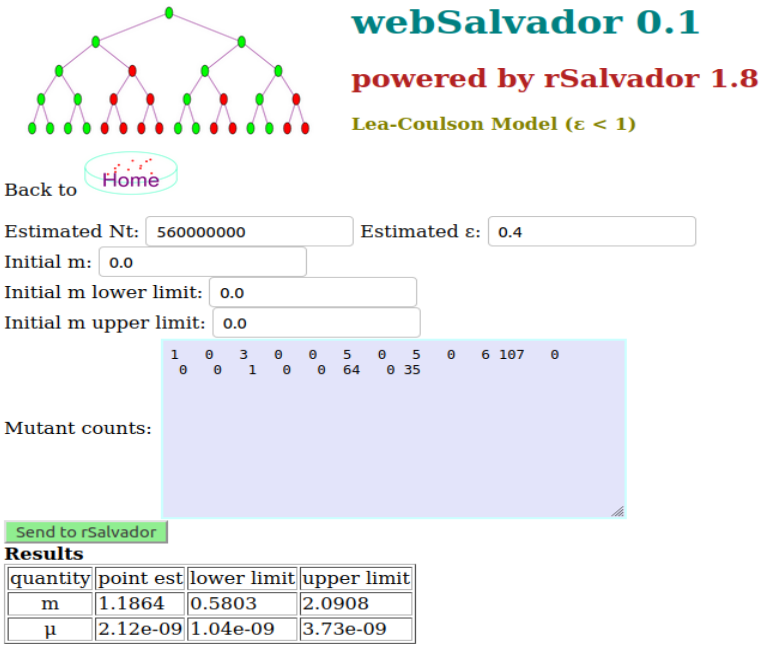


Figure 3: Classic analysis of experiment 16 of Luria and Delbrück.

Now consider experiment 16 of Luria and Delbrück [6], which was first analyzed in a likelihood based method by Zheng [11]. In this experiment, a portion of 0.08ml of each of the 20 0.2ml-cultures was plated. Hence, the plating efficiency is $\epsilon = 0.4$. In addition, it was determined that each whole culture contained about 5.6×10^8 cells. Figure 3 shows how this experiment can be analyzed by webSalvador. The output array gives an estimated mutation rate of 2.12×10^{-9} mutations per cell division, with a 95% CI for the mutation rate being $[1.04 \times 10^{-9}, 3.73 \times 10^{-9}]$.

4. Mutation rate comparison

webSalvador is the first web tool that offers likelihood ratio tests for mutation rate comparison based on fluctuation assay data. This feature is important because sample sizes (the numbers of parallel cultures) were small in most published fluctuation experiments. Agresti's advice [1, p.107] is relevant in the present context: "Although the Wald test is adequate for large samples, the likelihood-ratio test is more powerful and more reliable for sample sizes often used in practice."

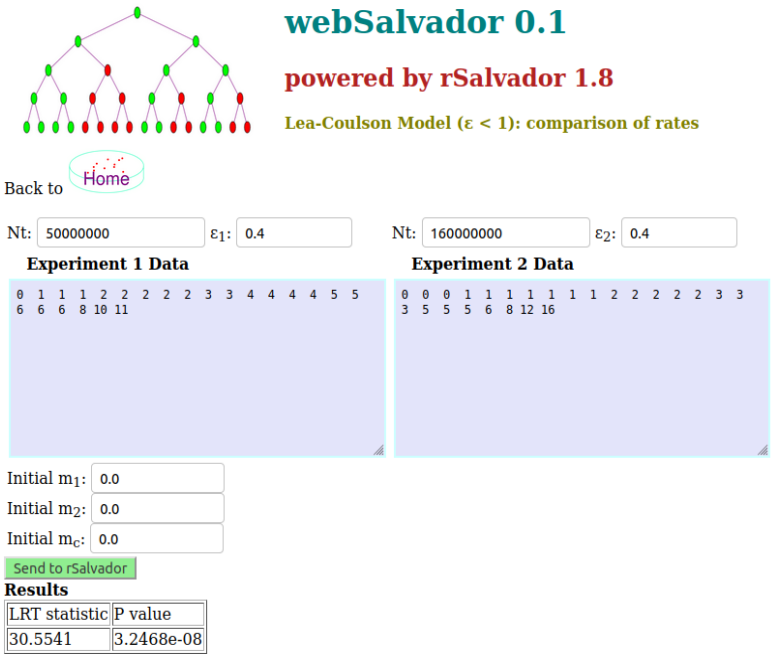


Figure 4: Comparison of mutation rates in two of the experiments of Werngren and Hoffner.

The following example shows how the mutation rates in Experiment 2 and Experiment 5 in the antibiotic resistance study of Werngren and Hoffner [9] are compared using the likelihood ratio test. As explained in Zheng [13], the plating efficiencies in both experiments were 0.4. In Experiment 2 (referred to as experiment 1 in webSalvador) $N_t = 5.0 \times 10^7$, and in Experiment 5 (referred to as experiment 2 in webSalvador) $N_t = 1.6 \times 10^8$. The output

array gives a test statistics of 30.55, and a p-value of 3.21×10^{-8}. (See Figure 4.) In view of the small p-value, there is statistical evidence that the two mutation rates are different. Finally, investigators should be cautious about the Mann-Whitney test, as in general it is not applicable to comparison of mutation rates [13, p.352].

5. Mutation rate fold change

Inference about mutation rate fold change is an alternative way to compare mutation rates. A long-standing roadblock to its widespread use is a lack of algorithms for computing CIs for mutation rate fold change. A profile likelihood approach has been proposed recently [16], and webSalvador uses this approach to construct CIs.

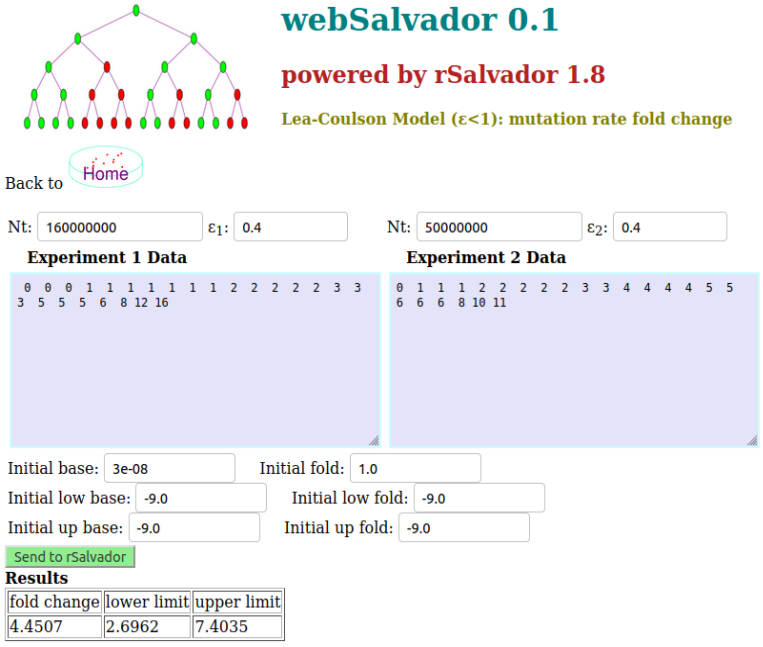


Figure 5: Inference about mutation rate fold change regarding mutation rates in two of the experiments of Werngren and Hoffner.

In webSalvador parlance, a mutation rate fold change is obtained by dividing the mutation rate in "experiment 2" divided by the mutation rate in "experiment 1." The mutation rate

in the first experiment is called the base rate. In practice, investigators often want a fold change to be larger than unity. Hence, they should name as "experiment 2" the experiment having the larger mutation rate. For instance, in the example of the preceding section, one would name experiment 5 as "experiment 1." As shown in Figure 5, the estimated mutation rate fold change is 4.45, and the corresponding 95% CI is [2.70, 7.40].

Note that the user can specify six initial values. Three initial values are for the base rate and its CI (the first column), and three other initial values are for the fold change and its CI (the second column). In this example, it is sufficient to choose initial base rate to be $3 \times 10-8$, and the initial fold change to be 2.0. The other initial values are left to their defaults (signified by the number -9.0).

6. The Mandelbrot-Koch model that accounts for differential fitness

The classic Luria-Delbrück distribution, which is due to Lea and Coulson [3], assumes that mutants and wild-type cells grow at the same rate. This assumption was removed by work of Mandelbrot [7] and of Koch [5], and the resulting mutant distribution is called the MK distribution. The MK distribution has a fitness parameter w defined as the ratio of the growth rate of mutants (numerator) to the growth rate of wild-type cells (denominator). If a fitness assay (aka a competition assay) is separately conducted to measure the mutants' relative fitness, the investigator may incorporate an estimate of w into the data analysis process.

Figure 6 shows how to use webSalvador to estimate the mutation rate in the Demerec experiment when a hypothetical estimate of w (0.9) is available. As expected, this slight elevated the estimated mutation rate because the mutants are assumed to have a slight

growth disadvantage.

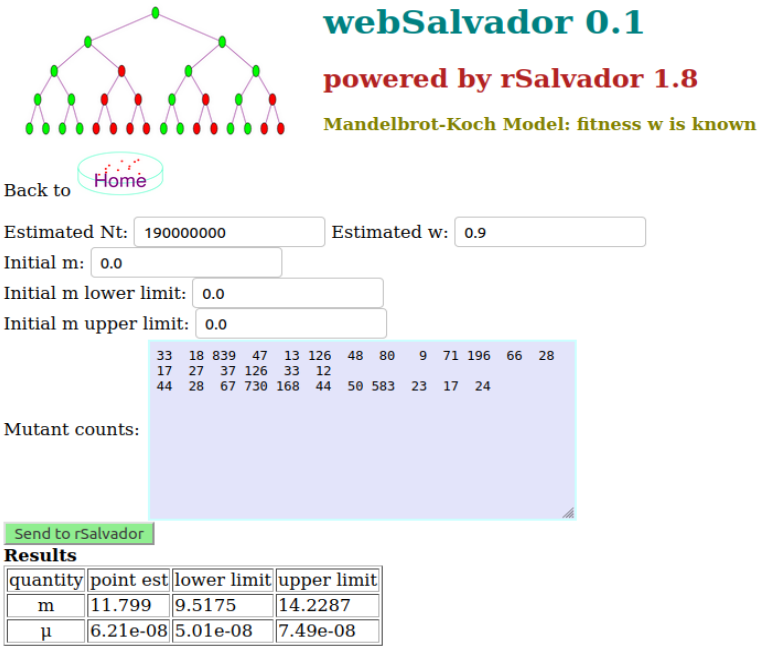


Figure 6: Using the Mandelbrot-Koch model to estimate the mutation rate in the Demerec experiment. A hypothetical estimate of the mutants' relative fitness ($w = 0.9$) is incorporated into the analysis.

7. The gamma mixture model that accounts for variation in N_t

There has been considerable confusion abut whether variation in N_t can seriously affect mutation rate estimation. A recent study [14] has found that the impact is largely negligible unless the variation is unusually large. Variation in N_t is often measured by the coefficient of variation (CV), which is defined as the ratio of the standard deviation to the mean. The gamma mixture model, discovered in 2011 [12], modifies the classic Luria-Delbrück distribution to accommodate variation in N_t. Assuming complete plating, investigators can employ the gamma mixture model to analyze fluctuation assay data. In the following example, using a hypothetical coefficient variation (CV) of 0.2, webSalvador reanalyzes the

Demerec data. Now a new estimate of the mutation rate is 5.93×10^{-8} (Figure 7), which is quite close to the estimate obtained by classic analysis which assumes a constant value of N_t across the cultures. As expected, the 95% CI is slightly wider due to the additional source of uncertainty.

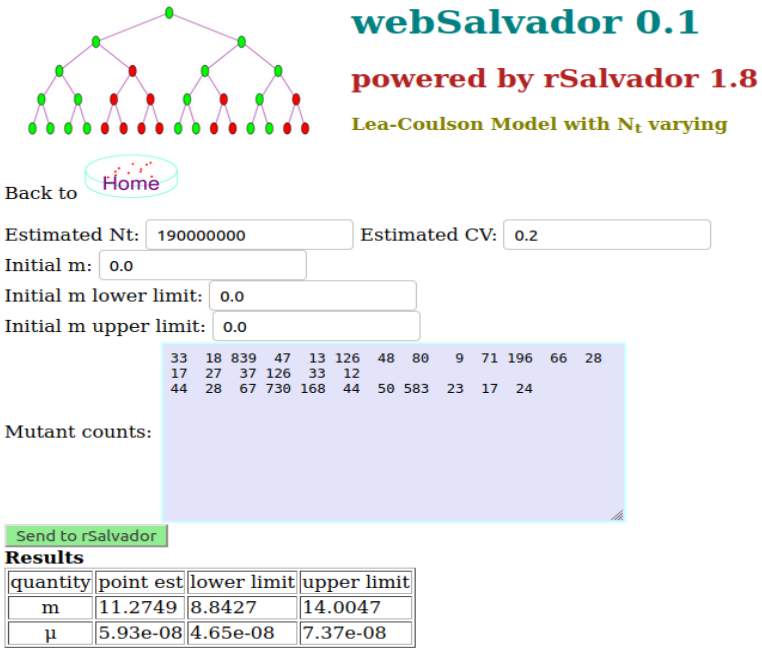


Figure 7: Using the gamma mixture model to estimate the mutation rate in the Demerec experiments. The coefficient of variation for N_t is assumed to be 0.2.

Acknowledgement

Special thanks to Eric Zheng for creating the build and deploy configuration.

References

1. Agresti A. 2007 An Introduction to Categorical Data Analysis, second edition, Wiley-Interscience.

2. Demerec M. 1945 Production of staphylococcus strains resistant to various concentrations of penicillin, Proc. Natl. Acad. Sci. USA 31:16-24.

3. Lea EA, Coulson CA. 1949 The distribution of the numbers of mutants in bacterial populations. J. Genetics 49:264-285.

4. Foster PL. 2006 Methods for determining spontaneous mutation rates. Methods Enzymol 409:195-213.

5. Koch AL. 1982 Mutation and growth rates from Luria-Delbrück fluctuation tests. Mutation Research 95:129-143.

6. Luria SE, M Delbrück D. 1943 Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491-511.

7. Mandelbrot B. 1974 A population birth-and-mutation process, I: Explicit distributions for the number of mutants in an old culture of bacteria, J Appl Prob 11:437-444.

8. Pawitan Y. 2001 In all likelihood: Statistical modeling and inference using likelihood. Clarendon Press, Oxford.

9. Werngren J, Hoffner SE. 2003 Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. Journal of Clinical Microbiology 41:1520-1524.

10. Zheng Q. 2002 Statistical and algorithmic methods for fluctuation analysis with SALVADOR as an implementation, Math Biosci 176:237-252.

11. Zheng Q. 2008 A note on plating efficiency in fluctuation experiments. Mathematical Biosciences 216:150-153.

12. Zheng Q. 2011 A Bayesian two-level model for fluctuation assay. Genetica 139:1409-1416.

13. Zheng Q. 2016 Comparing mutation rates under the Luria-Delbrück protocol. Genetica 144:351-359.

14. Zheng Q. 2016 A second look at the final number of cells in a fluctuation experiment. Journal of Theoretical Biology 401:54-63.

15. Zheng Q. 2017 rSalvador: an R package for the fluctuation experiment. G3:Gene|Genome|Genetics 7:3849–3856.

16. Zheng Q. 2021 New approaches to mutation rate fold change in Luria-Delbrück fluctuation experiments, *Mathematical Biosciences*, 335:108572.