Check for updates

BRIEF REPORT

# REVISED The number of neutral mutants in an expanding Luria-Delbrück population is approximately Fréchet [version 2; peer review: 2 approved]

Steven A. Frank [iD]

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, 92697-2525, USA

## Abstract

**Background:** A growing population of cells accumulates mutations. A single mutation early in the growth process carries forward to all descendant cells, causing the final population to have a lot of mutant cells. When the first mutation happens later in growth, the final population typically has fewer mutants. The number of mutant cells in the final population follows the Luria-Delbrück distribution. The mathematical form of the distribution is known only from its probability generating function. For larger populations of cells, one typically uses computer simulations to estimate the distribution.
**Methods:** This article searches for a simple approximation of the Luria-Delbrück distribution, with an explicit mathematical form that can be used easily in calculations.
**Results:** The Fréchet distribution provides a good approximation for the Luria-Delbrück distribution for neutral mutations, which do not cause a growth rate change relative to the original cells.
**Conclusions:** The Fréchet distribution apparently provides a good match through its description of extreme value problems for multiplicative processes such as exponential growth.

## Keywords

Population genetics, probability distributions, extreme value distributions

This article is included in the Genomics and Genetics gateway.

**Open Peer Review**

**Approval Status** ✔ ✔

|  | 1 | 2 |
|---|---|---|
| version 2 (revision) 03 Mar 2023 |  |  |
| version 1 04 Nov 2022 | ✔ view | ✔ view |

1. **Qi Zheng**, Texas A&M University School of Public Health, College Station, USA

2. **Pavol Bokes**, Comenius University, Bratislava, Slovakia

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Steven A. Frank (safrank@uci.edu)

**Author roles: Frank SA**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Suppose a single cell expands exponentially to a population of size $N$, with a mutation rate of $u$ per cell division. The number of mutant cells, $m$, in the final population depends on the number of mutations that occur and when those mutations occur. For example, a single mutation in the final round of cell division is limited to one cell. By contrast, a single mutation transmitted to one of the daughters in the first cellular division may occur in approximately one-half of the final population.

The distribution of the number mutants, $m$, is known as the Luria–Delbrück distribution[1]. That distribution is widely used to estimate the mutation rate. The distribution also arises when studying the amount of mutational mosaicism within multicellular individuals[2–4].

Currently, for experiments with a small number of mutational events, one typically calculates the distribution with a probability generating function[5,6]. However, that approach becomes numerically inaccurate for larger numbers of mutational events, in which case the distribution is calculated by computer simulation.

This article shows that the Fréchet distribution provides a good approximation for the number of neutral mutants. In particular, the probability that the number of mutants, $m$, is less than $z$ is approximately

$$F(z) = \text{Prob}(m \leq z) = \exp\left(-\left(\frac{z-\beta}{s}\right)^{-\alpha}\right), \qquad (1)$$

in which $\exp(z) = e^z$ is the exponential function. The probability of being in the upper tail, $m > z$, is $1 - F(z)$. The three parameters set the shape, $\alpha$, the scale, $s$, and the minimum value, $\beta$, such that $z, m > \beta$.

This form of the Fréchet distribution has three parameters. I found that the following parameterization matches closely the Luria–Delbrück process for neutral mutations

$$\alpha = e/2$$

$$s = eNu$$

$$\beta = Nu \log\left(Nu\, e^{-(1+\alpha)}\right)$$

in which $e$ is the base of the natural logarithm. This parameterization depends on the single parameter, $Nu$, the final population size times the mutation rate.

Figure 1 shows the good fit. Two aspects of mismatch occur. First, the number of mutants is discrete, whereas the Fréchet is continuous. As $Nu$ declines to one, significant amounts of probability mass concentrate at particular mutant number values, causing discrepancy between the distributions. Nonetheless, the Fréchet remains a good approximation.

Second, the lower tail of the Luria–Delbrück process spreads to lower values than the Fréchet. One can see this mismatch most clearly in the figure for $Nu \geq 100$.

This mismatch may occur because the Luria–Delbrück process transitions from a highly stochastic process in earlier cellular generations to a nearly deterministic accumulation of mutations in later cellular generations, when the larger population size reduces the coefficient of variation in the number of new mutations. The Fréchet applies most closely to the earlier generations for the following reasons.

In an expanding population, the earliest mutation strongly influences the final number of mutants. An early mutant carries forward to all descendant cells in an expanding mutant
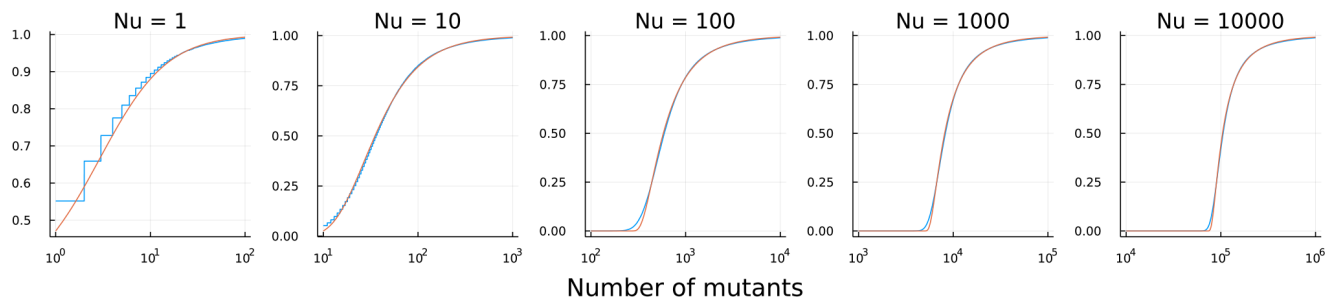


**Figure 1. Cumulative distribution of the number of neutral mutants in an expanding population.** Each population begins with one cell and grows to $N$ cells. Mutation occurs at rate $u$. Blue curves show the distribution from a computer simulation using the simu. cultures command of the R package rSalvador[7]. Orange curves show the Fréchet distribution in Equation 1. In rSalvador, I used sample sizes of $10^6$ or $10^7$, values of $Nu$ varying as shown above the plots, and values of $N$ ranging from $10^6$ to $10^{10}$. The Julia software code to produce this figure is available from Zenodo[8]. The input data for calculating the empirical Luria-Delbrück CDF is also available from Zenodo[9].

clone. If we start with the final cells and then look back through the cellular generations toward the original progenitor, the mutation with the most extreme time from the end toward the beginning tends to dominate the final mutant number.

The extreme value of a temporal extent often has a Gumbel distribution. In this case, once the mutation arises, it increases multiplicatively by cell division to affect the final mutation count. Substituting the extreme Gumbel time for its multiplicative consequence provides a common way to observe a Fréchet probability pattern.

Prior mathematical work also supports the Fréchet approximation. Kessler and Levine[10] showed that the Luria–Delbrück distribution converges to a Landau distribution for large $Nu$, in which the Landau distribution is a special case of the Lévy $\alpha$-stable distribution. However, the Landau distribution does not have a closed-form expression for its probability or cumulative distribution functions.

Separately, Simon[11] showed the close match between the Lévy $\alpha$-stable distribution and the Fréchet distribution. That match of a Lévy distribution to the Fréchet

distribution had not previously been associated with the Luria–Delbrück distribution. The Fréchet parameterization in this article provides a simple expression that can be used to develop further theory and applications of the Luria–Delbrück process.

---

## Data availability

### Underlying data

The input data for calculating the empirical Luria-Delbrück CDF:

Zenodo: Empirical CDF for Luria-Delbrück distribution from rSalvador package. https://doi.org/10.5281/zenodo.7075655[9].

## Software availability

The Julia software code used to produce Figure 1:

Source code available from: https://github.com/evolbio/FrechetLD

Archived source code at time of publication: https://doi.org/10.5281/zenodo.7255050[8]

License: MIT

---

### References

1. Zheng Q: **Progress of a half century in the study of the Luria-Delbrück distribution.** *Math Biosci.* 1999; **162**(1–2): 1–32.
   **PubMed Abstract** | **Publisher Full Text**

2. Otto SP, Hastings IM: **Mutation and selection within the individual.** *Genetica.* 1998; **102–103**(1–6): 507–524.
   **PubMed Abstract** | **Publisher Full Text**

3. Frank SA: **Somatic mosaicism and cancer: inference based on a conditional Luria-Delbrück distribution.** *J Theor Biol.* 2003; **223**(4): 405–412.
   **PubMed Abstract** | **Publisher Full Text**

4. Iwasa Y, Nowak MA, Michor F: **Evolution of resistance during clonal expansion.** *Genetics.* 2006; **172**(4): 2557–2566.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

5. Ma WT, Sandri GH, Sarkar S: **Analysis of the Luria-Delbrück distribution using discrete convolution powers.** *J Appl Probab.* 1992; **29**(2): 255–267.
   **Publisher Full Text**

6. Zheng Q: **Estimation of rates of non-neutral mutations when bacteria are exposed to subinhibitory levels of antibiotics.** *Bull Math Biol.* 2022;

**84**(11): 131.
**PubMed Abstract** | **Publisher Full Text**

7. Zheng Q: **rSalvador: an R package for the fluctuation experiment.** *G3 (Bethesda).* 2017; **7**(12): 3849–3856.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

8. Frank SA: **evolbio/FrechetLD: F1000 (1.0.1).** [Software], Zenodo. 2022.
   **http://www.doi.org/10.5281/zenodo.7255050**

9. Frank SA: **Empirical CDF for Luria-Delbrück distribution from rSalvador package (1.0.0).** [Dataset]. *Zenodo.* 2022.
   **http://www.doi.org/10.5281/zenodo.7075656**

10. Kessler DA, Levine H: **Large population solution of the stochastic Luria-Delbruck evolution model.** *Proc Natl Acad Sci U S A.* 2013; **110**(29): 11682–11687.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

11. Simon T: **Comparing Fréchet and positive stable laws.** *Electron J Probab.* 2014; **19**: 1–25.
    **Publisher Full Text**

# Open Peer Review

## Current Peer Review Status: ✔ ✔

---

**Version 1**

Reviewer Report 20 February 2023

✔ **Pavol Bokes**
[1] Department of Applied Mathematics and Statistics, Comenius University, Bratislava, Slovakia
[2] Department of Applied Mathematics and Statistics, Comenius University, Bratislava, Slovakia

The paper compares the empirical (simulational) Luria-Delbruck mutant-number distribution to a Frechet distribution. The advantage of the Frechet distribution over other known approximations, e.g. the skewed alpha-stable distributions, is that it possesses a closed-form cumulative distribution function (cdf), see Equation (1). The parameters alpha, s, and beta of the Frechet distribution are set by the author as specific functions of the population-wide mutation rate N*u. Figure 1 visually demonstrates a solid agreement between the Frechet distribution and the empirical Luria-Delbruck distribution.

Similarly to right-skewed alpha-stable distributions, the Frechet distribution has a heavy right tail and a light left tail (for $z <$ beta the density is zero). It follows from (1) that the complementary cdf decays as $1/z^{alpha}$, where alpha has been set by the author to exp(1)/2. In addition to the log-linear plots of Figure 1, it would be interesting to look closer at the power laws of the theoretical and empirical complementary cdfs e.g. using a log-log plot.

Overall, the paper is well written, presents sound ideas, and develops an interesting approximation to the Luria-Delbruck mutant-number distribution.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Mathematical biology, stochastic modelling, gene expression, differential equations

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

> Author Response ( *F1000Research Advisory Board Member* ) 20 Feb 2023
> **Steven Frank**
>
> Thank you, I appreciate these comments. I agree that looking closely at the upper tail in a log-log plot would provide additional insight about the frequency of rare but potentially important events. However, to achieve a good computational estimate for the true cdf of the assumed process would require some new analyses to obtain precise estimates of the numerical error of the computation. That error could potentially be significant for the very rare upper tail events that would be the focus of such analysis. As applications arise that require a good estimate of the upper tail, the analyses and calculations would be a worthwhile new project.
>
> *Competing Interests:* No competing interests were disclosed.

Reviewer Report 23 November 2022

https://doi.org/10.5256/f1000research.139979.r155112

✔ **Qi Zheng**
[1] Department of Epidemiology and Biostatistics, Texas A&M University School of Public Health, College Station, TX, USA
[2] Department of Epidemiology and Biostatistics, Texas A&M University School of Public Health, College Station, TX, USA

In his brief report the author presents an interesting approximation of the Luria-Delbruck

distribution, which microbiologists use to help determine microbial mutation rates in the laboratory. Specifically, equation (1) in the brief report is an approximation of the cumulative probability. If $Pk$ denotes the probability of $k$ mutants, the author implicitly defines the cumulative probability $c_n$ as $c_n = \sum_{k=0}^{n-1} Pk$.

The author's key finding is that $c_n \approx F(n)$, where $F$ is defined by equation (1) in the brief report. Note that the approximation in (1) is valid for any $n > \beta$. However, as pointed out by the author, the approximation works well only for values of $n$ that are noticeably larger than $\beta$. I have conducted a number of computer experiments and confirmed the numerical results in the brief report. The approximation is theoretically interesting, and it may stimulate further theoretical developments. Thus, the paper merits indexing.

I have a minor comment. There appears to be a typo in equation (1) in the brief report. If Prob($m < z$) is changed to Prob($m \leq z$), the correlative change in the definition of $c_n$ will make $c_n$ conform to the accepted definition of the cumulative probability. (That is, $c_n = \sum_{k=0}^{n} Pk \cdot$ ) More importantly, this may make the approximation more accurate for small $n$. Consider the case $m = 10$ (The symbol $m$ here is the same as the symbol $Nu$ in the brief report). Table 1 shows results obtained by using the revised definition, while Table 1A shows corresponding results obtained by using the original definition. In both tables, "error" refers to the following quantity:

$$error = \frac{|exact - Fréchet|}{exact} \times 100\%$$

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

***Competing Interests:*** No competing interests were disclosed.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response ( *F1000Research Advisory Board Member* ) 20 Feb 2023

**Steven Frank**

Thank you for the careful reading. With regard to the comment about m < z versus m <= z, the calculations to make figure 1 used m <= z for the empirical distribution, as recommended by the reviewer. For the theoretical continuous Frechet the numerical values are the same
for the two cases. However, I agree that the notation in the original version of the manuscript is misleading. I will post a revised version that uses m <= z, as recommended.

*Competing Interests:* No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research