# Statistics of estimating relative fitness from competition assays based on viable count data

## Robert J Clegg and Jan-Ulrich Kreft

The development of this statistical method was prompted by the evolution experiment and competition assays to measure fitness carried out by Feng Dong et al. in our lab:

Feng Dong, Ana C. Quevedo, Xiang Wang, Eugenia Valsami-Jones, Jan-Ulrich Kreft (2020). Experimental Evolution of *Pseudomonas putida* under Silver Ion versus Nanoparticle Stress. <https://doi.org/10.1101/2020.09.17.302794>

See the preprint for further information on the study.

<https://www.biorxiv.org/content/10.1101/2020.09.17.302794v2>

The idea of the method is to simulate the experimental procedure to derive estimates of fitness.

For each population i, we start with a test tube of volume Vt (known) containing Nt,i (unknown) cells. From this, a sample of volume Vst (known) is taken and diluted Dt-fold (known): this is then plated and the colonies counted to give xt,i (known). From all this, we see that xst,i is drawn from a Binomial distribution:

.

Note that Nt,i are the sole unknowns here, and so we estimate them following (Blumenthal and Dahiya, 1981; Feldman and Fox, 1968).

Note also that plating and counting more than once could reduce the uncertainty in Nt,i, but that this is often impractical.

Another sample, volume V0 (known), is taken from each test tube and transferred into a flask of volume Vf (known). The numbers of cells in this sample will be . The uncertainty in N0,i is not considered in the current script that assumes the whole tube is poured into the flask. The culture in the flask is then incubated for time T and the number of cells of each population is assumed to grow exponentially with rate ri (unknown) to and so

.

A sample of the flask is then taken, plated and counted as before to estimate the xf,i. We denote the sample volume to be Vsf, the dilution Df, and the colony counts as xsf,I:

.

We then estimate Nf,i and the uncertainty in our estimate in the same manner as for the Nt,i.

The uncertainty in the growth rates ri is currently not considered but could be included following (Katz et al., 1978) and papers citing it but uncertainty is in p rather than in n.

Once we have estimates of N0,i and Nf,i, together with estimates of their uncertainty, we can calculate relative fitness.

See the Python code “sampling\_confidence\_interval.py” for further information.

**Table 1.** Descriptions of symbols used.

|  |  |
| --- | --- |
| **Symbol** | **Description** |
| \*0 | Index for time 0 in the flask |
| \*f | Index for the final time in the flask |
| \*s | Index for a sample |
| \*t | Index for the test tube |
| D\* | Dilution factor |
| N\*,i | Unknown number of cells of population i |
| ri | Growth rate of population i |
| V\* | Volume of a tube, flask, or sample |
| x\*,i | Number of colonies of population i counted on a plate |

## References

Blumenthal, S. and Dahiya, R. C. (1981) ‘Estimating the Binomial Parameter n’, *Journal of the American Statistical Association*, vol. 76, no. 376, pp. 903–909 [Online]. DOI: 10.2307/2287586.

Feldman, D. and Fox, M. (1968) ‘Estimation of the Parameter n in the Binomial Distribution’, *Journal of the American Statistical Association*, vol. 63, no. 321, pp. 150–158 [Online]. DOI: 10.1080/01621459.1968.11009230.

Katz, D., Baptista, J., Azen, S. P. and Pike, M. C. (1978) ‘Obtaining confidence intervals for the risk ratio in cohort studies’, *Biometrics*, vol. 34, no. 3, pp. 469–474 [Online]. DOI: 10.2307/2530610.