

Sensitivity Analysis of Computational Models for solving Paradox

Jurij Nastran, Beno Sircelj, Matjaz Gams
"Jozef Stefan" Institute, Jamova cesta 39
Ljubljana, Slovenia
jurij.nastran@gmail.com
beno.sircelj@gmail.com
matjaz.gams@ijs.si

Drago Bokal
University of Maribor, Koroska cesta 46
Maribor, Slovenia
drago.bokal@um.si



ABSTRACT: *Given the estimated number of stars and planets in our galaxy, the probability of existence of intelligent civilizations seems high. The first to indicate this was the Drake equation with the assumed default parameters. Yet, the actual observations have yet to reflect those expectations. This discrepancy corresponds to the so-called Fermi Paradox. Although many key factors about the likelihood of alien civilizations still remain largely unknown, new methods of estimating the probability are being proposed. Some of them use probability distributions and the Monte Carlo methods. In this paper we recalculate one of those - the Sandberg interpretation of the Drake equation, analyze the difference between the methods, their strengths and weaknesses. In the conclusion, we find that the probability distribution better reflects our ignorance about the properties of alien environments than the dot-product method.*

In our opinion, there are several ways to further improve the computational model based on the Drake equation using the AI methods, thus eliminating the problem with too small probabilities and introducing 3D views.

What makes these analysis relevant, is not only the number of estimated civilizations in our galaxy and a probability that we encounter them in the near future. More important, these models enable estimation of the life-span of the human civilization. Unfortunately, there is a considerably high probability that it will be quite short.

Keywords: Drake Equation, Fermi Paradox, Monte Carlo, Probability Distribution, Extraterrestrial Intelligence

Received: 5 August 2018, Revised 1 November 2018, Accepted 10 November 2018

DOI: 10.6025/jnt/2019/10/1/18-24

© 2019 DLINE. All Rights Reserved

1. Introduction

There are billions of stars in the observable universe. Hundreds of millions are broadly estimated to be in our galaxy alone. If there is at least a modest chance of intelligent life emerging on a given planet, surely there should be at least some number of alien life in the relative vicinity, yet we see none. We are apparently the only one in our part of the universe even though we were able to expand our search quite successfully in recent decades [5].

Why is there no evidence of other civilizations in our galaxy despite the sheer number of planets we now know of? This is the question physicist Enrico Fermi first asked in 1950 and is known as the Fermi paradox. Fermi was particularly effective when dealing with estimates of ill-defined physical phenomena. However, he did not coin the first universally accepted equation for an estimate of the number of civilizations.

1.1 Drake Equation

Probably the best known equation for an estimate of the number of detectable civilizations in the Milky Way was composed by Frank Drake [3], phrased as a product of seven factors:

$$N = R_* f_p n_e f_i f_c L$$

The parameters are: R_* which is the rate of star formation per year, f_p is the fraction of stars with planets, n_e is the number of Earth-like (or otherwise habitable) planets per a star that has planets, f_i is the fraction of habitable planets with actual life, f_c is the fraction of life-bearing planets that develop intelligence, f_c is the fraction of intelligent civilizations that are detectable and L is the average longevity of such civilizations. Finally N is the number of detectable civilizations.

The Drake equation is used to directly estimate the number of civilizations or as an analysis tool for various components in our galaxy. Most importantly, it can also be used to estimate the lifespan of our civilization (L). The Drake equation can provide exact numbers given proper parameters (i.e. factors), but the problem is that several factors in the equation are not well determined either by observations or with laboratory models. By assuming different values for them, say f_i – probability of life on a habitable planet, results vary a lot [1].

1.2 Point Estimates

When trying to estimate the number of civilizations in the galaxy, a point estimate is often used for each of the seven parameters of the Drake equation. This provides an exact numerical value. If we take estimates based on the distributions in Sandberg's paper [6], we get: $R_* = 10$, $f_p = 0.3$, $n_e = 0.3$, $f_i = 0.5$, $f_i = 0.03$, $f_c = 0.1$, $L = 10^6$ which gives us N around 270. Drake with colleagues originally estimated that $L = N$ and probably between 1000 and 100.000.000. Current best estimates[4] differ from several hundreds to several millions for the civilization's lifespan, and from being alone to several millions civilizations in our galaxy.

Based on actual observations, it seems quite likely that the optimists overestimate their factors: If an advanced civilization appeared somewhere in the galaxy before, moving with the speed of one percent of light speed, it would reach all parts of our galaxy in less than 20 million years. This is a tiny fraction of the lifespan of our galaxy, which is as old as our universe, i.e. around 13.5 billion years, with the perimeter around 100 - 200.000 light years.

2. Setting The Parameters and Computational Models

Sandberg and colleagues [6] suggested that using point estimates to solve the Drake equation is too wild a guess, providing only one number. In their paper they suggest an approach that models each parameter by its distribution, thus the computation results in a probability distribution. They used a Monte Carlo method for calculating the distribution of the final result. During each iteration, they sampled from distributions to obtain point estimates to use in Drake equation, which results can be used to generate final distribution of N .

2.1 A Toy Model

To show how using distributions for calculations leads to results that differ from those obtained by simply multiplying point estimates, Sandberg and colleagues introduced a simplified toy model. In their toy model to demonstrate the differences, there are 9 parameters (f_1, f_2, \dots), which if multiplied together determine probability of ETI (extraterrestrial intelligence) on a single star.

Each of those parameters can obtain values from an interval [0, 0.2], with an average of 0.1. In their case-example, the point estimate of each factor is set to 0.1, the same as the average from the interval. For 100 billion stars (as in our galaxy) the computation gives 100 intelligent civilizations. But if instead of using a point estimate we sample from a uniform distribution with an average of 0.1, we get the results indicating that there is 21.45% chance that we are alone in the galaxy.

2.2 Recomputing the Sandberg interpretation of Drake equation

Sandberg et al. applied probability distribution as a way of recalculating the Drake equation, but instead of a uniform distribution presented in Subsection 2.1, they used the probability distribution for each factor obtained from scientific literature - a range and the type of the distribution. They defined a parameter “log-uncertainty” of a parameter X ($LU[X] = \log(\max(X)/\min(X))$) as an estimate of the number of orders of magnitude of the current uncertainty of parameter X . Consequently, the factors were defined in the following way:

- Star formation rate R_* is fairly well constrained by astronomical data and ranges over a maximum of 5 orders of magnitude given other galaxies. Our uncertainty about this parameter is from 2 to 16 solar masses, $LU[R_*] = 0.9$.
- Fraction of stars with planets f_p is also pretty well known and is about 1 with $LU[f_p] = 1$.
- The estimates for n_e , which is number of habitable plants, range from $<10^{-12}$ in rare earth arguments to >1 when taking non terrestrials like icy moons into account. Sandberg proposed $LU[n_e] = 12$. Post-2000 literature estimates cover smaller orders of magnitude so they postulated earth-like planet as rocky planet within habitable zone and assumed $LU[n_e] \approx 2$.
- The parameters with the most uncertainty are f_i and f_l . f_l (probability of life) is modeled as a physical transition that occurs at some rate per unit time per unit volume of a suitable prebiotic substrate. The probability on a habitable planet with volume V , time period t and abiogenesis rate λ is $f_l = 1 - e^{-\lambda V t}$. They take log-uniform distributions of t with $LU[t] = 3$ and range from 10^7 to 10^{10} , V with $LU[V] \geq 20$ range from 10^{-35} to 10^{15} . They use log-normal distribution for λ with a mean of 1 and σ of several orders of magnitude. In the paper they do not specify the exact number used for σ , just that its very big. When we recreated their experiment we tested several values for sigma and 200 gave the most similar results.
- Based on the literature, the parameter f_c , which is the fraction of planets which develop civilization, is between 0.001 and 1.
- The final factor L , is longevity of a space-communicating civilization in years and is in the range from 100 to 10^{10} , which is the current estimate of the age of the universe.

All parameter(factor) distributions are listed in Table 1.

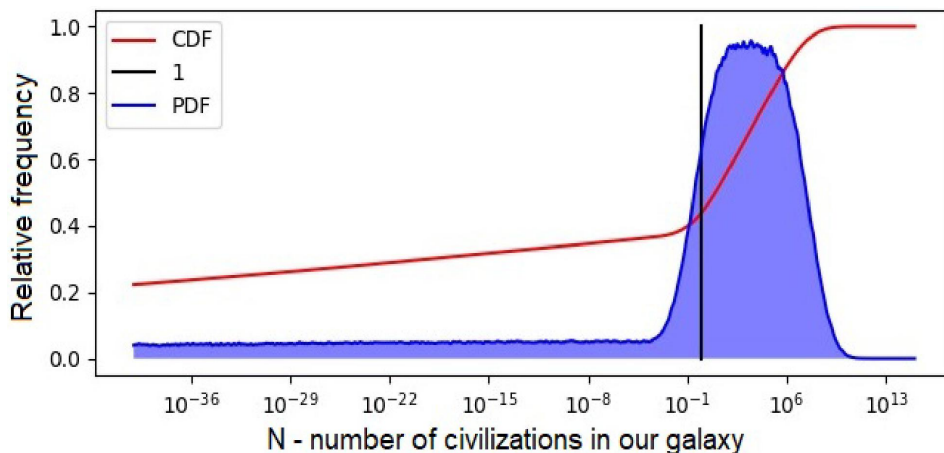


Figure 1. Recomputation of the Sandberg approach: probability density function and cumulative density function of N computed with Monte Carlo method

With this set of parameters we managed to obtain a distribution for N , displayed in Figure 1, quite similar to the one by Sandberg and colleagues. The two axes in Figure 1 represent N and its corresponding frequency with blue color. A vertical black line marks N equal to 1. The blue graph is therefore PDF (probability density function). It is scaled so that the highest value is 1 to t on the same graph as the red line which is CDF (cumulative density function). From Figure 1 it can be observed that the probability of us being alone in our galaxy is about one half, e.g. since the red graph reaches 0.5 around N equal to 1.

| Parameter | Distribution |
|-----------|--|
| R_* | log-uniform from 1 to 100 |
| f_p | log-uniform from 0.1 to 1 |
| n_e | log-uniform from 0.1 to 1 |
| f_l | log-normal rate, described above |
| f_i | log-uniform from 0.001 to 1 |
| f_c | log-uniform from 0.01 to 1 |
| L | log-uniform from 100 to 10,000,000,000 |

Table 1. Summary of the current knowledge about the parameters of the Drake equation

3. Analysis of the Sandberg’s Computational Model

In this section, we present our initial analysis of the Sandberg et al. computational approach. We tested various issues only on the toy model as it is simpler and better reflects the issues with the computational model. We hypothesized that if one parameter is assigned a number close to zero, the whole product will be a very low number no matter what the other parameters are. For example, the blue graph in Figure 1 might give a misleading impression that the majority of the possibilities is on the right side of the black line indicating 1. However, due to the logarithmic x scale, it is about the same possible that there is only 1 civilization in our galaxy compared to 2 or more, as the red graph shows. The probabilities on the left of the black line indicating 1 are therefore significantly higher. The closer to 0, the higher, meaning lots of N s are strangely close to 0. From the AI field we are familiar with this problem from the analysis of the naive Bayer theorem: if one of the factors is 0 or close to 0, it should better be modified. We tried to analyze this hypothesis by calculating the same results by using a different number of parameters and by using ranges with small offset to move away from zero. Namely, if any of the parameters is zero, then there is no intelligent civilization in that galaxy.

3.1 Effect of Small Values with Multiple Parameters

To study the effect of small values of parameters/factors in the product, we varied the number of parameters. To normalize the result, we adjusted the number of stars so that the average value of N according to the distribution obtained by the Monte Carlo analysis is still 100 as it is in the original case of 9 parameters and 1011 stars. So the number of stars is 10^{2+Np} where Np is the number of parameters. All the parameters are in the range $[0, 0.2]$. The graph showing 3 parameters and 9 parameters can be seen in Figure 2. The solid lines are PDFs and the dotted lines are CDFs. The PDF is scaled like before so that the maximum value is 1. In both cases the average value is still 100. With green dot we mark the probability that there is no civilizations in the toy galaxy. Note that the size of our toy galaxy also varies with the number of parameters to enable comparison. The dot is on the CDF at value of one civilization in the galaxy. It can be seen that the value at 9 parameters is higher than the value at 3 parameters. Therefore, the more parameters, the more probability of being alone. Or in other words: the more uniformly distributed parameters one introduces in a product, the more likely small values of N if parameter values are in $[0, 1]$.

We tested this phenomenon on multiple number of parameters ranging from 1 to 9 and the result can be seen in Figure 3. We can see that by increasing the number of parameters the PDF widens, thus increasing the possibility of us being alone.

3.2 Modifying Small Values with Epsilon

We tried to minimize the effect of parameter/factor values being too close to zero, which seemingly spoils the calculation, by introducing an offset ϵ . Instead of the range $[0, 0.2]$ we now try the same experiments with range $[0+\epsilon, 0.2-\epsilon]$. The offset on the

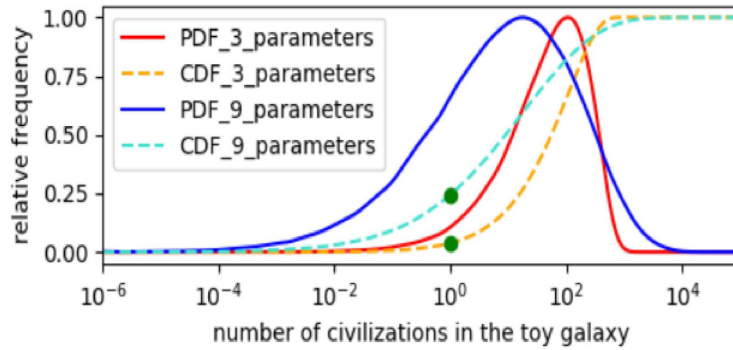


Figure 2. Graphs for Toy model with 3 parameters and 9 parameters

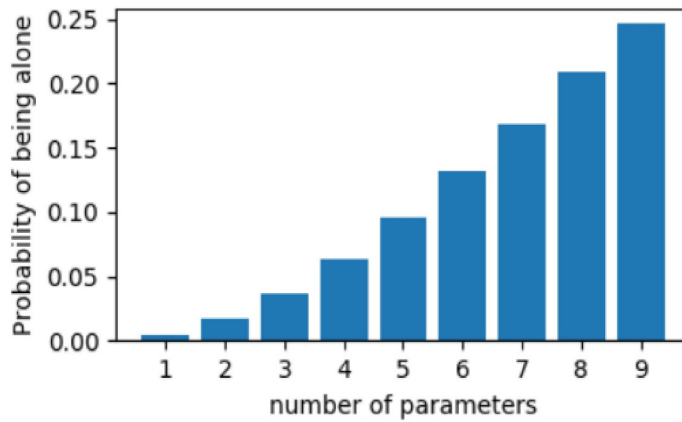


Figure 3. Probability of there being no civilizations depending on the number of parameters (1-9)

left side ensures that the minimum value is increased and the offset on the right side is there to keep the same average. As we can see in Figure 4 for [0.02, 0.18], the computed N changes significantly. In Figure 5, the same effect is demonstrated for values from 0.00 to 0.040 by a blue line. By increasing the ϵ , the curve becomes narrower thus chances of us being alone decrease.

3.3 Effects of log-uniform Distributions

The toy model had all its parameters distributed uniformly. We recreated the same experiments, but with log uniform distribution of parameters, which is the same type of distribution as in the Sandberg paper. In Figure 5 one can see that the effect of adding epsilon in the log distribution modifies the graph even more. Figure 6 represent another analysis with the toy problem and various epsilons, using the log uniform distribution. This is another indication that the small values of parameters introduced by probability distribution by Sandberg strongly influence the final result.

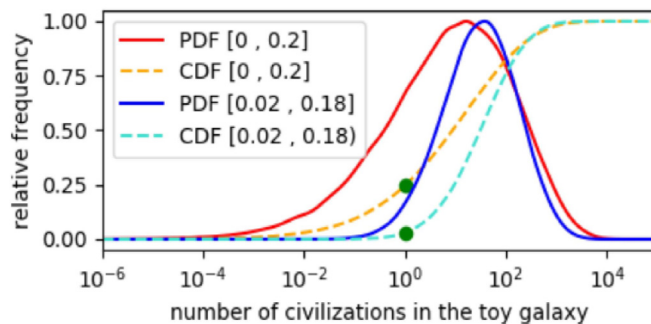


Figure 4. The effect of ϵ on the probability distribution

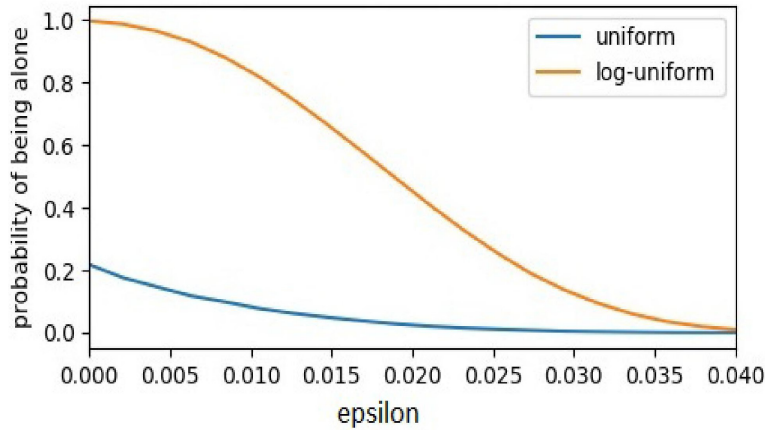


Figure 5. Effect of ϵ is enhanced with log-uniform distributions

4. Discussion: AI-based Modifications

The proposed and to a certain extent already tested AI-based modifications are to be fully described in the submission of an SCI paper. Some of the ideas will be presented live at the paper presentation. Here, a couple of hints are presented here.

The first observation when reanalyzing the Sandberg approach is that it is significantly more informative than point estimates. Unlike providing just one number based on assumptions, it shows the whole probability distribution, i.e. all possible combinations of values of parameters. On the second thought, the computational model has a weakness - when multiplying with zero or very small values. In the Drake equation, several factors are multiplied together, but originally, none of them was very close to 0. However, when dealing with probability distributions, numbers close to 0 can appear and as a result N becomes very small. We presented this effect in the Toy problem with uniform distributions, and later with log uniform distributions where the effect intensified.

In AI, there have been similar problems when multiplying with small values and solutions. For example, several analyses of the Naive Bayes were performed for the case when one of the factors was zero. One of the first analysis was by Cestnik [2].

By modifying the Cestnik or Laplace approach for the Sandberg method, we managed to introduce important modifications.

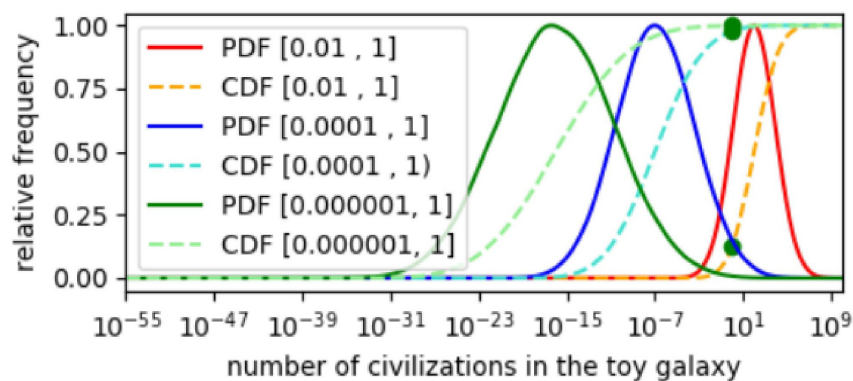


Figure 6. How ϵ affects distributions in log-space

5. Conclusion and Discussion

We first showed how the number of ETIs was estimated using the original Drake equation, and then with the advanced approach by Sandberg and colleagues. Our re-computation of the Sandberg method yielded nearly the same results.

When reanalyzing the Sandberg approach, one issue emerged - multiplying with small values. We analyzed this phenomena and found out that with increasing number of parameters the probability gets closer to zero. We also introduced some corrections to the ranges of parameters.

While it is unlikely that the computing mechanisms will solve the Fermi problem on its own, they can provide better understanding of the current and future observations.

The computational models and universe observations progress, increasing our knowledge and narrowing the unknowns in the estimation of the number of civilizations and the lifespan of our civilization. However, until we meet another civilization or none for a long time or, as the third and most tragic option - our civilization decays, there is still a long way to go.

Acknowledgments

We would like to thank Filip Talimdzioski for his advice and cooperation.

References

- [1] Achenbac, J., Essick, P. (2000). Life Beyond Earth. National Geographic, 197 (1) 24-51.
- [2] Cestnik. B. (1990). Estimating probabilities: a crucial task in machine learning. *In*: ECAI, volume 90, p. 147-149, 1990.
- [3] Drake, F. (2015). The Drake Equation: Estimating the Prevalence of Extraterrestrial Life Through the Ages. Cambridge University Press.
- [4] Drake, F., Sobel, D. (1992). Is anyone out there?: The scientific search for extraterrestrial intelligence. Delacorte Press.
- [5] Gribbin, J. (2018). Are Humans Alone in the Milky Way? Scientific American, (September).
- [6] Sandberg, A., Drexler, E., Ord. T. (2018). Dissolving the Fermi paradox. arXiv preprint arXiv:1806.02404, 2018.