A Probabilistic Search for Extraterrestrial Life
A Tale of Priors

Grateful Hippos
University of Rochester, Rochester, NY
(Dated: January 10, 2022)

The search for extraterrestrial life has so far yielded no results. In the absence of data, experts in the field have so far relied on statistical methods and real-world arguments to constrain the probability of the existence of extraterrestrial life. In this paper, two such analyses are discussed, which adopt fundamentally different perspectives to answer the same question - are we alone in the universe? Under a Bayesian approach, it is found that the posterior probability of the existence of life is heavily informed by our priors. A number of priors are tested and their effects on the posterior are discussed qualitatively. The merits and demerits of the analyses are discussed, as is the future scope of the search for extraterrestrial life.

INTRODUCTION

Despite decades of hopeful searching, the pursuit of discovering life beyond Earth has been met with a continually disappointing result: nothing. However, this discouraging result has not dissuaded researchers from placing estimates on the existence of extraterrestrial life in the universe.

One way of simply quantifying our ignorance is through the Drake equation. Proposed by Frank Drake in 1961, this equation aims to provide us with an estimate of the number of civilizations with which humans could communicate.

$$N = R_* \cdot f_p \cdot n_e \cdot f_l \cdot f_i \cdot f_c \cdot L \tag{1}$$

The terms are described below:

Term	Description
N	number of civilizations with which humans could communicate
R_*	average rate of star formation
f_p	fraction of stars that have planets
n_e	mean number of planets that could support life per star with planets
f_l	fraction of life-supporting planets that develop life
f_i	fraction of planets with life where life develops intelligence
f_c	fraction of intelligent civilizations that develop communication
L	mean length of time that civilizations can communicate

The first few terms of the Drake equation- R_*, f_p, and n_e - have been meticulously constrained using real data obtained from experiments within the last few decades [1][2][3]. However, the remaining terms are currently unknown; worse, with only a single datum for an intelligent civilization - the Earth - any hope for an experimental constraint on these values is still far-fetched. While there exist many possible ways to approach this problem, in our paper we will be focusing on statistical arguments - namely, a frequentist and a Bayesian approach - to discuss the terms f_l and possibly f_i.

For the proposed problem, frequentists would be expected to formulate a "counting problem", such as rerunning the timeline of the Earth n times and see how many of those trials result in abiogenesis. As for a Bayesian school of thought, they would rely on Bayes' theorem:

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M}) \times P(\mathcal{M})}{P(\mathcal{D})} \tag{2}$$

where $P(\mathcal{M}|\mathcal{D})$ is the posterior probability of the model given data, $P(\mathcal{D}|\mathcal{M})$ is the likelihood of the data assuming the model to be correct, and $P(\mathcal{M})$ is our prior belief in the model. The Bayesian approach for our discussion would be predicting the probability of abiogenesis in the event we get data, by choosing an appropriate likelihood and setting realistic priors on the problem. However, this Bayesian method poses the question that given we have only the Earth as our data point, can we make a claim that there is life outside? How can we choose priors that take into account our state of ignorance while being unbiased and uninformed?
In our attempts of discussing these questions, we explore two papers: first, a Bayesian analysis focusing on the probability per unit time of abiogenesis, which in turn affects the probability of life arising [4]; and a second that defines a frequentist probability of existence of life by setting a Bayesian prior on its value. [5].

ON SPIEGEL AND TURNER'S ANALYSIS

Spiegel and Turner's (S&T hereafter) Bayesian analysis of the astrophysical implications of life's early emergence on Earth is considered a pioneering effort in interpreting the factor f_L in Drake's equation using purely statistical motivations. They argue that the development of life on a planet is a simple Poisson process that occurs between the times t_{min} and t_{max}, and is a function of the probability per unit time of abiogenesis λ [4]. Therefore, the probability of life arising n times in a time $t_{min} < t < t_{max}$ is:

$$P[\lambda, n, t] = e^{-\lambda(t-t_{min})} \frac{\lambda(t-t_{min})^n}{n!} \tag{3}$$

This choice of model can be justified as the , the conditions on a young planet preclude the development of life for a time period of t_{min} after its formation. Furthermore, if the planet remains lifeless until t_{max} has elapsed, it will remain lifeless thereafter as well because conditions no longer permit life to arise. [4]

Since S&T are only interested in the probability of life arising on a planet *at all*, only consider the case $n \geq 1$ i.e. life arises at least once is considered. This means that the above expression can be simplified to be one minus the probability that life *never* occurs on a planet, that is:

$$P_{\text{life}} = 1 - e^{-\lambda(t - t_{min})} \tag{4}$$

This term P_{life} is directly proportional to the factor f_l from Drake's equation. Also, this clearly shows that P_{life} depends on λ, the probability per unit time of abiogenesis on a planet. This probability "rate" could take up any value– if λ is large, it implies that there is a high probability that abiogenesis occurred early on in the planet's lifespan, and vice versa. Since there is no present estimate for this probability rate λ, S&T aim to analyze this problem with a Bayesian lens and attempt to define a posterior probability distribution in λ. S&T focus their analysis on using Bayes' Theorem to define a posterior probability for λ i.e. the probability that λ takes on a certain value, given some data. However, we currently have no data that proves or disproves the existence of extraterrestrial life. Therefore, our posterior will be heavily informed by our prior belief in a particular value of λ. The question we then seek to answer is, is there a way to make the prior as uninformed as possible while still being responsive to the possibility of existence of extraterrestrial life? S&T consider three choices for these priors - uniform in λ, uniform in λ^{-1}, and uniform in $\log \lambda$ spanning a range from some λ_{min} to λ_{max}.

It is noted that the values of λ_{min} and λ_{max} have been chosen somewhat arbitrarily [6]: S&T provide no reasoning for choosing $\lambda_{max} = 10^3$ Gyr^{-1}, and there is no physical reason as to why it could not be a few orders of magnitude higher. Similarly, they provide three options for λ_{min}: 10^{-22} Gyr^{-1}, 10^{-11} Gyr^{-1}, and 10^{-3} Gyr^{-1}, corresponding to life occurring once in the observable universe, once in our galaxy, and once per 200 stars. As we shall see in later sections, the analysis is very sensitive to both the choice of priors and the range of λ considered.

ON BRIAN LACKI'S ANALYSIS

Lacki's analysis takes the probability of intelligent life evolving as a frequentist probability and introduces a "Log-Log prior" as a way of containing its value. [5]. Lacki seeks to attack both f_l and f_i simultaneously by defining the probability that life evolves at a suitable site to be $\mathcal{P}_{ETI}$ which is given by:

$$\mathcal{P}_{ETI} = f_l \cdot f_i \tag{5}$$

Because the number of requirements necessary for life to appear is unknown, the order of $\mathcal{P}_{ETI}$ is also unknown. Lacki explains:

> "Suppose there are N conditions that must be fulfilled for aliens to appear, and [...] each is independent of each [...with...] probability 1/2. Then $\mathcal{P}_{ETI} = 2^{-N}$[...]. If N is uncertain at the order of magnitude level, then [...] $\mathcal{P}_{ETI}$ is also uncertain at the order of magnitude." [5]

Lacki argues that a standard Jeffreys prior, one that is uniform in $\log(\mathcal{P}_{ETI})$, would be unsuitable as it would not account for this uncertainty in the order of magnitude. Instead, he suggests a prior which is constant in $\log(\log(\mathcal{P}_{ETI}))$, which he calls the "Log-Log" prior [5]. Justifying the prior is especially crucial because as discussed by S&T, the prior dominates the posterior due to the lack of data [4].

The Log-Log prior is meant to account for the highest degree of variability in the range of possible values for a parameter (p) and in an marginalized PDF, takes the form:

$$\ln(1 - \ln(p)), \quad p \in (0, 1) \tag{6}$$

One notes that the Log-Log prior is undefined at $p = 0$, however, Lacki resolves this issue by considering a maximal combinatoric bound i.e. the total number of ways to arrange the observable universe; this is at most $e^{3\times10^{122}}$ [7]. Lacki seems to assert that a subset of the number of possible arrangements of the universe must be configurations that capable of supporting life (such as the case with Earth) and thus considers this the largest possible number of "conditions" for life to occur. While not the same, this is similar to suggesting life on Earth would not have occurred unless each object in observable universe were in their exact location when life began.

Ultimately, the Log-Log prior places the highest weight upon life, requiring an extraordinarily large number of conditions being needed for intelligent life to appear. The remainder of Lacki's paper investigates ways to place optimistic bounds on extraterrestrial life (specifically surrounding the number of possible genome sequences) and despite taking these heavily optimistic considerations, Lacki concludes that there is only Bayesian confidence of 18% that there is extraterrestrial intelligence in the observable universe.

Lacki's analysis is very robust; he considers a very broad and minimally informed range of values for bounds on $\mathcal{P}_{ETI}$. However, one of the most immediate critiques is that the considered bounds for $\mathcal{P}_{ETI}$

are arguably too broad. Lacki even concedes this, saying that at this bound, the prior probability that aliens even exist is on the order of $\mathcal{P}_{ETI} \sim e^{10^{-122}}$. Alternative bounds could be on the total number of planets in the observable universe or even the number of ways to arrange all planets in the universe.

DISCUSSION

The Importance of Priors in Both Analyses

As discussed in an earlier section, the absence of any data implies that priors will heavily inform our posterior in this problem. We illustrate this quantitatively in Fig.1, where it is evident that different priors on λ lead to vastly different shapes of the posterior probability density function (PDF). We recreate the prior and posterior PDFs considered by S&T in their analysis [4], and also attempt to apply Lacki's log-log prior to S&T's λ for demonstrative purposes.

One may recall from S&T's analysis that the choice of λ_{min} and λ_{max} was arbitrary. Therefore, we need to consider our choice of priors based on not only the shape of the priors, but also how they respond to the change in the range of λ and their degree of bias.

Uniform Priors

S&T first consider two types of uniform priors on λ - one uniform in λ, and another uniform in λ^{-1} (or an inverse uniform prior) [4].

However, they note that both these priors are considered highly informed in standard Bayesian terminology. That is, the uniform and inverse uniform priors assume, without evidence, that very large or small values of λ are more likely, respectively. For instance, assuming a uniform prior on λ would imply our confidence that it is a hundred times less likely that λ is less than 10^{-3} Gya^{-1} than it is less than 0.1 Gya^{-1}. Of course, we have no basis to claim this type of scale variance without any evidence. Fig. 1 shows how uniform shows perverse support for the the larger values of λ as discussed above, the converse is true for inverse uniform prefers the smaller values of λ.

We may also consider the effects of changing the range of λ on these priors. In Fig. 2, we see that as we decrease λ_{min} from 10^{-3} Gya^{-1} to 10^{-10} Gya^{-1} while keeping λ_{max} constant, the shape of the inverse-uniform curves changes (the uniform has been ommitted as it is also defined in terms of its range). Since our priors are so heavily dependent on the unknown range, they are unsuitable for this problem.

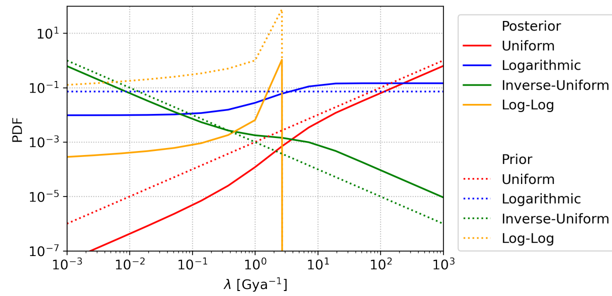

Figure 1. PDF of λ with uniform, logarithmic, inverse-uniform, and log-log priors. We follow S&T's recommended values $\lambda_{min} = 10^{-3}$ Gya^{-1} and $\lambda_{max} = 10^{3}$ Gya^{-1}

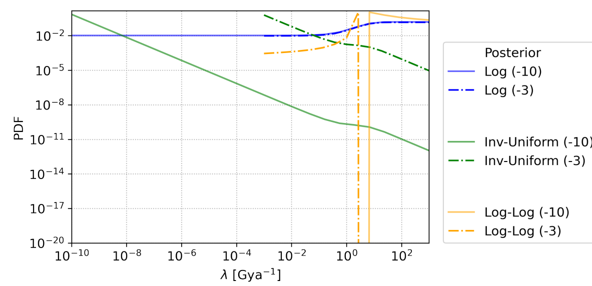

Figure 2. PDF of λ with logarithmic, inverse-uniform, and log-log priors, this time assuming $\lambda_{min} = 10^{-10}$ Gya^{-1} and keeping $\lambda_{max} = 10^{3}$ Gya^{-1}

Logarithmic Prior

An alternate approach S&T consider is choosing a prior that is uniform in logλ [4]. This solves some of the drawbacks of the uniform priors considered earlier. For example, in Bayesian statistics, this prior is considered a relatively uninformed one, as it asserts that we have no prior information that informs us of even the order of magnitude of λ. Due to the scale-invariant nature of the log prior, S&T consider it the most appropriate for this particular analysis. Furthermore, we found that this prior is invariant under the change of the range of λ, as demonstrated in Fig. 2.

However, another expert Kipping notes that a logarithmic prior is an improper one, as the probability of λ over a finite range tends to zero. Additionally, Kipping disagrees with S&T's fixed upper bound $\lambda_{max} = 10^3$ Gya^{-1}. Instead, he shows that in order for the logarithmic prior to be fair and unbiased, λ_{max} will have to be expressed as a function of λ_{min}. This means that the choice of prior would be very subjective rather than objective, and the prior cannot be uniquely defined. Furthermore, he shows that the prior is extremely sensitive to the choice of λ_{min}, which exposes serious gaps in even this choice of prior [6].

Log-Log Prior

Lacki does not apply the log-log prior on λ, but with a fundamentally different quantity $\mathcal{P}_{ETI}$. We, however do it to in Figs. 1 and 2 to demonstrated the importance of the choice of priors and note some interesting properties of this prior when applied on λ. We notice that the log-log prior behaves asymptotically as $\log \lambda \rightarrow 1$. However, it has an interesting property not displayed by the other priors considered by S&T, that the shape of the prior and posterior is essentially the same, just translated along the vertical axis. We, however, note that the log-prior goes through dramatic changes in shape when we change the range; making it unsuitable for the analysis used by S&T.

A more intuitive way of comparing the log-log prior to the logarithmic prior is by using S&T's parameter λ and transform it into the $P_{\rm life}$ space using Eq. 4. Since the functional form of the log-log prior can only be applied to a probability and not a rate, we can further apply this to $P_{\rm life}$. The PDF of the priors obtained in both cases is given in Fig. 3. As expected from Lacki's analysis, the log-log prior places a large weight at low probabilities of abiogenesis due to the extraordinarily high number of conditions needed for the evolution of intelligent life. On the other hand, the logarithmic prior places high weights on the extremes of $P_{\rm life}$, suggesting that either life is widespread in the universe or not common at all [4][6].

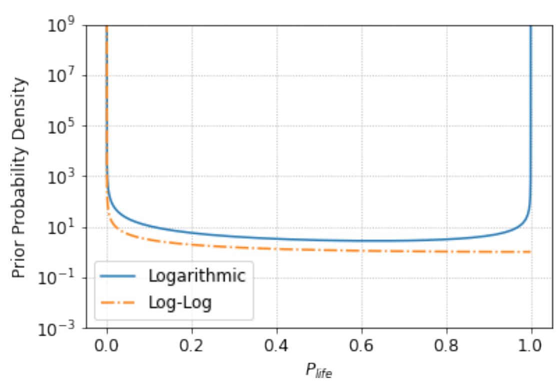

Figure 3. The prior PDFs for P_{life} (Eq. 4) assuming a logarithmic and log-log prior (assuming $t_{\rm min} = 0.5$ Gya^{-1}, as mentioned in S&T [4]).

Comparing the Two Analyses

Lacki and S&T are asking two related yet fundamentally different question. S&T are exploring different priors on the rate of abiogenesis denoted by λ and how each of them might affect the value of $P_{\rm life}$. On the other hand, Lacki's analysis is an attempt to use the fact that Earth is the only planet observed with life, and explore the probability space by considering probabilities that wide range $(e^{-3\times10^{122}}, 1)$ [5].

S&T have a Bayesian understanding of the underlying factors that drive f_l itself, whereas Lacki treats $\mathcal{P}_{ETI}$ as a combinatorial factor which is simply given by Size of the candidate space^{-1}, where the size of the candidate space considers the places we might expect to find life but have not found any.

S&T's approach seems more nuanced as it provides a better understanding and control of the assumptions made in considering how λ affects $P_{\rm life}$. Lacki's approach seems to dodge the major question of how priors on underlying factors (λ in this case) may affect $\mathcal{P}_{ETI}$. However, since we have no experimental information about how the priors on these factors should behave, Lacki's analysis may be correct in not assuming anything a priori on $\mathcal{P}_{ETI}$.

An interesting approach going ahead might be to use the probabilities derived from different priors as considered in S&T and then using Lacki's method to find a confidence interval on $\mathcal{P}_{ETI}$ for abiogenesis guided by different priors.

CONCLUSIONS

From where we stand right now, it appears that the discussion of the existence of life is rather inconclusive as we only have one data point: Earth. Despite the many contentions in the discussion of life beyond Earth, however, one thing is for certain and agreed upon - even one more data point will substantially change the way we think about and look at this problem. As discussed many times in these papers, the dominant nature of the priors on the posterior is a consequence of the lack of data we possess now. However, that being said, it is realistic to expect that a lot may change within our lifetime as the search for exoplanets is more widely practiced. With the development of space telescopes, like Kepler and K2, and techniques like radial velocity and transit photometry, we have been able to identify over 4000 exoplanets [8]. Now, with the help of transit spectroscopy, we are also able to learn about the atmospheres of exoplanets. Since the presence of life and the presence of abundant atmospheric O_2 occur simultaneously, learning about these exoplanetary atmospheres will hopefully provide us with potential biosignatures to search for. Just like how within the past half a century or so we have managed to nail the first three terms in the Drake equation, we are hopeful to make such progress in the next terms. Nevertheless, exciting things await in the future of this discussion, and hopefully, we will be a part of that excitement in the coming years.

[1] Robert C. Kennicutt and Neal J. Evans. Star formation in the milky way and nearby galaxies. *Annual Review of Astronomy and Astrophysics*, 50(1):531–608, Sep 2012.

[2] A. Cassan, D. Kubas, J.-P. Beaulieu, M. Dominik, K. Horne, J. Greenhill, J. Wambsganss, J. Menzies, A. Williams, U. G. Jørgensen, and et al. One or more bound planets per milky way star from microlensing observations. *Nature*, 481(7380):167–169, Jan 2012.

[3] C. H. Lineweaver. The galactic habitable zone and the age distribution of complex life in the milky way. *Science*, 303(5654):59–62, Jan 2004.

[4] David S. Spiegel and Edwin L. Turner. Bayesian analysis of the astrobiological implications of life's early emergence on earth. *Proceedings of the National Academy of Sciences*, 109(2):395–400, 2012.

[5] Brian C Lacki. The log log prior for the frequency of extraterrestrial intelligences. *arXiv preprint arXiv:1609.05931*, 2016.

[6] David Kipping. An objective bayesian analysis of life's early start and our late arrival. *Proceedings of the National Academy of Sciences*, 117(22):11995–12003, 2020.

[7] Chas A Egan and Charles H Lineweaver. A larger estimate of the entropy of the universe. *The Astrophysical Journal*, 710(2):1825, 2010.

[8] NASA. Exoplanet catalog. 2021.