Acta Astronautica ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Acta Astronautica

journal homepage: www.elsevier.com/locate/actaastro

SETI and SEH (Statistical Equation for Habitables)[☆]

Claudio Maccone [a,b,*]

[a] Technical Director, International Academy of Astronautics (IAA)
[b] Co-Chair, SETI Permanent Study Group of the IAA

ARTICLE INFO

Article history:
Received 22 December 2009
Received in revised form
13 May 2010
Accepted 5 June 2010

Keywords:
Habitable planets
Statistics
Dole Equation
Drake Equation

ABSTRACT

The statistics of habitable planets may be based on a set of ten (and possibly more) astrobiological requirements first pointed out by Stephen H. Dole in his book "Habitable planets for man" (1964). In this paper, we first provide the statistical generalization of the original and by now too simplistic Dole equation. In other words, a product of ten positive numbers is now turned into the product of ten positive random variables. This we call the SEH, an acronym standing for "Statistical Equation for Habitables".

The mathematical structure of the SEH is then derived. The proof is based on the central limit theorem (CLT) of Statistics. In loose terms, the CLT states that the sum of any number of independent random variables, each of which may be ***arbitrarily*** distributed, approaches a Gaussian (i.e. normal) random variable. This is called the Lyapunov form of the CLT, or the Lindeberg form of the CLT, depending on the mathematical constraints assumed on the third moments of the various probability distributions. In conclusion, we show that

(1) The new random variable N_{Hab}, yielding the number of habitables (i.e. habitable planets) in the Galaxy, follows the ***lognormal*** distribution. By construction, the mean value of this lognormal distribution is the total number of habitable planets as given by the statistical Dole equation. But now we also derive the standard deviation, the mode, the median and all the moments of this new lognormal N_{Hab} random variable.
(2) The ten (or more) astrobiological factors are now positive random variables. The probability distribution of each random variable may be ***arbitrary***. The CLT in the so-called Lyapunov or Lindeberg forms (that both do not assume the factors to be identically distributed) allows for that. In other words, the CLT "translates" into our SEH by allowing an arbitrary probability distribution for each factor. This is both astrobiologically realistic and useful for any further investigations.
(3) An application of our SEH then follows. The (average) ***distance between any two nearby habitable planets*** in the Galaxy may be shown to be inversely proportional to the cubic root of N_{Hab}. Then, in our approach, this distance becomes a new random variable. We derive the relevant probability density function, apparently previously unknown and dubbed "Maccone distribution" by Paul Davies in 2008.
(4) ***Data Enrichment Principle.*** It should be noticed that ***ANY*** positive number of random variables in the SEH is compatible with the CLT. So, our generalization allows for many more factors to be added in the future as long as more refined scientific knowledge about each factor will be known to the scientists.

[☆] This paper was presented during the 60th IAC in Daejeon.

[*] Mail address: Via Martorelli 43, 10155 Torino (Turin), Italy. Tel.: +39 0 112055387.
E-mail addresses: clmaccon@libero.it, claudio.maccone@iaamail.org
URL: http://www.maccone.com

0094-5765/$ - see front matter
doi:10.1016/j.actaastro.2010.06.010

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

This capability to make room for more future factors in the SEH we call the "Data Enrichment Principle", and we regard it as the key to more profound future results in the fields of Astrobiology and SETI.

(5) A practical example is then given of how our SEH works numerically. We work out in detail the case where each of the ten random variables is uniformly distributed around its own mean value as given by Dole back in 1964 and has an assumed standard deviation of 10%. The conclusion is that the average number of habitable planets in the Galaxy should be around 100 million ± 200 million, and the average distance in between any couple of nearby habitable planets should be about 88 light years ± 40 light years.

(6) Finally, we match our SEH results against the results of the Statistical Drake Equation that we introduced in our 2008 IAC presentation. As expected, the number of currently communicating ET civilizations in the Galaxy turns out to be much smaller than the number of habitable planets (about 10,000 against 100 million, i.e. one ET civilization out of 10,000 habitable planets). And the average distance between any two nearby habitable planets turns out to be much smaller than the average distance between any two neighboring ET civilizations: 88 light years vs. 2000 light years, respectively. This means an ET average distance about 20 times higher than the average distance between any couple of adjacent habitable planets.

1. Introduction to SETI

SETI is an acronym for "Search for Extra Terrestrial Intelligence".

SETI is a comparatively new branch of scientific research that began only in 1959.

The goal of SETI is to ascertain whether Alien Civilizations exist in the universe, how far from us they exist, and possibly how much more advanced than us they may be.

As of 2010, the only physical tools we know that could help us get in touch with Aliens are the electromagnetic waves that an Alien Civilization could emit and we could detect. This forces us to use the largest radiotelescopes on Earth for SETI research because the higher our collecting area of electromagnetic radiation, the higher our sensitivity, i.e. the further in space we can probe. Yet, even by using the largest radiotelescopes we have on Earth (the 305 m dish at Arecibo, for instance) we cannot search for Aliens beyond, say, a few hundred light years away. This is a very, very small amount of space around us within our Galaxy, the Milky Way, that is about a hundred thousand light years in diameter. Thus, current SETI can cover only a very tiny fraction of the Galaxy, and it is not surprising that in the past 50 years of SETI searches ***NO*** extraterrestrial civilization was discovered. Quite simply, we did not get far enough!

This demands the construction of much more powerful and radically new radiotelescopes. Rather than big and heavy metal dishes, whose mechanical problems hamper SETI research too much, we are now turning to "software radiotelescopes". These are made up by a large number of small dishes (ATA=Allen telescope array, and ALMA=Atacama large millimeter/submillimeter array) or even just of simple dipoles (LOFAR=low frequency array) using state-of-the-art electronics and very high-speed computing that can outperform the classical radiotelescopes in many regards. The final dream in this field is the SKA (=square kilometer array), currently being designed and expected to be completed around 2020.

2. The key question: how far are they?

But still, the key question remains: how far are they?

Or, more correctly, how far do we expect the ***NEAREST*** extraterrestrial civilization to be from the Solar System in the Galaxy?

This question was first faced in a scientific manner back in 1961 by the same scientist who also was the first experimental SETI radio astronomer ever: the American, Frank Donald Drake (born 1930). He first considered the shape and size of the Galaxy where we are living: the Milky Way. This is a spiral galaxy measuring some 100,000 light years in diameter and some 1600 light years in thickness of the Galactic Disk at half-way from its center. That is:

(1) The diameter of the Galaxy is (about) 100,000 light years (abbreviated ly), i.e., its radius, R_{Galaxy}, is about 50,000 ly.

(2) The thickness of the Galactic Disk at half-way from its center, h_{Galaxy}, is about 1600 ly. Then

(3) The volume of the Galaxy may then be approximated as the volume of the corresponding cylinder, i.e.

$$V_{Galaxy} = \pi\, R_{Galaxy}^2\, h_{Galaxy}. \tag{1}$$

(4) Now consider the sphere around us having as radius the half distance in between us and the nearest ET Civilization. Its volume is given by

$$V_{Our\_Sphere} = \frac{4}{3}\pi \left(\frac{\text{ET\_Distance}}{2}\right)^3 \tag{2}$$

In the last equation, we had to divide the distance "ET_Distance" between ourselves and the nearest ET

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

civilization by 2 because we are now going to make the unwarranted assumption that ***all ET Civilizations are equally spaced from each other in the Galaxy!*** This is a crazy assumption, clearly, and should be replaced by more scientifically grounded assumptions as soon as we know more about our Galactic neighborhood. At the moment, however, this is the best guess that we can make, and so we shall take it for granted, although we are aware that this is a weak point in the reasoning.

Furthermore, let us denote by N the total number of civilizations now living in the Galaxy, including ourselves. Of course, this number N is unknown. We only know that $N \geq 1$ since one civilization does at least exist!

Having thus assumed that ET civilizations are *uniformly spaced in the Galaxy*, we can then write down the proportion:

$$\frac{V_{Galaxy}}{N} = \frac{V_{Our\_Sphere}}{1}. \tag{3}$$

That is, upon replacing both (1) and (2) into (3):

$$\frac{\pi R^2_{Galaxy} h_{Galaxy}}{N} = \frac{\frac{4}{3}\pi (\text{ET\_Distance}/2)^3}{1}. \tag{4}$$

The last equation contains two unknowns: N and ET_Distance, and so we do not know which one it is better to solve for.

However, we may suppose that, by resorting to the (rather uncertain) knowledge that we have about the Evolution of the Galaxy through the last 10 billion years or so, we might somehow compute an approximate value for N.

Then, we may solve (4) for ET_Distance thus obtaining the (average) distance between any pair of neighboring civilizations in the Galaxy (Distance Law)

$$\text{ET\_Distance}(N) = \frac{\sqrt[3]{6\, R^2_{Galaxy}\, h_{Galaxy}}}{\sqrt[3]{N}} = \frac{C}{\sqrt[3]{N}}, \tag{5}$$

where the positive constant C is defined by

$$C = \sqrt[3]{6\, R^2_{Galaxy}\, h_{Galaxy}} \approx 28{,}845 \text{ light years}. \tag{6}$$

Eqs. (5) and (6) are the starting point to understand the origin of the Drake equation that we discuss in detail in Section 3 of this paper (see also Ref. [2]).

Let us just complete this section by pointing out three different numerical cases of the distance law (5):

(1) We know that we exist, so N may not be smaller than 1, i.e., $N \geq 1$. Suppose then that we are alone in the Galaxy, i.e., that $N=1$. Then the distance law (5) yields as distance to the nearest civilization from us just the constant C, i.e., 28,845 light years. This is about the distance in between ourselves and the center of the Galaxy (i.e. the Galactic Bulge). Thus, this result seems to suggest that, if we do not find any extraterrestrial civilization around us in these outskirts of the Galaxy where we live, we should look around the Galactic Center first. And this is indeed what is happening, i.e., many SETI searches are actually pointing the antennas towards the Galactic Center, looking for beacons (see, for instance Ref. [1]).

(2) Suppose next that $N=1000$, i.e. there are about a thousand extraterrestrial communicating civilizations in the whole Galaxy right now. Then the distance law (5) yields an average distance of 2885 light years. This is a distance that most radiotelescopes in Earth may not reach for SETI searches right now: hence the need to build larger radiotelescopes, like ALMA, LOFAR and the SKA.

(3) Suppose finally that $N=1{,}000{,}000$, i.e., there are a million communicating civilizations now in the Galaxy. Then the distance law (5) yields an average distance of 288 light years. This is within the (upper) range of distances that our current radiotelescopes may reach for SETI searches, and that justifies all SETI searches that have been done so far in the first fifty years of SETI (1960–2010).

In conclusion, interpolating the above three special cases of N, we may say that the distance law (5) yields the following key diagram of the average ET distance vs. the assumed number of communicating civilizations, N, in the Galaxy right now (Fig. 1). This we call the ***Distance Law***, i.e., the average distance (plot along the vertical axis in light years) versus the ***number*** of communicating civilizations ***assumed*** to exist in the Galaxy right now.

3. Computing *N* by virtue of the Drake equation (1961)

In the previous section, the problem of finding how close the nearest ET civilization may be was "solved" by reducing it to the computation of N, the total number of extraterrestrial civilizations now existing in this Galaxy. In this section the famous Drake equation is described, which was proposed back in 1961 by Frank Donald Drake to estimate the numerical value of N.

N can be written as the product or multiplication of a number of factors, each a kind of filter, every one of which must be sizable for there to be a large number of civilizations:

Ns,	the number of stars in the Milky Way Galaxy;
fp,	the fraction of stars that have planetary systems;
ne,	the number of planets in a given system that are ecologically suitable for life;
fl,	the fraction of otherwise suitable planets on which life actually arises;
fi,	the fraction of inhabited planets on which an intelligent form of life evolves;
fc,	the fraction of planets inhabited by intelligent beings on which a communicative technical civilization develops; and
fL,	the fraction of planetary lifetime graced by a technical civilization.

Written out, the equation reads

$$N = Ns \cdot fp \cdot ne \cdot fl \cdot fi \cdot fc \cdot fL \tag{7}$$

All of the *f*s are fractions, having values between 0 and 1; they will pare down the large value of *Ns*. To derive N we must estimate each of these quantities (see also Ref. [3]).

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

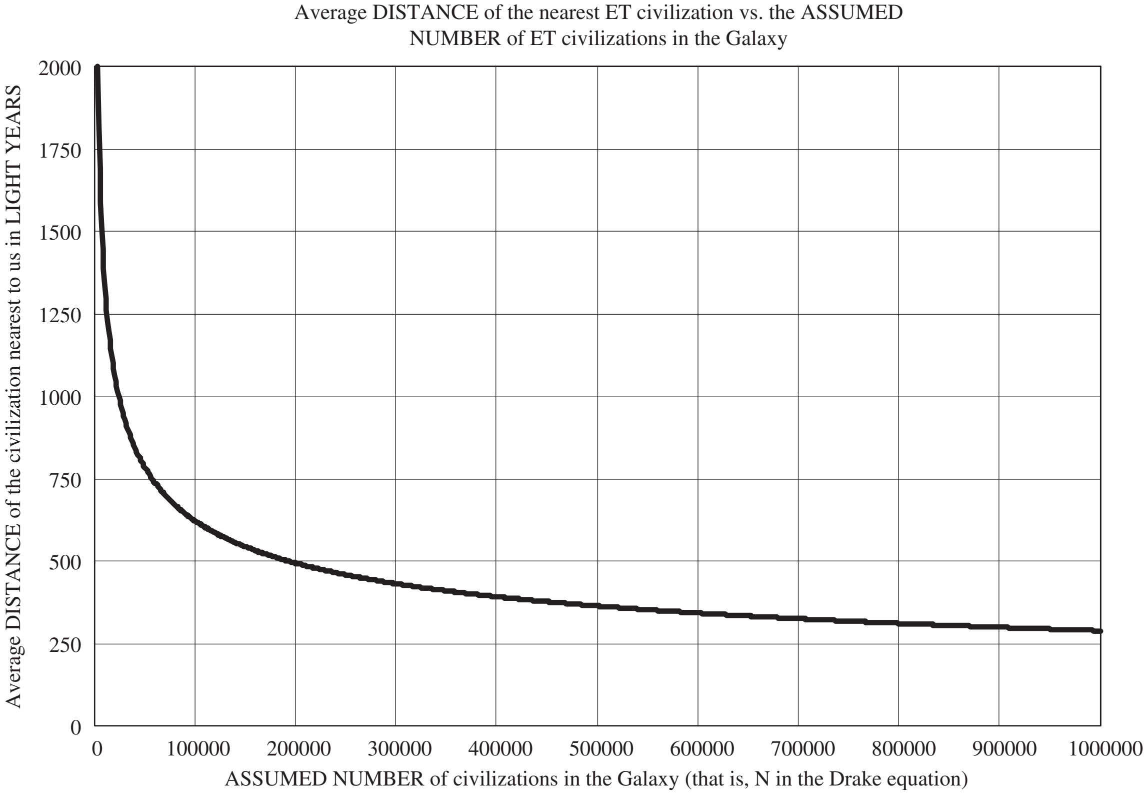

Fig. 1. ***Distance Law***, i.e., the average distance (plot along the vertical axis in light years) versus the number of communicating civilizations ***assumed*** to exist in the Galaxy right now.

4. The Drake equation is over-simplified

In the nearly fifty years (1961–2009) elapsed since Frank Drake proposed his equation, a number of scientists and writers tried to find out which numerical values of its seven independent variables are more realistic in agreement with our present-day knowledge. Thus there is a considerable amount of literature about the Drake equation nowadays, and, as one can easily imagine, the results obtained by the various authors largely differ from one another. In other words, the value of N, that various authors obtained by different assumptions about the astronomy, the biology and the sociology implied by the Drake equation, may range from a few tens (in the pessimist's view) to some million or even billion in the optimist's opinion. A lot of uncertainty is thus affecting our knowledge of N as of 2010. In all cases, however, the final result about N has always been a sheer number, i.e., a positive integer number ranging from 1 to million or billion. This is precisely the aspect of the Drake equation that this author regarded as "too simplistic" and improved mathematically in his paper #IAC-08-A4.1.4, entitled "The Statistical Drake Equation" and presented on October 1st, 2008, at the 59th International Astronautical Congress (IAC) held in Glasgow, Scotland, UK, September 29th–October 3rd, 2008, that is Refs. [4,5].

5. The statistical Drake equation by Maccone (2008)

We start by an example.

Consider the first independent variable in the Drake equation (7), i.e., Ns, the number of stars in the Milky Way Galaxy. Astronomers tell us that *approximately* there should be about 350 billion stars in the Galaxy. Of course, nobody has counted (or even seen in the photographic plates) *all* the stars in the Galaxy! There are too many practical difficulties preventing us from doing so: just to name one, the dust clouds that do not allow us to see even the Galactic Bulge (i.e. the central region of the Galaxy) in the visible light (although we may "see it" at radio frequencies like the famous neutral hydrogen line at 1420 MHz). So, it does not make any sense to say that $Ns = 350 \times 10^9$, or, say (even worse) that the number of stars in the Galaxy is (say) 354,233,321,123, or similar fanciful exact integer numbers. That is just silly and non-scientific. Much more scientific, on the contrary, is to say that the number of stars in the Galaxy is 350 billion plus or minus, say, 50 billion (or whatever values the astronomers may regard as more appropriate, since this is just an example to let the reader understand the difficulty).

Thus, it makes sense to ***replace each of the seven independent variables in the Drake equation (7) by a***

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

mean value (350 billion, in the above example) plus or minus a certain standard deviation (50 billion, in the above example).

By doing so, we have made a great step ahead: we have abandoned the too-simplistic equation (7) and replaced it by something more sophisticated and scientifically more serious: the ***statistical*** Drake equation. In other words, we have transformed the classical and simplistic Drake equation (7) into an advanced statistical tool for the investigation of a host of facts hardly known to us in detail. In other words still:

(1) We replace each independent variable in (7) by a ***positive random variable***, labelled D_i (from Drake).
(2) We assume that the *mean value* of each D_i is the same numerical value previously attributed to the corresponding algebraic independent variable in (7).
(3) But now we also ***add a standard deviation*** σ_{D_i} on each side of the mean value, that is provided by the knowledge gathered by scientists in each discipline encompassed by each D_i.

Having so done, the next question is:

How can we find out the ***probability distribution*** for each D_i?

For instance, shall that be a Gaussian, or what?

This is a difficult question, for nobody knows, for instance, the probability distribution of the number of stars in the Galaxy, not to mention the probability distribution of the other six variables in the Drake equation (7).

There is a brilliant way to get around this difficulty, though.

We start by excluding the Gaussian because each variable in the Drake equation is a ***positive*** (or, more precisely, a non-negative) random variable, while the Gaussian applies to *real* random variables only. So, the Gaussian is out. Then, one might consider the large class of well-studied and positive probability densities called "the gamma distributions," but it is then unclear why one should adopt the gamma distributions and not any other one. The solution to this apparent conundrum comes from Shannon's Information Theory and a theorem that he proved in 1948: "The probability distribution having maximum entropy (=uncertainty) over any ***finite*** range of real values is the ***uniform*** distribution over that range".

So, at this point, we assume that each of the seven D_i in (7) is a uniform random variable, whose mean value and standard deviation is known by the scientists working in the respective field (let it be astronomy, or biology, or sociology). Notice that, for such a uniform distribution, the knowledge of the mean value μ_{D_i} and of the standard deviation σ_{D_i} automatically determines the RANGE of that random variable in between its lower (called a_i) and upper (called b_i) limits: in fact these limits are given by the equations:

$$\begin{cases} a_i = \mu_{D_i} - \sqrt{3}\,\sigma_{D_i} \\ b_i = \mu_{D_i} + \sqrt{3}\,\sigma_{D_i} \end{cases} \quad (8)$$

(the "surprising" factor $\sqrt{3}$ in the above equations comes from the definitions of mean value and standard deviation: please see Eqs. (12), (15) and (17) in Refs. [4,5] for the relevant proof). So the uniform distribution of each random variable D_i is perfectly determined by its mean value and standard deviation, and so are all its other properties.

The next problem is the following:

OK, since we now know everything about each uniformly distributed D_i, what is the probability distribution of N, given that N is the product (7) of all the D_i?

In other words, not only do we want to find the analytical expression of the probability density function of N, but we also want to relate its mean value μ_N to all mean values μ_{D_i} of the D_i, and its standard deviation σ_N to all standard deviations σ_{D_i} of the D_i.

This is a difficult problem.

It occupied the author's mind for no less than about ten years (1997–2007).

It is actually an *analytically unsolvable* problem, in that, to the best of this author's knowledge, it is *impossible* to find an analytic expression for any *finite product* of uniform random variables D_i. This result is proven in Section 2–3.3 of Refs. [4,5] (unfortunately!).

6. Solving the statistical Drake equation by virtue of the central limit theorem (CLT) of statistics

The solution to the problem of finding the analytical expression for the probability density function of N in the statistical Drake equation was found by this author only in September 2007. The key steps are the following:

(1) Take the natural logs of both sides of the statistical Drake equation (7). This changes the product into a sum.
(2) The mean values and standard deviations of the logs of the random variables D_i may all be expressed analytically in terms of the mean values and standard deviations of the D_i.
(3) Recall the Central Limit Theorem (CLT) of Statistics, stating that (loosely speaking) if you have a *sum* of independent random variables, each of which is *arbitrarily distributed* (hence, also including uniformly distributed), then, when the number of terms in the sum increases indefinitely (i.e. for a sum of random variables infinitely long)… *the sum random variable tends to a Gaussian.*
(4) Thus, the natural log of N tends to a Gaussian.
(5) Thus, ***N tends to the lognormal distribution.***
(6) The mean value and standard deviations of this lognormal distribution of N may all be expressed analytically in terms of the mean values and standard deviations of the logs of the D_i already found previously.

This result is fundamental.

All the relevant equations are summarized in Table 1. This table is actually the same as Table 2 of the author's original paper (Ref. [4]) IAC-08-A4.1.4, entitled "The Statistical Drake Equation" and presented by him at the

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

Table 1
Summary of the properties of the lognormal distribution that applies to the random variable N=number of ET communicating civilizations in the Galaxy.

Random variable	N=number of communicating ET civilizations in Galaxy
Probability distribution	Lognormal
Probability density function	$f_N(n) = \frac{1}{n}\,\frac{1}{\sqrt{2\pi}\sigma}\,e^{-(\ln(n)-\mu)^2/(2\sigma^2)} \quad (n \ge 0)$
Mean value	$\langle N \rangle = e^{\mu}\,e^{\sigma^2/2}$
Variance	$\sigma_N^2 = e^{2\mu}\,e^{\sigma^2}\,(e^{\sigma^2}-1)$
Standard deviation	$\sigma_N = e^{\mu}\,e^{\sigma^2/2}\sqrt{e^{\sigma^2}-1}$
All the moments, i.e. *k*th moment	$\langle N^k \rangle = e^{k\mu}\,e^{k^2\,\sigma^2/2}$
Mode (= abscissa of the lognormal peak)	$n_{\text{mode}} \equiv n_{\text{peak}} = e^{\mu}\,e^{-\sigma^2}$
Value of the Mode Peak	$f_N(n_{\text{mode}}) = \frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\mu}\,e^{\sigma^2/2}$
Median (= fifty–fifty probability value for N)	$\text{median} = m = e^{\mu}$
Skewness	$\frac{K_3}{(K_4)^{3/2}} = (e^{\sigma^2}+2)\sqrt{\frac{e^{-6\mu}\,e^{-3\sigma^2}}{(e^{\sigma^2}-1)^5\,(e^{3\sigma^2}+3e^{2\sigma^2}+6e^{\sigma^2}+6)^3}}$
Kurtosis	$\frac{K_4}{(K_2)^2} = e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-6$
Expression of μ in terms of the lower (a_i) and upper (b_i) limits of the Drake **uniform** input random variables D_i	$\mu = \sum_{i=1}^{7}\langle Y_i \rangle = \sum_{i=1}^{7}\frac{b_i[\ln(b_i)-1]-a_i[\ln(a_i)-1]}{b_i-a_i}$
Expression of σ^2 in terms of the lower (a_i) and upper (b_i) limits of the Drake **uniform** input random variables D_i	$\sigma^2 = \sum_{i=1}^{7}\sigma_{Y_i}^2 = \sum_{i=1}^{7}\left(1-\frac{a_i b_i[\ln(b_i)-\ln(a_i)]^2}{(b_i-a_i)^2}\right)$

International Astronautical Congress (IAC) held in Glasgow, UK, on October 1st, 2008. This original paper is reproduced just the same in Ref. [5].

To sum up, not only we have found that N approaches the completely known lognormal distribution for an *infinity* of factors in the statistical Drake equation (7), but we also paved the way to further applications by removing the condition that the number of terms in the product (7) must be *finite.*

This possibility of ***adding any number of factors in the Drake equation (7)*** was not envisaged, of course, by Frank Drake back in 1961, when “summarizing” the evolution of life in the Galaxy in *seven* simple *steps.* But today, the number of factors in the Drake equation should already be increased: for instance, there is no mention in the original Drake equation of the possibility that asteroidal impacts might destroy the life on Earth at any time, and this is because the demise of the dinosaurs at the K/T impact had not been yet understood by scientists in 1961, and was so only in 1980!

In practice, we are suggesting ***increasing*** the number of factors as much as necessary in order to get better and better estimates of N as long as our scientific knowledge increases. This we call the “Data Enrichment Principle” and believe should be the next important goal in the study of the statistical Drake equation.

Finally, we wish to provide a numerical example explaining how the statistical Drake equation works in the practice. This will be done in the next section.

7. An example explaining the statistical Drake equation

To understand how things work in practice for the statistical Drake equation, please consider Input Table 1. It is made up of three columns:

(1) The first column on the left lists the seven input sheer numbers that also become…
(2) The mean values (middle column).
(3) Finally the last column on the right lists the seven input standard deviations.

The bottom line is the classical Drake equation (7). We see that, for this particular set of seven inputs, the classical Drake equation (i.e. the product of the seven numbers) yields a total of 3500 communicating extraterrestrial civilizations existing in the Galaxy right now.

The statistical Drake equation, however, provides a much more articulated answer than just the above sheer number N=3500. In fact, a MathCad code written by this author and capable of performing all the numerical calculations required by the statistical Drake equation for a given set of seven input mean values plus seven input standard deviations, yields for N the lognormal distribution (thin curve) plotted in Fig. 2. We see immediately that the peak of this thin curve (i.e. the mode) falls at about $n_{\text{mode}} \equiv n_{\text{peak}} = e^{\mu}\,e^{-\sigma^2} \approx 250$ (this is Eq. (99) of Refs. [4,5]), while the median (fifty–fifty value splitting the lognormal density in two parts with equal undergoing areas) falls at about $n_{\text{median}} \equiv e^{\mu} \approx 1740$. These seem to be smaller values than N=3500 provided by the classical Drake equations, but it is a wrong impression due to a poor “intuitive” understanding of what statistics is! In fact, neither the mode nor the median are the “really important” values: the really important value for N is the *mean value!* Now if you look at the thin curve in Fig. 2 (i.e. the lognormal distribution arising from the Central Limit Theorem), you see that this curve has a *long tail on the right!* In other words, it does *not* immediately go down to nearly zero beyond the peak of the mode. Thus, when you actually compute the mean value, you should not be too surprised to find out that it equals $\langle N \rangle = e^{\mu}e^{\sigma^2/2} \approx 4589.559 \sim 4590$ communicating civilizations now in the Galaxy. This is the important number, and it is *higher* than the 3500 provided by the classical Drake equation. Thus, in conclusion, ***the statistical extension of the classical Drake equation that we made increases our hopes to find an extraterrestrial civilization!***

Even more so our hopes are increased when we go on to consider the standard deviation associated with the mean value 4590. In fact, the standard deviation is given by equation (97) of Refs. [4,5]. This yields $\sigma_N = e^{\mu}\,e^{\sigma^2/2}\sqrt{e^{\sigma^2}-1} = 11{,}195$ and so the expected number of N may actually be even much higher than the 4590 provided by the mean value alone! The “upper limit of the 1-sigma confidence interval” (as statisticians call it), i.e. the sum 4590+11195=15,785, yields a higher number still!

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

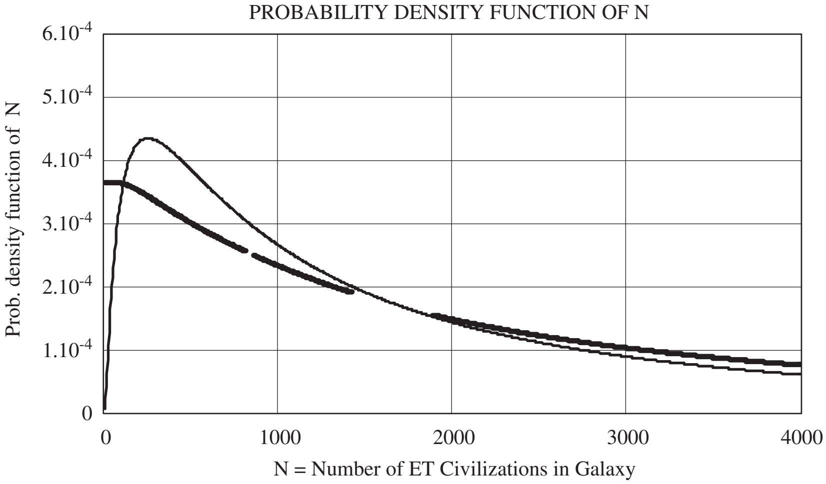

Fig. 2. Comparing the two probability density functions of the random variable N found:

(1) at the end of Section 3.3 in a purely numerical way and without resorting to the CLT at all (thick curve), and
(2) analytically using the CLT and the relevant lognormal approximation (thin curve).

Finally, the reader should not worry about the thick curve depicted in Fig. 2: it is just the *numerical* solution of the statistical Drake equation for a *finite* number of 7 input factors. Fig. 2 actually shows that this curve "is well interpolated" by the lognormal distribution (thin curve), i.e., by the neat analytical expression provided by the Central Limit Theorem for an *infinite* number of factors in the Drake equation. That is, in conclusion, Fig. 2 visually shows that taking 7 factors or an infinity of factors "is almost the same thing" already for a value as small as 7.

8. Finding the probability distribution of the ET-distance by virtue of the statistical Drake equation

Having solved the statistical Drake equation by finding the lognormal distribution, we are now in a position to solve the *ET-distance* problem by resorting to statistics again, rather than just to the purely deterministic Distance Law (5), as we did in Section 2. This is "scientifically more serious" than just the purely deterministic Distance Law (5) inasmuch as the new statistical Distance Law will yield a ***probability density for the Distance***, with the relevant mean value and standard deviation, of course. In other words, the Distance Law (5), now itself becomes a random variable whose probability distribution, mean value and standard deviation must be computed by "replacing" (so to say) into (5) the fact that N is now known to follow the lognormal distribution. This was done by this author in September 2007 also, and is mathematically described in detail in Section 7 of Refs. [4,5].

So, the important new result is the ***probability density for the distance, that the well-known physicist Paul Davies dubbed "the Maccone distribution" and whose equation reads***

$$f_{\text{ET\_Distance}}(r) = \frac{3}{r}\,\frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\frac{\left(\ln\left[\frac{6\,R_{Galaxy}^2\,h_{Galaxy}}{r^3}\right]-\mu\right)^2}{2\,\sigma^2}} \tag{9}$$

and holds for $r \geq 0$. This is Eq. (114) of Refs. [4,5].

Starting from this equation, the author computed the ***mean value of the random variable ET_Distance***

$$\langle \text{ET\_Distance} \rangle = C\,e^{-\mu/3}\,e^{\sigma^2/18} \tag{10}$$

which is Eq. (119) of Refs. [4,5], and finally the ***ET_Distance standard deviation***

$$\sigma_{\text{ET\_Distance}} = C\,e^{-\mu/3}e^{\sigma^2/18}\sqrt{e^{\sigma^2/9}-1} \tag{11}$$

which is Eq. (123) of Refs. [4,5]. Of course, all other descriptive statistical quantities, such as moments, cumulants, etc. can be computed upon starting from the probability density (9), and the result is Table 2 hereafter, that is Table 3 of Refs. [4,5].

Finally, we wish to complete this section, as well as this "easy introduction to the statistical Drake equation," by pointing out the numerical values that Eqs. (10) and (11) yield for the Input Table 1. They are, respectively:

$$r_{mean\_value} = C\,e^{-\mu/3}\,e^{\sigma^2/18} \approx 2{,}670 \ \text{light years} \tag{12}$$

which is Eq. (153) of Ref. [4], and

$$\sigma_{\text{ET\_Distance}} = C\,e^{-\mu/3}\,e^{\sigma^2/18}\sqrt{e^{\sigma^2/9}-1} \approx 1309 \ \text{light years} \tag{13}$$

which is Eq. (154) of Refs. [4,5].

It is actually clarifying to draw the graph of the ET_Distance probability density (9), that is Fig. 3.

From Fig. 3 we see that the probability of finding extraterrestrials is practically zero up to a distance of

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

Table 2
Summary of the properties of the probability distribution that applies to the random variable ET_Distance yielding the (average) distance between any two neighboring communicating civilizations in the Galaxy.

Random variable	ET_Distance between any two neighboring ET civilizations in Galaxy assuming they are UNIFORMLY distributed throughout the whole Galaxy volume.
Probability distribution	Unnamed (Paul Davies suggested “Maccone distribution”)
Probability density function	$f_{\text{ET\_distance}}(r) = \frac{3}{r}\,\frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\left(\ln\left[\frac{6R_{Galaxy}^2\,h_{Galaxy}}{r^3}\right]-\mu\right)^2/(2\,\sigma^2)}$
Numerical constant C related to the Milky Way size	$C = \sqrt[3]{6R_{Galaxy}^2\,h_{Galaxy}} \approx 28{,}845$ light years
Mean value	$\langle \text{ET\_Distance} \rangle = C\,e^{-\mu/3}e^{\sigma^2/18}$
Variance	$\sigma^2_{\text{ET\_Distance}} = C^2 e^{-\frac{2}{3}\mu}e^{\sigma^2/9}\left(e^{\sigma^2/9}-1\right)$
Standard deviation	$\sigma_{\text{ET\_Distance}} = C\,e^{-\mu/3}e^{\sigma^2/18}\sqrt{e^{\sigma^2/9}-1}$
All the moments, i.e. kth moment	$\langle \text{ET\_Distance}^k \rangle = C^k e^{-k(\mu/3)}e^{k^2(\sigma^2/18)}$
Mode (= abscissa of the lognormal peak)	$r_{\text{mode}} \equiv r_{peak} = Ce^{-\mu/3}e^{-\sigma^2/9}$
Value of the Mode Peak	Peak Value of $f_{\text{ET\_Distance}}(r) = \equiv f_{\text{ET\_distance}}(r_{\text{mode}}) = \frac{3}{C\sqrt{2\pi}\,\sigma}e^{\mu/3}e^{\sigma^2/18}$
Median (= fifty–fifty probability value for N)	$\text{median} = m = Ce^{-\mu/3}$
Skewness	$\frac{K_3}{(K_4)^{3/2}} = \frac{e^{-\mu}(e^{\sigma^2/2}-3e^{5\sigma^2/18}+2e^{\sigma^2/6})}{C^3(e^{8\sigma^2/9}-4e^{5\sigma^2/9}-3e^{4\sigma^2/9}+12e^{\sigma^2/3}-6e^{2\sigma^2/9})^{3/2}}$
Kurtosis	$\frac{K_4}{(K_2)^2} = e^{4\sigma^2/9}+2e^{\sigma^2/3}+3e^{2\sigma^2/9}-6$
Expression of μ in terms of the lower (a_i) and upper (b_i) limits of the Drake **uniform** input random variables D_i	$\mu = \sum_{i=1}^{7}\langle Y_i \rangle = \sum_{i=1}^{7}\frac{b_i[\ln(b_i)-1]-a_i[\ln(a_i)-1]}{b_i-a_i}$
Expression of σ^2 in terms of the lower (a_i) and upper (b_i) limits of the Drake **uniform** input random variables D_i	$\sigma^2 = \sum_{i=1}^{7}\sigma_{Y_i}^2 = \sum_{i=1}^{7}\left(1-\frac{a_i b_i[\ln(b_i)-\ln(a_i)]^2}{(b_i-a_i)^2}\right)$

Input Table 1
Input values (i.e. mean values and standard deviations) for the seven Drake uniform random variables D_i. The first column on the left lists the seven input sheer numbers that also become the mean values (middle column). Finally the last column on the right lists the seven input standard deviations. The bottom line is the classical Drake equation (7).

$Ns{:}350 \times 10^9$	$\mu_{Ns}=Ns$	$\sigma_{Ns}=1 \times 10^9$
$fp = \frac{50}{100}$	$\mu_{fp}=fp$	$\sigma_{fp} = \frac{10}{100}$
$ne=1$	$\mu_{ne}=ne$	$\sigma_{ne} = \frac{1}{\sqrt{3}}$
$fl = \frac{50}{100}$	$\mu_{fl}=fl$	$\sigma_{fl} = \frac{10}{100}$
$fi = \frac{20}{100}$	$\mu_{fi}=fi$	$\sigma_{fi} = \frac{10}{100}$
$fc = \frac{20}{100}$	$\mu_{fc}=fc$	$\sigma_{fc} = \frac{10}{100}$
$fL = \frac{10000}{10^{10}}$	$\mu_{fL}=Fl$	$\sigma_{fL} = \frac{10}{100}$
$N{:}=Ns \cdot fp \cdot ne \cdot fl \cdot fi \cdot fc \cdot fL$	$N=3500$	

about 500 light years from Earth. Then it starts increasing with the increasing distance from Earth, and reaches its maximum at

$$r_{\text{mode}} \equiv r_{peak} = C\,e^{-\mu/3}\,e^{-\sigma^2/9} \approx 1933 \text{ light years.} \tag{14}$$

This is the most likely value of the distance at which we can expect to find the nearest extraterrestrial civilization.

It is *not*, as we said, the mean value of the probability distribution (9) for $f_{\text{ET\_Distance}}(r)$. In fact, the probability density (9) has an infinite tail on the right, as clearly shown in Fig. 3, and hence its mean value must be higher than its peak value. As given by (10) and (12), its mean value is $r_{mean\_value} = Ce^{-\mu/3}e^{\sigma^2/18} \approx$ ***2670 light years. This is the mean (value of the) distance at which we can expect to find extraterrestrials.***

After having found the above two distances (1933 and 2670 light years, respectively), the next natural question that arises is: “what is the range, back and forth around the mean value of the distance, within which we can expect to find extraterrestrials with “the highest hopes?”. The answer to this question is given by the notion of standard deviation that we already found to be given by (11) and (13),

$$\sigma_{\text{ET\_Distance}} = C\,e^{-\mu/3}e^{\sigma^2/18}\sqrt{e^{\sigma^2/9}-1} \approx 1309 \text{ light years.}$$

More precisely, this is the so-called 1-sigma (distance) level. Probability theory then shows that the nearest extraterrestrial civilization is expected to be located within this range, i.e. within the two distances of $(2670-1309)=1361$ light years and $(2670+1309)=3979$ light years, with probability given by the integral of $f_{\text{ET\_Distance}}(r)$ taken in between these two lower and upper limits, that is:

$$\int_{1361 \text{ light years}}^{3979 \text{ light years}} f_{\text{ET\_Distance}}(r)dr \approx 0.75 = 75\%. \tag{15}$$

In plain words: with 75% probability, the nearest extraterrestrial civilization is located in between the distances of 1361 and 3979 light years from us, having assumed the input values to the Drake Equation given by Input Table 1. If we change those input values, then all the numbers change again, of course.

9. The “Data Enrichment Principle” as the best CLT consequence upon the statistical Drake equation (any number of factors allowed)

As a fitting climax to all the statistical equations developed so far, let us now state our “Data Enrichment Principle.” It simply states that “The higher the number of factors in the statistical Drake equation, the better.”

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

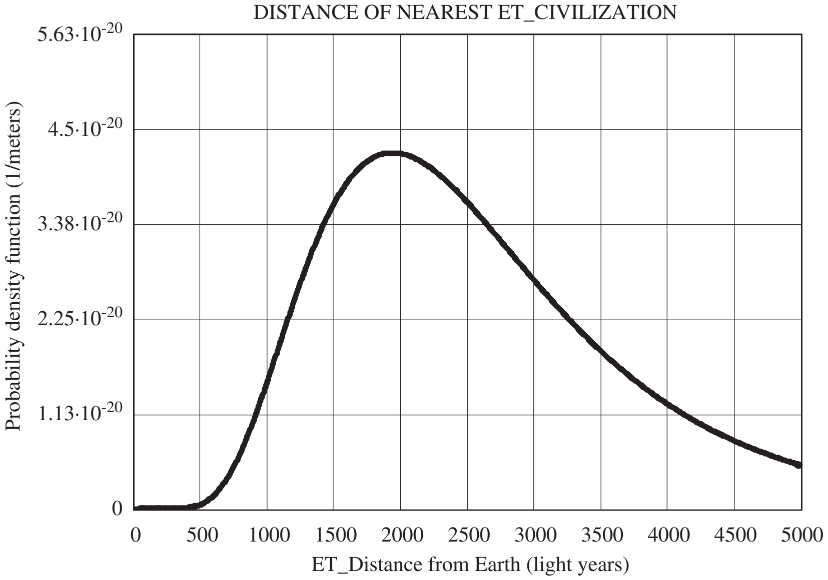

Fig. 3. This is the probability of finding the nearest extraterrestrial civilization at the distance r from Earth (in light years) if the values assumed in the Drake Equation are those shown in Input Table 1. The relevant probability density function $f_{\mathrm{ET\_Distance}}(r)$ is given by Eq. (9). Its mode (peak abscissa) equals 1933 light years, but its mean value is higher since the curve has a long tail on the right: the mean value equals in fact 2670 light years. Finally, the standard deviation equals 1309 light years: *this is good news for SETI, inasmuch as the nearest ET civilization might lie at just* 1 sigma=2670−1309=1361 light years *from us*.

Put in this simple way, it simply looks like a new way of saying that the CLT lets the random variable $Y=\ln(N)$ approach the normal distribution when the number of terms in the product (7) approaches infinity. And this is the case, indeed.

However, *our "Data Enrichment Principle" has more profound methodological consequences* that we cannot explain now, but hope to describe more precisely in one or more coming papers.

10. Habitable Planets for Man

Let us now change topics completely!

Rather than seeking for ETs in the Galaxy, we now seek for habitable planets for man in the Galaxy. How many are there? And how far from us is the nearest such a habitable planets?

These topics seem to have been faced "seriously" for the first time in 1964 by Stephen H. Dole, then with the Rand Corporation.

Back in 1964, only three years had elapsed since Frank Drake had made known his now famous Drake equation. Dole learned the lesson of the Drake equation perfectly, and in his now famous book entitled "Habitable Planets for Man" (Ref. [6]) ***Dole used the same mathematical structure as the Drake equation (7) in order to find the number of habitable planets for man in the Galaxy***. In other words, on page 82 of his book, he wrote the same mathematical thing as the Drake equation, but he applied it to habitable planets. Nowadays Dole's 1964 book can be freely downloaded from the Rand Corporation web site. The equation on page 82 we shall call "the classical Dole equation".

As we can see from Dole's book, the classical Dole equation is made up by TEN factors (instead of SEVEN factors as in the Drake equation):

$$N_{Hab} = Ns \cdot Pp \cdot Pi \cdot PD \cdot PM \cdot Pe \cdot PB \cdot PR \cdot PA \cdot PL. \qquad (16)$$

Here N_{Hab} is the total number of habitable planets for man in the Galaxy, and it is given by the product of the following TEN input numbers:

(1) *Ns* is the number of stars in the suitable mass range 0.35–1.43 solar masses (this is Dole's assumption about to the mass of "habitable stars").
(2) *Pp* is the probability that a given star has planets in orbit around it.
(3) *Pi* is the probability that the inclination of the planet's equator is correct for its orbital distance.
(4) *PD* is the probability that at least one planet orbits within an ecosphere.
(5) *PM* is the probability that the planet has a suitable mass, 0.4 to 2.35 Earth masses (again, this is Dole's assumption in this regard).
(6) *Pe* is the probability that the planet's orbital eccentricity is sufficiently low.
(7) *PB* is the probability that the presence of a second star has not rendered the planet uninhabitable.
(8) *PR* is the probability that the planet's rate of rotation is neither too fast nor too slow.
(9) *PA* is the probability that the planet is of the proper age.
(10) *PL* is the probability that, all astronomical conditions being proper, life has developed on the planet.

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

11. The statistical Dole equation

It is now natural to rename the above ten input variables of the classical Dole equation (16) as follows:

$$\begin{cases} D_1 = Ns \\ D_2 = Pp \\ D_3 = Pi \\ D_4 = PD \\ D_5 = PM \\ D_6 = Pe \\ D_7 = PB \\ D_8 = PR \\ D_9 = PA \\ D_{10} = PL \end{cases} \tag{17}$$

so that our classical Dole equation may be simply rewritten as

$$N_{Hab} = \prod_{i=1}^{10} D_i. \tag{18}$$

We now let (18) undergo exactly the same changes that we applied to the classical Drake equation (7). In other words:

(1) All the input variables on the right-hand side of (18) now become *positive random variables.*

(2) All these random variables are supposed to be *uniformly distributed* with assigned mean values μ_{D_i} and standard deviations σ_{D_i}. It can then be shown that assigning them actually amounts to assigning the lower and upper limits (a_i and b_i, respectively) of each uniform random variable D_i.

(3) As a consequence of these assumptions, the total number of habitable planets in the Galaxy, N_{Hab}, also becomes a random variable, that we already know to be lognormally distributed from our previous similar work about the Drake equation.

Thus, we may now call (18) the *statistical Dole equation*.

It is true that the classical Drake equation (7) and the classical Dole equation (16) have a different number of factors (7 and 10, respectively), but… frankly speaking, who cares? This perfectly in line with what we did already for the Drake equation, and so ***the number of factors in both (7) and (16) is totally irrelevant, thanks to the central limit theorem!***

12. The number of Habitable Planets for Man in the Galaxy follows the lognormal distribution

We now just repeat the same arguments developed for the Drake equation to immediately conclude that: ***the total number of habitable planets in the Galaxy follows the lognormal distribution given in Table 1.***

13. The distance between any two nearby habitable Planets follows the Maccone distribution

Again, we now just repeat the same arguments developed for the Drake equation to immediately conclude that ***the distance between any two nearby habitable planets follows the Maccone distribution given in Table 2.***

14. A numerical example: A some hundred million habitable planets exist in the Galaxy!

We just need to complete this paper by giving a numerical example of how our Statistical Dole equation (18) works, and this we will do in the present section.

Consider Input Table 2. This is in principle comparable to Input Table 1 for the Statistical Drake equation. In fact, the arguments developed by Dole in Chapter 5 of Ref. [6] do provide the mean values of each D_i, but only such mean values, and not the relevant standard deviations, of course.

To set up a working example of the Statistical Dole Equation, however, we must assign the ten standard deviations also, that were not given by Dole and are unknown to this author from the current scientific literature about these matters.

No problem. In order to cut short, this author thus simply assigned the value of 1/10 (i.e. 10%) to each of the ten standard deviations listed in Input Table 2, and Input Table 2 is now complete.

Having assumed all the values listed in Input Table 2 as the input values, a new (unpublished) MathCad code was created by this author for the Statistical Dole Equation. For the input values of Input Table 2, this code yielded the results described hereafter.

First of all, the lognormal probability density for the random variable N_{Hab} is shown in Fig. 4. We see that the peak (i.e. the mode) corresponds to about ten million planets, but the tail is rather long.

Input Table 2

Input values (i.e. mean values and standard deviations) for the ten Dole uniform random variables D_i. The first column on the left lists the ten input sheer numbers that also are the mean values (middle column). The last column on the right lists the ten input standard deviations. The bottom line is the classical Dole equation (16). So, the number of habitable planets in the Galaxy, given by the classical Dole equation just as a sheer number, is 35 million 171 hundred thousand and 930.

$Ns = 60448.10^8$	$\mu_{Ns} := Ns$	$\sigma_{Ns} = 1 \times 10^7$
$Pp = 1.0$	$\mu_{Pp} = \mathrm{Pp}$	$\sigma_{Pp} = \frac{10}{100}$
$Pi = 0.81$	$\mu_{Pi} = Pi$	$\sigma_{Pi} = \frac{10}{100}$
$PD = 0.63$	$\mu_{PD} = PD$	$\sigma_{PD} = \frac{10}{100}$
$PM = 0.19$	$\mu_{PM} = PM$	$\sigma_{PM} = \frac{10}{100}$
$Pe = 0.94$	$\mu_{Pe} = Pe$	$\sigma_{Pe} = \frac{10}{100}$
$PB = 0.95$	$\mu_{PB} = PB$	$\sigma_{PB} = \frac{10}{100}$
$PR = 0.9$	$\mu_{PR} = PR$	$\sigma_{PR} = \frac{10}{100}$
$PA = 0.7$	$\mu_{PA} = PA$	$\sigma_{PA} = \frac{10}{100}$
$PL = 1$	$\mu_{PL} = PL$	$\sigma_{PL} = \frac{10}{100}$

$N_{Hab} = Ns \cdot Pp \cdot Pi \cdot PD \cdot PM \cdot Pe \cdot PB \cdot PR \cdot PA \cdot PL$

$N_{Hab} := 3.5171930508624 \times 10^7$

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

To quantify these remarks, let us first point out that the author's MathCad code yields the following numerical values for the two parameters μ and σ given by the last two rows in both Tables 1 and 2:

$$\begin{cases} \mu_{Hab} = 1.76{,}268{,}289{,}631{,}314 \times 10^{1} \\ \sigma_{Hab} = 1.27{,}010{,}132{,}908{,}265 \times 10^{0}. \end{cases} \tag{19}$$

Then, the mean value of the random variable N_{Hab}, given by the fourth row in Table 1, is given by

$$\langle N_{Hab} \rangle = e^{\mu_{Hab}}\, e^{\sigma_{Hab}^{2}/2} = 1.012 \times 10^{8} \approx 100 \text{ million}. \tag{20}$$

In other words, our statistical (and thus more serious, scientifically speaking) treatment of the Dole equation yields 100 million expected habitable planets in the Galaxy.

This figure is higher than the 35 million given by the classical Dole equation, and much higher than the value of the mode (10 million) shown by the lognormal curve in Fig. 5.

The last result, stating that there are about 100 million habitable planets in the Galaxy, is of course good news for the future "human conquest of the Galaxy" (if there will ever be one!), since it raises to 100 million the expected number of "Earths" to land on!

But what about the standard deviation around the mean value given by (20)? Table 1, row 6, shows that such a standard deviation of the random variable N_{Hab} is given by

$$\sigma_{N_{Hab}} = e^{\mu_{Hab}}\, e^{\sigma_{Hab}^{2}/2} \sqrt{e^{\sigma_{Hab}^{2}} - 1} = 2.0 \times 10^{8} \approx 200 \text{ million}. \tag{21}$$

In other words, the standard deviation of the number of habitable planets is 200 million. And so, with probability 1-sigma, we might expect the actual number of habitable planets to rise up 100 million plus 200 million=300 million.

Finally, the median (fifty–fifty probability of the lognormal distribution shown in Fig. 4) yields a value of

$$\text{median} = m = e^{\mu_{Hab}} = 4.521 \times 10^{7} \approx 45 \text{ million}. \tag{22}$$

15. Distance (Maccone) distribution of the nearest habitable planet to us according to the previous numerical inputs

Next comes the ***distance*** distribution of the nearest habitable planet to us (of course under the easy hypothesis that the distribution of habitable planets in the Galaxy is *uniform*). Well, from the third row of Table 2 it follows that the relevant probability density is given by the Maccone distribution, and this is plotted in Fig. 5.

The mean value of the Maccone distribution is given by the fifth row in Table 2, that is, for the data given by the Input Table 2:

$$\langle \text{Hab\_Distance} \rangle = C\, e^{-\mu_{Hab}/3}\, e^{\sigma_{Hab}^{2}/18} = 8.8 \times 10^{1} \text{ ly} \approx 88 \text{ ly}. \tag{23}$$

The relevant standard deviation is given by the seventh row in Table 2, and reads

$$\sigma_{\text{Hab\_Distance}} = C\, e^{-\mu_{Hab}/3}\, e^{\sigma_{Hab}^{2}/18} \sqrt{e^{\sigma_{Hab}^{2}/9} - 1} = 3.9 \times 10^{1} \text{ly} \approx 40 \text{ ly}. \tag{24}$$

Thus, with probability 1 sigma, it should not be hopeless to expect a detection of a habitable planet even at, say, just 88 – 40=48 ~ 50 light years from us.

16. Comparing the statistical Dole and Drake equations : number of Habitable Planets vs. number of ET civilizations in this Galaxy

It is now appropriate to make a comparison between the number of habitable planets and the number of expected ET civilizations in the Galaxy.

In other words, we want the "get the feeling" of the numbers that we have worked out in this paper just to see if the comparison among them "makes sense".

This we can do by putting on a same table

1) the mean value and standard deviation of the total number of both habitable planets and ET civilizations, and

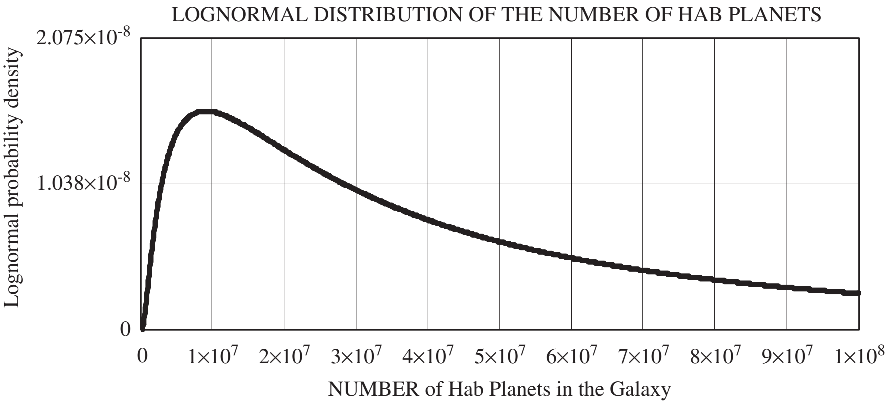

Fig. 4. The lognormal probability density of the overall NUMBER of habitable planets in the Galaxy as described in Stephen H. Dole's book "Habitable Planets for Man", first edition published in 1964 (Ref. [6]), and implemented by assigning a 10% standard deviation to all the ten input random variables listed in Input Table 2.

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

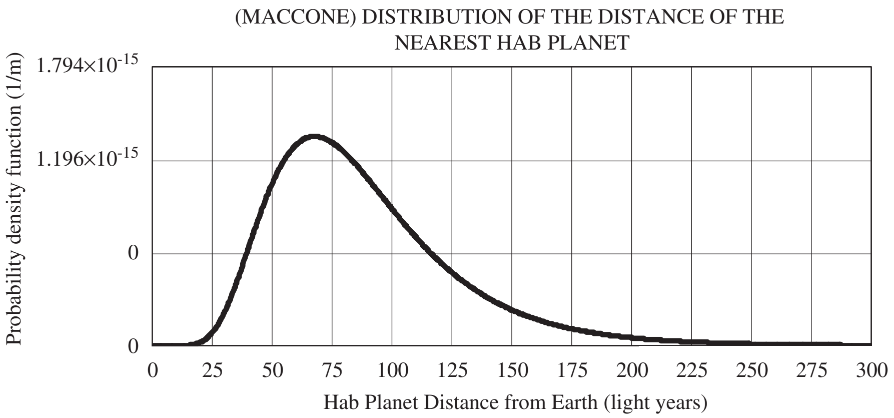

Fig. 5. The Maccone probability distribution of the distance of the nearest habitable planet to us in the Galaxy for the data of the Input Table 2 assumed as inputs to the statistical Dole equation (18). A glance to this plot immediately reveals that it is "hopeless" to expect to detect a habitable planet at distances smaller than 25 light years from us, since the value of the Maccone distribution is practically zero at such distances. Thus, future Interstellar Spacecraft designers should keep this lower bound in mind wished they land on habitable planets, rather than just on "any Planet". Also, the curve reaches its peak (mode) at about 67 light years from us, its mode (fifty–fifty probability) at about 80 light years and, above all, its mean value at 88 light years from us. The relevant standard deviation turns out to be about 40 light years, since the distribution tail is rather "short".

Table 3
Comparing the results of the Statistical Dole and Drake equation found by inputting to them the Input Tables 2 and 1, respectively.

	Statistical Dole equation	Statistical Drake equation
Mean Value of the *Total Number* of	Habitable planets in the Galaxy ~ 100 million	ET civilizations in the Galaxy ~ 4590
Standard Deviation of the *Total Number* of	Habitable planets in the Galaxy ~ 200 million	ET civilizations in the Galaxy ~ 11195
Mean Value of the *Distance* of	Nearest habitable planet ~ 88 light years	Nearest ET civilization ~ 2670 light years
Standard Deviation of the *Distance* of	Nearest habitable planet ~ 40 light years	Nearest ET civilization ~ 1309 light years

2) the mean value and standard deviation of their respective distances from us (of course, under the hypothesis that both of them are uniformly scattered throughout the Galaxy).

The result is Table 3, clearly showing that how much "more rare" the ET civilizations are with respect to the habitable planets. Roughly, one has:

$$\frac{\langle N_{Hab} \rangle}{\langle N_{ET} \rangle} = \frac{100\ million}{4950} \approx 20{,}202 \tag{25}$$

so that the habitable planets seem to 20,000 more frequent than ET civilizations, or, if you wish, only one ET civilization emerges out of 20,000 habitable planets.

As for the distances, the ratio is the other way round:

$$\frac{\langle \text{Hab\_Distance} \rangle}{\langle \text{ET\_Distance} \rangle} = \frac{2670\ \text{ly}}{88\ \text{ly}} \approx 30.340 \tag{26}$$

meaning that ETs are, on the average, 30 times further out that habitable planets.

And all these results, however, are just statistical, of course!

17. SEH, the "Statistical Equation for Habitables" is just the statistical Dole equation

So far we have referred to (18) as to the Statistical Dole equation. In view of further improvements in the mathematical analysis of this equation, however, it appears to be suitable to rename it "SEH", an acronym standing for ***"Statistical Equation for Habitables"***. This will be clear in the future papers by the author, where a number possibly higher than ten will be the new number of independent, uniform random variables describing the equation inputs.

These topics have to be deferred to a further paper, though.

18. Conclusions

We have sought to extend both the classical Drake and Dole equations to let them encompass statistics and probability.

This approach appears to pave the way to future, more profound investigations intended not only to associate "error bars" to each factor in the equation, but especially to increase the number of factors themselves. In fact, this seems to be the only way to incorporate into the equations more and more new scientific information as soon as it becomes available. In the long run, our Statistical equations might just become a huge computer code, growing in size and especially in the depth of the scientific information it contained. It would thus be Humanity's first "Encyclopaedia Galactica."

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010

Unfortunately, to extend the Drake equation to Statistics, it was necessary to use a mathematical apparatus that is more sophisticated than just the simple product of seven numbers.

When this author had the honour and privilege to present his results at the SETI Institute on April 11th, 2008, in front of an audience also including Professor Frank Drake, he felt he had to add these words of apology to him: "My apologies, Frank, for disrupting the beautiful simplicity of your equation."

Acknowledgements

The author is grateful to the leader of the Tau Zero Foundation, Marc Millis, and to Drs. Hal Puthoff and Eric Davis of the Institute for Advanced Study at Austin for the reading the author's paper #IAC-08-A4.1.4 and letting him have their appreciated comments. Also, the author is grateful to Drs. Jill Tarter, Paul Davies, Seth Shostak, Doug Vakoch, Tom Pierson, Carol Oliver, Paul Shuch and Kathryn Denning for attending his first presentation ever about these topics at the "Beyond" Center of the University of Arizona at Phoenix on February 8, 2008. He also would like to thank Dr. Dan Werthimer and his School of SETI experts at the Berkeley Space Sciences Laboratory for keeping alive the interplay between experimental and theoretical SETI.

But the greatest "thanks" go of course to the Teacher to all of us: Professor Frank D. Drake, whose equation opened to Humanity a new way of thinking about the past and the future of Humans in the Galaxy.

References

[1] Benford Gregory, Benford Jim, Benford Dominic, Cost optimized interstellar beacons: SETI, arXiv.org web site, 22 October 2008.

[2] Jeffrey Bennet, Shostak Seth, in: Life in the Universe, second edition, Pearson—Addison Wesley, San Francisco, 2007 see in particular p. 404.

[3] Carl Sagan, in: Cosmos, Random House, New York, 1983 See in particular, pp. 298–302.

[4] Claudio Maccone, The Statistical Drake Equation, paper #IAC-08-A4.1.4 presented on October 1, 2008, at the 59th International Astronautical Congress (IAC) held in Glasgow, Scotland, UK, September 29–October 3 2008.

[5] Claudio Maccone, The Statistical Drake Equation, in press, Acta Astronautica, 2010, in the Proceedings of the First IAA Symposium on "Searching for Life Signatures" held at UNESCO, Paris, September 22–26, 2008.

[6] Stephen H. Dole, Habitable planets for man, first edition, 1964, © 1964 by the RAND Corporation, Library of Congress Catalogue Card Number 64-15992. See in particular p. 82, i.e. the beginning of Chapter 5, entitled "Probability of Occurrence of habitable planets".

Please cite this article as: C. Maccone, SETI and SEH (Statistical Equation for Habitables), Acta Astronautica (2010), doi:10.1016/j.actaastro.2010.06.010