New Evo-SETI results about civilizations and molecular clock

Claudio Maccone

International Academy of Astronautics (IAA) and Istituto Nazionale di Astrofisica (INAF), Via Martorelli, 43 – Torino (Turin) 10155, Italy. E-mails: clmaccon@libero.it and claudio.maccone@iaamail.org

Abstract: In two recent papers (Maccone 2013, 2014) as well as in the book (Maccone 2012), this author described the Evolution of life on Earth over the last 3.5 billion years as a lognormal stochastic process in the increasing number of living Species. In (Maccone 2012, 2013), the process used was 'Geometric Brownian Motion' (GBM), largely used in Financial Mathematics (Black-Sholes models). The GBM mean value, also called 'the trend', always is an exponential in time and this fact corresponds to the so-called 'Malthusian growth' typical of population genetics. In (Maccone 2014), the author made an important generalization of his theory by extending it to lognormal stochastic processes having an *arbitrary* trend $m_L(t)$, rather than just a simple exponential trend as the GBM have.

The author named 'Evo-SETI' (Evolution and SETI) his theory inasmuch as it may be used not only to describe the full evolution of life on Earth from RNA to modern human societies, but also the possible evolution of life on exoplanets, thus leading to SETI, the current Search for ExtraTerrestrial Intelligence. In the Evo-SETI Theory, the life of a living being (let it be a cell or an animal or a human or a Civilization of humans or even an ET Civilization) is represented by a *b*-lognormal, i.e. a lognormal probability density function starting at a precise instant *b* ('birth') then increasing up to a peak-time *p*, then decreasing to a senility-time *s* (the descending inflexion point) and then continuing as a straight line down to the death-time *d* ('finite *b*-lognormal').

(1) Having so said, the present paper describes the further mathematical advances made by this author in 2014–2015, and is divided in two halves: Part One, devoted to new mathematical results about the History of Civilizations as *b*-lognormals, and

(2) Part Two, about the applications of the Evo-SETI Theory to the Molecular Clock, well known to evolutionary geneticists since 50 years: the idea is that our EvoEntropy grows *linearly* in time just as the molecular clock.

  (a) Summarizing the new results contained in this paper: In Part One, we start from the History Formulae already given in (Maccone 2012, 2013) and improve them by showing that it is possible to determine the *b*-lognormal not only by assigning its birth, senility and death, but rather by assigning birth, *peak* and death (BPD Theorem: no assigned senility). This is precisely what usually happens in History, when the life of a VIP is summarized by giving birth time, death time, and the date of the peak of activity in between them, from which the senility may then be calculated (approximately only, not exactly). One might even conceive a *b*-scalene (triangle) probability density just centred on these three points (*b, p, d*) and we derive the relevant equations. As for the uniform distribution between birth and death only, that is clearly the minimal description of someone's life, we compare it with both the *b*-lognormal and the *b*-scalene by comparing the Shannon Entropy of each, which is the measure of how much information each of them conveys. Finally we prove that the Central Limit Theorem (CLT) of Statistics becomes a new 'E-Pluribus-Unum' Theorem of the Evo-SETI Theory, giving formulae by which it is possible to find the *b*-lognormal of the History of a Civilization *C* if the lives of its Citizens C_i are known, even if only in the form of birth and death for the vast majority of the Citizens.

  (b) In Part Two, we firstly prove the crucial Peak-Locus Theorem for any given trend $m_L(t)$ and not just for the GBM exponential. Then we show that the resulting Evo-Entropy grows exactly linearly in time if the trend is the exponential GMB trend.

  (c) In addition, three Appendixes (online) with all the relevant mathematical proofs are attached to this paper. They are written in the Maxima language, and Maxima is a symbolic manipulator that may be downloaded *for free* from the web.

In conclusion, this paper further increases the huge mathematical spectrum of applications of the Evo-SETI Theory to prepare Humans for the first Contact with an Extra-Terrestrial Civilization.

PART 1: NEW RESULTS ABOUT CIVILIZATIONS IN EVO-SETI THEORY

Introduction

Two mathematical papers were published by this author in 2013 and 2014, respectively:

(1) 'SETI, Evolution and Human History Merged into a Mathematical Model', *International Journal of Astrobiology*, vol. 12, issue (3), pp. 218–245 (2013) (this will be called (Maccone 2013) in the sequel of the current paper) and

(2) 'Evolution and Mass Extinctions as Lognormal Stochastic Processes', *International Journal of Astrobiology*, vol. 13, issue (4), p. 290–309 (2014) (this will be called (Maccone 2014) in the sequel of the current paper).

They provide the mathematical formulation of the 'Evo-SETI Theory', standing for 'a unified mathematical Theory of Evolution and SETI'.

Hoverer, the calculations required to prove all Evo-SETI results are lengthy, and this circumstance may unfortunately 'scare' potential readers that would love to understand Evo-SETI, but do not want to face all the calculations. To get around this obstacle, the three Appendixes at the end of this paper are a printout of all the analytical calculations that this author conducted by the Maxima symbolic manipulator, especially to prove the Peak-Locus Theorem described in Section 'Peak-Locus Theorem'. It is interesting to point out that the Macsyma symbolic manipulator or 'computer algebra code' (of which Maxima is a large subset) was created by NASA at the Artificial Intelligence Laboratory of MIT in the 1960s to check the equations of Celestial Mechanics that had been worked out by hand by a host of mathematicians in the previous 250 years (1700–1950). Actually, those equations might have contained errors that could have jeopardized the Moon landings of the Apollo Program, and so NASA needed to check them by computers, and Macsyma (nowadays Maxima) did a wonderful job. Today, everyone may download Maxima *for free* from the website http://maxima.source forge.net/. The Appendixes of this paper are written in Maxima language and the conventions apply of denoting the input instructions by (%i [equation number]) and the output results by (%o[equation number]), as we shall see in a moment.

In conclusion, in order to allow non-mathematically trained readers to appreciate this unified vision of how life developed on Earth over the last 3.5 billion years, a 'not-too mathematical' summary of the content of these two papers is now provided, also enabling readers to grasp the wide spectrum of Evo-SETI applications.

A simple proof of the *b*-lognormal's pdf

This paper is based on the notion of a *b*-lognormal, just as are (Maccone 2013, 2014). To let this paper be self-contained in this regard, we now provide an easy proof of the *b*-lognormal

equation as a probability density function (pdf). Just start from the well-known Gaussian or normal pdf

$$\text{Gaussian\_or\_normal}(x\,;\mu,\sigma) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}. \qquad (1)$$

This pdf has two parameters:

(1) μ turns out to be the mean value of the Gaussian and the abscissa of its peak. Since the independent variable x may take up any value between $-\infty$ and $+\infty$, i.e. it is a real variable, so μ must be real too.

(2) σ turns out to be the standard deviation of the Gaussian and so it must be a positive variable.

(3) Since the Gaussian is a pdf, it must fulfil the normalization condition

$$\int_{-\infty}^{\infty} \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}\,dx = 1 \qquad (2)$$

and this is the equation we need in order to 'discover' the *b*-lognormal. Just perform in the integral (2) the substitution $x = \ln t$ (where ln is the natural log). Then (2) is turned into the new integral

$$\int_{0}^{\infty} \frac{e^{-(\ln t-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma t}\,dt = 1. \qquad (3)$$

But this (3) may be regarded as the normalization condition of another random variable, ranging 'just' between zero and $+\infty$, and this new random variable we call 'lognormal' since it 'looks like' a normal one except that x is now replaced by $\ln t$ and t now also appears at the denominator of the fraction. In other words, the lognormal pdf is

$$\begin{cases} \text{lognormal}(t\,;\mu,\sigma) = \dfrac{e^{-(\ln(t)-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma \cdot t}, \\ \text{holding for } 0 \le t < \infty,\, -\infty < \mu < \infty,\, \sigma \ge 0. \end{cases} \qquad (4)$$

Just one more step is required to jump from the 'ordinary lognormal' (4) (i.e. the lognormal starting at $t = 0$) to the *b*-lognormal, that is the lognormal starting at any positive instant $b > 0$ ('*b*' stands for 'birth'). Since this simply is a shifting along the time axis from 0 to the new time origin $b > 0$, in mathematical terms it means that we have to replace t by $(t - b)$ everywhere in the pdf (4). Thus, the *b*-lognormal pdf must have the equation

$$\begin{cases} b\_\text{lognormal}(t\,;\mu,\sigma,b) = \dfrac{e^{-(\ln(t-b)-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma \cdot (t-b)} \\ \qquad \text{holding for } t \ge b \text{ and up to } t = \infty. \end{cases} \qquad (5)$$

The *b*-lognormal (5) is called 'three-parameter lognormal' by statisticians, but we prefer to call it *b*-lognormal to stress its biological meaning described in the next section.

Defining 'life' in the Evo-SETI Theory

The first novelty brought by our Evo-SETI Theory is our definition of life as a 'finite *b*-lognormal in time', extending from the time of birth (*b*) to the time of death (*d*) of the living
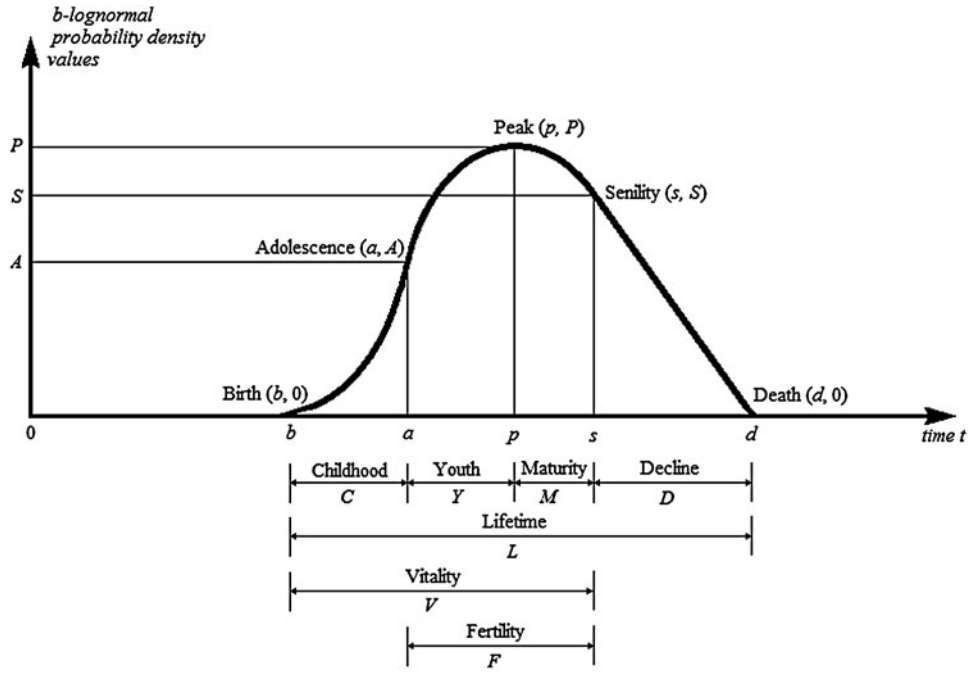
Fig. 1. 'Life' in the Evo-SETI Theory is a 'finite b-lognormal' made up by a lognormal pdf between birth b and senility (descending inflexion point) s, plus the straight tangent line at s leading to death d.

creature, let it be a cell, an animal, a human, a civilization of humans or even an Extra-Terrestrial (ET) civilization. Figure 1 shows what we call a 'finite b-lognormal'.

On the horizontal axis is the time t ranging between b and d. But the curve on the vertical axis is actually made up by two curves:

(1) Between b and the 'senility' time s (i.e. the descending inflexion point of the curve) on the vertical axis are the positive numerical values taken up by the pdf (5), that we prefer to call 'infinite b-lognormal' to distinguish it from the 'finite b-lognormal' shown in Fig. 1.

(2) Between s and d the curve is just a straight line having the same tangent at s as the b-lognormal (5). We are not going to derive its equation since that would take too long, but its meaning is obvious: since nobody lives for an infinite amount of time, it was necessary to 'cut' the infinite b-lognormal (5) at the junction point s and continue it with a simple straight line finally intercepting the time axis at the death instant d. As easy as that.

History formulae

Having so defined 'life' as a finite b-lognormal, this author was able to show that, given one's birth b, death d and (somewhere in between) one's senility s, then the two parameters μ (a real number) and σ (a positive number) of the b-lognormal (5) are given by the two equations

$$\begin{cases} \sigma = \dfrac{d - s}{\sqrt{d - b}\sqrt{s - b}}, \\ \mu = \ln(s - b) + \dfrac{(d - s)(b + d - 2s)}{(d - b)(s - b)}. \end{cases} \tag{6}$$

These were called 'History Formulae' by this author for their use in Mathematical History, as shown in the next section. The mathematical proof of (6) is found in (Maccone 2013, p. 227–231) and follows directly from the definition of s (as descending inflexion point) and d (as interception between the descending tangent straight line at s and the time axis). In previous versions of his Evo-SETI Theory, the author gave an apparently different version of the History Formulae (6) reading

$$\begin{cases} \sigma = \dfrac{d - s}{\sqrt{d - b}\sqrt{s - b}}, \\ \mu = \ln(s - b) + \dfrac{2s^2 - (3d + b)s + d^2 + b\,d}{(d - b)s - b\,d + b^2}. \end{cases} \tag{7}$$

This simply was because he had not yet factorized the fraction of the second equation (with apologies).

Death formula

One more interesting result discovered by this author, and firstly published by him in 2012 (Maccone 2012, Chapter 6, equation (6.30), p. 163) is the following 'Death Formula' (its proof is obtained by inserting the History Formulae (6) into the equation for the peak abscissa, $p = b + e^{\mu - \sigma^2}$):

$$d = \frac{s + b \cdot \ln((p - b)/(s - b))}{\ln((p - b)/(s - b)) + 1}. \tag{8}$$

This formula allows one to compute the death time d if the birth time b, the peak time p and the senility time s are known. The difficulty is that, while b and p are usually well known, s is *not* so, thus jeopardizing the practical usefulness of the Death Formula (8).

Birth–Peak–Death (BPD) theorem

This difficulty of estimating s for any b-lognormal led the author to discover the following BPD theorem that he only obtained on April 4, 2015, and presented here for the first time.

Ask the question: can a given b-lognormal be entirely characterized by the knowledge of its birth, peak and death only? Yes is the answer, but no *exact* formula exists yielding s in terms of (b, p, d) only.

Proof. Start from the exact Death Formula (8) and expand it into a Taylor series with respect to s around p and, say, to order 2. The result given by Maxima is

$$d = p + 2(s - p) - \frac{3(s-p)^2}{2(b-p)} + \cdots. \qquad (9)$$

Equation (9) is quadratic equation in s that, once solved for s, yields the second-order approximation for s in terms of (b, p, d)

$$s = \frac{\sqrt{2}\sqrt{-p^2 + (3d-b)p - 3bd + 2b^2} + p + 2b}{3}. \qquad (10)$$

In the practice, equation (10) is a 'reasonable' numeric approximation yielding s as a function of (b, p, d), and is certainly much better that the corresponding first-order approximation given by the linear equation

$$d = p + 2(s - p) + \cdots \qquad (11)$$

whose solution simply is

$$s = \frac{p + d}{2}, \qquad (12)$$

i.e., s (to first approximation) simply is the middle point between p and d, as geometrically obvious.

However, if one really wants a better approximation than the quadratic one (10), it is possible to expand the Death Formula (8) into a Taylor series with respect to s around p to third order, finding

$$d = p + 2(s - p) + \frac{3(s-p)^2}{2(p-b)} + \frac{5(s-p)^3}{6(p-b)^2} + \cdots \qquad (13)$$

Equation (13) is a cubic (i.e. third-degree polynomial) in s that may be solved for s by virtue of the well-known Cardan (Girolamo Cardano 1501–1576) formulae that we will not repeat here since they are exact but too lengthy to be reproduced in this paper.

As a matter of fact, it might even be possible to expand the Death Formula (8) to fourth order in s around p that would lead to the fourth-degree algebraic equation (a quartic) in s

$$d = p + 2(s - p) + \frac{3(s-p)^2}{2(p-b)} + \frac{5(s-p)^3}{6(p-b)^2}$$
$$+ \frac{(s-p)^4}{2(p-b)^3} + \cdots \qquad (14)$$

and then solve equation (14) for s by virtue of the exact four formulae of Lodovico Ferrari (1522–1565) (he was Cardan's pupil) that are huge and occupy a whole page each one. However, this game may not go on forever: the fifth-degree

algebraic equation is *not* solvable by virtue of radicals and so we must stop with degree 4.

Then there is the problem of finding which one, out of the three (Cardan) or four (Ferrari) roots numerically is 'the right one'. This author thus wrote a Maxima code given here as #1 Appendix to this paper where he solved several cases of finding s from (b, p, d) related to the important Fig. 2 of this paper.

In other words, the *inputs* to Table 1 of this paper were (b, p, d) and *not* (b, s, d), as the author had always done previously, for instance in deriving the whole of Chapter 7 of (Maccone 2012) back in 2012. This improvement is remarkable since it allowed a fine-tuning of Table 1 with respect to all similar previous material. In other words still, '*it is easier to assign birth, peak and death rather than birth, senility and death*'. That's why the Theorem described in this section was called BPD Theorem.

The reader is invited to ponder over Appendix 1 as the key to all further, future developments in Mathematical History.

Mathematical history of nine key civilizations since 3100 BC

The author called (6) the 'History Formulae' since in (Maccone 2013, p. 231–235), equations (6), with the numerical values provided there, allowed him to draw the b-lognormals of eight leading civilizations in Western History: Greece, Rome, Renaissance Italy, and the Portuguese, Spanish, French, British and American (USA) Empires.

Please notice that:

(1) The data in Table 1 and the resulting b-lognormals in Fig. 2 are *experimental* results, meaning that we just took what described in History textbooks (with a lot of words) and translated that into the simple b-lognormals shown in Fig. 2. In other words, a new branch of knowledge was forged: we love to call it 'Mathematical History'. More about this in future papers.

(2) The envelope of all the above b-lognormals 'looks like' a simple exponential curve. In Fig. 2, two such exponential envelopes were drawn: the one going from the peak of Ancient Greece (the Pericles age in Athens, cradle of Democracy) to the peak of the British Empire (Victorian age, the age of Darwin and Maxwell) and to the peak of the USA Empire (Moon landings in 1969–1972). This notion of b-lognormal envelope will later be precisely quantified in our 'Peak-Locus Theorem'.

(3) It is now high time to introduce a 'measure of evolution' namely a function of the three parameters μ, σ and b accounting for the fact that 'the experimental Fig. 2 clearly shows that, the more the time elapses, the more highly peaked, and narrower and narrower, the b-lognormals are'. In (Maccone 2013, p. 238–243), this author showed that the requested measure of evolution is the (Shannon) *entropy*, namely the entropy of each infinite b-lognormal that fortunately has the simple equation

$$H_{\text{infinite}\_b-\text{lognormal}}(\mu, \sigma) = \ln\left(\sqrt{2\pi}\sigma\right) + \mu + \frac{1}{2}. \qquad (15)$$

Table 1. *Birth, peak, decline and death times of nine Historic Western Civilizations (3100 BC - 2035 AD), plus the relevant peak heights. They are shown in Fig. 2 as nine b-lognormal pdfs.*

|	b = birth time	p = peak time	s = decline = senility time	d = death time	P = peak ordinate
Ancient Egypt	3100 BC Lower and Upper Egypt unified. First Dynasty.	1154 BC Luxor and Karnak temples edified by Ramses II by 1260 BC.	689 BC Assyrians invade Egypt in 671 BC, leave 669 BC.	30 BC Cleopatra dies: last Hellenistic queen.	8.313×10^{-4}
Ancient Greece	776 BC First Olympic Games, from which Greeks compute years.	438 BC Pericles' Age. Democracy peak. Arts and Science peak. Aristotle.	293 BC Alexander the Great dies in 323. Hellenism starts in Near East.	30 BC Cleopatra dies: last Hellenistic queen.	2.488×10^{-3}
Ancient Rome	753 BC Rome founded. Italy seized by Romans by 270 BC, Carthage and Greece by 146 BC, Egypt by 30 BC. Christ born around the year 0.	117 AD Rome at peak: Trajan in Mesopotamia. Christianity preached in Rome by Saints Peter and Paul against slavery by 69 AD.	273 AD Aurelian builds new walls around Rome after Military Anarchy, 235–270 AD.	476 AD Western Roman Empire ends. Dark Ages start in West. Not in East.	2.193×10^{-3}
Renaissance Italy	1250 Frederick II dies. Middle Ages end. Free Italian towns start Renaissance.	1497 Renaissance art and architecture. Birth of Science. Copernican revolution (1543).	1562 Council of Trent ends in 1563. Catholic and Spanish rule.	1660 Cimento Academy ended. Bruno burned at stake in 1600. Galileo dies in 1642.	5.749×10^{-3}
Portuguese Empire	1419 Madeira island discovered. African coastline explored by 1498.	1716 Black slave trade to Brazil at its peak. Millions of blacks enslaved or killed.	1822 Brazil independent, other colonies retained.	1999 Last colony, Macau, lost to Republic of China.	3.431×10^{-3}
Spanish Empire	1402 Canary islands are conquered by 1496. Columbus discovers America in 1492.	1798 Largest extent of Spanish colonies in America: California settled since 1769.	1844 Spanish fleet lost at Trafalgar in 1805.	1898 Last colonies lost to the USA: Philippines, Cuba and Puerto Rico.	5.938×10^{-3}
French Empire	1524 Verrazano first in New York bay. Cartier in Canada, 1534.	1812 Napoleon I dominates continental Europe and reaches Moscow.	1870 Napoleon III defeated. Third Republic starts. World Wars One and Two.	1962 Algeria lost as most colonies. Fifth Republic starts in 1958.	4.279×10^{-3}
British Empire	1588 Spanish Armada Defeated. British Empire's expansion starts.	1904 British Empire peak. Top British Science: Faraday, Maxwell, Darwin, Rutherford.	1947 After World Wars One and Two, India gets independent.	1974 Britain joins the EEC and loses most of her colonies.	8.447×10^{-3}
USA Empire	1898 Philippines, Cuba, Puerto Rico seized from Spain.	1972 Moon Landings, 1969–72: America leads the world.	2001 9/11 terrorist attacks: decline. Obama 2009.	2035 Singularity: computers ruling? Will the USA yield to China?	0.013
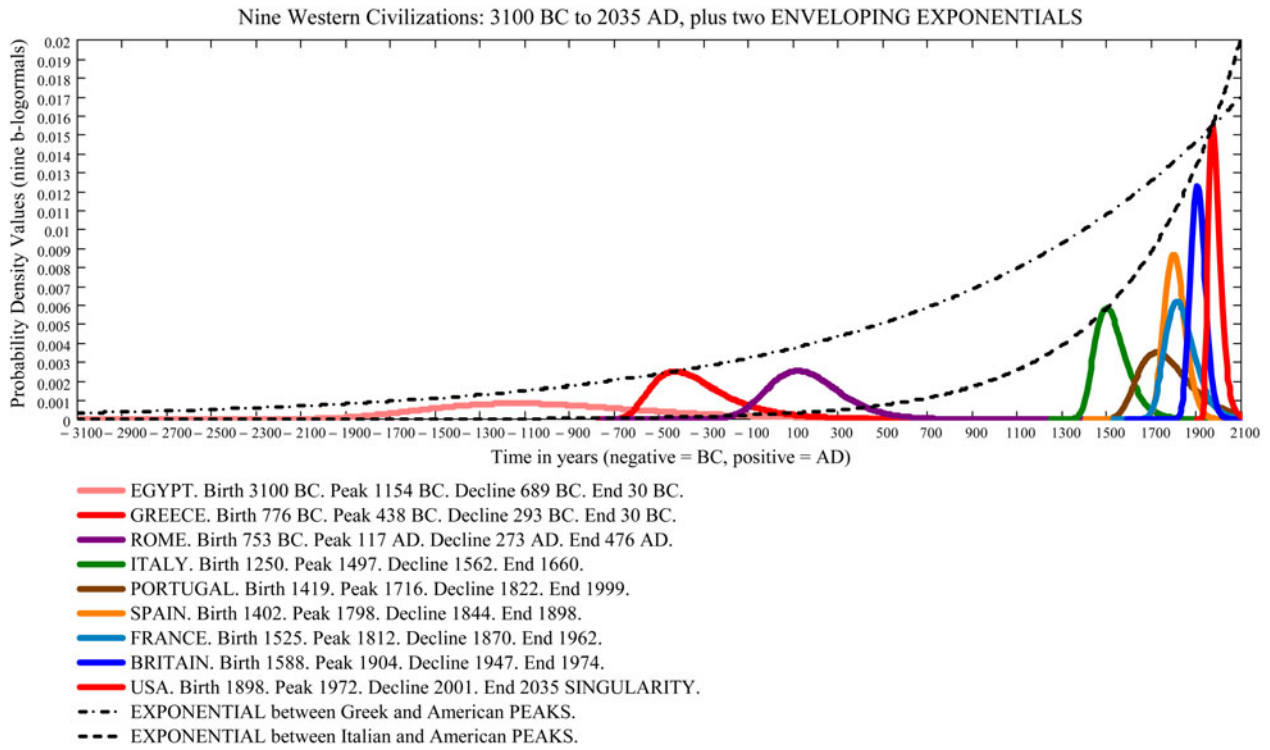
Fig. 2. The b-lognormals of nine Historic Western Civilizations computed thanks to the History Formulae (6) with the three numeric inputs for b, p and d of each Civilization given by the corresponding line in Table 1. The corresponding s is *derived* from b, p and d by virtue of the second-order approximation (10) provided by the solution of the quadratic equation in the BPD theorem.

The proof of this result was given in (Maccone 2013, p. 238–239). If measured in bits, as customary in Shannon's Information Theory, equation (15) becomes

$$H_{\text{infinite\_b-lognormal\_in\_bits}}(\mu, \sigma) = \frac{\ln(\sqrt{2\pi}\sigma) + \mu + 1/2}{\ln 2}. \quad (16)$$

This is the b-lognormal entropy definition that was used in (Maccone 2013, 2014) and we are going to use in this paper also. In reality, Shannon's entropy is a measure of the disorganization of an assigned pdf $f_X(x)$, rather than a measure of its organization. To change it into a measure of organization, we should just drop the minus sign appearing in front of the Shannon definition of entropy for any assigned pdf $f_X(x)$:

$$H = -\int_{-\infty}^{\infty} f_X(x) \cdot \ln f_X(x) \, dx \quad (17)$$

We will do so to measure Evolution of life on Earth over the last 3.5 billion years.

The final goal of all these mathematical studies is of course to 'prepare' the future of Humankind in SETI, when we will have to face other Alien Civilizations whose past may be the future for us.

b-Scalene (triangular) probability density

Having recognized that BPD (and *not* birth–senility–death) are the three fundamental instants in the lifetime of any living

creature, we are tempted to introduce a new pdf called b-scalene, or, more completely, b-scalene triangular pdf.

The idea is easy:

(1) The horizontal axis is the time axis, denoted by t.
(2) The vertical axis is denoted by y.
(3) The b-scalene pdf starts at the instant b 'birth'.
(4) The b-scalene pdf ends at the instant d 'death'.
(5) Somewhere in between is located the pdf peak, having the coordinates (p, P).
(6) The pdf between $(b, 0)$ and (p, P) is a straight line, hereafter called 'first b-scalene' (line).
(7) The pdf between (p, P) and $(d, 0)$ is a straight line, hereafter called 'second b-scalene' (line).

Let us now work out the equations of the b-scalene. First of all its normalization condition implies that the sum of the areas of the two triangles equals 1:

$$\frac{(p - b)P}{2} + \frac{(d - p)P}{2} = 1. \quad (18)$$

Solving equation (18) for P we get

$$P = \frac{2}{d - b}. \quad (19)$$

Then, the equations of the two straight lines making up the b-scalene pdf are found to be, respectively:

$$y = \frac{2(t - b)}{(d - b)(p - b)}, \quad \text{for} \quad b \leq t \leq p \quad (20)$$

and

$$y = \frac{2(t-d)}{(d-b)(p-d)}, \quad \text{for} \quad p \leq t \leq d. \tag{21}$$

Proof. The proofs of equations (20) and (21), as well as of all subsequent formulae about the b-scalene, are given in #2 Appendix to the present paper. We will simply refer to them with the numbers of the resulting equations in the Maxima code. Thus, equation (20) corresponds to (%o15) and equation (21) to (%o20). Also, it is possible to compute all moments (i.e. the kth moment) of the b-scalene immediately. In fact, Maxima yields ((%o27) and (%o28))

$$\langle b\_scalene^k \rangle = \frac{2[(d-b)p^{k+2} + (b^{k+2}-d^{k+2})p + bd^{k+2} - b^{k+2}d]}{(d-b)(k+1)(k+2)(p-b)(p-d)}. \tag{22}$$

Setting $k = 0$ into equation (22) yields of course the normalization condition (%o29)

$$\langle b\_scalene^0 \rangle = 1. \tag{23}$$

Setting $k = 1$ into equation (22) yields the mean value (%o30)

$$\langle b\_scalene \rangle = \frac{b+p+d}{3}. \tag{24}$$

Setting $k = 2$ into equation (22) yields the mean value of the square (%o32)

$$\langle b\_scalene^2 \rangle = \frac{b^2 + p^2 + d^2 + bp + dp + bd}{6}. \tag{25}$$

Then, subtracting the square of equation (24) into (25), one gets the b-scalene variance (%o34)

$$\sigma^2_{\text{b-scalene}} = \frac{(d-b)^2}{24}. \tag{26}$$

The square root of equation (26) is the b-scalene standard deviation (%o36)

$$\sigma_{\text{b\_scalene}} = \frac{d-b}{2^{\frac{3}{2}}\sqrt{3}}. \tag{27}$$

We could go on to find more descriptive statistical properties of the b-scalene, but we prefer to stop at this point. Much more important, in fact, is to compute the Shannon Entropy of the b-scalene. Equations (%i37) through (%o41) show that the Shannon Entropy of the b-scalene is given by

b_scalene_Shannon_ENTROPY_in_bits

$$= \frac{1 + \ln((d-b)^2/4)}{2\ln 2}. \tag{28}$$

This is a simple and important result. Since p does not appear in equation (28), the Shannon Entropy of the b-scalene is actually independent of where its peak is!

Also, one is tempted to make a comparison between the Entropy of the b-scalene and the Entropy of the UNIFORM distribution over the same interval $(d-b)$ This will be done in the next section.

Uniform distribution between birth and death

In the Evo-SETI Theory, the meaning of a uniform distribution over the time interval $(d-b)$ simply is 'we know nothing about that living being except when he/she/it was born (at instant b) and when he/she/it died (at instant d)'. No idea even about when the 'peak' p of his/her/its activity occurred. Thus, the uniform distribution is the *minimal* amount of information about the lifetime of someone that one might possibly have.

The pdf of the uniform distribution over the time interval $(d-b)$ is obviously given by the constant in time (%o43)

$$f_{\text{uniform\_b\_d}}(t) = \frac{1}{d-b}, \quad \text{for} \quad b \leq t \leq d. \tag{29}$$

It is immediately possible to compute all moments of the uniform distribution (%o44)

$$\int_b^d \frac{t^k}{d-b}\, dt = \frac{d^{k+1} - b^{k+1}}{(d-b)(k+1)}. \tag{30}$$

The normalization condition of equation (30) is obviously found upon letting $k = 0$.

The mean value is found by letting $k = 1$ into equation (30), (%o53), and is just the middle point between birth and death

$$\langle \text{uniform\_b\_d} \rangle = \frac{b+d}{2}. \tag{31}$$

The mean value of the square is found by letting $k = 2$ into equation (30) and reads (%o54)

$$\langle \text{uniform\_b\_d}^2 \rangle = \frac{b^2 + bd + d^2}{3}. \tag{32}$$

Subtracting the square of equation (31) into (32), we get the uniform distribution variance (%o58)

$$\sigma^2_{\text{uniform\_b\_d}} = \frac{(d-b)^2}{12}. \tag{33}$$

Finally, the uniform distribution standard deviation is the square root of (33) (%o59)

$$\sigma_{\text{uniform\_in\_b\_d}} = \frac{d-b}{2\sqrt{3}}. \tag{34}$$

We stop the derivation of the descriptive statistics of the uniform distribution at this point, since it is easy to find all other formulae in textbooks. Rather, we prefer to concentrate on the Shannon Entropy of the uniform distribution, that upon inserting the pdf (29) into the entropy definition (17), yields (%o62)

$$\text{Uniform\_distribution\_ENTROPY\_in\_bits} = \frac{\ln(d-b)}{\ln 2}. \tag{35}$$

Entropy difference between uniform and b-scalene distributions

We are now in a position to find out the '*Entropy Difference*' between the uniform and the b-scalene distributions.

Table 2. *Summary of the properties of the b-lognormal distribution that applies to the random variable C = History in time of a certain Civilization. This set of results we like to call 'E-Pluribus-Unum Theorem' of the Evo-SETI Theory.*

Random variable	C = History in time of a certain civilization
Probability distribution	b-lognormal with parameters μ, σ and b
Probability density function	$f_C(t) = \dfrac{e^{-[\ln(t-b)-\mu]^2/(2\sigma^2)}}{\sqrt{2\pi}\,\sigma(t-b)}$ with $t \geq b = \min b_i$
Mean value	$\langle C \rangle = b + e^{\mu+(\sigma^2)/2}$
Variance	$\sigma_C^2 = e^{2\mu}e^{\sigma^2}\left(e^{\sigma^2} - 1\right)$
Standard deviation	$\sigma_C = e^{\mu}e^{\sigma^2/2}\sqrt{e^{\sigma^2} - 1}$
Mode (= abscissa of the b-lognormal peak)	$t_{\text{mode}} = t_{\text{peak}} = p = b + e^{\mu-\sigma^2}$
Value of the Mode Peak	$f_C(t_{\text{mode}}) = f_C(t_{\text{peak}}) = P = \dfrac{e^{(\sigma^2/2)-\mu}}{\sqrt{2\pi}\sigma}$
Median (= fifty-fifty probability value for C)	$\text{median\_of\_}C = b + e^{\mu}$
Skewness	$\dfrac{K_3}{K_2^{3/2}} = \sqrt{e^{\sigma^2} - 1}\left(e^{\sigma^2} + 2\right)$
Kurtosis	$\dfrac{K_4}{K_2^2} = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6$
Expression of μ in terms of the birth instants b_i and death instants d_i of the *uniform* input random variables C_i	$\mu = \sum_{i=1}^{N} \langle Y_i \rangle = \sum_{i=1}^{N} \left(\dfrac{d_i \ln(d_i) - b_i \ln(b_i)}{d_i - b_i} - 1 \right)$
Expression of σ^2 in terms of the birth instants b_i and death instants d_i of the *uniform* input random variables C_i	$\sigma^2 = \sum_{i=1}^{N} \sigma_{Y_i}^2 = \sum_{i=1}^{N} \left(1 - \dfrac{b_i d_i [\ln(d_i) - \ln(b_i)]^2}{(d_i - b_i)^2} \right)$

Subtracting equation (28) into (35) one gets (%o71)

$$\frac{\ln(d - b)}{\ln 2} - \frac{1 + \ln((d - b)^2/4)}{2\ln 2}$$
$$= \frac{\ln(d - b)}{\ln 2} - \frac{1}{2\ln 2} - \frac{2[\ln(d - b) - \ln 2]}{2\ln 2} \qquad (36)$$
$$= 1 - \frac{1}{2\ln 2} = 0.27865247955552.$$

One may say that in passing from just knowing birth and death to knowing birth, peak and death, one has reduced the uncertainty by 0.27865247955552 bits, or, if you prefer, the Shannon Entropy has been reduced by an amount of 0.27865247955552 bits. Again in a colourful language, if you just know that Napoleon was born in 1769 and died in 1821, and then add that the peak occurred in 1812 (or at any other date), than you have added 0.27865247955552 bits of information about his life.

Readers might now wish to ponder over statements like the last one about Napoleon in order build up a Mathematical Theory of History, simply called Mathematical History.

We stop here now, but some young talent might wish to develop these ideas much more in depth, disregarding all criticism and just being bold, bold, bold,….

'Equivalence' between uniform and *b*-lognormal distributions

One more 'crazy idea' suggested by the Evo-SETI Theory is the 'equivalence' between uniform and lognormal distributions, as described in #2 Appendix.

The starting point is to equate the two mean values and the two standard deviations of these two distributions and then… see what comes out!

So, just equate the two mean values first, i.e. just equate equation (31) and the well-known mean value formula for the lognormal distribution (see Table 2, fourth line) (%o2)

$$\frac{b + d}{2} = e^{(\sigma^2/2)+\mu}. \qquad (37)$$

Similarly, we equate the uniform standard deviation (34) and the lognormal standard deviation (see Table 2, sixth line) and get (%o3)

$$\frac{d - b}{2\sqrt{3}} = e^{(\sigma^2/2)+\mu}\sqrt{e^{\sigma^2-1}}. \qquad (38)$$

A glance to equations (37) and (38) shows that we may eliminate μ upon dividing equation (38) by (37), and that yields the resolving equation in σ (%o4)

$$\frac{d - b}{\sqrt{3}(b + d)} = \sqrt{e^{\sigma^2-1}}. \qquad (39)$$

After a few steps, equation (39) may be solved for the exponential, yielding (%05)

$$e^{\sigma^2} = \frac{4(b^2 + bd + d^2)}{3(b + d)^2} \qquad (40)$$

and finally, taking logs

$$\sigma^2 = \ln\left[\frac{4(b^2 + b\,d + d^2)}{3(b + d)^2}\right]. \qquad (41)$$

Taking the square root, equation (41) becomes

$$\sigma = \sqrt{\ln\left[\frac{4(b^2 + bd + d^2)}{3(b + d)^2}\right]}. \qquad (42)$$

Then, inserting equation (41) into (38) and solving the resulting equation for μ, one finds for μ (%o10)

$$\mu = \ln\left[\frac{\sqrt{3}(b+d)^2}{4\sqrt{b^2+bd+d^2}}\right]. \tag{43}$$

In conclusion, we have proven that, if we are given just the birth and death times of the life of anyone, this uniform distribution between birth and death may be converted into the 'equivalent' lognormal distribution starting at the same birth instant and having the two parameters μ and σ given by, respectively

$$\begin{cases} \mu = \ln\left[\dfrac{\sqrt{3}(b+d)^2}{4\sqrt{b^2+bd+d^2}}\right], \\[3mm] \sigma = \sqrt{\ln\left[\dfrac{4(b^2+bd+d^2)}{3(b+d)^2}\right]}. \end{cases} \tag{44}$$

One may also invert the system of two simultaneous equations (44). In fact, multiplying equation (37) by 2 and equation (38) by $2\sqrt{3}$ and then summing, b disappears and one is left with the d expression (%o13)

$$d = e^{(\sigma^2/2)+\mu}\left(1 + \sqrt{3}\sqrt{e^{\sigma^2-1}}\right). \tag{45}$$

Similarly, multiplying equation (37) by 2 and equation (38) by $2\sqrt{3}$ and then subtracting, d disappears and one is left with the b expression (%o14)

$$b = e^{(\sigma^2/2)+\mu}\left(1 - \sqrt{3}\sqrt{e^{\sigma^2-1}}\right). \tag{46}$$

In conclusion, the inverse formulae of equation (44) are

$$\begin{cases} b = e^{(\sigma^2/2)+\mu}\left(1 - \sqrt{3}\sqrt{e^{\sigma^2-1}}\right) \\[2mm] d = e^{(\sigma^2/2)+\mu}\left(1 + \sqrt{3}\sqrt{e^{\sigma^2-1}}\right). \end{cases} \tag{47}$$

Let us now find how much the Shannon Entropy changes when we replace the lognormal distribution to the uniform distribution between birth and death. We already know that the uniform distribution entropy is the largest possible entropy, and is given by equation (35). Then, we only need to know that the lognormal entropy is given by the expression (%o16) (for the proof, see, for instance, (Maccone 2012), Chapter 30, p. 685–687)

$$\text{lognormal\_entropy\_in\_bits} = \frac{\ln(\sqrt{2\pi}\,\sigma) + \mu + 1/2}{\ln 2} \tag{48}$$

Inserting equations (44) into (48) a complicated expression would be found (%o17) that we will not re-write here. Also the uniform entropy (35) may be rewritten in terms of μ and σ by inserting equation (44) into it, and the result is (%o20). At this point we may subtract the lognormal entropy to the uniform entropy and so find out how much information we 'arbitrarily inject into the system' if we replace the uniform pdf by the lognormal pdf. The result is given by (%o22) and reads

uniform_pdf_ENTROPY − lognormal_pdf_ENTROPY

$$= \frac{\ln\left[12(e^{\sigma^2}-1)\right] + 2\mu + \sigma^2}{2\ln 2} - \frac{\ln(\sqrt{2\pi}\sigma) + \mu + 1/2}{\ln 2} \tag{49}$$

$$= \frac{\ln\left[6(e^{\sigma^2}-1)/\pi\,\sigma^2\right] + \sigma^2 + 1}{2\ln 2}.$$

Notice that, rather unexpectedly, equation (49) is independent of μ. Numerically, we may get an idea about equation (49) in the limit case when σ → 0, then finding

$$\lim_{\sigma \to 0}\frac{\ln\left[6(e^{\sigma^2}-1)/\pi\,\sigma^2\right] + \sigma^2 + 1}{2\ln 2} = \frac{\ln(6/\pi)+1}{2\ln 2}$$

$$= -0.254 \text{ bits}. \tag{50}$$

Not too a big numeric error, apparently.

b-lognormal of a civilization's history as CLT of the lives of its citizens

This and the following sections of Part 1 are most important since they face mathematically the finding of the *b*-lognormal of a certain Civilization in time, like any of the Civilizations shown in Fig. 2. We claim that the *b*-lognormal of a Civilization History is obtained by applying the CLT of Statistics to the lifetimes of the millions of Citizens that make and made up for that Civilization in time.

Though this statement may appear rather obvious, the mathematics is not so, and we are going to explain it from scratch right now.

Then:

(1) Denote by C the random variable (in time) yielding the History of that Civilization in time. In the end, the pdf of C will prove to be a *b*-lognormal and we will derive this fact as a consequence of the CLT of Statistics.

(2) Denote by C_i the random variable (in time) denoting the lifetime of the *i*th Citizen belonging to that Civilization. We do not care about the actual pdf of the random variable C_i: it could be just uniform between birth and death (in this case, C_i is the lifetime of a totally anonymous guy, as the vast majority of Humans are, and certainly cells are too, and so forth for other applications). Or, on the contrary, it could be a *b*-scalene, as in the example about Napoleon, born 1769, died 1821, with peak in 1812, or this pdf could be anything else: no problem since the CLT allows for *arbitrary* input pdf*s*.

(3) Denote by N the total number of individuals that made up and are making up and will make up for the History of that Civilization over its total existence in time, let this time be years or centuries or millions or even billions of years (for ET Civilizations, we suppose!). In general, this positive integer number N is going to be very large: thousands or millions or even billions, like that fact that Humans nowadays number about 7.3 billion people. In the practice, we may well suppose that N approaches infinity, i.e. $N \to \infty$,

which is precisely the mathematical condition requested to apply the CLT of Statistics, as we shall see in a moment.

(4) Then consider the statistical equation

$$C = \prod_{i=1}^{N} C_i. \tag{51}$$

This we shall call 'the Statistical Equation of each Civilization' (abbreviated SEC).

What is the meaning of this equation?

Well, if we suppose that all the random variables C_i are '*statistically independent of each other*', then equation (51) is the '*Law of Compound Probability*', well known even to beginners in statistical courses. And the lifespans of Citizens C_i almost certainly are independent of each other in time: dead guys may hardly influence the life of alive guys!

(5) Now take the logs of equation (51). The product is converted into a sum and the new form of our SEC is

$$\ln C = \sum_{i=1}^{N} \ln C_i. \tag{52}$$

(6) To this equation (52) we now apply the CLT. In loose terms, the CLT states that 'if you have a sum of a number of independent random variables, and let the number of terms in the sum approach infinity, then, regardless of the actual probability distribution of each term in the sum, the overall sum approaches the normal (i.e. Gaussian) distribution'.

(7) And the mean value of this Gaussian equals the sum of the mean values of the $\ln C_i$, while the variance equals the sum of the variances of the $\ln C_i$. In equations, one has

$$\ln C = \text{normally\_distributed\_random\_variable} \tag{53}$$

with mean value given by

$$\mu = \langle \ln C \rangle = \sum_{i=1}^{N} \langle \ln C_i \rangle \tag{54}$$

and variance given by

$$\sigma_{\ln C}^2 = \sum_{i=1}^{N} \sigma_{\ln C_i}^2. \tag{55}$$

(8) Let us now '*invert*' equation (53), namely solve it for C. To do so, we must recall an important theorem that is proved in probability courses, but, unfortunately, does not seem to have a specific name. It is the *transformation law* (so we shall call it, see for instance (Papoulis & Pillai 2002, p. 130–131)) allowing us to compute the pdf of a certain new random variable Y that is a known function $Y = g(X)$ of another random variable X having a known pdf. In other words, if the pdf $f_X(x)$ of a certain random variable X is known, then the pdf $f_Y(y)$ of the new random variable Y, related to X by the functional relationship

$$Y = g(X) \tag{56}$$

can be calculated according to the following rules:

(a) First, invert the corresponding non-probabilistic equation $y = g(x)$ and denote by $x_i(y)$ the various real roots resulting from this inversion.

(b) Second, take notice whether these real roots may be either finitely- or infinitely-many, according to the nature of the function $y = g(x)$.

(c) Third, the pdf of Y is then given by the (finite or infinite) sum

$$f_Y(y) = \sum_i \frac{f_X(x_i(y))}{|g'(x_i(y))|}, \tag{57}$$

where the summation extends to all roots $x_i(y)$ and $|g'(x_i(y))|$ is the absolute value of the first derivative of $g(x)$ where the ith root $x_i(y)$ has been replaced instead of x.

Going now back to (53), in order to invert it, i.e. in order to find the pdf of C, we must apply the general transformation law (57) to the particular transformation

$$y = g(x) = e^x. \tag{58}$$

That, upon inversion, yields the *single* root

$$x_1(y) = x(y) = \ln y. \tag{59}$$

On the other hand, differentiating equation (58) one gets

$$g'(x) = e^x \tag{60}$$

and

$$g'(x_1(y)) = e^{\ln y} = y, \tag{61}$$

where equation (60) was already used in the last step. So, the general transformation law (57) finally yields just the lognormal pdf in y for the random variable C, the time History of that Civilization:

$$f_C(y) = \sum_i \frac{f_x(x_i(y))}{|g'(x_i(y)|} = \frac{1}{|y|} f_Y(\ln(y))$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma y} e^{-(\ln(y)-\mu)^2/(2\sigma^2)}, \quad \text{for} \quad y > 0. \tag{62}$$

with μ given by equation (54) and σ given by equation (55). This is a very important result to understand the History of Civilizations mathematically: we now see why, for instance, all Civilizations shown in Fig. 2 are *b*-lognormals in their Historic development!

The pdf (62) actually is a *b*-lognormal, rather than just an ordinary lognormal starting at zero. In fact, the instant *b* at with it starts may not be smaller than the birth instant of the first (Historically!) individual of the population. Thus, the true *b*-lognormal pdf of the C Civilization is

$$f_C(t) = \frac{e^{-[\ln(t-b)-\mu^2]/(2\sigma^2)}}{\sqrt{2\pi}\,\sigma\,(t-b)} \quad \text{with} \quad t \geq b = \min b_i. \tag{63}$$

The very important special case of C_i uniform random variables: E-Pluribus-Unum Theorem

This author has discovered new, important and rather simple equations for the particular case where the input variables C_i

are uniformly distributed between birth and death, namely, the pdf of each C_i is

$$f_{C_i}(t) = \frac{1}{d_i - b_i}, \quad \text{for} \quad b_i \leq t \leq d_i. \tag{64}$$

In equation (64) b_i is the instant when he/she/it was born, and d_i is the instant when he/she/it died. We may not know them at all: just think of the millions of Unknown Soldiers died in World War One and in all wars (billions?). But that will not prevent us from doing the mathematics of equation (64).

Our primary goal now is to find the pdf of the random variable $Y_i = \ln C_i$ as requested by equation (52). To this end, we must apply again the transformation law (57), this time applied to the transformation

$$y = g(x) = \ln(x). \tag{65}$$

Upon inversion, equation (65) yields the *single* root

$$x_1(y) = x(y) = e^y. \tag{66}$$

On the other hand, differentiating equation (65) yields

$$g'(x) = \frac{1}{x} \tag{67}$$

and

$$g'(x_1(y)) = \frac{1}{x_1(y)} = \frac{1}{e^y}, \tag{68}$$

where equation (66) was already used in the last step. Then, by virtue of the uniform pdf (64), the general transformation law (57) finally yields

$$f_{Y_i}(y) = \sum_i \frac{f_x(x_i(y))}{|g'(x_i(y))|} = \frac{1}{d_i - b_i} \cdot \frac{1}{|1/e^y|}$$
$$= \frac{e^y}{d_i - b_i}. \tag{69}$$

In other words, the requested pdf of $Y_i = \ln C_i$ is

$$f_{Y_i}(y) = \frac{e^y}{d_i - b_i}, \quad \text{for} \quad \ln b_i \leq y \leq \ln d_i. \tag{70}$$

These are the probability density functions of the natural logs of all the uniformly distributed C_i random variables. Namely, in the colourful language of the applications of the Evo-SETI Theory, equation (70) is the pdf of all UNKNOWN FORMS OF LIFE, about which we only known when each of them was born and when it died.

Let us now check that the pdf (70) fulfils indeed its normalization condition

$$\int_{\ln b_i}^{\ln d_i} f_{Y_i}(y)\mathrm{d}y = \int_{\ln b_i}^{\ln d_i} \frac{e^y}{d_i - b_i}\mathrm{d}y$$
$$= \frac{e^{\ln d_i} - e^{\ln b_i}}{d_i - b_i} = \frac{d_i - b_i}{d_i - b_i} = 1. \tag{71}$$

Next we want to find the mean value and standard deviation of each Y_i, since they play a crucial role for future developments. The mean value of the pdf (70) is given by either of

the following alternative forms:

$$\langle Y_i \rangle = \int_{\ln(b_i)}^{\ln(d_i)} y \cdot f_{Y_i}(y)\mathrm{d}y = \int_{\ln(b_i)}^{\ln(d_i)} \frac{y \cdot e^y}{d_i - b_i}\mathrm{d}y$$
$$= \frac{d_i[\ln(d_i) - 1] - b_i[\ln(b_i) - 1]}{d_i - b_i}$$
$$= \frac{d_i \ln(d_i) - b_i \ln(b_i)}{d_i - b_i} - 1 \tag{72}$$
$$= \frac{\ln[(d_i)^{d_i}/(b_i)^{b_i}]}{d_i - b_i} - 1$$
$$= \ln\left[\frac{(d_i)^{d_i/(d_i - b_i)}}{(b_i)^{b_i/(d_i - b_i)}}\right] - 1.$$

This is thus *the mean value of the natural log of all the uniformly distributed random variables C_i (just to use a few of the above equivalent forms). Thus, the whole Civilization is a b-lognormal with the following parameters:*

$$\mu = \sum_{i=1}^N \langle Y_i \rangle = \sum_{i=1}^N \left(\frac{d_i \ln(d_i) - b_i \ln(b_i)}{d_i - b_i} - 1\right)$$
$$= \sum_{i=1}^N \left(\ln\left[\frac{(d_i)^{d_i/(d_i - b_i)}}{(b_i)^{b_i/(d_i - b_i)}}\right] - 1\right)$$
$$= \sum_{i=1}^N \ln\left[\frac{(d_i)^{d_i/(d_i - b_i)}}{(b_i)^{b_i/(d_i - b_i)}}\right] - N \tag{73}$$
$$= \ln\left[\prod_{i=1}^N \frac{(d_i)^{d_i/(d_i - b_i)}}{(b_i)^{b_i/(d_i - b_i)}}\right] - N.$$

The last form of μ shows that the exponential of μ is

$$e^\mu = e^{-N} \prod_{i=1}^N \frac{(d_i)^{d_i/(d_i - b)}}{(b_i)^{b_i/(d_i - b)}}. \tag{74}$$

In order to find the variance also, we must first compute the mean value of the square of Y_i, that is

$$\langle Y_i^2 \rangle = \int_{\ln(b_i)}^{\ln(d_i)} y^2 \cdot f_{Y_i}(y)\mathrm{d}y = \int_{\ln(b_i)}^{\ln(d_i)} \frac{y^2 \cdot e^y}{d_i - b_i}\mathrm{d}y$$
$$= \frac{d_i[\ln^2(d_i) - 2\ln(d_i) + 2] - b_i[\ln^2(b_i) - 2\ln(b_i) + 2]}{d_i - b_i}. \tag{75}$$

The *variance* of $Y_i = \ln(C_i)$ is now given by equation (75) minus the square of equation (73), that, using the first form of equation (73) and after a few reductions, yields:

$$\sigma_{Y_i}^2 = \sigma_{\ln(C_i)}^2 = 1 - \frac{b_i d_i[\ln(d_i/b_i)]^2}{(d_i - b_i)^2}$$
$$= 1 - \left[\ln\left\{\left(\frac{d_i}{b_i}\right)^{\frac{\sqrt{b_i d_i}}{d_i - b_i}}\right\}\right]^2. \tag{76}$$

Whence, using the first form of equation (76) and taking the square root, yields the standard deviation of Y_i

$$\sigma_{Y_i} = \sigma_{\ln(D_i)} = \sqrt{1 - \frac{b_i d_i[\ln(d_i) - \ln(b_i)]^2}{(d_i - b_i)^2}}. \tag{77}$$

Like the μ given by equation (73), equation (76) also may be rewritten in a few alternative forms. For instance

$$\sigma_Y^2 = \sum_{i=1}^{N} \sigma_{Y_i}^2 = \sum_{i=1}^{N} \sigma_{\ln(C_i)}^2$$

$$= \sum_{i=1}^{N} \left(1 - \frac{b_i \, d_i [\ln(d_i/b_i)]^2}{(d_i - b_i)^2} \right) \qquad (78)$$

$$= N - \sum_{i=1}^{N} \left[\ln \left\{ \left(\frac{d_i}{b_i} \right)^{\frac{\sqrt{b_i d_i}}{d_i - b_i}} \right\} \right]^2 .$$

We stop at this point, for we feel we have really proven a new theorem, yielding the b-lognormal in time of the History of any Civilization. This new theorem deserves a new name. We propose to call it by the Latin name of 'E-Pluribus-Unum' Theorem.

Indeed, 'E-Pluribus-Unum' stands for 'Out of Many, just One', and this was the official motto of the USA from 1782 to 1956, when replaced by 'In God we trust' (probably in opposition to atheists views then supported by the Soviet Union). In this author's view, 'E-Pluribus-Unum' adapts well to what we have described mathematically in the first part of this paper about Civilizations in Evo-SETI Theory.

PART 2: NEW RESULTS ABOUT MOLECULAR CLOCK IN EVO-SETI THEORY

Darwinian evolution as a Geometric Brownian Motion (GBM)

In (Maccone 2013, p. 220–227), this author 'dared' to re-define Darwinian Evolution as 'just one particular realization of the stochastic process called GBM in the increasing number of Species living on Earth over the last 3.5 billion years'.

Now, the GBM mean value is the simple exponential function of the time

$$m_{GBM}(t) = A \, e^{Bt} \qquad (79)$$

with A and B being the positive constants. Thus, A equals $m_{GBM}(0)$, the number of Species living on Earth right now, and

$$m_{GBM}(ts) = 1 \qquad (80)$$

represents the first 'living Species' (call it RNA ?) that started life on Earth at the 'initial instant' ts ('time of start'). In (Maccone 2013, 2014) we assumed that life started on Earth 3.5 billion years ago, that is

$$ts = -3.5 \times 10^9 \text{ years} \qquad (81)$$

and that the number of Species living on Earth nowadays is 50 million

$$A = 50 \times 10^6 . \qquad (82)$$

Consequently, the two constants A and B in equation (79) may be exactly determined as follows:

$$\begin{cases} A = 50 \times 10^6 \\ B = -\dfrac{\ln(A)}{ts} = 1.605 \times 10^{-16} \text{ s}^{-1}. \end{cases} \qquad (83)$$

Please note that these two numbers are to be regarded as *experimental constants* (*valid for Earth only*), just like the acceleration of gravity $g = 9.8 \text{ m s}^{-2}$, the solar constant, and other Earthly constants.

Also, some paleontologists claim that life on Earth started earlier, say 3.8 billion years ago. In this case, equations (83) is to be replaced by the slightly different

$$\begin{cases} A = 50 \times 10^6 \\ B = -\dfrac{\ln(A)}{ts} = 1.478 \times 10^{-16} \text{ s}^{-1}. \end{cases} \qquad (84)$$

but B did not change much, and so we will keep equations (83) as the right values as it was done in (Maccone 2013, 2014).

Figure 3 shows two realizations of GBM revealing 'at a glance' the exponential increasing mean value of this lognormal stochastic process (see (Maccone 2013, p. 222–223) for more details, and (Maccone 2014, p. 291–294) for a full mathematical treatment).

Assuming GBM as the 'curve' (a fluctuating one!) representing the increasing number of Species over the last 3.5 billion years has several advantages:

(1) It puts on a firm mathematical ground the intuitive notion of a 'Malthusian' exponential growth.

(2) It allows for Mass Extinctions to have occurred in the past history of life on Earth, as indeed it was the case. Mass Extinctions in the Evo-SETI Theory are just times when the number of living Species 'decreased very much' from its exponential mean value, for instance going down by 70% just 250 million years ago, but not going down to zero, otherwise we would not be living now. In (Maccone 2014) this author did more modelling about Mass Extinctions.

(3) After what we just said, the two curves called 'upper' and 'lower standard deviation curve' are clearly playing a major role in Evo-SETI Theory. They represent the average departure of the actual number of living Species from their exponential mean value, as shown in Figure 4. In (Maccone 2014, p. 292–293), the author proved that the upper (plus sign) and lower (minus sign) standard deviation curves of GBM (above and below the mean value exponential (79), respectively), are given by the equations

$$\text{upper\_\&\_lower\_std\_curves\_of\_GBM}(t)$$

$$= m_{GBM}(t) \cdot \left[1 \pm \sqrt{e^{\sigma_{GBM}^2 (t - ts)} - 1} \right]. \qquad (85)$$

The new constant σ_{GBM} appearing in equation (85) (not to be confused with the simple σ of the b-lognormal (5)) is provided by the *final conditions* affecting the GBM at the final instant of its motion, namely zero (=now) in our conventions. Denoting by A the current number of Species on Earth, as we did in equations (79) and (82), and by δA the standard deviation around A nowadays (for instance, we assumed A to be equal to 50 million but we might add an uncertainty of, say, ±10 million Species around that value), then the σ_{GBM} in equation (85) is given by

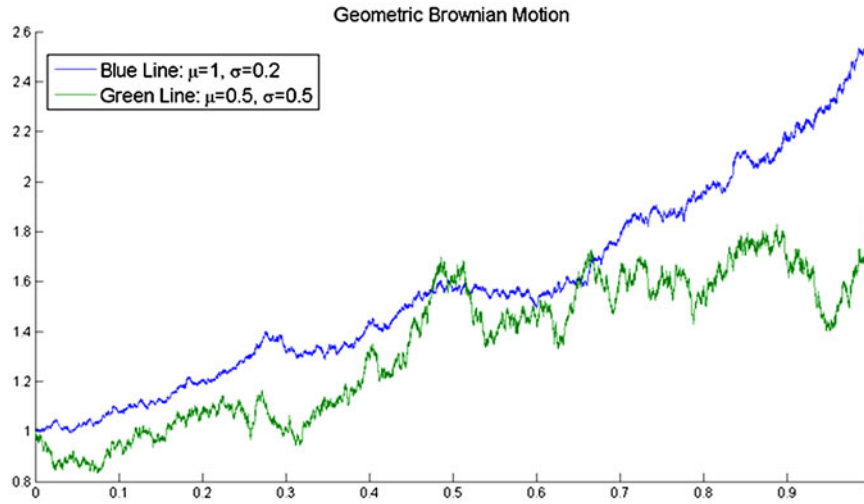$$\sigma_{GBM} = \frac{\sqrt{\ln[1 + (\delta A/A)^2]}}{\sqrt{-ts}} . \qquad (86)$$

Fig. 3. Two realizations (i.e. actual instances) of Geometric Brownian Motion taken from the GBM Wikipedia site http://en.wikipedia.org/wiki/Geometric_Brownian_motion . Please keep in mind that the name 'Brownian Motion' is *incorrect* and misleading: in fact, physicists and mathematicians mean by 'Brownian Motion' a stochastic process whose pdf is Gaussian, i.e. normal. But this is *not* the case with GBM, whose pdf is lognormal instead. This incorrect denomination seems to go back to the Wall Street financial users of the GBM.

A leap forward: for any assigned mean value $m_L(t)$ we construct its lognormal stochastic process

A profound message was contained in (Maccone 2014) for all future applications of lognormal stochastic processes (both GBM and other than GBM): for any assigned at will mean value function of the time $m_L(t)$, namely for any *trend*, we are able to find the equations of the lognormal process that has exactly that mean value, i.e. that trend !

This author was so amazed by this discovery (that he made between September 2013 and January 2014) that he could not give a complete account of it when he published (Maccone 2014) available in Open Access since October 2014. Thus, the present new paper is a completion of (Maccone 2014), but also is a leap forward in another unexpected direction: the proof that the Molecular Clock, well-known to geneticists for more than 50 years, may be derived mathematically as a consequence of the Evo-SETI Theory.

Completing (Maccone 2014): letting $M_L(t)$ there be replaced everywhere by $m_L(t)$, the assigned trend

In (Maccone 2014) this author started by considering the general lognormal process $L(t)$ whose pdf is the lognormal

$$L(t)\text{\_pdf}(n\,;M_L(t),\sigma_L,t) = \frac{e^{-[\ln(n)-M_L(t)]^2/(2\sigma_L^2(t-ts))}}{\sqrt{2\pi}\,\sigma_L\sqrt{t-ts}\,n} \tag{87}$$

$$\text{with}\begin{cases} n \geq 0, \\ t \geq ts, \end{cases} \text{and}\begin{cases} \sigma_L \geq 0, \\ M_L(t) = \text{arbitrary function of } t. \end{cases}$$

Equation (87) also is the starting point of all subsequent calculations in the #3 Appendix, where it has the number (%o6). Notice that the positive parameter σ_L in the pdf (87) is denoted sL in the #3 Appendix, simply because Maxima did not allow

us to denote it σ_L for Maxima-language reasons too long to explain! Also, $m_L(t)$ is more simply denoted $m(t)$ in the #3 Appendix, and $M_L(t)$ is more simply denoted $M(t)$.

The mean value, i.e. the trend, of the process $L(t)$ is an *arbitrary* (and continuous) function of the time denoted by $m_L(t)$ in the sequel. In equations, that is, one has, by definition

$$m_L(t) = \langle L(t) \rangle. \tag{88}$$

In other words, we analytically compute the following integral, yielding the mean value of the pdf (87), getting (for the proof, see (%o5) and (%o6) in the #3 Appendix)

$$\begin{aligned} m_L(t) &\equiv \int_0^\infty n \cdot \frac{e^{-[\ln(n)-M_L(t)]^2/(2\sigma_L^2(t-ts))}}{\sqrt{2\pi}\,\sigma_L\sqrt{t-ts}\,n}\,dn \\ &= e^{M_L(t)}e^{(\sigma_L^2/2)(t-ts)}. \end{aligned} \tag{89}$$

This is (%o8) in the #3 Appendix, and from now on, we will drop the usual sentence 'in the #3 Appendix' and just report the #3 Appendix equation numbers corresponding to the equation numbers in this paper.

We have thus discovered the following crucial *mean value formula*, holding good for the general lognormal process $L(t)$ inasmuch as the function $M_L(t)$ is arbitrary, and so is the trend $m_L(t)$ (%o9)

$$m_L(t) = e^{M_L(t)}\,e^{(\sigma_L^2/2)(t-ts)}. \tag{90}$$

This was done by the author in (Maccone 2014) already, p. 292, equation (3). But at that time this author failed to *invert* (90), i.e. to solve it for $M_L(t)$, with the result (%o10):

$$M_L(t) = \ln[m_L(t)] - \frac{\sigma_L^2}{2}\,(t-ts). \tag{91}$$

Equation (91) shows that it is always possible to get rid of $M_L(t)$ by substituting equation (91) into any equation
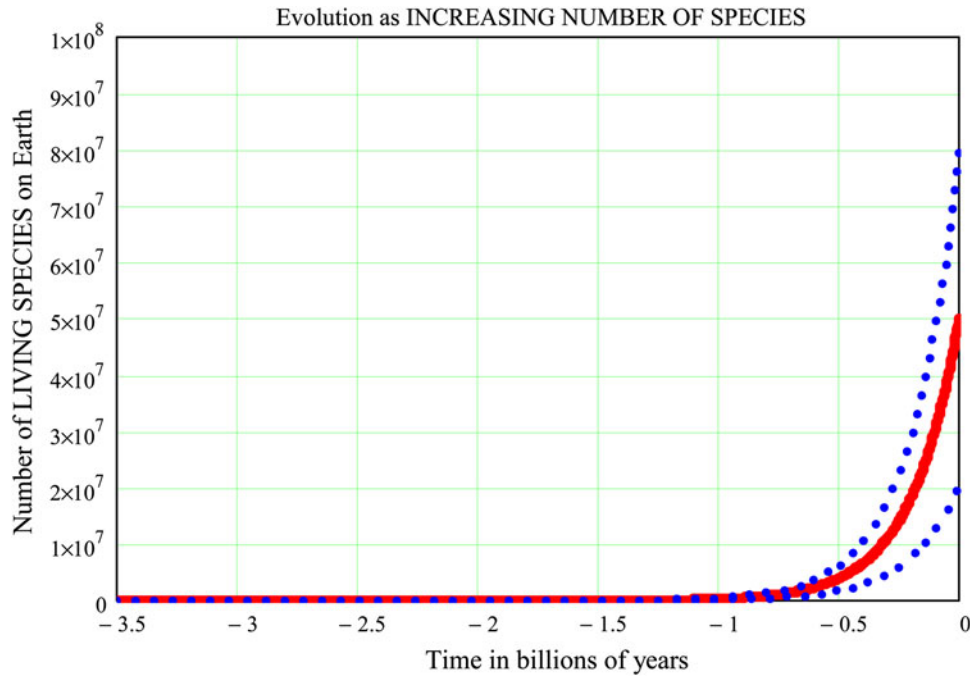
Fig. 4. Darwinian evolution as the increasing number of living Species on Earth between 3.5 billion years ago and now. The red solid curve is the mean value of the GBM stochastic process, given by equation (79), while the blue dot–dot curves above and below the mean value are the two standard deviation upper and lower curves, given by equation (105). The 'Cambrian Explosion' of life that started around 542 million years ago, approximately marks in the above plot 'the epoch of departure from the time axis for all the three curves, after which they start climbing up more and more'. Notice also that the starting value of living Species 3.5 billion years ago is ONE by definition, but it 'looks like' zero in this plot since the vertical scale (which is the true scale here, not a log scale) does not show it. Notice finally that nowadays (i.e. at time $t = 0$) the two standard deviation curves have exactly the same distance from the middle mean value curve, i.e. 30 million living Species above or below the mean value of 50 million Species. These are assumed values that we used just to exemplify the GBM mathematics: biologists might assume other numeric values.

containing $M_L(t)$ and appearing in (Maccone 2014). In other words, one may re-express all results of Maccone 2014) in terms of the trend function $m_L(t)$ **only**, justifying the idea 'you give me the trend $m_L(t)$ and I'll give you all the equations of the lognormal process $L(t)$ for which $m(t)$ is the trend'.

An immediate consequence of (91) is found by letting $t = ts$ ((‰o11) and (‰o12)):

$$M_L(ts) = \ln[m_L(ts)], \quad \text{i.e.} \quad m_L(ts) = e^{M_L(ts)}. \quad (92)$$

For instance, equation (8) on p. 292 of (Maccone 2014) yields the σ_L in terms of both the initial input data (ts, Ns) and final input data $(te, Ne, \delta Ne)$:

$$\sigma_L = \frac{\sqrt{\ln[e^{2M_L(ts)} + (\delta Ne)^2 (Ns/Ne)^2] - 2M_L(ts)}}{\sqrt{te - ts}} \quad (93)$$

Well, this equation simplifies dramatically once equation (92) and the initial condition (equation (5) on p. 292 of Maccone (2014)) (‰o13)

$$m_L(ts) = Ns \quad (94)$$

are taken into account. In fact, a few steps starting from equation (93) show that, by virtue of equations (92) and (94), it

reduces to (‰o31)

$$\sigma_L = \frac{\sqrt{\ln[1 + (\delta Ne/Ne)^2]}}{\sqrt{te - ts}} \quad (95)$$

Of course, the corresponding GBM special case of equation (95) is (86), obtained by letting $(te = 0, Ne = A)$ into equation (95).

Also, equation (95) may be formally rewritten as follows:

$$\sigma_L^2 = \frac{\ln[1 + (\delta Ne/Ne)^2]}{(te - ts)} = \ln\left\{\left[1 + \left(\frac{\delta Ne}{Ne}\right)^2\right]^{1/(te - ts)}\right\}. \quad (96)$$

Taking the exponential of equation (96), one thus gets a yet unpublished equation that we shall use in a moment

$$e^{\sigma_L^2 (t - ts)} = \left[1 + \left(\frac{\delta Ne}{Ne}\right)^2\right]^{(t - ts)/(te - ts)}. \quad (97)$$

Going back to the general lognormal process $L(t)$, in (Maccone 2014), Tables 1–3, we also proved that the moment of order k (with $k = 0, 1, 2, \ldots$) of the $L(t)$ process is given by

$$\langle [L(t)]^k \rangle = [m_L(t)]^k e^{k(k-1)\sigma_L^2 (t - ts)/2} \quad (98)$$

The mathematical proof of this key result by virtue Maxima is given in the #3 Appendix, equations (‰i16) through (‰o21).

Table 3. *Summary of the most important mathematical properties of the lognormal stochastic process $L(t)$.*

Lognormal stochastic process $L(t)$	In terms of $m_L(t)$ and $\sigma_L(t)$:	In terms of $m_L(t)$ and of: (1) the two initial inputs (ts, Ns). (2) plus the three final inputs (te, Ne, δNe):
b-lognormal pdf starting at ts (%o6)	$L(t)\_\text{pdf}(n; M_L(t), \sigma_L, t) = \dfrac{e^{-[\ln(n)-M_L(t)]^2/(2\sigma_L^2(t-ts))}}{\sqrt{2\pi}\sigma_L\sqrt{t-ts}\,n}$ with $\begin{cases} n \geq 0, \\ t \geq ts, \end{cases}$ and $\begin{cases} \sigma_L \geq 0, \\ M_L(t) = \text{arbitrary function of } t. \end{cases}$	$\sigma_L = \dfrac{\sqrt{\ln[1+(\delta Ne/Ne)^2]}}{\sqrt{te-ts}}$ that is (%o31)
Mean value $\langle L(t)\rangle$ also called Trend	$m_L(t) = e^{M_L(t)}e^{(\sigma_L^2/2)(t-ts)}$	$m_L(t)$ (arbitrarily assigned, i.e. known)
Variance	$[m_L(t)]^2\left[e^{\sigma_L^2(t-ts)}-1\right]$ that is (%o25)	$[m_L(t)]^2\left\{\left[1+\left(\dfrac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}-1\right\}$
Standard Deviation	$m_L(t)\sqrt{e^{\sigma_L^2(t-ts)}-1}$ that is (%o27)	$m_L(t)\sqrt{\left[1+\left(\dfrac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}-1}$
All moments, i.e. k-th moment	$\langle L(t)\rangle^k = [m_L(t)]^k e^{k(k-1)\sigma_L^2(t-ts)/2}$ that is (%o20)	$\langle L(t)\rangle^k = [m_L(t)]^k\left[1+\left(\dfrac{\delta Ne}{Ne}\right)^2\right]^{k(k-1)(t-ts)/(2(te-ts))}$
Upper Standard Deviation Curve	$m_L(t)\left[1+\sqrt{e^{\sigma_L^2(t-ts)}-1}\right]$ that is (%o34)	$m_L(t)\left\{1+\sqrt{\left[1+\left(\dfrac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}-1}\right\}$
Lower Standard Deviation Curve	$m_L(t)\left[1-\sqrt{e^{\sigma_L^2(t-ts)}-1}\right]$ that is (%o42)	$m_L(t)\left\{1-\sqrt{\left[1+\left(\dfrac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}-1}\right\}$
Mode (= peak abscissa)	$n_\text{mode} = n_\text{peak} = \dfrac{m_L(t)}{e^{(3/2)\sigma_L^2(t-ts)}}$ that is (%o61)	$n_\text{mode} = n_\text{peak} = \dfrac{m_L(t)}{[1+(\delta Ne/Ne)^2]^{3(t-ts)/(2(te-ts))}}$
Ordinate of the peak = Mode ordinate	$f_{L(t)}(n_\text{mode}) = \dfrac{1}{m_L(t)\sqrt{2\pi}\sigma_L\sqrt{t-ts}\,e^{\sigma_L^2(t-ts)}}$ that is (%o65)	$f_{L(t)}(n_\text{mode}) = \dfrac{1}{m_L(t)\sqrt{2\pi}\sigma_L\sqrt{t-ts}\,[1+(\delta Ne/Ne)^2]^{(t-ts)/(te-ts)}}$
Median (= fifty-fifty probability)	$\dfrac{m_L(t)}{\sqrt{e^{\sigma_L^2(t-ts)}}}$ that is (%o57)	$\dfrac{m_L(t)}{\sqrt{[1+(\delta Ne/Ne)^2]^{(t-ts)/(te-ts)}}}$
Skewness	$\dfrac{K_3}{(K_2)^{3/2}} = \left[e^{\sigma^2(t-ts)}+2\right]\sqrt{e^{\sigma^2(t-ts)}-1}$	$\left\{\left[1+\left(\dfrac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}+2\right\}\sqrt{\left[1+\left(\dfrac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}-1}$
Kurtosis	$\dfrac{K_4}{(K_2)^2} = e^{4\sigma^2(t-ts)}+2e^{3\sigma^2(t-ts)}+3e^{2\sigma^2(t-ts)}-6$	Same with $e^{\sigma_L^2(t-ts)} = \left[1+\left(\dfrac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}$

A new discovery, presented in this paper for the first time, is that, by virtue of equation (97), equation (98) may be directly rewritten in terms of the boundary conditions (ts, te, Ne, δNe):

$$\langle [L(t)]^k\rangle = [m_L(t)]^k\left[1+\left(\frac{\delta Ne}{Ne}\right)^2\right]^{k(k-1)(t-ts)/(2(te-ts))}. \quad (99)$$

For $k = 0$, both equations (98) and (99) yield the normalization condition of $L(t)$:

$$\langle [L(t)]^0\rangle = 1. \quad (100)$$

For $k = 1$ both equations (98) and (99) yield the mean value again

$$\langle [L(t)]^1\rangle = \langle L(t)\rangle = m_L(t). \quad (101)$$

But for $k = 2$ (the mean value of the square of $L(t)$) the novelties start. In fact, equation (99) yields

$$\langle [L(t)]^2\rangle = [m_L(t)]^2\left[1+\left(\frac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}. \quad (102)$$

Since the variance of $L(t)$ is given by the mean value of its square minus the square of its mean value, subtracting the square of equation (101) into (102) yields

$$\text{Variance of } L(t) = [m_L(t)]^2\left\{\left[1+\left(\frac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)}-1\right\}. \quad (103)$$

The square root of equation (103) is of course the standard deviation of $L(t)$:

$$\text{Standard\_Deviation\_of\_}L(t)$$

$$= m_L(t) \cdot \sqrt{\left[1 + \left(\frac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)} - 1}. \tag{104}$$

This is a quite important formula for all future applications of our general lognormal process $L(t)$ to the Evo-SETI Theory.

Even more important for all future graphical representations of the general lognormal process $L(t)$ is the formula yielding the upper (plus sign) and lower (minus sign) standard deviation curves as two functions of t. It follows immediately from the mean value $m_L(t)$ plus or minus the standard deviation (104):

$$\text{Two\_Standard\_Deviation\_CURVES\_of\_}L(t)$$

$$= m_L(t) \cdot \left\{1 \pm \sqrt{\left[1 + \left(\frac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)} - 1}\right\}. \tag{105}$$

Just to check that our results are correct, from (105) one may immediately verify that:

(1) Letting $t = ts$ in (105) yields $m_L(ts)$ as the value of both curves. But this is the same value as the mean value at $t = ts$ also. Thus, at $t = ts$ the process $L(t)$ starts with probability one, since all three curves are at the just the same point.

(2) Letting $t = te$ in (105) yields

$$\text{Two\_Standard\_Deviation\_CURVES\_at\_}te$$

$$= m_L(te) \cdot \left\{1 \pm \frac{\delta Ne}{Ne}\right\} = Ne \pm \delta Ne \tag{106}$$

where $m_L(te) = Ne$ was used in the last step. This result is correct inasmuch as the two curves intercept the vertical line at $t = te$ exactly at those two ordinates.

(3) Letting $\delta Ne = 0$ in (105) makes the two curves coincide with the mean value $m_L(t)$, and that is correct.

(4) As a matter of terminology, we add that the factor

$$\sqrt{e^{\sigma_L^2 (t-ts)} - 1} = \sqrt{\left[1 + \left(\frac{\delta Ne}{Ne}\right)^2\right]^{(t-ts)/(te-ts)} - 1} \tag{107}$$

is called 'coefficient of variation' by statisticians since it is the ratio between the standard deviation and the mean value for all time values of the $L(t)$ process, and in particular at the end time $t = te$, when it equals $\delta Ne/Ne$.

(5) Finally, we have summarized the content of this important set of mathematical results in Table 3.

Peak-Locus Theorem

The Peak-Locus Theorem is a new mathematical discovery of ours playing a central role in the Evo-SETI theory. In its most general formulation, it holds good for any lognormal process $L(t)$ and any arbitrary mean value $m_L(t)$, as we show in this section.

In words, and utilizing the simple example of the Peak-Locus Theorem applied to GBMs, the Peak-Locus Theorem states what shown in the Fig. 5: the family of all b-lognormals 'trapped' between the time axis and the growing exponential of the GBMs (where all the b-lognormal peaks lie) can be *exactly* (i.e. without any numerical approximation) described by three equations yielding the three parameters $\mu(p)$, $\sigma(p)$ and $b(p)$ as three functions of the peak abscissa, p, only.

In equations, the Peak-Locus Theorem states that the family of b-lognormals having each its peak exactly located *on* the mean value curve (88), is given by the following three equations, specifying the parameters $\mu(p)$, $\sigma(p)$ and $b(p)$, appearing in the b-lognormal (5) as three functions of the single 'independent variable' p, i.e. the abscissa (i.e. the time) of the b-lognormal's peak:

$$\begin{cases} \mu(p) = \dfrac{e^{\sigma_L^2 \cdot p}}{4\pi[m_L(p)]^2} - p\dfrac{\sigma_L^2}{2} \\[2ex] \sigma(p) = \dfrac{e^{(\sigma_L^2/2) \cdot p}}{\sqrt{2\pi}\, m_L(p)} \\[2ex] b(p) = p - e^{\mu(p) - [\sigma(p)]^2}. \end{cases} \tag{108}$$

This general form of the Peak-Locus Theorem is proven in the Appendix by equations (%i66) through (%o82). The remarkable point about all this seems to be the exact separability of all the equations involved in the derivation of equations (108), a fact that was unexpected to this author when he discovered it around December 2013. And the consequences of this new result are in the applications:

(1) For instance in the 'parabola model' for Mass Extinctions that was studied in Section 10 of (Maccone 2014).

(2) For instance to the Markov–Korotayev Cubic that was studied in Section 12 of (Maccone 2014; Markov & Korotayev 2007, 2008).

(3) And finally in the many stochastic processes having each a Cubic mean value that are just the natural extension into statistics of the deterministic Cubics studied by this author in Chapter 10 of his book 'Mathematical SETI' (Maccone 2012). But the study of the Entropy of all these Cubic Lognormal Processes has to be differed to a future research paper.

Notice now that, in the particular case of the GBMs having mean value $e^{\mu_{\text{GBM}}(t-ts)}$ with $\mu_{\text{GBM}} = B$, and starting at $ts = 0$ with $N_0 = Ns = Ne = A$, the Peak-Locus Theorem (108) boils down to the simpler set of equations

$$\begin{cases} \mu(p) = \dfrac{1}{4\pi A^2} - Bp \\[2ex] \sigma = \dfrac{1}{\sqrt{2\pi}\, A} \\[2ex] b(p) = p - e^{\mu(p) - \sigma^2}. \end{cases} \tag{109}$$

In this simpler form, the Peak-Locus Theorem was already published by the author in Maccone (2012), while its most general form (108) is now proven in detail.
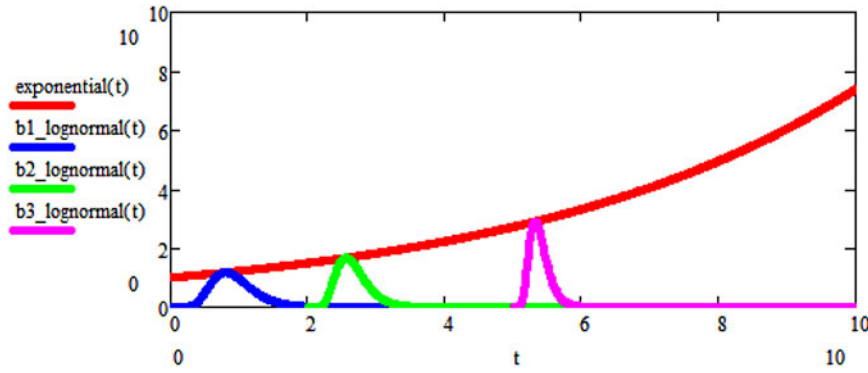
Fig. 5. GBM exponential as the geometric LOCUS OF THE PEAKS of *b*-lognormals. Each *b*-lognormal is a lognormal starting at a time (*b* = birth time) larger than zero and represents a different SPECIES that originated at time *b* of Evolution. That is CLADISTICS in our Evo-SETI Model. It is evident that, the more the generic 'Running *b*-lognormal' moves to the right, its peak becomes higher and higher and narrower and narrower, since the area under the *b*-lognormal always equals 1 (normalization condition). Then, the (Shannon) ENTROPY of the running *b*-lognormal is the DEGREE OF EVOLUTION reached by the corresponding SPECIES (or living being, or civilization, or ET civilization) in the course of Evolution.

Proof. Let us firstly call 'Running *b*-lognormal' (abbreviated 'RbL') the generic *b*-lognormal of the family, starting at *b*, having peak at *p* and having the variable parameters $\mu(p)$ and $\sigma(p)$. Then the starting equation of the Peak-Locus Theorem (108) is the #3 Appendix equation (‰73) expressing the fact that the peak of the RbL equals the mean value (90), where, however, the old independent variable *t* must be replaced by the new independent variable *p* of the RbL, that is

$$\frac{e^{\sigma^2/2-\mu}}{\sqrt{2\pi}\sigma} = e^{M_L(p)}e^{(\sigma_L^2/2)(p-ts)}. \tag{110}$$

This equation may 'surprisingly' be separated into the following two simultaneous equations (‰74)

$$\begin{cases} e^{(\sigma^2/2)-\mu} = e^{(\sigma_L^2/2)p} \\ \dfrac{1}{\sqrt{2\pi}\sigma} = e^{M_L(p)}e^{-(\sigma_L^2/2)ts}. \end{cases} \tag{111}$$

There are two advantages brought in by this separation of variables: in the upper equation (111) the exponentials 'disappear' yielding (‰79)

$$\frac{\sigma^2}{2} - \mu = \frac{\sigma_L^2}{2}\,p \tag{112}$$

while the lower equation (111) is in σ only, and thus it may be solved for σ immediately (‰76)

$$\sigma = \frac{e^{(\sigma_L^2/2)ts}e^{-M_L(p)}}{\sqrt{2\pi}}. \tag{113}$$

We may now get rid of *M(p)* in equation (113) by replacing it by virtue of (91), getting, after a few steps and rewriting *p* instead of *t*, (‰78)

$$\sigma = \frac{e^{(\sigma_L^2/2)p}}{\sqrt{2\pi}\,m_L(p)}. \tag{114}$$

which is just the middle equation (108).

Finally, equation (112) may be solved for μ (‰79).

$$\mu = \frac{\sigma^2}{2} - \frac{\sigma_L^2}{2}\,p \tag{115}$$

so that, inserting equation (114) into (115), the final expression of μ is found also (‰80)

$$\mu = \frac{e^{\sigma_L^2 p}}{4\pi[m_L(p)]^2} - \frac{\sigma_L^2}{2}p. \tag{116}$$

Our general Peak-Locus Theorem (108) has thus been proven completely.

tsGBM and GBM sub-cases of the Peak-Locus Theorem

The general Peak-Locus Theorem proved in the previous section includes, as sub-cases, many particular forms of the arbitrary mean-value function $m_L(t)$. In particular, we now want to consider two of them:

(1) The tsGBM, i.e. the GBM starting at any given time *ts*, like the origin of life on Earth, that started at *ts* = −3.5 billion years ago.
(2) The 'ordinary' GBM, used in the Mathematics of Finances, starting at *ts* = 0. Clearly, the ordinary GBM is, in its turn, a sub-subcase of the tsGBM.

Then, the tsGBM is characterized by the equation

$$m_{\text{tsGBM}}(p) = m_{\text{tsGBM}}(ts)e^{B(p-ts)} \tag{117}$$

having set in agreement with equations (79) and (90), (‰84)

$$B = \frac{\sigma_{\text{tsGBM}}^2}{2}. \tag{118}$$

One may determine the numeric constant *B* in terms of both the initial and final conditions of the tsGBM by replacing into equation (117) *p* by *te* (the end-time, i.e. the time of the final condition) and then solving equation (117) for *B* (‰88)

$$B = \frac{\ln(m_{\text{tsGBM}}(te)/m_{\text{tsGBM}}(ts))}{te - ts}. \tag{119}$$

In the Evo-SETI Theory we assume ts to be the time of the 'beginning of life', when there was only one living Species (the first one, probably RNA, at 3.5 billion year ago on Earth, but we do not know at what time on exoplanets) and so we have

$$m_{\text{tsGBM}}(ts) = 1. \tag{120}$$

Then equation (117) reduces to

$$m_{\text{tsGBM}}(p) = e^{B(p-ts)} \tag{121}$$

and equation (119) reduces to

$$B = \frac{\ln(m_{\text{tsGBM}}(te))}{te - ts}. \tag{122}$$

As we already did in the section 'Death Formula', we assume the number of living Species on Earth nowadays (i.e. at $te = 0$) to be equal to 50 million, namely $m(te) = 50$ million. Then equation (122) reduces to equation (83), as it must be.

Finally, the ordinary GBM subcase of tsGBM and sub-subcase of $L(t)$ is characterized by equation (118) and by

$$m_{\text{GBM}}(t) = Ae^{Bt}. \tag{123}$$

Then, inserting both equations (118) and (123) into the general Peak-Locus Theorem (108), the latter yields equation (109) as shown by (%o95). This is the 'old' Peak-Locus Theorem, firstly discovered by this author in late 2011 and already published by him in Chapter 8 of his 2012 book 'Mathematical SETI' of 2012, p. 218–219.

For more applications of the Peak-Locus Theorem to polynomial mean values, see Maccone (2014) (p. 294–308).

Shannon entropy of the running *b*-lognormal

The Shannon Entropy of the Running *b*-lognormal is the key to measure the 'disorganization' of what that running *b*-lognormal represents, let it be a Species (in Evolution) or a Civilization (in Human History) or even an Alien Civilization (in SETI).

As it is well known, the Shannon Entropy (17) (measured in bits) of the Running *b*-lognormal having its peak at time p and the three parameters μ, σ, b is given by (%o96) (for the proof of this key mathematical result, please see Chapter 30 of the author's book 'Mathematical SETI', p. 685–687, the idea behind the proof is to expand the log of the Shannon Entropy of the *b*-lognormal, so that the calculation is split into three integrals, each of which may actually be computed exactly):

$$H = \frac{\ln(\sqrt{2\pi}\sigma) + \mu + 1/2}{\ln 2}. \tag{124}$$

Having so said, the next obvious step is to insert the μ and σ given by the Peak-Locus Theorem (108) into (124). After a few steps, we thus obtain the Shannon Entropy of the Running *b*-lognormal (see (%o97) and (%o98)):

$$H(p, m_L(p)) = \frac{e^{\sigma_L^2 \cdot p}}{4\pi \ln 2 \cdot [m_L(p)]^2} - \frac{\ln(m_L(p))}{\ln 2} + \frac{1}{2\ln 2}. \tag{125}$$

This is the fundamental Shannon Entropy H of the Running *b*-lognormal for *any* given mean value $m_L(t)$. Notice that H is a function of the peak abscissa p in two ways:
(1) Directly, as in the term $e^{\sigma_L^2 \cdot p}$, and
(2) Through the assigned mean value $m_L(p)$.

Introducing our... Evo-Entropy(*p*) measuring how much a life form has evolved

The Shannon Entropy was introduced by Claude Shannon (1916–2001) in 1948 in his seminal work about Information Theory, dealing of course with telecommunications, channel capacities and computers. But... *we need something else to measure 'how evolved' a life form is: we need a positive function of the time starting at zero at the time ts of the origin of life on a certain planet, and then increasing (rather than decreasing).*

This new function is easily found: it is just the Shannon Entropy (17) WITHOUT THE MINUS SIGN IN FRONT OF IT (so as to make it an increasing function, rather than a decreasing function) and WITH THE NUMERIC VALUE $-H(ts)$ SUBTRACTED, so as it starts at zero at the initial instant ts.

This new function of p we call EVO-ENTROPY (Evolution Entropy) and its mathematical definition is thus simply (see (%o101) and (%o102)):

$$\begin{aligned} \text{EvoEntropy}&(p, m_L(p)) \\ &= -H(p, m_L(p)) + H(ts, m_L(ts)). \end{aligned} \tag{126}$$

In some previous papers by this author about Evo-SETI Theory, the Evo-Entropy (126) was called 'Evo-Index' (Index of Evolution) or with other similar names, but we now prefer to call it Evo-Entropy to make it clear that it is just the Shannon Entropy with the sign reversed and with value zero at the origin of life.

Next we compute the actual expression of Evo-Entropy as a function of the only variable p, the Running *b*-lognormal peak. To this end, we must first get the expression of (125) at the initial time ts. It is (%o99)

$$\begin{aligned} H(ts, m_L(ts)) = \frac{e^{p \cdot ts}}{4\pi \ln 2 \cdot [m_L(ts)]^2} - \frac{\ln(m_L(ts))}{\ln 2} \\ + \frac{1}{2\ln 2}. \end{aligned} \tag{127}$$

Subtracting equation (127) into (125) with the minus sign reversed, we get the for the final (126) form of our Evo-Entropy (%o100)

$$\begin{aligned} \text{EvoEntropy}(p, m_L(p)) = \frac{1}{4\pi \ln 2} \left\{ \frac{e^{\sigma_L^2 \cdot ts}}{[m_L(ts)]^2} - \frac{e^{\sigma_L^2 \cdot p}}{[m_L(p)]^2} \right\} \\ + \frac{\ln(m_L(p)/m_L(ts))}{\ln 2}. \end{aligned} \tag{128}$$

The Evo-Entropy (128) is thus made up by two terms:
(1) The term

$$\frac{1}{4\pi \ln 2} \left\{ \frac{e^{\sigma_L^2 \cdot ts}}{[m_L(ts)]^2} - \frac{e^{\sigma_L^2 \cdot p}}{[m_L(p)]^2} \right\} \tag{129}$$
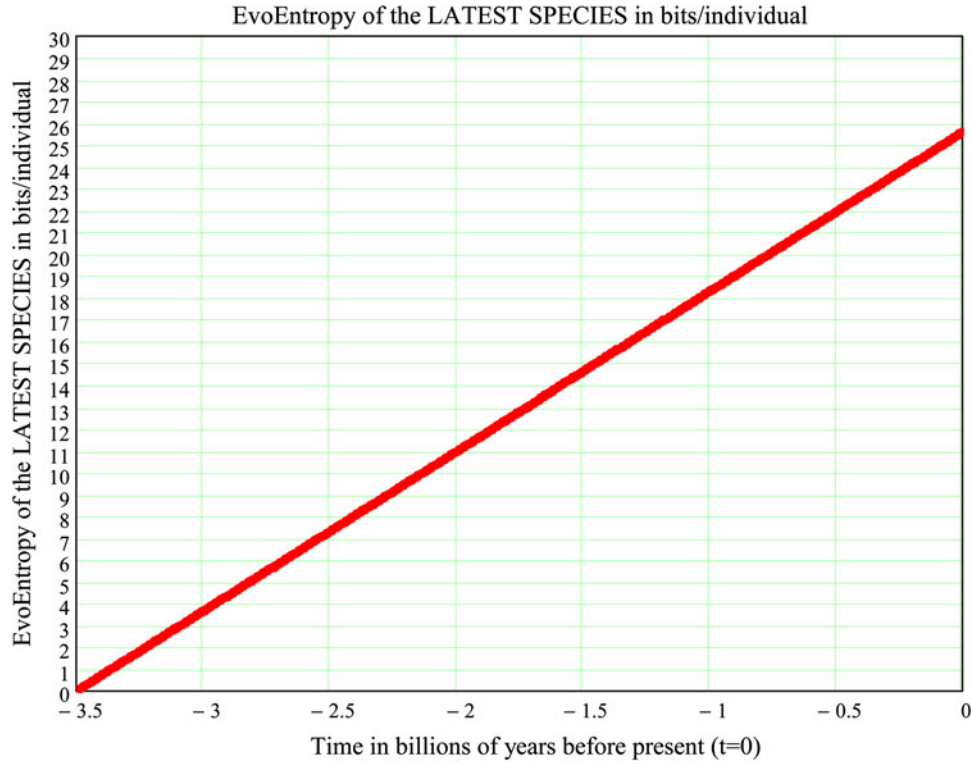
Fig. 6. EvoEntropy (in bits per individual) of the latest species appeared on Earth during the last 3.5 billion years. This shows that a Man (i.e. the leading Species nowadays) is 25.575 bits more evolved than the first form of life (call it RNA?) 3.5 billion years ago.

we shall call the NON-LINEAR PART of the Evo-Entropy (128), while

(2) The term

$$\frac{\ln(m_L(p)/m_L(ts))}{\ln 2} \tag{130}$$

we shall call the LINEAR PART of the Evo-Entropy (128), as we explain in the next section.

The Evo-Entropy(p) of tsGBM increases exactly linearly in time

Consider again the tsGBM defined by equation (117) with (118). If we insert equation (117) into the EvoEntropy (128), then two dramatic simplifications occur:

(1) The non-linear term (129) *vanishes*, inasmuch as it reduces to

$$\frac{1}{4\pi\ln 2[m_L(ts)]^2}\left\{e^{\sigma_L^2\cdot ts} - \frac{e^{\sigma_L^2\cdot p}}{e^{\sigma_L^2\cdot p}\cdot e^{-\sigma_L^2\cdot ts}}\right\} = 0. \tag{131}$$

(2) The linear term (130) simplifies, yielding (%o104)

$$\frac{\ln(m_L(p)/m_L(ts))}{\ln 2} = \frac{\ln(m_L(ts)e^{B(p-ts)}/m_L(ts))}{\ln 2}$$
$$= \frac{B}{\ln 2}(p - ts). \tag{132}$$

In other words, the Evo-Entropy of tsGBM simply is the LINEAR function of the Running b-lognormal peak p

$$\text{EvoEntropy}_{\text{tsGBM}}(p) = \frac{B}{\ln 2}(p - ts). \tag{133}$$

This is a great result! And it was already envisioned back in 2012 in Chapter 30 of the author's book 'Mathematical SETI' when he found that the Evo-Entropy *difference* between two Civilizations 'with quite different levels of technological development' (like the Aztecs and the Spaniards in 1519) is given by the equation

$$\text{EvoEntropy}_{\text{tsGBM}}(p_2 - p_1) = \frac{B}{\ln 2}(p_2 - p_1). \tag{134}$$

(see equation (30.29) on p. 693 of that book, where the old minus sign in front of the Shannon Entropy still ruled because this author had not yet 'dared' to get rid of it, as he did now in the new definition (126) of EvoEntropy.

But what is the graph of this famous linear increase of Evo-Entropy ? It is given by Fig. 6.

So, we have discovered that the tsGBM Entropy in our Evo-SETI model and the Molecular Clock (see Nei (2013) and Nei & Kumar (2000)) are the same linear time function, apart for multiplicative constants (depending on the adopted units, such as bits, seconds, etc.). This conclusion appears to be of key importance to understand 'where a newly discovered exoplanet stands on its way to develop LIFE'.

Conclusions

More and more exoplanets are now being discovered by astronomers either by observations from the ground or by virtue of space missions, like 'CoRot', 'Kepler', 'Gaia' and other future space missions.

As a consequence, a recent estimate sets at 40 billion the number of Earth-sized planets orbiting in the habitable zones of Sun-like stars and red dwarf stars within the Milky Way Galaxy.

With such huge numbers of 'possible Earths' in sight, Astrobiology and SETI are becoming research fields more and more attractive to a number of scientists.

Mathematically innovative papers like this one, revealing an unsuspected relationship between the Molecular Clock and the Entropy of b-lognormals in Evo-SETI Theory, should thus be welcome.

Acknowledgements

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1473550415000506.

References

Felsenstein, J. (2004). *Inferring Phylogenies*. Sinauer Associates Inc., Sunderland, Massachusetts.

Maccone, C. (2012). *'Mathematical SETI', a 724-Pages book jointly published Praxis (Chichester, UK) and Springer (Heidelberg and Berlin) in the Fall of 2012*. ISBN, ISBN-10: 3642274366 | ISBN-13: 978–3642274367 | Edition.

Maccone, C. (2013). SETI, evolution and human history merged into a mathematical model. *Int. J. Astrobiol.* **12**(3), 218–245.

Maccone, C. (2014). Evolution and mass extinctions as lognormal stochastic processes. *Int. J. Astrobiol.* **13**(4), 290–309.

Markov, A. & Korotayev, A. (2007). Phanerozoic marine biodiversity follows a hyperbolic trend. *Palaeoworld* **16**(4), 311–318.

Markov, A. & Korotayev, A. (2008). Hyperbolic growth of marine and continental biodiversity through the Phanerozoic and community evolution. *J. Gen. Biol.* **69**(3), 175–194.

Nei, M. (2013). *Mutation-Driven Evolution*. Oxford University Press, Oxford, UK.

Nei, M. & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Papoulis, A. & Pillai, S.U. (2002). *Probability, Random Variables and Stochastic Processes*, 4rth edn. Tata McGraw-Hill, New Delhi, ISBN 0-07-048658-1.