

# Bernoulli, Darwin, and Sagan: the probability of life on other planets

D. Kim Rossmo

Texas State University, 601 University Drive, San Marcos, Texas 78666, USA e-mail: [krossmo@txstate.edu](mailto:krossmo@txstate.edu)

**Abstract:** The recent discovery that billions of planets in the Milky Way Galaxy may be in circumstellar habitable zones has renewed speculation over the possibility of extraterrestrial life. The Drake equation is a probabilistic framework for estimating the number of technological advanced civilizations in our Galaxy; however, many of the equation's component probabilities are either unknown or have large error intervals. In this paper, a different method of examining this question is explored, one that replaces the various Drake factors with the single estimate for the probability of life existing on Earth. This relationship can be described by the binomial distribution if the presence of life on a given number of planets is equated to successes in a Bernoulli trial. The question of exoplanet life may then be reformulated as follows – given the probability of one or more independent successes for a given number of trials, what is the probability of two or more successes? Some of the implications of this approach for finding life on exoplanets are discussed.

Received 26 November 2015, accepted 14 March 2016, first published online 25 April 2016

**Key words:** astrobiology, Bernoulli trials, exoplanets, extraterrestrial life probability, probability analysis.

## Introduction

The recent discovery that most stars likely have companion planets has renewed speculation over the possibility of extraterrestrial life (Sagan 1980; Cassan *et al.* 2012). While most of these worlds will not be life-sustaining – we know of only one out of eight in the Solar System – others, such as Kepler-452b, appear much more promising (Jenkins *et al.* 2015). NASA estimates that 20 billion planets<sup>1</sup> in the Milky Way Galaxy are in circumstellar habitable zones (Bonfils *et al.* 2011).

The prevalence of habitable exoplanets and the probability of life emerging on one of those worlds are two key factors in the Drake equation (2015), the standard framework for estimating the number of technologically advanced civilizations in the Galaxy. The formula is a probabilistic argument composed of seven factors relevant to contact with alien life (see the Appendix). While billions of habitable planets significantly increase this possibility, many of the equation's other component probabilities, including the chances of life developing on a habitable planet, are either unknown or have large error intervals.

In this paper, we explore a different method of examining this question, one that replaces the various Drake factors with the single estimate for the probability of life existing on Earth. Some of the implications of this approach for finding life on exoplanets are discussed.

<sup>1</sup> Depending on the method used, estimation of the number of solar masses in our Galaxy ranges from 100 to 400 billion. Observations by NASA's Kepler Space Telescope suggest about 10% of these have potentially habitable Earth-sized planets (Overbye 2015).

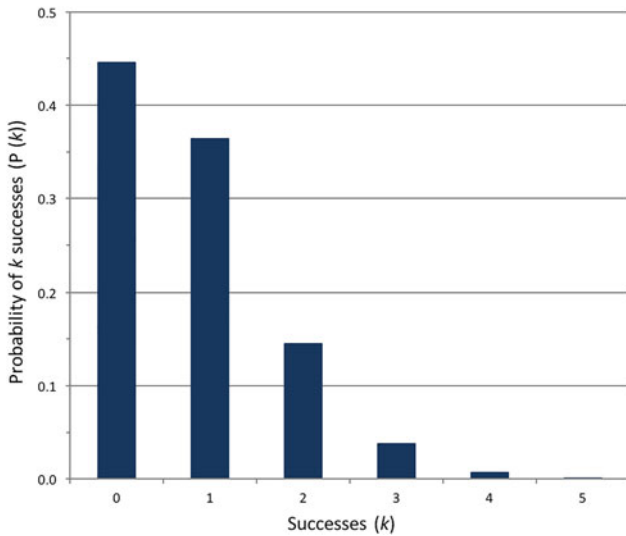
## How likely is it that life formed on one and only one planet?

The likelihood of extraterrestrial life can be redefined in terms of the probability of a known event, the existence of life on at least one planet (i.e. Earth). This relationship can be described by the binomial distribution if the presence of life on a habitable planet is treated as a success in a Bernoulli trial. Named for 17th-century Swiss mathematician Jacob Bernoulli, a Bernoulli trial has only two possible outcomes, such as heads or tails in a coin flip, which can be equated to success or failure. The sum of the probabilities of these two outcomes ( $p$  and  $q$ ) must equal 1. The binomial distribution is the discrete probability distribution of the number of successes ( $k$ ) in a sequence of independent Bernoulli trials. The distribution has two parameters: the probability of a success ( $p$ ); and the number of trials ( $n$ ). For example, Fig. 1 shows the probabilities for 0, 1, 2, 3, 4 and 5 successes for 40 trials given a success probability of 0.02.

If the presence of life on a habitable planet in the Milky Way Galaxy is treated as a Bernoulli trial success, then the binomial distribution can be used to determine the probability of  $k$  successes. While estimates exist for the number of habitable planets in the Galaxy ( $n$ ), the probability of a success ( $p$ ) is unknown<sup>2</sup>. However, we can use the binomial distribution to replace this parameter with the probability that life exists on at least one planet ( $\alpha$ )<sup>3</sup>. The question of exoplanet life can then be reformulated as follows – given the probability of one

<sup>2</sup> The value of  $p$  depends on the definition of 'habitable planet'; the more restrictive the definition (e.g. a requirement the planet has water), the higher the probability of life formation (Ward & Brownlee 2000).

<sup>3</sup> The weak anthropic principle warns that in order to avoid a selection bias, the probability of life emerging specifically on Earth, or any other given planet, cannot be equated to the probability of life forming on at



**Fig. 1.** Binomial distribution (with parameters  $p = 0.02$ ,  $n = 40$ ).

or more independent successes for a given number of trials, what is the probability of two or more successes?

The probability mass function for the binomial distribution is (Evans *et al.* 2000):

$$f(k; n, p) = P_L(k) = \binom{n}{k} p^k q^{n-k}$$

for  $k = 0, 1, 2, 3, \dots, n$  (number of habitable planets in the Galaxy), where  $p$  is the probability of success (life forming on a given habitable planet);  $q = 1 - p$ , the probability of failure (life not forming on a given habitable planet);  $n$ , the number of trials (total number of habitable planets);  $P_L(k)$ , the probability of  $k$  successes (life on  $k$  planets), for a given  $p$  and  $n$ ; and  $\binom{n}{k}$ , the binomial coefficient  $= \frac{n!}{k!(n-k)!}$ .

For  $k = 0$  (life exists on no planets), then  $\binom{n}{k} = 1$ , and

$$P_L(0) = q^n.$$

For  $k = 1$  (life exists on only one planet), then  $\binom{n}{k} = n$ , and:

$$P_L(1) = npq^{n-1}.$$

For  $k \geq 2$  (life exists on more than one planet):

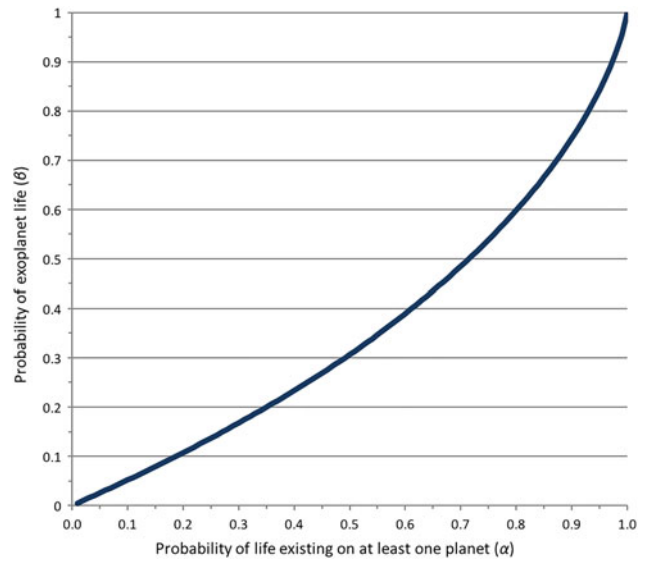
$$\begin{aligned} P_L(\geq 2) &= 1 - P_L(0) - P_L(1) = 1 - q^n - npq^{n-1} \\ &= 1 + (n-1)q^n - nq^{n-1}. \end{aligned}$$

The probability of life existing on at least one planet in the Galaxy<sup>4</sup> is therefore:

$$P_L(\geq 1) = 1 - q^n = \alpha.$$

least one planet in the Galaxy (Carter 2011);  $p$  and  $\alpha$  are therefore not equivalent.

<sup>4</sup> The fact that life exists on at least one planet in the Galaxy (i.e. Earth) does not mean the probability of such an outcome is 1, just as a single roll of a pair of dice resulting in a 7 does not mean the probability of such a result is 1 (it is actually 1/6).



**Fig. 2.** Probability of exoplanet life by probability of life existing on at least one planet (for  $n = 10\,000$  habitable planets).

Thus,

$$\begin{aligned} q &= (1 - \alpha)^{1/n}, \\ P_L(0) &= 1 - \alpha, \\ P_L(1) &= n[\alpha + (1 - \alpha)^{(n-1)/n} - 1], \\ P_L(\geq 2) &= \alpha - n[\alpha + (1 - \alpha)^{(n-1)/n} - 1]. \end{aligned}$$

As we know that life exists on Earth, it is necessary to calculate the conditional probability of life on more than one planet given that life exists on at least one planet. A conditional probability is defined as:

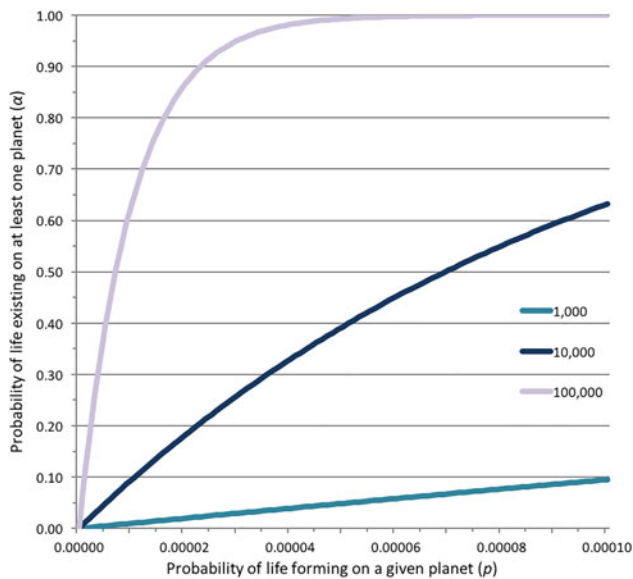
$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

We therefore need to adjust our estimate by dividing by  $1 - q^n$  (or  $\alpha$ ):

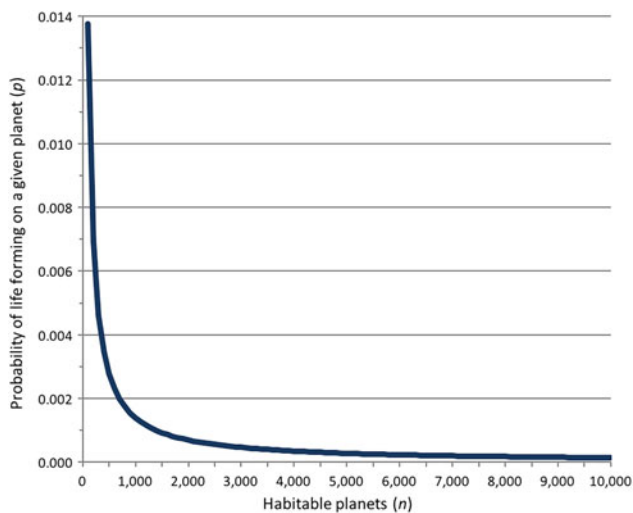
$$P_L(\geq 2|k \neq 0) = 1 - n + \frac{n}{\alpha}[1 - (1 - \alpha)^{(n-1)/n}] = \beta.$$

The above equation allows us to compare the probability of life on more than one planet ( $\beta$ ) to the probability of life on at least one planet ( $\alpha$ ). Figure 2 shows the relationship between these two probabilities for 10 000 habitable planets<sup>5</sup>. The ratio  $\beta/\alpha$  increases from 0.5 to 1 as  $\alpha$  grows from very low values to certainty (i.e. as  $q^n$  approaches 0). The relationship is nonlinear; for example,  $\beta/\alpha = 0.50$  for  $\alpha = 0.01$ , 0.61 for  $\alpha = 0.50$ , 0.72 for  $\alpha = 0.75$ , and 0.83 for  $\alpha = 0.90$ . While all we know for certain is that  $0 < \alpha < 1$ , a Bayesian argument can be made for  $\alpha$  falling between 0.5 and 1 given the existence of life on Earth (Eddy 2004). At the midpoint of this range ( $\alpha = 0.75$ ), for a conservative estimate of 10 000 habitable planets  $\beta = 0.54$ , meaning that exoplanet life is more probable than not.

<sup>5</sup> While this relationship becomes increasingly linear with large values of  $n$ , overall it depends little on the number of habitable planets.



**Fig. 3.** Probability of life existing on at least one planet by probability of life forming on a given planet (for  $n = 1000, 10\,000$  and  $100\,000$  habitable planets).



**Fig. 4.** Probability of life forming on a given planet by number of habitable planets (for  $\alpha = 0.75$ ).

The value of  $\beta$  is a function of  $\alpha$  and  $n$ , and  $\alpha$  in turn is a function of  $p$  and  $n$ . With enough planets, even very low life probabilities translate into a likelihood of life on more than one planet. Figure 3 shows the relationship between  $\alpha$  and  $p$  for 1000, 10 000 and 100 000 habitable planets. For 20 billion habitable planets, a  $p$  value as low as  $7 \times 10^{-11}$  still produces an  $\alpha$  value of 0.75 ( $\beta > 0.5$ ); for 1 billion habitable planets, the necessary  $p$  value is  $1.4 \times 10^{-9}$ . As the number of planets increases, the value of  $p$  necessary to produce a given value of  $\alpha$  decreases and vice versa (Fig. 4). Table 1 lists  $\beta$  values and the necessary  $p$  values and numbers of planets for various  $\alpha$  probabilities.

**Table 1.** Probability of exoplanet life

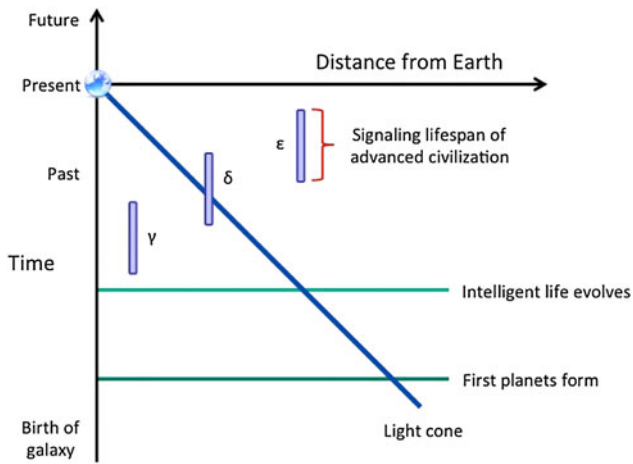
Probability of life existing on at least one planet ( $\alpha$ )	Number of habitable planets ( $n$ )	Probability of life forming on a given planet ( $p$ )	Probability of exoplanet life ( $\beta$ )
0.90	100	0.0227628	0.741189
	1000	0.0022999	0.743862
	10 000	0.0002302	0.744128
0.75	100 000	0.0000230	0.744154
	100	0.0137673	0.534684
	1000	0.0013853	0.537581
0.50	10 000	0.0001386	0.537870
	100 000	0.0000139	0.537899
	100	0.0069075	0.304445
0.10	1000	0.0006929	0.306613
	10 000	0.0000693	0.306829
	100 000	0.0000069	0.306850
0.01	100	0.0010531	0.051256
	1000	0.0001054	0.051705
	10 000	0.0000105	0.051750
0.001	100 000	0.0000011	0.051755
	100	0.0001005	0.004967
	1000	0.0000101	0.005012
0.0001	10 000	0.0000010	0.005016
	100 000	0.0000001	0.005017

## Evolution

Life falls along a continuum of advancement: microorganisms, multicellular organisms, animals, intelligent beings, societies capable of interstellar communication. As higher life must evolve from lower life, the probability of any type of life emerging on a planet is higher than the probability for intelligent life (Sterelny & Griffiths 1999). Therefore, the more advanced the life form, the smaller its  $p$  and  $\alpha$  values.

Life first emerged on Earth 3.4 billion years ago, 1.1 billion years after its formation; humans, however, appeared only in the past 200 000 years. Given sufficient time, is it inevitable that basic life forms will evolve into intelligent life? Once simple molecules develop the ability to replicate, natural selection suggests those that adapt well will replace those that do not (Darwin 1859). However, the outcome of that process is not inevitable. Evolutionary factors such as initial conditions, environment, and random mutations all influence outcomes. Initial conditions and environment dictate what traits are favoured and thus shape natural selection. Mutations are also important, but as a rule more so in small than large populations.

Intelligent life may not be inevitable, and might not even be probable (Ward & Brownlee 2000). Dennett (1995) discusses ‘good tricks’ in design space – solutions that natural selection has repeatedly found. Vision, for example, has evolved independently in multiple lineages, suggesting this is a good enough trick to justify the expectation that natural selection will most often realize it, at least where light is present. While there are several traits that fall into this category, advanced intelligence may not be one of them; it apparently failed to evolve in dinosaurs, the dominant terrestrial vertebrates for over 100 million years. Although intelligence has emerged in a number of taxa, if we define higher intelligence as something at the human level,



**Fig. 5.** Spatial-temporal communication range of extraterrestrial advanced civilizations.

the only evidence available to us is that it evolved in one species on one planet. Furthermore, the most intelligent life forms might not be the most adaptive for survival (Gould 1996). Consequently, complex life on other planets may not possess higher intelligence as we currently understand it.

### Communication limits

The Fermi paradox is the supposed contradiction between the high probability some have proposed for the existence of intelligent alien life and our lack of evidence of such life (Crawford 2000). However, we can only learn of advanced civilizations on exoplanets through communication, which is limited by the speed of light. The Milky Way Galaxy is large, roughly 100 000 light years in diameter. Even after a millennium, a transmission from Earth would only reach a small fraction of its stars; a response would then take another 1000 years to return.

Extraterrestrial civilizations much older than ours might exist, and signals from even very distant planets could have reached Earth if sent far enough in the past. As a star ages; however, its size and temperature change and so its habitable zone moves. Consequently, habitable planets – and any advanced civilizations they may host – have a limited lifespan. In some cases, a planet may not exist long enough for higher life with interstellar communication abilities to evolve.

A signal from an advanced alien civilization has to intersect Earth's light cone to reach us. This means: (1) intelligent life must have formed on an exoplanet and evolved to the point of having a method of interstellar communication; (2) their planet must be close enough to Earth for their messages to have had time to reach us; and (3) their civilization must not have died out too long in the past given their distance from Earth (if they did, their signals would have arrived before we developed the technology to recognize them).

Figure 5 (a two-dimensional picture of a four-dimensional configuration) illustrates these requirements. It depicts three planets with advanced civilizations, one that intersects

Earth's light cone ( $\delta$ ), and two that do not ( $\epsilon$  and  $\gamma$ ). Of the latter, one planet ( $\epsilon$ ) is so distant that we have not yet received its messages. The civilization of the other planet ( $\gamma$ ) disappeared long ago, and its messages, which reached Earth in our far past, were ignored.

The proportion of the Galaxy's history that has fallen within our planet's light cone since we developed a signal detection capability is extremely small. Moreover, the shorter the time a civilization engages in interstellar communication, the less likely its signals will be detected. How long such communication might last is unknown as the longevity of intelligent life, including our own, is uncertain. Unfortunately, the risks of self-destruction and environmental ruin appear to increase with growth and development (Diamond 2005; Doomsday Clock 2012).

Given these spatial and temporal limits, communication from an advanced civilization in our Galaxy may not be detected by us for a long time to come. Moreover, with a typical distance of 3 billion light years between galaxies, human life could well have disappeared before any signals from outside the Milky Way Galaxy ever have a chance to reach us.

### Conclusion

The Bernoulli trial method reframes the likelihood of alien life in terms of the probability that life emerged on at least one planet (i.e. Earth). The value assigned to  $\alpha$  can be considered a level of belief, a starting point for considering the question of whether other life forms exist in the Galaxy.

This approach assumes the binomial distribution is appropriate for modelling the probability of life emerging on a planet in our Galaxy and that Earth is not particularly unique (the mediocrity principle). The discussion further supposes limits imposed by the current laws of physics cannot be overcome, life requires a habitable (as currently defined) planet, and that extraterrestrial civilizations have not found traversable wormholes or engaged in interstellar migration. While these assumptions may not be all correct, they do reflect our current state of knowledge.

Even if the existence of intelligent life elsewhere in the Galaxy is likely, its detection in the near future is far from certain due to the distances involved. Of course, assuming it were possible to communicate with intelligent aliens, it may still not be wise to do so (Ambassador for Earth 2006). Our own world's history of what happens when an advanced civilization encounters a less developed one has not been particularly good.

### Acknowledgements

I would like to thank Dr Steven Le Comber, School of Biological and Chemical Sciences, Queen Mary University of London, Dr Lucia Summers, School of Criminal Justice, Texas State University, and the anonymous reviewer for their helpful comments.

## References

- Ambassador for Earth: Is it time for SETI to reach out to the stars? (2006). *Nature* **443**, 606.
- Bonfils, X. *et al.* (2011). The HARPS search for southern extra-solar planets XXXI. The M-dwarf sample. *Astron. Astrophys.* **526**, A112.
- Carter, B. (2011). Large number coincidences and the anthropic principle in cosmology. *Gen. Relat. Gravit.* **43**, 3225–3233.
- Cassan, A. *et al.* (2012). One or more bound planets per Milky Way star from microlensing observations. *Nature* **481**, 167–169.
- Crawford, I.A. (2000). Where are they? Maybe we are alone in the galaxy after all. *Sci. Am.* **283**(1), 38–43.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
- Dennett, D.C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster, New York.
- Diamond, J.M. (2005). *Collapse: How Societies Choose to Fail or Succeed*. Viking, New York.
- Doomsday Clock (2012). Bulletin of the Atomic Scientists. <http://www.thebulletin.org/content/doomsday-clock/timeline>.
- Eddy, S.R. (2004). What is Bayesian statistics? *Nature Biotechnology* **22**, 1177–1178.
- Evans, M., Hastings, N. & Peacock, B. (2000). *Statistical Distributions*. 3rd edn. John Wiley & Sons, New York.
- Gould, S.J. (1996). *Full House: The Spread of Excellence from Plato to Darwin*. Harmony Books, New York.
- Jenkins, J.M. *et al.* (2015). Discovery and validation of Kepler-452b: a 1.6-R<sub>⊕</sub> super Earth exoplanet in the habitable zone of a G2 star. *Astron. J.* **150**, 56.
- Overbye, D. (2015). Astronomers have found an analogue for the Earth. The New York Times 24 July, p. A20.
- Sagan, C. (1980). *Cosmos*. Random House, New York.
- Sterelny, K. & Griffiths, P.E. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. University of Chicago Press, Chicago.
- The Drake equation (2015). SETI Institute. <http://www.seti.org/drakeequation>.
- Ward, P.D. & Brownlee, D. (2000). *Rare Earth: Why Complex Life is Uncommon in the Universe*. Copernicus Books, New York.

## Appendix

The Drake equation is:

$$N = R^* f_p n_e f_i f_c L,$$

where  $N$ , is the number of detectible extraterrestrial civilizations in the Milky Way Galaxy;  $R^*$ , the mean rate of star formation per year in the Galaxy;  $f_p$ , the fraction of the above that have planets;  $n_e$ , the mean number of the above that can support life;  $f_i$ , the fraction of the above that go on to develop life;  $f_c$ , the fraction of the above that develop intelligent life;  $f_c$ , the fraction of the above that send detectable signs of their existence into space;  $L$ , the length of time for which the above are sent into space.

The value of  $p$  for the probability of any type of life forming on a given habitable planet is equal to  $f_i$ , while the value of  $p$  for the probability of intelligent life forming on a given habitable planet is equal to  $f_i f_c$ .