On the Fermi paradox, playing with unknown probabilities, and statistics of phylogenetic trees

David Aldous

18 March 2024
Woodinville High School Math Club

Time permitting, I will talk about 3 different topics.

- Humphrey Huang suggested a talk on the **Fermi paradox** – a topic that I find very interesting as "organized logic", but with little actual mathematics.

The other two are topics you could actually participate in.

- Statistical properties of phylogenetic trees (you could help with my research).
- Assessing probabilities of future real-world events (a game you yourself can play).

A very natural question is

(*) Is there extraterrestrial life anywhere in the universe?

One can look for signs of life, which has been done under the name SETI (Search for Extraterrestrial Intelligence). Wikipedia has a good article on SETI. The bottom line: no signs have been found yet ......

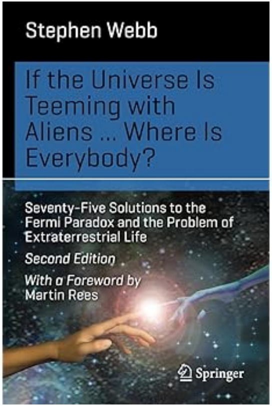
Note the shift from **life** to **intelligence**.

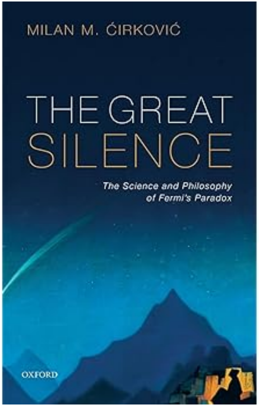
How would we know if there is very primitive life somewhere far away? In the 1950s it was realized that we could detect radio/TV signals from hypothetical aliens at our tech level, or we might observe SF-style macro-engineering such as Dyson spheres by far advanced aliens. It was then thought we would be unable to detect very primitive life somewhere far away. Nowadays we might in principle infer presence of primitive life, if we observed an exoplanet with oxygen in its atmosphere.

Anyway, let me formulate the Fermi paradox as follows.

> **The Universe is very big and very old; given there is a human technological civilization on Earth, why don't we see evidence of technologically advanced extraterrestrial civilizations?**

This formalization allows us to devise and organize possible explanations. Here are two books on the Fermi paradox, one "popular science" and the other "professional science". The first suggests 75 possibilities.


Stephen Webb
If the Universe Is Teeming with Aliens ... Where Is Everybody?
Seventy-Five Solutions to the Fermi Paradox and the Problem of Extraterrestrial Life
Second Edition
With a Foreword by Martin Rees
Springer


MILAN M. ĆIRKOVIĆ
THE GREAT SILENCE
The Science and Philosophy of Fermi's Paradox
OXFORD

Two strategies for thinking about this.

- What do we know about how humans came to exist – how likely was that? – based on known science.
- Thinking outside the “science” box – *more things in heaven and Earth than are dreamt of in our (science) philosophy*.

Let's consider just our own galaxy. We know

- (roughly) number of stars
- (more roughly) number of planets
- (even more roughly) number of Earth-size planets in liquid-water zone.

Because we don't know how life on Earth got started, and we don't know what those exoplanets are like, the first thing we really **don't** know is

$N_{suitable}$ = number of planets in galaxy which are potentially (at formation) suitable for life.

But suppose we have such a planet. We can consider successive probabilities, such as

- (p_1) Simple life (e.g. single cell)
- (p_2) Complex life (e.g. simple plant)
- (p_3) Animal-level intelligence (e.g. dinosaurs)
- (p_4) Human-level intelligence
- (p_5) 21st century human level technology
- (p_6) "A sufficiently advanced technology that is indistinguishable from magic".

The point is that (one version of the Drake equation)

$$N_{tech} = N_{suitable} \times p_1 \times p_2 \times p_3 \times p_4 \times p_5$$

gives the expected number of **tech civs** that have arisen sometime in the galaxy.

We have no idea what these probabilities are (people have made lots of silly guesses – ignore them) so it may seem we have gotten nowhere. But it does lead to two alternative possibilities.

$$N_{tech} = N_{suitable} \times p_1 \times p_2 \times p_3 \times p_4 \times p_5$$

It would be a weird coincidence if $N_{tech} \approx 1$, so there are two alternate possibilities.

- $N_{tech} \ll 1$ (our galaxy won the lottery; only Earth has a tech civ in this galaxy)
- $N_{tech} \gg 1$ (A whole bunch of tech civs have arisen somewhere in the galaxy; either disappeared without trace or still present but undetectable to us).

This is the best starting point for discussing the Fermi paradox.

The possibility $N_{tech} \ll 1$ could only be rejected if we had some positive evidence of extraterrestrial life, or if we had some very detailed understanding of the route from formation of Earth to human tech civ so that we could confidently assess probabilities of some steps.

The other possibility is

- $N_{tech} \gg 1$ (A whole bunch of tech civs have arisen somewhere in the galaxy; either disappeared without trace or still present but undetectable to us).

This is more fun to consider, because it relates to science fiction (SF) – what might non-human tech civs be like?

The **space opera** style of SF envisages aliens at a similar tech level, which would be a very implausible coincidence, and envisages FTL travel, which is outside the known laws of physics. One can speculate what aliens might do within known physics, though this involves imagining the motivations of aliens. Some ideas:

(Wikipedia) **von Neumann probes**. Von Neumann argued that the most effective way of performing large-scale mining operations such as mining an entire moon or asteroid belt would be by self-replicating spacecraft, taking advantage of their exponential growth.

It has been theorized that a self-replicating starship utilizing relatively conventional theoretical methods of interstellar travel and speeds limited to an "average cruising speed" of 0.1c., could spread throughout a galaxy the size of the Milky Way in as little as half a million years.

The point is that **it would need only one civ to do this**; but we don't see any evidence in our solar system. So this argues against the possibility $N_{tech} \gg 1$.

(Wikipedia) The **dark forest** hypothesis is the conjecture that many alien civilizations exist throughout the universe, but they are both silent and hostile, maintaining their undetectability for fear of being destroyed by another hostile and undetected civilization. In this framing, it is presumed that any space-faring civilization would view any other intelligent life as an inevitable threat, and thus destroy any nascent life that makes itself known. As a result, the electromagnetic spectrum would be relatively quiet, without evidence of any intelligent alien life, as in a "dark forest" filled with "armed hunter(s) stalking through the trees like ghosts".

The general point is that, to survive a million years, a civ would need to be extremely risk-averse. So it would need to take precautions against all, however unlikely-seeming, possibilities. Civs that don't take precautions would fail in one of many possible ways, and not leave any traces.

Another possibility is

$N_{tech} \gg 1$ **but tech civs do not survive long enough to make their mark.**

This is a common SF theme – imagine we find evidence of lost alien civs.

- nuclear war or ecological collapse
- conflict with another alien civ
- "sublime" to some state of existence outside the universe observable to us
- ......

The Great Filter

This is a very speculative line of thought, due to Robin Hanson. Consider the product

$$Npq \tag{1}$$

where

- N is the number of Earth-like (loosely, and at formation) planets in the galaxy
- p is the chance that, on such a planet, an intelligent species at a technological level comparable to ours will arise at some time
- q is the chance that such a species would survive in such a way as to be observable (via communication or exploration) to other galactic species for an appreciable length of time.

The point is that Npq indicates (after some more time scaling) the number of *other* intelligent species we expect to observe in the galaxy. Because we don't observe any, we conclude *prima facie* (treating absence of evidence as evidence of absence) that it cannot be true that $Npq \gg 1$. Since it would be a bizarre coincidence if $Npq \approx 1$, we should conclude (*according to this argument*) that $Npq \ll 1$ and so humans are most likely to be the only technological species in the galaxy.

A paradoxical argument: we should hope to not find other life in our solar system.

Human beings did not create the Universe or direct the course of evolution, so N and p are not our responsibility. But q, as applied to us (i.e. will our species leave its mark on the galaxy?) is presumably under our control. Viewing q very roughly as the chance that a hypothetical technological species arising across the galaxy 25 million years in the future would then be able to observe the then-current or previous existence of humans, being told that $q = 10^{-6}$ would be rather depressing. Depressing, because of the ways this might come about, for instance if humans soon become extinct, or change and cease to interact with the macroscopic physical universe. Knowing $q = 10^{-6}$ would be knowing that something like this is *almost certain* to happen.

Now having decided that Npq is small, implying pq is very small, the only way to avoid the depressing possibility of q being very small is to for p to be very small.

The main conclusion of this argument is that the easier it was for life to evolve to the present stage, the bleaker the future chances of humanity probably are.

- Thinking outside the “science” box – *more things in heaven and Earth than are dreamt of in our (science) philosophy*.

What if reality is different from what we think?

.

This is fun (or silly?) to imagine ...

1. (Wikipedia) **Solipsism** is the philosophical idea that only one's mind is sure to exist. ......solipsism holds that knowledge of anything outside one's own mind is unsure; the external world and other minds cannot be known and might not exist outside of one's own mind.

2. A supernatural creator of the universe, implictly with some special interest in humans. It would be their choice whether to create aliens also.

3. Universe created yesterday with false memories of previous days.

4. Are you sleeping and dreaming right now? One can do a simplistic Bayes analysis by saying

- you are awake for 16 hours, asleep and dreaming for 1 hour per day.
- therefore chance $1/17$ that you are dreaming now.

Most people reject that conclusion. But some believe

5. (Wikipedia) The **simulation hypothesis** proposes that what humans experience as the world is actually a simulated reality, such as a computer simulation in which humans themselves are constructs.

The argument (ignoring any invented numbers) rests on comparing:
(i) A hypothetical universe in which aliens simulate an arbitrarily large number ("many") of virtual universes
(ii) our one apparent real universe
and then argues:
many is more than one, so the former is more likely.

To me, this is ridiculous. It's like the awake/asleep argument, which is not convincing, but now with made-up numbers.

Once one envisages hypothetical universes one can replace (ii) by (iii) an arbitrarily large number ("many") of hypothetical universes in which we would be real, not simulated, inhabitants. And we now have no way to compare many with many.

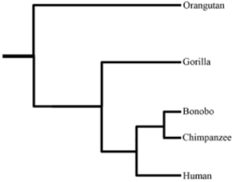
Other thoughts

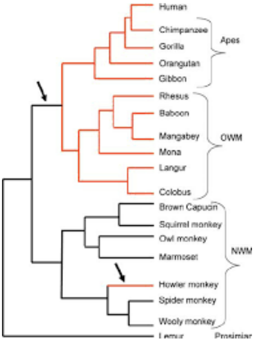
Maybe our form of **self-conscious** intelligence is rare, or we don't understand what intelligence is?

If cats had human-level intelligence would they build starships?

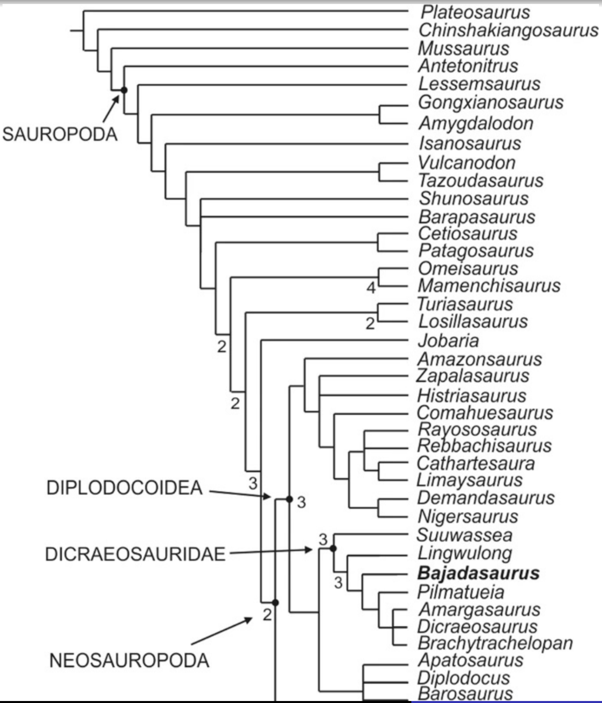
What if the immune system (rather than the brain) had became self-aware?

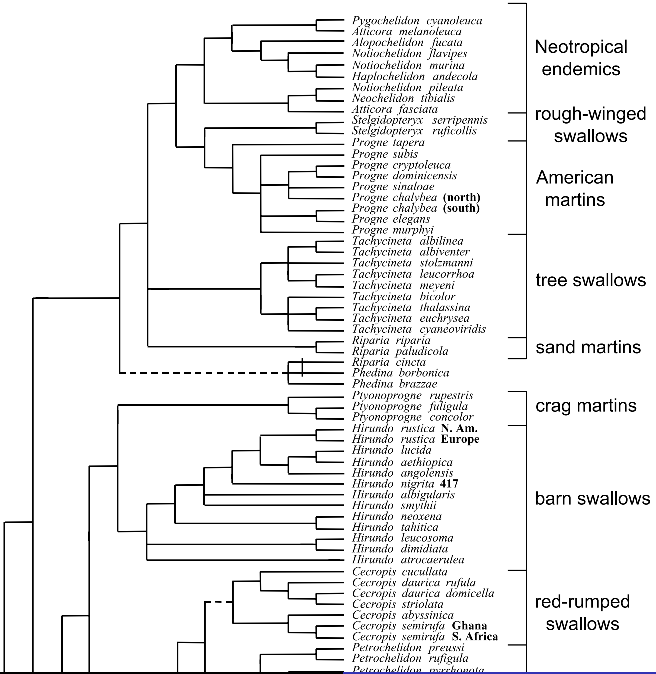
Topic 2: A little project – statistics of phylogenetic trees

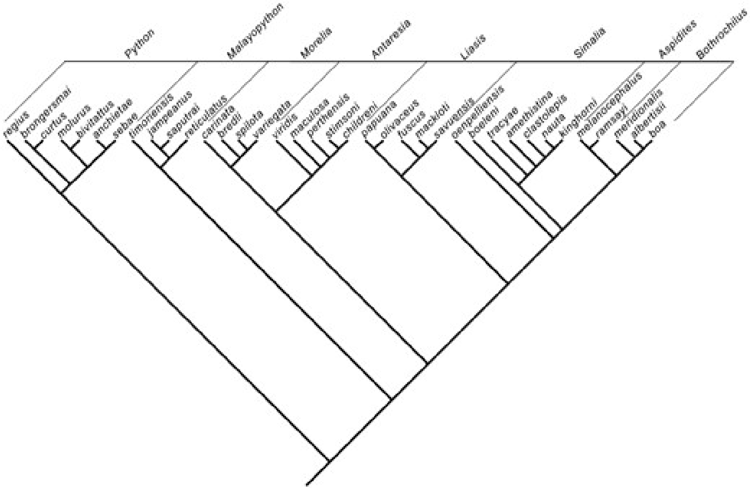




Easy to find many examples via Google Images.


Plateosaurus
Chinshakiangosaurus
Mussaurus
Antetonitrus
Lessemsaurus
Gongxianosaurus
Amygdalodon
Isanosaurus
Vulcanodon
Tazoudasaurus
Shunosaurus
Barapasaurus
Cetiosaurus
Patagosaurus
Omeisaurus
Mamenchisaurus
Turiasaurus
Losillasaurus
Jobaria
Amazonsaurus
Zapalasaurus
Histriasaurus
Comahuesaurus
Rayososaurus
Rebbachisaurus
Cathartesaura
Limaysaurus
Demandasaurus
Nigersaurus
Suuwassea
Lingwulong
Bajadasaurus
Pilmatueia
Amargasaurus
Dicraeosaurus
Brachytrachelopan
Apatosaurus
Diplodocus
Barosaurus
SAUROPODA
DIPLODOCOIDEA
DICRAEOSAURIDAE
NEOSAUROPODA
4
2
2
2
3
3
3
3
2


Pygochelidon cyanoleuca
Atticora melanoleuca
Alopochelidon fucata
Notiochelidon flavipes
Notiochelidon murina
Haplochelidon andecola
Notiochelidon pileata
Neochelidon tibialis
Atticora fasciata
Stelgidopteryx serripennis
Stelgidopteryx ruficollis
Progne tapera
Progne subis
Progne cryptoleuca
Progne dominicensis
Progne sinaloae
Progne chalybea (north)
Progne chalybea (south)
Progne elegans
Progne murphyi
Tachycineta albilinea
Tachycineta albiventer
Tachycineta stolzmanni
Tachycineta leucorrhoa
Tachycineta meyeni
Tachycineta bicolor
Tachycineta thalassina
Tachycineta euchrysea
Tachycineta cyaneoviridis
Riparia riparia
Riparia paludicola
Riparia cincta
Phedina borbonica
Phedina brazzae
Ptyonoprogne rupestris
Ptyonoprogne fuligula
Ptyonoprogne concolor
Hirundo rustica N. Am.
Hirundo rustica Europe
Hirundo lucida
Hirundo aethiopica
Hirundo angolensis
Hirundo nigrita 417
Hirundo albigularis
Hirundo smythii
Hirundo neoxena
Hirundo tahitica
Hirundo leucosoma
Hirundo dimidiata
Hirundo atrocaerulea
Cecropis cucullata
Cecropis daurica rufula
Cecropis daurica domicella
Cecropis striolata
Cecropis abyssinica
Cecropis semirufa Ghana
Cecropis semirufa S. Africa
Petrochelidon preussi
Petrochelidon rufigula
Neotropical endemics
rough-winged swallows
American martins
tree swallows
sand martins
crag martins
barn swallows
red-rumped swallows


Python
Malayopython
Morelia
Antaresia
Liasis
Simalia
Aspidites
Bothrochilus
regius
brongersmai
curtus
molurus
bivittatus
anchietae
sebae
timoriensis
ampeanus
sapufrai
reticulatus
carinata
bredli
spilota
variegata
viridis
maculosa
perthensis
stimsoni
childreni
papuana
olivaceus
fuscus
mackloti
savuensis
oenpelliensis
boeleni
fracyae
amethistina
clastolepis
nauta
kinghorni
melanocephalus
ramsayi
meridionalis
albertisii
boa

When we hear the word “data” we think of numbers, but data can also be other math structures like these trees.

These trees are all different, but one can look at statistics of trees. One thing that's noticeable is that they have **uneven splits** and no-one really knows why. That is, the mathematical models one might use do not reproduce the observed property

when the lineage of a large clade, with say m species, splits, the median size of the smaller sub-clade is roughly $m^{1/2}$.

I invented a model which does have that property. The model makes predictions about frequencies of small fringe trees.

Research Project: Look at many examples and get data on these frequencies.

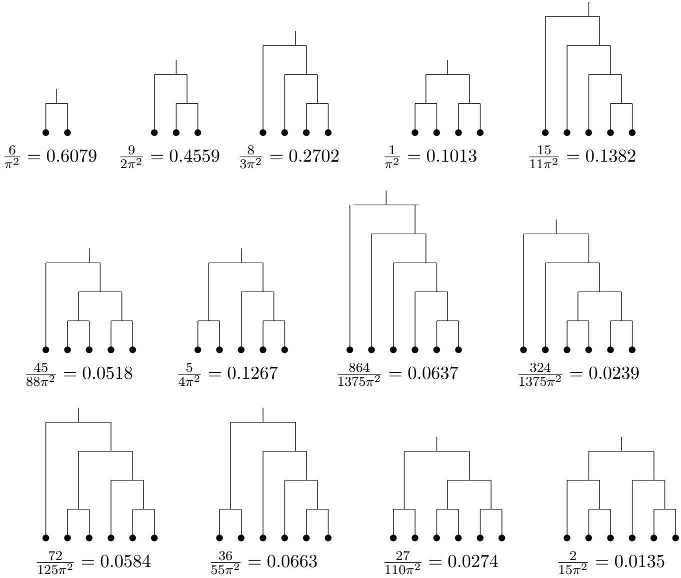


Figure 5: Proportions of leaves in clades of a given shape, for each shape with $2-6$ leaves in the fringe process.

Topic 3: Assessing probabilities of future real-world events,

In a **prediction tournament**, contestants state **probabilities** (not yes/no) of future geopolitical events. Here are some questions asked currently on `gjopen.com`.

• Before 13 June 2024, will the Federal Reserve announce a cut in the US federal funds rate?

• Will there be a lethal confrontation between the national military forces and/or law enforcement of Russia and those of a NATO member state resulting in at least three fatalities before 1 September 2024?

• Will China, Russia, and/or the US detonate a nuclear device in their respective territories before 1 October 2024?

• Will NASA's Europa Clipper mission be launched successfully before 2 November 2024?

[Make prediction on `gjopen`]

Different from ("all your own work") school exam . More like a civil jury – assess all the evidence and opinions that you can find.

How can we assess someone's ability? We do what Gauss said 200 years ago – use **mean square error** MSE. An event is a 0 - 1 variable; if we predict 70% probability then our score on that question is the "squared error":
(if event happens) $(1.0 - 0.70)^2 = 0.09$
(if event doesn't happen) $(0.0 - 0.70)^2 = 0.49$

Your **tournament score** is the sum of scores on each question. As in **golf** one seeks a **low** score. Also as in golf, in a *tournament* all contestants address the same questions; it is not a single-elimination tournament as in tennis.

Suppose there are n events. Write S for your "tournament score" when the (unknown) true probabilities of the n events are $(p_i, 1 \leq i \leq n)$ and you predict $(q_i, 1 \leq i \leq n)$.

- Your **actual** score depends on the (q_i) and the event outcomes.
- Your **expected** score depends on the (q_i) and the (p_i).

A short calculation shows

$$\mathbb{E}S = \sum_i p_i(1 - p_i) + n\sigma^2 \quad \text{where} \tag{2}$$

$$\sigma^2 := n^{-1} \sum_i (q_i - p_i)^2$$

σ^2 is your MSE (mean squared error) in **assessing the probabilities**. So for contestants A and B

$$n^{-1}\mathbb{E}(S_A - S_B) = \sigma_A^2 - \sigma_B^2$$

and so **in the long run** we can tell who is the more accurate forecaster **without knowing true probabilities**.

Extensive data, e,g. from IARPA-sponsored prediction tournaments over 2013-2017, shows that some individuals consistently get better scores than others. The natural interpretation is that some individuals are better than others at assessing true probabilities.

This setting relates to a philosophical question: we can understand what probability means in the context of repeatable events like coin-tossing or dice-throwing, but what about these unique future events?

People have argued about this for centuries. Our analysis is based on the assumption that events have unknown "true" probabilities; it is hard to explain the IARPA data under any other philosophical viewpoint.