A Bayesian Analysis of the Astrobiological Implications of the Rapid Emergence of Life on the Early Earth*

Ed Turner
Princeton University, Department of Astrophysical Sciences
and
University of Tokyo, IPMU

Harvard University, Statistics Department – November 15, 2011

*Collaboration with David Spiegel (Princeton University Observatory)
(+ thanks to former Princeton undergraduate Carl Boettiger’07)
Preface

• Field: Astrobiology
• Content: Mostly statistics with small amounts of biology & geology but *almost no astronomy*
• Motivation 1: Clarify the implications of the early appearance of life on Earth for *biology elsewhere*
• Motivation 2: A worked example of a Bayesian approach to a hand-waving, intuitive, ambiguous, judgment-call, *small-number-statistics* problem with some parallels to *anthropic arguments*
Fundamental Questions of Astrobiology

• Does extraterrestrial life exist?
• What is its nature?
• How common is it?

The overwhelmingly most practical and promising approaches to these questions are empirical, searches for life on bodies in the Solar System and beyond. But that is not our topic today.
A Possible-in-Principle Calculation

Given

- $P_{\text{abiogenesis}}(\text{physical conditions})$ per unit volume and per unit time, and
- prevalence of such suitable abiotic environments in the Universe

a straightforward statistical estimate could be made.

But, alas, we are quite ignorant of both and have no immediate prospects of remedying either situation.
A Potential Finesse

• Exoplanet studies strongly suggest, but have not yet proven, that planets resembling the Earth in a very general/crude way (mass, primary star, orbit, composition etc) are reasonably common.

• Life arose very quickly on the early Earth.

• This suggests that $P_{\text{abiogenesis}}$ ($\text{early-Earth-like conditions}$) per unit planet per unit age of the Earth is not extremely small.

• Thus, simple/primitive extraterrestrial life (at least) is not too rare.  ➔ roughly the current “consensus” view
How quickly is very quickly?

• Earth formed at 4.54 Ga (Ga = $10^9$ yrs ago, like redshift)
• Water was reasonably abundant by ~4.3 Ga
• Likely sterilization by the LHB and subsequent global volcanic resurfacing around 3.85-4.0 Ga
• Highly controversial isotopic indicators ($^{13}$C depletion) of metabolic activity in the oldest known surviving rocks, 3.7-3.82+ Ga
• Wide-spread “probable fossils” of microbes & macroscopic microbial biofilms (stromatolites) in sediments formed ~3.5 Ga
• Definite, highly evolved macro-fossils ~3.2 Ga
Stromatolites
(Van Zuilen 2006)
Abiogenesis almost certainly occurred on Earth significantly earlier than the time of the most ancient life detected to date.

Thus, “very quickly” must mean within a few hundred million years but it could have been much faster.
How small is extremely small? (1)

- The basic molecular chemistry of all terrestrial life is essentially identical and very complex.
- Two classes of macro-molecules, proteins and nucleic acids, play central roles currently, but it is imagined that RNA alone might have sufficed for an earlier form of life (the “RNA world”).
- The building blocks of both proteins (amino acids) and nucleic acids (nucleotides) are a set of small molecules which form bonds in an arbitrary order to create the long polymers which are these macro-molecules.
How small is *extremely small*? (2)

- Both amino acids and nucleotides are produced in reasonable abundance by physical (abiotic) chemical reactions that plausibly (empirically in some cases) commonly occur in nature (GOOD!).

- However, the polymers in question are huge, typically 100s to 100s of millions of “building blocks” long, with even minimally functional ones requiring ~100 nucleotides or amino acids.

- The probability of any specific one arising randomly is then *factorially small*, *i.e.*, $\approx 10^{-100s}$ or less (BAD!).

- Thus, “*extremely small*” $\Rightarrow$ no extraterrestrial life
Abiogenesis: Stochastic Lego Construction?

“the time from soup to bugs may have been far less than 10 million years. Making life may be a rapid operation – a key observation supporting our contention that life may be very common in the Universe.”
Irrational Exuberance

“Personally, given the ubiquity and propensity of life to flourish wherever it can, I would say that, my own personal feeling is that the chances of life on this planet are 100%. I have almost no doubt about it.”

- Steven Vogt speaking to L.-J. Zgorski (NSF) in regard to Gl 581g, a reported exoplanet with $M_{\text{sin}}(i) \approx 3 M_\odot$ orbiting within its primary’s HZ
Sensible (but not exactly right)

“It is too glib to claim that if the origin of life took place on Earth immediately after the end of the heavy bombardment, then the origin of life must be ‘easy,’ so that the prospects for life elsewhere are increased”

Does the Rapid Appearance of Life on Earth Suggest that Life Is Common in the Universe?

CHARLES H. LINEWEAVER\textsuperscript{1,2} and TAMARA M. DAVIS\textsuperscript{1}

ABSTRACT

It is sometimes assumed that the rapidity of biogenesis on Earth suggests that life is common in the Universe. Here we critically examine the assumptions inherent in this if-life-evolved-rapidly-life-must-be-common argument. We use the observational constraints on the rapidity of biogenesis on Earth to infer the probability of biogenesis on terrestrial planets with the same unknown probability of biogenesis as the Earth. We find that on such planets, older than \textasciitilde 1 Gyr, the probability of biogenesis is $>13\%$ at the 95\% confidence level. This quantifies an important term in the Drake Equation but does not necessarily mean that life is common in the Universe. Key Words: Biogenesis—Drake Equation. Astrobiology 2, 293–304.

N.B. – A formal \textit{frequentist} analysis with a misleading result
Start from the Bayesian’s $F=ma$

\[ P[A|B] = \frac{P[B|A] \times P_{\text{prior}}[A]}{P[B]} \, . \]

A = the model
B = the data
P[A|B] = probability of the model given the data = the posterior
(what we want!)
P[B|A] = probability of the data given the model = the likelihood
(what we can obtain from a “forward” calculation)
P_{\text{prior}}[A] = a priori probability the model is true/correct = the prior
(often requires “arbitrary” judgment calls, source of controversy)
P[B] = probability of the data
(typically unknown, but only needed for normalization)
A Uniform Rate (Poisson) Model

\( \lambda = \text{rate of abiogenesis per Gyr per Earth-like planet} \)

\( t = \text{age of the Earth (like expansion factor)} \)

\( \tau_{\text{sterile}} = \text{earliest time at which life could develop} \)

\( \tau_{\text{max}} = \text{latest time at which life could develop} \)

\( n = \text{number of independent abiogenesis events} \)

\[
P[\lambda, n, t] = P_{\text{Poisson}}[\lambda, n, t] = e^{-\lambda(t-\tau_{\text{sterile})}} \frac{\{\lambda(t - \tau_{\text{sterile}})\}^n}{n!}
\]
Two “Anthropic” Constraints

• Life must have arisen on Earth ≥ one times or we would not be here to carry out the calculation

\[ P_{\text{life}} = 1 - P_{\text{Poisson}}[\lambda, 0, t] = 1 - e^{-\lambda(t - \tau_{\text{sterile}})} \]

• Life must have arisen early enough on Earth to allow us to evolve by the present

\[ t_{\text{emerge}} < t_0 - \tau_{\text{evolve}} \]

\[ t_{\text{emerge}} = \text{Earth’s age at abiogenesis}, \quad t_0 = \text{its present age} \quad \& \quad \tau_{\text{evolve}} = \text{time for evolution of astrobiologists} \]
The Likelihood Term

\[ P[B|A] = \frac{1 - \exp[-\lambda(t_{\text{emerge}} - \tau_{\text{sterile}})]}{1 - \exp[-\lambda(\tau_{\text{req}} - \tau_{\text{sterile}})]} \]

where \( \tau_{\text{req}} \equiv \min[t_0 - \tau_{\text{evolve}}, \tau_{\text{max}}] \)

In the limit of extremely small \( \lambda \)

\[ P[B|A] \approx \frac{t_{\text{emerge}} - \tau_{\text{sterile}}}{\tau_{\text{req}} - \tau_{\text{sterile}}} \]

In the limit of extremely large \( \lambda \)

\[ P[B|A] \approx 1 \]
The Bayes Factor/Ratio/Evidence for Model Selection

\[ R \equiv \frac{\text{P}[\text{data}|\text{large } \lambda]}{\text{P}[\text{data}|\text{small } \lambda]} \approx \frac{\Delta t_2}{\Delta t_1}. \]

where

\[ \Delta t_1 \equiv t_{\text{emerge}} - \tau_{\text{sterile}} \quad \text{and} \quad \Delta t_2 \equiv \tau_{\text{req}} - \tau_{\text{sterile}}. \]

Rule of thumb: \( R < 10 \) \( \Rightarrow \) barely worth mentioning
The Prior Factor – Problematic!

\[ P_{\text{prior}}[A] = f_{\lambda}[\lambda] f_{\tau_s}[\tau_{\text{sterile}}] f_{\tau_m}[\tau_{\text{max}}] f_{\tau_e}[\tau_{\text{evolve}}] \]

- **Uniform:**
  - Informative=BAD

- **Inverse Uniform:**
  - Informative=BAD

- **Log Uniform:**
  - Uniformative=GOOD
  - but improper=not so good

\[ \lambda_{\text{max}} = 1000 \text{ Gyr}^{-1} \]

\[ \lambda_{\text{min}}: \ 10^{-22} \text{ Gyr}^{-1}, \ 10^{-11} \text{ Gyr}^{-1}, \ \text{and} \ 10^{-3} \text{ Gyr}^{-1} \]
Use δ-function estimates for the $f_\tau$ terms

Table 1: Models of $t_0 = 4.5$ Gyr-Old Planets

<table>
<thead>
<tr>
<th>Model</th>
<th>$\tau_{\text{sterile}}$ (Gyr)</th>
<th>$t_{\text{emerge}}$ (Gyr)</th>
<th>$\tau_{\text{max}}$ (Gyr)</th>
<th>$\tau_{\text{evolve}}$ (Gyr)</th>
<th>$\tau_{\text{req}}$ (Gyr)</th>
<th>$\Delta t_1$ (Gyr)</th>
<th>$\Delta t_2$ (Gyr)</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical</td>
<td>0.5</td>
<td>0.51</td>
<td>10</td>
<td>1</td>
<td>3.5</td>
<td>0.01</td>
<td>3.00</td>
<td>300</td>
</tr>
<tr>
<td>Conservative$_1$</td>
<td>0.5</td>
<td>1.3</td>
<td>1.4</td>
<td>2</td>
<td>1.4</td>
<td>0.80</td>
<td>0.90</td>
<td>1.1</td>
</tr>
<tr>
<td>Conservative$_2$</td>
<td>0.5</td>
<td>1.3</td>
<td>10</td>
<td>3.1</td>
<td>1.4</td>
<td>0.80</td>
<td>0.90</td>
<td>1.1</td>
</tr>
<tr>
<td>Optimistic</td>
<td>0.5</td>
<td>0.7</td>
<td>10</td>
<td>1</td>
<td>2.5</td>
<td>0.20</td>
<td>3.00</td>
<td>15</td>
</tr>
</tbody>
</table>

Hypothetical = extremely quick terrestrial abiogenesis (speculative)
Conservative = slow terrestrial abiogenesis consistent with the data
Optimistic = quick terrestrial abiogenesis suggested by the data
Turn the Bayes crank ➔
posterior distributions
Optimistic

Probability Density

Uniform
Log (-3)
InvUnif (-3)

Posterior: Solid
Prior: Dashed

$\log_{10}[\lambda]$ ($\lambda$ in Gyr$^{-1}$)
Cumulative Probability

Optimistic

Uniform
Log (-3)
InvUnif (-3)

Posterior: Solid
Prior: Dashed

\( \log_{10} \lambda \) (\( \lambda \) in Gyr\(^{-1} \))
Cumulative Probability

Optimistic

Posterior: Solid
Prior: Dashed

Uniform
Log (-3)
Log (-11)
Log (-22)
InvUnif (-3)
InvUnif (-11)
InvUnif (-22)
Bounds on $\log_{10}[h]$ ($h$ in Gyr$^{-1}$)

- Median Value of $\lambda$
- 1-$\sigma$ Lower Bound
- 2-$\sigma$ Lower Bound

Optimistic ($\lambda_{\text{max}} = 10^3$ Gyr$^{-1}$)
Cumulative Probability

- Uniform
- Log (-3)
- Log (-11)
- Log (-22)
- InvUnif (-3)
- InvUnif (-11)
- InvUnif (-22)

Posterior: Solid
Prior: Dashed

Hypothetical

\[ \log_{10} [\lambda] \quad (\lambda \text{ in Gyr}^{-1}) \]
Conservative

- Uniform
- Log (-3)
- Log (-11)
- Log (-22)
- InvUnif (-3)
- InvUnif (-11)
- InvUnif (-22)

Posterior: Solid
Prior: Dashed
Summary

• Paleobiological & geological evidence favors a very fast development of life on the early Earth.
• If so, this datum implies a best estimate value of $\lambda$ very roughly of order $\sim1$ Gyr$^{-1}$ or greater.
• However, even extremely small values of $\lambda$, that nearly exclude life elsewhere in the observable universe, are consistent with the datum.
• Stronger intuitive or formal conclusions about $\lambda$ are the consequence of an informative prior and not an implication of the datum.
Conclusions

A Bayesian fan of extraterrestrial life should be encouraged by, but not highly confident based on, the rapid emergence of life on the early Earth.

A straightforward, but careful, Bayesian analysis can yield transparent and well founded results even in situations involving (very) small number statistics, anthropic considerations, intuitive arguments, uncertain data etc.
Mars, logarithmic prior

Cumulative Probability

\[ \log_{10}[h] \quad (h \text{ in Gyr}^{-1}) \]

Posterior: Solid
Prior: Dashed
Mars, logarithmic prior

Posterior: Solid
Prior: Dashed
Probability Density

Posterior: Solid
Prior: Dashed

Conservative

Uniform
Log (-3)
Log (-11)
Log (-22)
InvUnif (-3)
InvUnif (-11)
InvUnif (-22)
PDF

(uneform prior)

\[ \log_{10} [\lambda] \quad (\lambda \text{ in Gyr}^{-1}) \]
CDF

(uniform prior)
PDF

(logarithmic prior)

\[ \log_{10}[\lambda] \text{ (} \lambda \text{ in Gyr}^{-1}) \]
CDF

(logarithmic prior)
PDF

(inverse uniform prior)

\( \log_{10}[\lambda] \) (\( \lambda \) in Gyr\(^{-1} \))

\( \tau_{\text{evolve}} \) (Gyr)
CDF

Inverse uniform prior
Bounds on $\log_{10}[h]$ ($h$ in Gyr$^{-1}$)

- Median Value of $\lambda$
- 1-$\sigma$ Lower Bound
- 2-$\sigma$ Lower Bound

Optimistic ($\lambda_{\text{min}} = 10^{-11}$ Gyr$^{-1}$)