Journal of Theoretical Biology 437 (2018) 222–224

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Letter to editor

A note on random catalytic branching processes

Mike Steel[a,*], Stuart Kauffman[b]

[a] Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand
[b] Affiliate Professor Institute for Systems Biology Emeritus Professor Biochemistry and Biophysics University of Pennsylvania, PA, USA

ARTICLE INFO

Article history:
Received 13 May 2017
Revised 29 August 2017
Accepted 22 October 2017

Keywords:
Protocell
Birth-death process
Extinction
Catalysis
Lateral gene transfer
Coupling
Gambler's ruin

ABSTRACT

A variety of evolutionary processes in biology can be viewed as settings where organisms 'catalyse' the formation of new types of organisms. One example, relevant to the origin of life, is where transient biological colonies (e.g. prokaryotes or protocells) give rise to new colonies via lateral gene transfer. In this short note, we describe and analyse a simple random process which models such settings. By applying theory from general birth-death processes, we describe how the survival of a population under catalytic diversification depends on interplay of the catalysis rate and the initial population size. We also note how such process can also be viewed within the framework of 'self-sustaining autocatalytic networks'.

1. Introduction

The process of lateral gene transfer is evident today in the pangenome size of many bacterial 'species' (including *Escherichia coli*) where the core genome shared by most individuals is often very small compared with the number of genomes across sampled individuals (McInerney et al., 2017). In particular, gene transfer can lead to new combinations of genes this can allow for new species to form (*c.f.* Papke and Gogarten (2012)). Very early in evolution, gene transfer events that generate new combinations of genes were also essential in order to bridge the evolutionary gap from the first genes to the first genomes in populations of protocells (Koonin, 2014; Koonin and Martin, 2005; Weiss et al., 2016; Woese, 2002).

In this note, we consider modelling this process (and others described shortly) as a type of random birth-death process, in which birth events are 'catalysed' by other individuals. In the setting of early life in the presence of lateral gene transfer, an 'individual' refers to a colony of genetically identical protocells (a 'species'), a birth event refers to a 'species' A giving rise to an additional new species A' (with a different combination of genes) by the lateral gene transfer of certain genes from a protocell of a third species B into a protocell in A (this protocell then subdivides repeatedly to form the additional colony A' that is genetically different from either A or B). In this way, a lateral gene transfer can be viewed formally as a 'catalyst' for the formation of a new 'species' from the existing pool of 'species' (i.e. the process does not destroy species B, much as a catalyst is not used up in a chemical reaction). A death event refers to a 'species' (i.e. a colony of protocells having a given combination of genes) dying out. This process is illustrated in Fig. 1.

Viewed as a simple stochastic model, this process is similar to classical linear birth-death Yule processes, but with an important difference: the birth rate per individual depends on the existing population size (the more colonies with distinct genomes are present, the more opportunity for transfer events to form new combinations there are (thus, the process is stochastically different from a standard constant birth-death model of cladogenesis). As we describe below, the population of colonies is either certain to die out or else it may survive and grow indefinitely with a positive probability that depends on three key parameters: the rate at which colonies die out, the initial number of colonies, and a function that describes the rate of lateral gene transfer.

The model we study is also motivated by four recent papers which consider different biological processes involving the formation of new entities by a process of catalysis involving other extant members of the set. These are the study by Gatti et al. (2017) in which biodiversity in ecology is viewed as an autocatalytic process, the paper by Montévil and Mossio (2015) on constraint closure in biological organisation (viewed here as a random process in which constraints are treated more abstractly as 'catalysts'), the recent paper Zeravcic and Brenner (2017) on the emergence and exponential growth of catalytic cycles with colloidal spheres, and the mod-

* Corresponding author.
E-mail address: mike.steel@canterbury.ac.nz (M. Steel).

https://doi.org/10.1016/j.jtbi.2017.10.024
0022-5193/

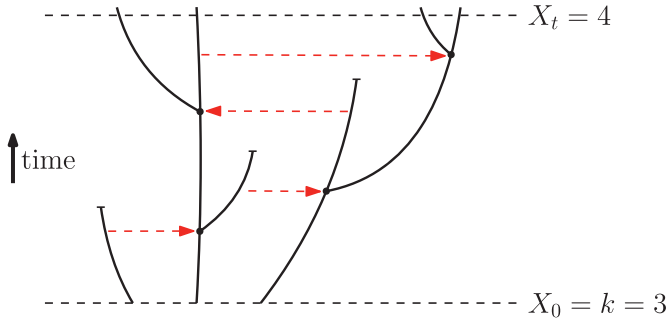

Fig. 1. An example of an catalytic branching processes, starting with three items, and involving four catalysation events (horizontal dashed arrows), each of which leads to a new items arising. For example, in lateral transfer in early life, a dashed arrow refer to the transfer of genes from a protocell of one colony into a protocell of a different colony leading to a protocell with a new combination of genes that can then form go on to produce a new, genetically distinct colony. There are also three loss (extinction or death) events. At time t, there are four extant items.

elling of autocatalytic networks in human cognition in Gabora and Steel (2017).

1.1. Model description

We consider a continuous-time Markovian random process that starts with $k \geq 1$ items at time $t = 0$ and proceeds as follows. At each given instant t, if there are $X_t = n$ items present, then each one of these items randomly and independently catalyses the creation of a new item from an existing (different) item at rate $\varphi(n)$. Here, $\varphi(n)$ is assumed to be a monotone increasing function. In addition, we will assume that each item can disappear independently from the system (i.e. die) at a constant rate μ. An example of this generic process is illustrated in Fig. 1. The assumption that φ is monotone is motivated by the assumption that an item has higher probability of forming a new lineage, if there are more items available to act as a potential catalyst. We call such a process $(X_t, t \geq 0)$ a *catalytic branching process*.

This process can be described as a particular instance of a general birth and death process (see e.g. Section 6.2 of Allen (2003)), where $\lambda_i = i\varphi(i)$ and $\mu_i = i\mu$. Unless otherwise stated, we assume that $\varphi(n)$ is bounded above and so the monotone increasing function $\varphi(n)$ converges to a limit, which we will denote by K.

Note that if we regard the splitting events formally as one type of 'reaction' (with one reactant, one catalyst and two products), and the extinction events as another type of 'reaction' ($x \mapsto \emptyset$) and take the set of k items at time 0 as a 'food set' of 'molecule types', then the set of splitting event reactions forms a CAF (constructively autocatalytic, food-generated set), as defined in Mossel and Steel (2005), and thereby a RAF (reflexively autocatalytic and food-generated set) as defined in Hordijk and Steel (2004). In other words, a catalytic branching process is a special type of autocatalytic network as introduced in Kauffman (1986); 1993).

2. Properties of the model

The process we have described is guaranteed to either become extinct (i.e. with probability 1, there is a time T for which $X_t = 0$ for all $t > T$) or to tend in size to infinity as t grows. This follows from a very general result of P. Jagers concerning stochastic processes that can undergo extinction (Theorem 2 of Jagers (1992); see also Steel (2015)).

Moreover, (i) if $K \leq \mu$, extinction is certain and (ii) if $K > \mu$, there is a positive probability that the population size will grow indefinitely. Claims (i) and (ii) can be justified by a coupling argument based on two associated linear birth-death processes (and the classical result that such processes are certain to become extinct precisely if the birth rate is less or equal to the death rate; see, for example, Kendall (1948)). Claims (i) and (ii) also follows from more explicit results concerning the extinction probability for general birth-death processes. In particular, if we let $q_k = \lim_{t\to\infty} \mathbb{P}(X_t = 0 | X_0 = k)$ denote the probability that the process eventually becomes extinct, then applying Theorem 6.2 of Allen (2003) with $\lambda_i = i\varphi(i)$ and $\mu_i = i\mu$ gives the following explicit description for q_k:

$$q_k = \frac{\sum_{i=k}^{\infty} \mu^i \prod_{j=1}^{i} \varphi(j)^{-1}}{1 + \sum_{i=1}^{\infty} \mu^i \prod_{j=1}^{i} \varphi(j)^{-1}}.$$

From this, it can be shown that for $K > \mu$, the survival probability $p_k = 1 - q_k$ converges to 1 at an exponential rate with k. The following result makes this a little more precise.

Proposition 1. *For a catalytic branching process starting with $k \geq 1$ items at time 0 and with $\mu < K (= \lim_{n\to\infty} \varphi(n))$, the probability p_k that the process grows indefinitely (rather than becoming extinct) satisfies the inequality:*

$$p_k \geq 1 - \left(\frac{\mu}{\varphi(\ell)}\right)^{k-\ell},$$

where ℓ is any value for which $\varphi(\ell) > \mu$ and $k \geq \ell$.

Proof. Suppose that $k > \ell$, where $\varphi(\ell) = \lambda > \mu$. Let us first consider a linear birth-death process Y_t, starting with k individuals at time 0 and with birth and death rates λ and μ respectively (i.e. in the setting of general birth-death models, $\lambda_i = i\lambda, \mu_i = i\mu$). Let $p_{k,\ell}$ be the probability of the event $E_\ell^{(k)}$ that $Y_t \geq \ell$ for all $t \geq 0$. Now consider the discrete-time random walk W_r $(r = 0, 1, 2, 3 \ldots)$ on the set $0, 1, 2, \ldots$, that starts at $W_0 = k - \ell \geq 1$, and which records the changing values of $Y_t - \ell$, but which treats 0 as an absorbing state. W_r is then a simple random random walk for which a step to the right at any point $i \geq 1$ occurs with probability $p = \lambda/(\lambda + \mu)$, whereas a step to the left at point $i \geq 1$ occurs with probability $1 - p = \mu/(\lambda + \mu)$. It is now a classic result from the Markov chain theory of simple random walks (namely the (unlimited) 'Gambler's Ruin' problem) that the probability of never hitting zero, given that one starts from position $i \geq 1$, is given by $1 - (\mu/\lambda)^i$. Taking $i = k - \ell$ gives:

$$p_{k,\ell} = 1 - \left(\frac{\mu}{\varphi(\ell)}\right)^{k-\ell}.$$

We now apply a coupling argument for the original random catalytic process X_t. Note that X_t can be realised as the process obtained from Y_t by randomly and independently (i) disallowing (with a certain probability) birth events in Y_t when there are fewer than ℓ individuals, and (ii) introducing (with a certain probability) additional birth events in Y_t when Y_t becomes larger than ℓ. Now, conditional on $E_\ell^{(k)}$, the disallowing events of Type (i) do not occur, and the additional events of Type (ii) only serve to increase the size of X_t. Thus the probability that X_t is at least equal to ℓ for all $t \geq 0$ is bounded below by $\mathbb{P}(E_\ell^{(k)})$, and this probability is given by the expression above for $p_{k,\ell}$. □

Remarks

- Proposition 1 makes explicit how the survival and growth of biological populations of protocell colonies (having different combinations of genes) evolving under a simple catalysation process (involving lateral gene transfer) depends on the interplay between a sufficiently high lateral gene transfer rate and the size of the initial population k. Notice, in particular, that the role of k in the survival probability of the process is more delicate than for a classic linear birth-death model with parameters $\lambda = K$ and μ. In the latter case, when $K > \mu$, the convergence of

the survival probability to 1 as k (the founding population size) grows is described exactly by an expression of the form $1-q^k$ for all values of k. However, for a catalytic branching process with $K>\mu$, the survival probability may still be close to zero for large values of k, when φ is a slowly converging function to its limit, and $\varphi(k)<\mu$.

A further difference between a catalytic branching process and a linear birth-death process (with rates λ and μ) is that the latter process survives with a probability tending to 1 as $\lambda/\mu\to\infty$ and this holds even when the founding population has size 1. By contrast, for a catalytic branching process, the probability that the process survives depends on the interaction of the initial population size and the small n behaviour of $\varphi(n)$ even though the ratio K/μ might be arbitrarily large.

- When the process X_t survives, its expected size grows at a rate that is (at least) exponential in t. That is, $\mathbb{E}[X_t|X_t>0]\geq e^{\gamma t}$ for some $\gamma>0$. If $M=M(t)=\mathbb{E}[X_t]$ denotes the expected size of the population at time t in a catalytic branching process X_t, Jensen's inequality (applied using the convex function $x\mapsto x\varphi(x)$) gives the following differential inequality for M, which holds for all $t\geq 0$:

$$\frac{dM}{dt}\geq M(\varphi(M)-\mu).$$

This inequality ensures that if $K(=\lim_{n\to\infty}\varphi(n))>\mu$, then we can select a value $0<\gamma<K$ for which exponential growth occurs. Moreover, if $\lim_{n\to\infty}\varphi(n)=\infty$, the population may undergo explosive growth (i.e. reach infinite size in a finite time with positive probability, depending on the rate at which φ grows (Allen, 2003)).

Acknowledgment

We thank William Martin and Wim Hordijk for some helpful discussion and references, and two anonymous reviewers for helpful comments.

References

Allen, L.J.S., 2003. An introduction to stochastic processes with applications to biology. Pearson Prentice Hall.
Gabora, L., Steel, M., 2017. Autocatalytic networks in cognition and the origin of culture. J. Theor. Biol. 63, 617–638.
Gatti, R.C., Hordijk, W., Kauffman, S., 2017. Biodiversity is autocatalytic. Ecol. Model. 346, 70–76.
Hordijk, W., Steel, M., 2004. Detecting autocatalyctic, self-sustaining sets in chemical reaction systems. J. Theor. Biol. 227 (4), 451–461.
Jagers, P., 1992. Stabilities and instabilities in population dynamics. J. Appl. Probab. 29 (4), 770–780.
Kauffman, S.A., 1986. Autocatalytic sets of proteins. J. Theor. Biol. 19, 1–24.
Kauffman, S.A., 1993. The origins of order. Oxford University Press.
Kendall, D.G., 1948. On the generalized "birth-and-death" process. Ann. Math. Stat. 19 (1), 1–15.
Koonin, E.V., 2014. Carl woese's vision of cellular evolution and the domains of life. RNA Biol. 11 (3), 197–204. http://doi.org/10.4161/rna.27673.
Koonin, E.V., Martin, W., 2005. On the origin of genomes and cells within inorganic compartments. Trends Genet. 21 (12), 647–654.
McInerney, J.O., Nally, A., O'Connell, J.O., 2017. Why prokaryotes have pangenomes. Nat. Microbiol. 2, 17040.
Montévil, M., Mossio, M., 2015. Biological organisation as closure of constraints. J. Theor. Biol. 372, 179–191.
Mossel, E., Steel, M., 2005. Random biochemical networks and the probability of self-sustaining autocatalysis. J. Theor. Biol. 233 (3), 327–336.
Papke, R.T., Gogarten, J.P., 2012. How bacterial lineages emerge. Science 336, 45–46.
Steel, M., 2015. Reflections on the extinction-explosion dichotomy. Theor. Popul. Biol. 101, 61–66.
Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., Martin, W.F., 2016. The physiology and habitat of the last universal common ancestor. Nat. Microbiol. 1, 16116. doi: 10.1038/nmicrobiol.2016.116.
Woese, C.R., 2002. On the evolution of cells. Proc. Natl. Acad. Sci. USA 99, 8742–8747.
Zeravcic, Z., Brenner, M.P., 2017. Spontaneous emergence of catalytic cycles with colloidal spheres. Proc. Natl. Acad. Sci. USA 114, 4342–4347. doi: 10.1073/pnas.1611959114.