HOW TO CREATE AN OPEN PROBLEM

LOUIGI ADDARIO-BERRY

Abstract. This is an attempt to create a teaching aid for the RandNET session on designing open problems, to be held online on June 7, 2022. It has several open problems scattered throughout.

Contents

1. **Know your audience.** 1
1.1. Participant feedback 2
2. **Get to the point.** 3
3. **Guidance for preparing open problems.** 5
4. **How to find an open problem.** 7
4.1. Feedback from participants 9

1. Know your audience.

This is the first principle that came to mind when I was thinking about what makes a good open problem.

Here are three examples of open problems.

OP 1.1 (Covering $\mathbb{Z}/p\mathbb{Z}$ with translates). *Fix $\alpha \in [0,1]$. Suppose that $A \subseteq \mathbb{Z}/p\mathbb{Z}$ is a random subset of size p^α. How many translates of A are needed to cover all of $\mathbb{Z}/p\mathbb{Z}$?*

OP 1.2 (Statistical analysis of the game of life). *Game of life, introduced by John Conway, is the following cellular automaton. On the planar lattice, each square is a cell which can be occupied or empty. Each cell has eight neighbours. The transition rules are as follows: if an empty cell has 3 occupied neighbours, it will be occupied in the next step (otherwise it remains empty) if an occupied cell has 2 or 3 occupied neighbours, it remains occupied (otherwise it will be empty). Of course, these rules may be changed, and one can also consider probabilistic versions. Now, if we start with an iid configuration on a N by N torus, say that the cells are occupied with probability p and empty with $1-p$, we will end either in a fixed state or in a periodic orbit. My conjecture is that if p is large enough (larger than some small value δ), the density of the occupied cells in the final state (which is now a random variable since we start with an iid configuration) is concentrated for $N \to \infty$.*

OP 1.3 (Size of the largest cluster in subcritical hypercube percolation). *Consider bond percolation on the Hamming hypercube $\{0,1\}^m$, with each edge independently appearing with probability $p \in [0,1]$. For a vertex $v \in \{0,1\}^m$ write $\mathcal{C}(v)$ for the percolation cluster (connected component) containing v. Also, write $\mathcal{C}_1$ for the largest percolation cluster (with ties broken lexicographically, say, and with size measured by number of vertices).*

Date: May 21, 2022.

Define the critical probability $p_c = p_c(m)$ *to be the unique solution of*

$$\mathbb{E}_{p_c}|\mathcal{C}(0)| = 2^{m/3}.$$

Show that for $p = p_c(1-\varepsilon_m)$ *with* $\varepsilon_m \gg 2^{-m/3}$,

$$\frac{|\mathcal{C}_1|}{\varepsilon_m^{-2}\log(\varepsilon_m^3 2^m)} \xrightarrow{\mathbb{P}} 2.$$

(Relevant reference: Hulshof and Nachmias, 2019)

OP 1.4 (Hodge Conjecture). *Let X be a non-singular complex projective manifold. Then every Hodge class on X is a linear combination with rational coefficients of the cohomology classes of complex subvarieties of X.*

OK, four examples of open problems – I got overenthusiastic.

Discussion:
⋆ Which of these do you think are suitable for a RandNET workshop? Why or why not?
⋆ Do you understand the statements of the open problems?
⋆ Do you find the open problems compelling? Well-motivated? Why or why not?
⋆ Can you think of ways to *improve* the open problem statements?

The answers to these questions will of course vary from person to person! But it's worth reflecting on how you react to the different questions.

When *preparing* an open question, think about who you are preparing it for: who will be reading, thinking about, and working on it. This should affect how you describe it, how much detail and background you give, and what sort of motivation you provide. (The principle of taking the time to think about your audience goes a long way not just in coming up with open problems, but more generally in our line of work as researchers and educators.)

1.1. **Participant feedback.** Below is my attempt to synthesize some of the feedback from workshop participants on each of the four problems. As you'll see, some of the feedback is contradictory – different breakout rooms had different opinions. There's a lesson there too; open problems, and their write-ups, are not one-size-fits-all.

Covering $\mathbb{Z}/p\mathbb{Z}$ with translates

- More motivation could be helpful.
- A variant: what if only some translations are allowed? How does the answer change?
- There is no guess for the answer in the problem statement.
- The statement could be made more precise. Are we looking for a "with high probability" result? An expectation result? How does the answer depend on α?
- The problem seems very elementary – but is that deceptive?
- The problem feels connected to covering problems from additive combinatorics (but such problems are often "worst-case" rather than average case). Nonetheless, could techniques from that area be useful, or necessary?
- The problem gets straight to the point.

Statistical analysis of the game of life

- The problem seems long to state at first, but most of that is description of the setting.
- It feels difficult, particularly given how well-known the game of life is.
- It feels accessible; in particular one may try simulation to make predictions.
- Is there an analogous question for the infinite lattice?
- The slide is easier to understand than the textual explanation.

- The problem seems suitable for RandNET.
- Unclear how the problem fits into RandNET.

Size of the largest cluster in subcritical hypercube percolation

- It would be good to have more motivation; where does the 1/3 come from?
- The scaling factors pique curiosity.
- It's interesting because bond percolation is a very familiar setting.
- It is clearly stated and has all the necessary definitions.
- It would be nice to have a discussion of what happens away from criticality (e.g. when $p = p_c(1+\varepsilon)$ for fixed $\varepsilon > 0$ not depending on m.

Hodge conjecture

- Not suitable for RandNET.
- Too technical to be easily approachable.

2. Get to the point.

Here are two open problem descriptions I wrote, both several years ago. Which write-up do you think is better?

Open problem: Is it easier to hit a moving target? (2012). A result of Aldous says that for a reversible Markov chain, hitting times can be used to bound meeting times. More precisely, let $(X_t, t \geq 0)$ and $(Y_t, t \geq 0)$ be independent and identically distributed copies of some reversible continuous time Markov chain on a finite state space V. For $v \in V$ write

$$H_v = \inf(t \geq 0 : X_t = v),$$

and let

$$\tau_{\mathrm{H}} = \max_{u,v \in V} \mathbf{E}\left(H_v \mid X_0 = u\right)$$

be the *maximum expected hitting time*. Also,

$$M_{\mathrm{XY}} = \inf(t \geq 0 : X_t = Y_t)$$

for the meeting time of the two chains, and let

$$\tau_{\mathrm{M}} = \max_{u,v \in V} \mathbf{E}\left(M_{\mathrm{XY}} \mid X_0 = u, Y_0 = v\right).$$

Aldous [1][1] then shows that

$$\tau_{\mathrm{M}} \leq \tau_{\mathrm{H}}, \tag{2.1}$$

and that for *symmetric*[2] chains, for all $u, v \in V$ we have

$$\mathbf{E}\left(M \mid X_0 = u, Y_0 = v\right) = \frac{1}{2}\mathbf{E}\left(H_v \mid X_0 = u\right)$$

which implies that for symmetric chains, $\tau_{\mathrm{M}} = \tau_{\mathrm{H}}/2$.

We can think of (2.1) as saying that for reversible Markov chains, it is easier to hit a target if it is moving than if it is sitting still (think of Y_t as the "moving target"). I would like to know if a result similar in spirit to (2.1) can be stated when a third chain is introduced. More precisely, let $(Z_t, t \geq 0)$ be a third chain with the same distribution as $(X_t, t \geq 0)$ and $(Y_t, t \geq 0)$ but independent of the first two chains. In what follows assume the three chains are started from stationarity. Define $M_{\mathrm{XZ}}, M_{\mathrm{YZ}}$ as we defined M_{XY}, and let

$$M = \min(M_{\mathrm{XY}}, M_{\mathrm{XZ}}, M_{\mathrm{YZ}}).$$

[1] In Aldous' paper this result is stated with a constant greater than 1; as written here it appears as Proposition 14.5 in the unpublished book of Aldous and Fill.

[2] A chain with transition rate matrix Q is symmetric if for any $u, v \in V$ there is a bijection $\sigma : V \to V$ with $\sigma(u) = v$ such that for all $i, j \in V$, $q_{\sigma(i),\sigma(j)} = q_{i,j}$.

By symmetry

$$\mathbf{P}\{M = M_{\mathrm{XY}}\} = \frac{1}{3} + \frac{2}{3}\mathbf{P}\{X_0 = Y_0 = Z_0\} = \frac{1}{3} + \sum_{v \in V} \pi_v^3.$$

(For many natural chains the last sum is negligible and so this essentially says that the probability X and Y meet before either finds Z is at least $1/3$.) It seems plausible that turning Z into a "sitting duck" should make Z harder to find. In other words, writing $H_{\mathrm{XZ}} = \inf(t \geq 0 : X_t = Z_0)$ and defining H_{YZ} accordingly, perhaps it the case that

$$\mathbf{P}\{M_{\mathrm{XY}} \leq \min(H_{\mathrm{XZ}}, H_{\mathrm{YZ}})\} \geq \frac{1}{3} + \frac{2}{3}\mathbf{P}\{X_0 = Y_0 = Z_0\}\,. \tag{2.2}$$

However, I don't even know if $\mathbf{P}\{M_{\mathrm{XY}} \leq \min(H_{\mathrm{XZ}}, H_{\mathrm{YZ}})\}$ can be bounded away from zero independent of the chain.

OP 2.1. *Is there an absolute constant $\epsilon > 0$ such that for any reversible chains X and Y as above,*

$$\mathbf{P}\{M_{\mathrm{XY}} \leq \min(H_{\mathrm{XZ}}, H_{\mathrm{YZ}})\} \geq \epsilon\,?$$

By thinking of first running X and Y until their meeting time, then choosing the stationary point Z_0, we can rephrase the question as follows. Write $X[0,t)$ for the set of states visited by X before time t, and likewise for Y.

OP 2.2. *Is there an absolute constant $\epsilon > 0$ such that for any reversible chains X and Y as above,*

$$\mathbf{E}\left(\sum_{v \in X[0,M_{\mathrm{XY}}) \cup Y[0,M_{\mathrm{XY}})} \pi(v)\right) < 1 - \epsilon\,?$$

Perhaps we can even take $\epsilon = 1/3$?

Note that

$$\begin{aligned}\mathbf{E}\left(\sum_{v \in X[0,M_{\mathrm{XY}}) \cup Y[0,M_{\mathrm{XY}})} \pi(v)\right) &= \sum_{v \in V} \pi(v)\mathbf{P} v \in X[0, M_{\mathrm{XY}}) \cup Y[0, M_{\mathrm{XY}}) \\ &\leq 2\sum_{v \in V} \pi(v)\mathbf{P}\{H_v < M_{\mathrm{XY}}\}\,,\end{aligned}$$

the inequality by symmetry and subadditivity of probabilities, so one potential approach is to try to show that

$$\sum_{v \in V} \pi(v)\mathbf{P}\{H_v < M_{\mathrm{XY}}\} < \frac{1-\epsilon}{2},$$

for an absolute constant $\epsilon > 0$. I am not convinced the latter is always true but do not have a counterexample.

Open problem: Random recursive series-parallel graphs (2016). Build a random series-parallel graph as follows. Start from $G_1 = st$, a network with a single edge (unit resistor). At each step choose an edge e and replace it by two edges, in series with probability $1/2$, otherwise in parallel. At step n we have an n-edge network called G_n. Write $R_n{=}R(G_n)$ for the resistance at step n. (Hambly and Jordan [2] have considered a version of this process where the probability of a series edge is $p \neq 1/2$.)

OP 2.3. *Does R_n converge in distribution? Does it converge almost surely?*

Observation: If so then the limit R must be a non-negative random variable and must satisfy the following recursive distributional equation:

$$R \stackrel{\mathrm{d}}{=} B(R_1 + R_2) + (1 - B)R_1R_2/(R_1 + R_2)$$

Here B is Bernoulli(1/2) independent of R_1 and R_2. Note that this implies $\mathbb{E}R$ is either 0 or ∞.

Observation 2: Let (D_n) be the sequence of networks obtained by making the opposite replacements from those used to build the sequence (G_n). So, when G_n adds edges in series, D_n adds edges in parallel, and vice-versa. Then D_n is the planar dual of G_n and so we have $R(D_n) = 1/R(G_n)$.

It follows that if R_n converges almost surely then the limit R satisfies the distributional identity $R \stackrel{\mathrm{d}}{=} 1/R$. In combination with Observation 1, this implies that $\mathbb{E}R = \infty$.

Observation 3: We can couple the process (G_n) with the binary search tree insertion process in the obvious way. We end up with a sequence of trees (T_n) where each node has a label (s for series or p for parallel). Suppose the root has label s. Then the subtree of nodes connected to the root by a path with all nodes of label s is distributed as a critical branching process with Binomial$(2, 1/2)$ offspring distribution.

This gives us a different tree representation of the network, in terms of the number of "changes of series/parallel type" on the path to the root. In the latter representation the mean number of offspring of the root is infinite because it is distributed like the number of progeny of a critical branching process.

Observation 4: The series and parallel laws imply that the process (R_n) has the same law as the process described as follows.

Let F_1 be the fraction $1/1$. To form F_2 from F_1, replace the 1 in the denominator by either $1 + 1$ or by $1/(1 + 1)$, each with probability $1/2$. In general, F_n will be a continued fraction whose entries are all 1's, and there will be exactly n 1's which are in the denominator of their "parent fraction". To form F_{n+1} from F_n, choose one of these n 1's uniformly at random, then replace it by $1 + 1$ or by $1/(1 + 1)$, each with probability $1/2$.

Setting V_n equal to the value of F_n, then the processes (V_n) and (R_n) have the same laws. This perspective suggests new continued fraction dynamics which may not all have obvious interpretations in terms of random resistor networks, but for which the convergence question remains interesting. (Somewhat related reference: [3].)

Comments. Looking back on them, I find the second write-up, on series-parallel graphs, better than the first. It gets to an open problem statement quickly, and (I think) it's easy to understand. An open problem which takes a page of non-trivial setup before it can be stated can feel intimidating.

Discussion:
$\star$ $\star$ How would you improve or modify these open problem statements?

3. **Guidance for preparing open problems.**

When preparing and typing up your open problem, here are some points to keep in mind. (This is my first attempt at writing down such a list, so surely there are other good things to keep in mind, too!

- You are probably not preparing the open problem in a void. In the current situation, it is for a specific activity, the RandNET workshop this August. The workshop is only a week long, and the participants have a variety of backgrounds. Problems with a low "barrier to entry", in terms of not needing a huge amount of background knowledge or mathematical technology, are preferred!

- The presentations of the open problems should ideally be fully self-contained. However, links to relevant literature (whether to motivate the problem or to provide potentially useful background material), if it exists, are a good idea.
- You should propose an open problem for which you are happy to have a potentially large group of collaborators. Problems where you and one or two other people are already in the course of writing up a paper are probably not suitable. (Of course, *spin-offs* are a different matter - it's normal for an existing project to generate other questions, some of which may make good open problems.)
- It's good to have at least a vague idea of the level of accessibility of your problem, but this can be tricky. I have made the mistake of proposing both easy-sounding problems that are impossibly hard (ask me about lines vs cliques in visibility graphs) and tricky-sounding problems that are actually easy given existing technology (ask me about the dimension of the sumset of the zero set of Brownian motion).

  I think it's ideal to propose a problem that is in an area you're somewhat familiar with, so that you have a sense of whether your open problem is heading into dangerous ground. Ground can be dangerous because it's been overly cultivated and is no longer fertile, or because there are so many brambles and heavy undergrowth that it's hard to make forward progress.
- It's great to state problems that include *intermediate goals.* If you can, try to choose intermediate goals that you think distill some the key features or challenges of the problem. Here is an example of an open problem which is stated with some intermediate goals. (It also has several very nice special cases and variants: by varying the tail behaviour of S, lots of different limiting behaviours should be possible! But I'm not an expert so I won't say more.)

Random-replacement Pólya urns **.** This question has been asked to me [ ] independently by Bénédicte Haas and Batı Şengül: they both encountered this problem studying random trees (or branching processes). It seems to be a very natural generalisation of the classical Pólya-Eggenberger Pólya urn, yet not mentioned in the literature.

Consider the Pólya urn with initial composition vector $(\alpha_1, \ldots, \alpha_d)$ and replacement matrix $S\mathtt{Id}_d$ where S is an integer-valued random variable, and $\mathtt{Id}_d$ the d-dimensional identity matrix. Let us denote by $U(n) = (U_1(n), \ldots, U_d(n))$ the composition vector of the urn at time n. The process defined by

$$Z_n := \frac{U_n}{\alpha + \sum_{i=1}^n S_i},$$

where $\alpha = \sum_{i=1}^d \alpha_i$ is a martingale, and where $(S_n)_{n\geq 1}$ is a sequence of independent copies of S.

Questions:

(i) Warm-up: $(Z_n)_{n\geq 1}$ is bounded in L^2 and thus converges almost surely to a random vector $V = (V_1, \ldots, V_d)$.
(ii) Harder: fix $i \in \{1, \ldots, d\}$, what can be said about the law of V_i (in terms of the law of S or maybe only in terms of its expectation)?
(iii) Main goal: what can be said about the law about the random vector V itself?

A related question. If you embed the process into continuous time with exponential clocks, each colour behaves independently from the others according to a continuous time branching process $Y_i(t)$ with reproduction distribution $S + 1$ and such that $Y_i(0) = \alpha_i$. It is well known that $\mathtt{e}^{-t\mathbb{E}S} Y_i(t)$ is a martingale that converges almost surely to a limit W_i. And knowing the law of the $W_i's$ would permit to deduce the law of the $V_i's$ from the original problem.

Question:

(iv) What is the law of W_i?

What is known when S is deterministic. When S is deterministic, we know that:

(i) For all $i \in \{1, \dots, d\}$, V_i is a Beta-distributed random variable with parameters $\left(\frac{\alpha_i}{S}, \frac{\alpha-\alpha_i}{S}\right)$.

(ii) The random vector V follows a Dirichlet distribution of parameter $\left(\frac{\alpha_1}{S}, \dots, \frac{\alpha_d}{S}\right)$.

(iii) The random variable SW_i follows a Gamma law of parameters $\left(\frac{\alpha_i}{S}, 1\right)$.

Some references to start with. A good start to look at the Yule process (Question *iv*) is certainly Athreya and Ney's book [AthreyaNey]: in Chapter III, the case of super-critical Galton-Watson is treated and can be applied to the generalised Yule processes $Y_i(t)$. I suggest looking at the paper by Janson on triangular urns schemes [Janson06], especially Section 11 where the 2-colour deterministic case is treated. The appendix of [CMP14] and references therein can help (the appendix treats question (iii) in the deterministic case).

References

[AthreyaNey] K. B. Athreya and P E. Ney *Branching processes*, Ann. Math. Stat. **15** (1944), 182–189.

[Janson06] S. Janson, *Limit theorems for triangular urn schemes*, Probability Theory and Related Fields **134**(3) (2006), 417–452.

[CMP14] B. Chauvin, C. Mailler and N. Pouyanne, *Smoothing equations for large Pólya urns*, Journal of Theoretical Probability (to appear).

Discussion:

⋆ ⋆ What other principles do you think are good to keep in mind when preparing open problems?

4. How to find an open problem.

This is probably the hardest section for me to give guidance on. Lots of academics jealously guard their "good" open problems for their own students, because good problems can feel hard to come by! But mathematics is vast, and there are plenty of questions to go around. Here are some ideas for how to find a good open problem to ask.

(1) **Steal one from another mathematician.** Sounds easier than it is. The basic idea is simple: look in other people's papers, see if they state any conjectures or pose any questions, and then ask those.

This sounds tempting, because you are outsourcing the work. However, if you keep the guidance from Section 3 in mind, theft becomes a little more complicated. For one thing: simply pointing your audience to "Conjecture 6.1 in Alon *et al*, 2021" doesn't provide much in the way of motivation or explanation.

For another: many authors only state a conjecture if they've spent some time trying to prove it without success, which may indicate that the problem is quite hard.

Another sort of caveat: in probability, you will sometimes see papers which do the "nice" case of a problem (e.g. with Gaussian noise), leaving the general case (e.g. finite variance noise) as an open problem. That may be interesting, but it likely requires a fair bit of background, and is possibly quite technical. There are related caveats in random graphs and networks (replace "changing the noise" by "changing the network model").

There is nothing wrong with looking through the manuscripts of others in search of open problems, but it's a good idea to spend some time reflecting on the problems you find, before proposing them to others.

(2) **Collaborate.** One of the things I enjoy most about mathematics is working with others, so I'm embarrassed to say that I didn't think of listing this technique myself - Serte and Joost suggested adding it. It's a good one. It's very common for discussions to spark new ideas. To give yourself a jumping-off point, you could, for example, read a paper you're interested in (or pick a paper you've already read), and then try to explain the key ideas from it to a math friend.

Alternately, and perhaps even better: try to explain your own current research to a friend and see if that spurs new ideas. I often find that I get ideas for new research topics when preparing talks about it — trying to formulate the "point" of the work at a high level often makes me think about other angles on the subject, in exactly the sort of way that helps with open problem generation. (If you're a student, then one downside of doing this with your own research is that you may end up with a problem that's too close to your thesis topic, and perhaps not suitable for broad circulation.)

You can also of course just try to brainstorm open problems with a friend *without* doing any preparation; sometimes the minimalist approach is the best, particularly if it's the one you will actually bother to do!

(3) **Change the setting.** One of my mathematical friends and mentors, who is also a great problem poser, is Gábor Lugosi. Over the years, I've observed that one of his "tricks" for finding open problems is to take an existing result he likes, and ask if there is an analogous theorem in a different setting. Often, switching from $\mathbb{R}^d$ to trees, or vice-versa, is a good thing to try.

For example: when I was a student, Benjamini and Rossignol wrote a paper bounding the variance the effective resistance in planar random electrical networks. More specifically: take an $n \times n$ grid and give each edge effective resistance $a > 0$ with probability $1/2$ or $b > 0$ with probability $1/2$. Benjamini and Rossignol showed that the effective resistance from the left to the right side of the grid is super-concentrated (i.e. variance $= o(\text{expectation}^{1/2})$ as $n \to \infty$).

Gábor asked what happens if the $n \times n$ grid is replaced by a complete binary tree; this led us to write the paper "Effective resistance of random trees" with Nicolas Broutin.

Here is another example of "changing the setting", also from Gábor. In the *random recursive tree* model, we begin from a rooted tree T_1 consisting of a single node with label 1. To go from T_n to T_{n+1}, attach a node with label $n+1$ whose parent is uniformly selected among the existing nodes $\{1, \ldots, n\}$. Gábor proposed the following, spatial variant of that model.

OP 4.1 (Spatial random recursive trees (Gabor Lugosi, 2017))**.** *Study the following random geometric graph on the vertex set* $\{1, \ldots, n\}$*. Let* $X_1, \ldots, X_n$ *be independent points drawn uniformly in* $[0,1]^d$*. For* $i < j$*, there is an edge between vertex* i *and vertex* j *if and only if* X_i *is the nearest neighbor of* X_j *among* $X_1, \ldots, X_{j-1}$*. The obtain graph is a tree, rooted at* X_1*. What is the height of the tree? What is the total Euclidean length of the edges? What is the Euclidean diameter of the graph?*

There has already been at least one paper on this model, and Gábor tells me that Nicolas Curien and some collaborators have a paper in progress.

If you are trying to come up with this sort of "change the setting" question, try to find a rich setting. A *bad* setting is one which is quite complicated to describe and where there are only a few properties that the community cares about. A *good* setting is one that's easy to describe and where there are lots of properties whose study is easy to motivate.

NB. Of course it is fine, and even good, to spend some of your time working on problems that no one "cares about". But mathematics is a social activity, and it takes place within a community; if you're coming up with problems that you want to work on in a group setting, it's good to think of questions where other people will see the interest.

(4) **Add randomness to the initial condition.** This title is fairly self-explanatory; the question about the game of life, above, is an example of this trick. This one can be applied to make previously deterministic settings random, so suitable for probability workshops.

(5) **Add random dynamics.** Given a geometric object (e.g. a network model), there are (at least) two natural ways to add random dynamics:
- Do a random walk on the object.
- Do a random walk on the space the object is drawn from.

Take random maps as an example; let Q_n be a uniformly random quadrangulation with n faces. A natural question about the random walk on Q_n is its mixing time (see Lehereicy 2021); a natural question about the random walk on the space of quadrangulations is the mixing time of the edge-flip dynamics (see Caraceni and Stauffer 2020). In both of these specific cases, pinning down the mixing time is still open.

(6) **Steal one from a non-mathematician.** There is another way to steal open problems: look in non-mathematical literature. For example: there's a rich and very active interplay between mathematics and physics, particularly but not exclusively statistical physics. This sometimes takes the form of mathematicians trying to put the predictions of physicists on rigorous ground (which often can involve the development of very interesting new mathematics). If you're comfortable enough with the language used in physics papers, you can try looking at the physics literature and drawing inspiration from it. The area of population genetics is also great for finding meaningful real-world models which are begging for mathematical explanations and analysis.

A challenge to this approach is that sometimes it can be hard to figure out how to formulate a good mathematical open problem. But fundamentally I think this is a great way for mathematicians to spend some of their energy - trying to have fruitful dialogue with researchers from other areas. Treat the effort as a learning process; if you browse through an issue of Nature or Physical Review Letters and you don't find an open problem as a result, you've probably still learned something new!

Discussion:

⋆ Try to formulate open problems using the above formulas for some of your favourite models.

⋆ Think about other pluses and minuses of each of the above techniques.

4.1. **Feedback from participants.** Below is my attempt to synthesize some of the feedback from workshop participants about the "'how to find"/"how to present" open problems part of the workshop. (I hope I am not misrepresenting what anyone said! I do not take credit for the below ideas but I do take responsibility for any lack of clarity, imprecision, or mistakes in their descriptions.)

- The "how to find an open problem" ideas also feel like guidance for doing mathematical research more generally. With this in mind, in the hunt for open problems, one may try to apply some other sorts approaches to mathematical research.
  - As an easier problem (a subcase; special parameter values).
  - Ask a harder problem (perhaps this will lead you to find the "kernel" of the original problem).
  - Start from a technique, and then try to reverse-engineer a problem or family of problems from it.
- Some participants liked the "change the setting" approach because it directly gives motivation for the problem, and can also suggest what behaviour to expect.
- When choosing an open problem, one may wish to know that something will result from thinking about it. This is tricky advice to apply; more senior researchers often have developed intuition for their subject which allows them to make high-quality predictions of this sort. But other than that sort of intuitive confidence, it's hard to say how one can have confidence that a given problem will bear fruit.
- One proposal for an ideal sort of open problem (or way to prepare an open problem) for a RandNET-style workshop:
  - State the problem in 3 sentences, without too much technology
  - Then add motivation, context, graphics, examples.
- To be suitable for RandNET, open problems should have some randomness, and some network aspect to them.
- For some problems, it may be better for the poser not to say everything they know about a problem right at the start, in order to let newcomers to the problem approach it with fresh eyes.
- Relatedly, presenting a bunch of failed avenues may be demotivating (and maybe someone else will even make an approach succeed that the problem poser had ruled out as hopeless).
- In addition to looking at the science literature for inspiration, one may look at the world itself for inspiration! There are physical, environmental, social, technological phenomena of all sorts that may inspire our creativity as mathematicians, and which can lead to interesting and meaningful mathematical models, questions and theories.

Acknowledgements

Thank very much to Serte Donderwinkel and Joost Jorritsma for proposing the workshop, and for giving me very valuable feedback on these notes in advance of the workshop, and for doing a great job on communication and tech in advance of and during the workshop. They deserve much of the credit for this document – it wouldn't have been created without them.

References

[1] David J. Aldous, *Meeting times for independent Markov chains*, Stochastic Processes and their Applications, **38** (1991), 185–193.

[2] B. M. Hambly and Jonathan Jordan. A random hierarchical lattice: the series-parallel graph and its properties. *Adv. in Appl. Probab.*, 36(3):824–838, 2004. ISSN 0001-8678. doi: 10.1239/aap/1093962236. URL http://dx.doi.org/10.1239/aap/1093962236.

[3] Jonathan Jordan. Almost sure convergence for iterated functions of independent random variables. *Ann. Appl. Probab.*, 12(3):985–1000, 2002. ISSN 1050-5164. doi: 10.1214/aoap/1031863178. URL http://dx.doi.org/10.1214/aoap/1031863178.

5. Session plan

⋆ **8:00-8:05** Serte and Joost welcome everyone.
⋆ **8:05-8:15** My intro. (Point out to people that my advice isn't necessarily *good* or *universally useful* advice.) I ask the groups to consider the four open problem statements. Keep the discussion points in mind. Each group should come back with *something* to say to the whole meeting.
⋆ **8:15-8:25** Breakout discussion.
⋆ **8:25-8:40** Group discussion, each breakout group says something.
⋆ **8:40-8:55** I re-present the two longer open problems. Five minutes each: show how to present an open problem in the context a session with a lot of other people presenting. Keep it short and snappy.
⋆ **8:55-9:05** Water break.
⋆ **9:05-9:15** Breakout discussion: Understand and improve these open problem statements. Uncover new, related problems?
⋆ **9:15-9:30** Group discussion, each breakout group says something. (Questions OK!)
⋆ **9:30-9:40** Guidance for preparing open problems.
⋆ **9:40-9:50** How to find an open problem: go through the six bullet points.
⋆ **9:50-10** Questions/comments/buffer time.

⋆ **Notes on how the workshop went.**

- There were 40-50 participants, and there was not enough time for the full program I had planned.
- After Part 1 we only had 7 breakout rooms give feedback, in the interest of time.
- We skipped "Part 2" (8:40-8:55) and the associated breakout discussion/group discussion (9:05-9:30).
- The guidance for preparing open problems took about 25 minutes to go through. People seemed to like this part.
- We added a breakout discussion/participant feedback session after sections 3 and 4, which resulted in some quite valuable discussion. We started with the groups that hadn't had a chance to comment after Part 1, so that every breakout group had a chance to share their thoughts about some aspect of the overall discussion.