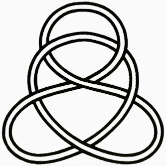
Exchangeability

The most basic probability concept you haven't heard of

Louigi Addario-Berry

February 5^2, $3^4 5^2$

McGill University



Thanks

Thanks to my postdoc coaches: Mohammad Farhat, June Vuong, Ian Waudby-Smith; to Miller Institute staff: Emily Birman, Donata Hubert, Hilary Jacobsen, Vrinda Khanna; and to all of you for being here.

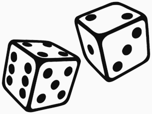
No Infinitesimals in Berkeley

Although this is a mathematical talk, I will try to respect George Berkeley's axiom:

No Infinitesimals in Berkeley

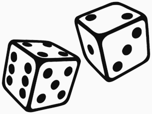
Although this is a mathematical talk, I will try to respect George Berkeley's axiom: "No reasoning about things whereof we have no idea. Therefore no reasoning about Infinitesimals."

Laws of large numbers



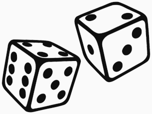
- The principle: over many repeated independent trials of a given experiment, a distribution reveals itself.

Laws of large numbers



- The principle: over many repeated independent trials of a given experiment, a distribution reveals itself.
- Example: if I roll a fair 6-sided die many times, the *average value* will be about $3.5 = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6)$.

Laws of large numbers



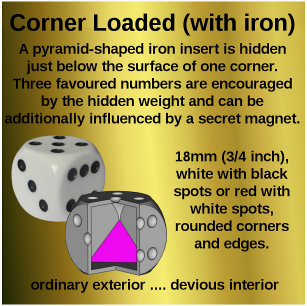
- The principle: over many repeated independent trials of a given experiment, a distribution reveals itself.
- Example: if I roll a fair 6-sided die many times, the *average value* will be about $3.5 = \frac{1}{6}(1+2+3+4+5+6)$.
- More strongly: each of the numbers $1, 2, 3, 4, 5, 6$ will come up approximately $1/6$ of the time in the long run.

Beyond independence

- Dice ordered from `trick-dice.com`

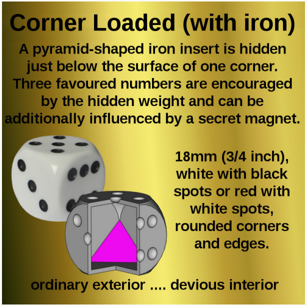
Beyond independence

- Dice ordered from `trick-dice.com`



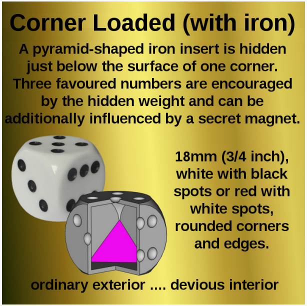
Beyond independence

- Dice ordered from `trick-dice.com`
- Two favour $\{1, 2, 3\}$ (low-bias), two favour $\{4, 5, 6\}$ (high-bias).



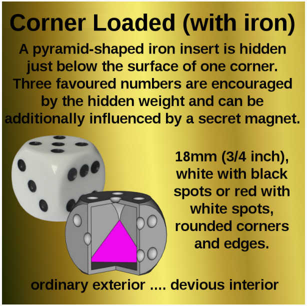
Beyond independence

- Dice ordered from `trick-dice.com`
- Two favour $\{1, 2, 3\}$ (low-bias), two favour $\{4, 5, 6\}$ (high-bias). (NB: the inclusion of the word “bias” renders this talk ineligible for NSF support.)



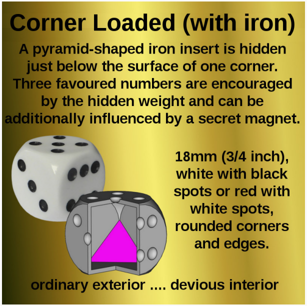
Beyond independence

- Dice ordered from `trick-dice.com`
- Two favour $\{1, 2, 3\}$ (low-bias), two favour $\{4, 5, 6\}$ (high-bias). (NB: the inclusion of the word “bias” renders this talk ineligible for NSF support.)
- I don’t know which are low-bias and which are high-bias.



Beyond independence

- Dice ordered from `trick-dice.com`
- Two favour $\{1, 2, 3\}$ (low-bias), two favour $\{4, 5, 6\}$ (high-bias). (NB: the inclusion of the word "bias" renders this talk ineligible for NSF support.)
- I don't know which are low-bias and which are high-bias.
- Let's gamble - but first...

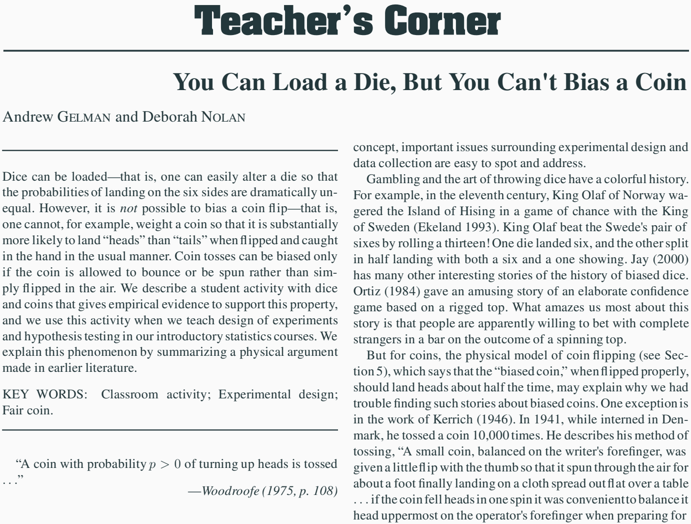


Before gambling

Why did I order loaded dice rather than biased coins?

Before gambling

Why did I order loaded dice rather than biased coins?



Teacher's Corner

You Can Load a Die, But You Can't Bias a Coin

Andrew GELMAN and Deborah NOLAN

Dice can be loaded—that is, one can easily alter a die so that the probabilities of landing on the six sides are dramatically unequal. However, it is *not* possible to bias a coin flip—that is, one cannot, for example, weight a coin so that it is substantially more likely to land "heads" than "tails" when flipped and caught in the hand in the usual manner. Coin tosses can be biased only if the coin is allowed to bounce or be spun rather than simply flipped in the air. We describe a student activity with dice and coins that gives empirical evidence to support this property, and we use this activity when we teach design of experiments and hypothesis testing in our introductory statistics courses. We explain this phenomenon by summarizing a physical argument made in earlier literature.

KEY WORDS: Classroom activity; Experimental design; Fair coin.

"A coin with probability $p > 0$ of turning up heads is tossed . . ."

—*Woodroofe (1975, p. 108)*

concept, important issues surrounding experimental design and data collection are easy to spot and address.

Gambling and the art of throwing dice have a colorful history. For example, in the eleventh century, King Olaf of Norway wagered the Island of Hising in a game of chance with the King of Sweden (Ekeland 1993). King Olaf beat the Swede's pair of sixes by rolling a thirteen! One die landed six, and the other split in half landing with both a six and a one showing. Jay (2000) has many other interesting stories of the history of biased dice. Ortiz (1984) gave an amusing story of an elaborate confidence game based on a rigged top. What amazes us most about this story is that people are apparently willing to bet with complete strangers in a bar on the outcome of a spinning top.

But for coins, the physical model of coin flipping (see Section 5), which says that the "biased coin," when flipped properly, should land heads about half the time, may explain why we had trouble finding such stories about biased coins. One exception is in the work of Kerrich (1946). In 1941, while interned in Denmark, he tossed a coin 10,000 times. He describes his method of tossing, "A small coin, balanced on the writer's forefinger, was given a little flip with the thumb so that it spun through the air for about a foot finally landing on a cloth spread out flat over a table . . . if the coin fell heads in one spin it was convenient to balance it head uppermost on the operator's forefinger when preparing for

Gambling, round 1

- Recall: one of the dice is $\{1, 2, 3\}$-biased, one is $\{4, 5, 6\}$-biased.

Gambling, round 1

- Recall: one of the dice is $\{1, 2, 3\}$-biased, one is $\{4, 5, 6\}$-biased.
- Let's start by choosing one of the dice uniformly at random.

Gambling, round 1

- Recall: one of the dice is $\{1, 2, 3\}$-biased, one is $\{4, 5, 6\}$-biased.
- Let's start by choosing one of the dice uniformly at random.
- I'll offer even odds; you can choose whether to bet on low, $\{1, 2, 3\}$, or high, $\{4, 5, 6\}$.

Gambling, round 2

- I'll offer even odds, but this time *I* choose whether to bet on low numbers or high numbers.

Independence is broken

- The rolls are correlated through the initial choice of which die to roll.

Independence is broken

- The rolls are correlated through the initial choice of which die to roll.
- A large number in the first roll makes it more likely that we picked a high-biased die.

Independence is broken

- The rolls are correlated through the initial choice of which die to roll.
- A large number in the first roll makes it more likely that we picked a high-biased die.
- This in turn makes it more likely that the second roll is high.

Exchangeability

- The rolls are not independent.

Exchangeability

- The rolls are not independent.
- However, they are exchangeable; any permutation of a given sequence of possible outcomes is equally likely.

Exchangeability

- The rolls are not independent.
- However, they are exchangeable; any permutation of a given sequence of possible outcomes is equally likely.
- For example: rolling 4, then 1, has the same probability as rolling 1, then 4.

Exchangeability

- The rolls are not independent.
- However, they are exchangeable; any permutation of a given sequence of possible outcomes is equally likely.
- For example: rolling 4, then 1, has the same probability as rolling 1, then 4. This is because whichever die we selected, the sequences $(4, 1)$ and $(1, 4)$ are equiprobable for that die.

Exchangeability

- The rolls are not independent.
- However, they are exchangeable; any permutation of a given sequence of possible outcomes is equally likely.
- For example: rolling 4, then 1, has the same probability as rolling 1, then 4. This is because whichever die we selected, the sequences $(4, 1)$ and $(1, 4)$ are equiprobable for that die.
- Similarly, the sequences $(3, 1, 4, 1, 5)$ and $(1, 1, 3, 4, 5)$ are equiprobable.

Recovering independence

Recovering independence

- If we somehow know the bias of the die (perhaps I secretly marked one of them in advance) then the correlation induced by the initial choice is removed.

Recovering independence

- If we somehow know the bias of the die (perhaps I secretly marked one of them in advance) then the correlation induced by the initial choice is removed.
- Problem: How can I figure out the bias of the die?

Recovering independence

- If we somehow know the bias of the die (perhaps I secretly marked one of them in advance) then the correlation induced by the initial choice is removed.
- Problem: How can I figure out the bias of the die?
- Solution: roll the selected die infinitely often, deduce the probability of each face.

Recovering independence

- If we somehow know the bias of the die (perhaps I secretly marked one of them in advance) then the correlation induced by the initial choice is removed.
- Problem: How can I figure out the bias of the die?
- Solution: roll the selected die infinitely often, deduce the probability of each face.
- The law of large numbers guarantees that, asymptotically, the probability distribution of the faces will reveal itself.

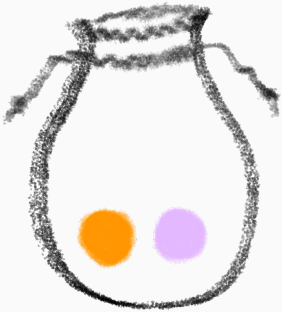
Recovering independence

- If we somehow know the bias of the die (perhaps I secretly marked one of them in advance) then the correlation induced by the initial choice is removed.
- Problem: How can I figure out the bias of the die?
- Solution: roll the selected die infinitely often, deduce the probability of each face.
- The law of large numbers guarantees that, asymptotically, the probability distribution of the faces will reveal itself.
- Conclusion: access to this asymptotic information allows one to recover independence from exchangeability.

Recovering independence

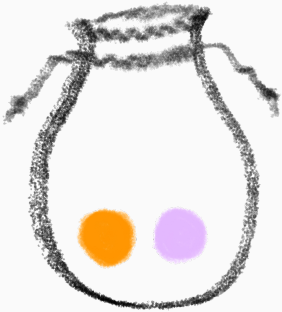
- If we somehow know the bias of the die (perhaps I secretly marked one of them in advance) then the correlation induced by the initial choice is removed.
- Problem: How can I figure out the bias of the die?
- Solution: roll the selected die infinitely often, deduce the probability of each face.
- The law of large numbers guarantees that, asymptotically, the probability distribution of the faces will reveal itself.
- Conclusion: access to this asymptotic information allows one to recover independence from exchangeability.
- Note: the asymptotic information is itself random as it depended on our initial, random choice of which die to roll.

Another exchangeability example: Pólya Urns



Another exchangeability example: Pólya Urns

- Random process indexed by positive integer times. Initial state: one orange ball, one lilac ball.



Another exchangeability example: Pólya Urns

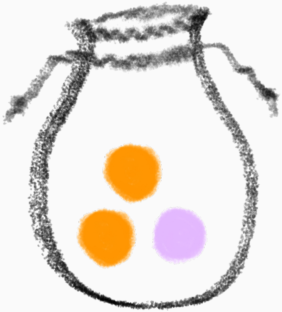
- Random process indexed by positive integer times. Initial state: one orange ball, one lilac ball.
- At each step: draw a ball from the urn uniformly at random; return it to the urn along with *another ball of the same colour*.

Another exchangeability example: Pólya Urns

- Random process indexed by positive integer times. Initial state: one orange ball, one lilac ball.
- At each step: draw a ball from the urn uniformly at random; return it to the urn along with *another ball of the same colour*.

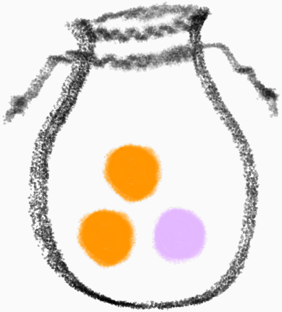
Another exchangeability example: Pólya Urns

- Random process indexed by positive integer times. Initial state: one orange ball, one lilac ball.
- At each step: draw a ball from the urn uniformly at random; return it to the urn along with *another ball of the same colour*.
- After n steps there are $n+1$ balls in the urn.



Another exchangeability example: Pólya Urns

- Random process indexed by positive integer times. Initial state: one orange ball, one lilac ball.
- At each step: draw a ball from the urn uniformly at random; return it to the urn along with *another ball of the same colour*.
- After n steps there are $n+1$ balls in the urn.
- Python simulation: link



Pólya Urns - exchangeability

- Suppose that at step n there are $k+1$ orange balls and $n-k$ lilac balls.

Pólya Urns - exchangeability

- Suppose that at step n there are $k + 1$ orange balls and $n - k$ lilac balls.
- This means we drew an orange ball k times, and a lilac ball $n - k - 1$ times.

Pólya Urns - exchangeability

- Suppose that at step n there are $k+1$ orange balls and $n-k$ lilac balls.
- This means we drew an orange ball k times, and a lilac ball $n-k-1$ times.
- The sequence of draws may have been

$$\Big(\underbrace{O, O, \ldots, O}_{k \text{ times}}, \overbrace{L, L, \ldots, L}^{n-k-1 \text{ times}}\Big)$$

Pólya Urns - exchangeability

- Suppose that at step n there are $k+1$ orange balls and $n-k$ lilac balls.
- This means we drew an orange ball k times, and a lilac ball $n-k-1$ times.
- The sequence of draws may have been

$$\Big(\underbrace{O, O, \ldots, O}_{k \text{ times}}, \overbrace{L, L, \ldots, L}^{n-k-1 \text{ times}}\Big)$$

or

$$\Big(\underbrace{L, L, \ldots, L}_{n-k-1 \text{ times}}, \overbrace{O, O, \ldots, O}^{k \text{ times}}\Big)$$

Pólya Urns - exchangeability

- Suppose that at step n there are $k+1$ orange balls and $n-k$ lilac balls.
- This means we drew an orange ball k times, and a lilac ball $n-k-1$ times.
- The sequence of draws may have been

$$\big(\underbrace{O, O, \ldots, O}_{k \text{ times}}, \overbrace{L, L, \ldots, L}^{n-k-1 \text{ times}}\big)$$

or

$$\big(\underbrace{L, L, \ldots, L}_{n-k-1 \text{ times}}, \overbrace{O, O, \ldots, O}^{k \text{ times}}\big)$$

or any other permutation of k 'O's and $n-k-1$ 'L's.

Pólya Urns - exchangeability

- Suppose that at step n there are $k+1$ orange balls and $n-k$ lilac balls.
- This means we drew an orange ball k times, and a lilac ball $n-k-1$ times.
- The sequence of draws may have been

$$\Big(\underbrace{O, O, \ldots, O}_{k \text{ times}}, \overbrace{L, L, \ldots, L}^{n-k-1 \text{ times}}\Big)$$

or

$$\Big(\underbrace{L, L, \ldots, L}_{n-k-1 \text{ times}}, \overbrace{O, O, \ldots, O}^{k \text{ times}}\Big)$$

or any other permutation of k ‘O’s and $n-k-1$ ‘L’s.

- Any such permutation is equally likely!

Pólya Urns - exchangeability

- Suppose that at step n there are $k+1$ orange balls and $n-k$ lilac balls.
- This means we drew an orange ball k times, and a lilac ball $n-k-1$ times.
- The sequence of draws may have been

$$\Big(\underbrace{O, O, \ldots, O}_{k \text{ times}}, \overbrace{L, L, \ldots, L}^{n-k-1 \text{ times}}\Big)$$

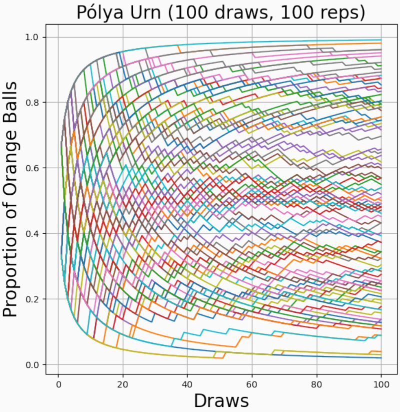
or

$$\Big(\underbrace{L, L, \ldots, L}_{n-k-1 \text{ times}}, \overbrace{O, O, \ldots, O}^{k \text{ times}}\Big)$$

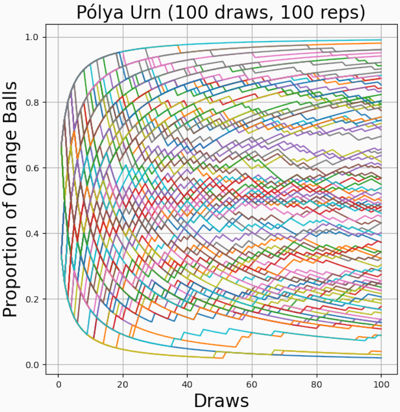
or any other permutation of k 'O's and $n-k-1$ 'L's.

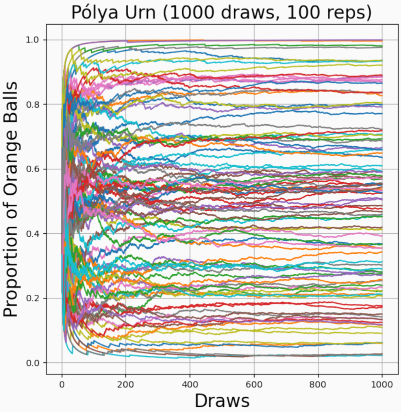
- Any such permutation is equally likely! (Requires an easy calculation.)

Pólya urn simulations and facts

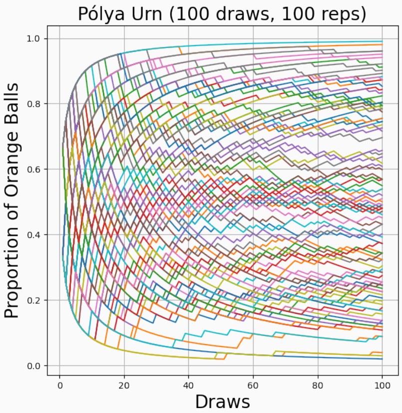


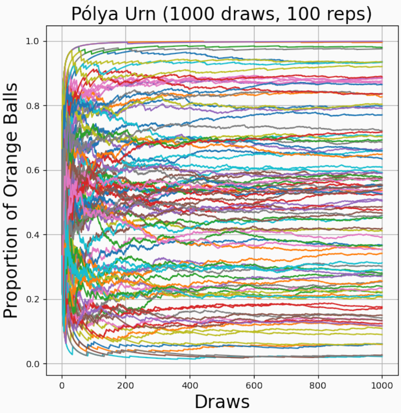
Pólya urn simulations and facts





Pólya urn simulations and facts

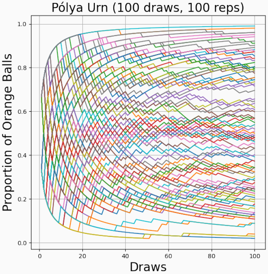


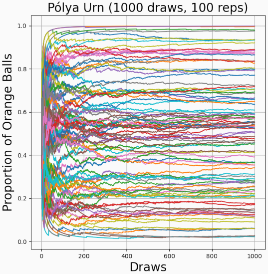


Theorem

Let X_n be the number of orange balls in the urn after n draws. Then X_n/n converges almost surely to a random variable U which is uniformly distributed on $[0, 1]$.

Pólya urn simulations and facts

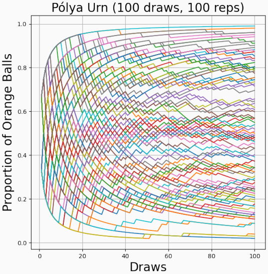


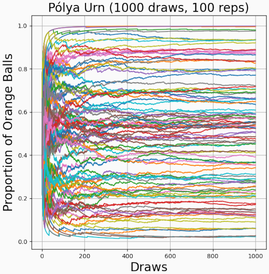


Theorem

Let X_n be the number of orange balls in the urn after n draws. Then X_n/n converges almost surely to a random variable U which is uniformly distributed on $[0, 1]$.

Pólya urn simulations and facts





Theorem

Let X_n be the number of orange balls in the urn after n draws. Then X_n/n converges almost surely to a random variable U which is uniformly distributed on $[0, 1]$. Moreover, conditionally given the limit U, the draws become independent with $\mathbf{Prob}(orange) = U$.

De Finetti's theorem

Theorem
The entries of any infinite sequence of exchangeable observations are conditionally independent relative to some latent variable.

De Finetti's theorem

Theorem

The entries of any infinite sequence of exchangeable observations are conditionally independent relative to some latent variable.

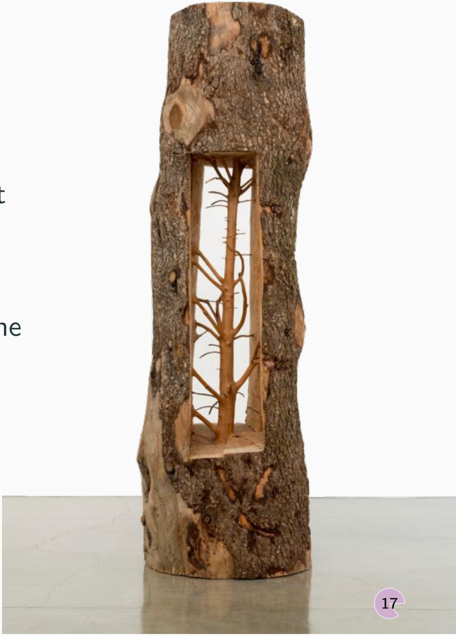
> *"A sequence of events is judged to be exchangeable if our subjective probability for each sequence is unaffected by the order of our observations. De Finetti brilliantly proved that this assumption is mathematically equivalent to acting as if the events are independent, each with some true underlying 'chance' of occurring, and that our uncertainty about that unknown chance is expressed by a subjective, epistemic probability distribution. This is remarkable: it shows that, starting from a specific, but purely subjective, expression of convictions, we should act as if events were driven by objective chances."* *David Spiegelhalter*

Network archaeology

- Consider a network structure that grows over time, starting from a small initial structure S.

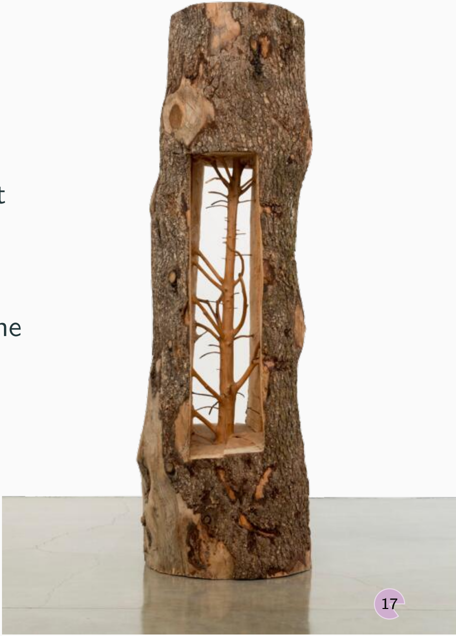
Network archaeology

- Consider a network structure that grows over time, starting from a small initial structure S.
- We observe it at some (large) time t and see a network G.



Network archaeology

- Consider a network structure that grows over time, starting from a small initial structure S.
- We observe it at some (large) time t and see a network G.
- Under what assumptions can we recover S from observing G?



Random recursive trees

- Initial structure S is a single node.

Random recursive trees

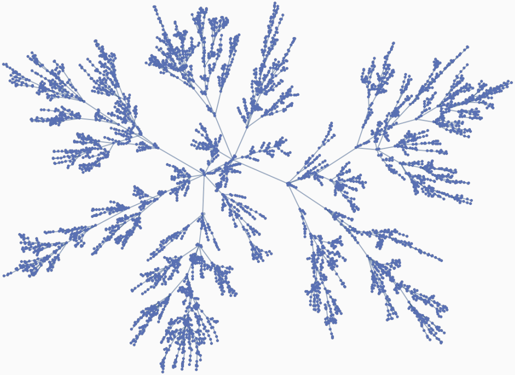
- Initial structure S is a single node.
- Growth rule: at step k a new node arrives and attaches to a uniformly random existing node.

Random recursive trees

- Initial structure S is a single node.
- Growth rule: at step k a new node arrives and attaches to a uniformly random existing node.
- After $n - 1$ steps the network G is a tree with n nodes.

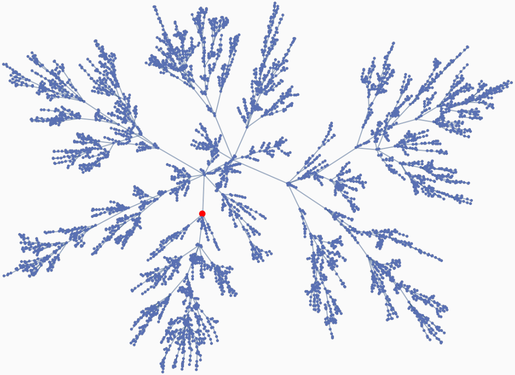
Random recursive trees

- Initial structure S is a single node.
- Growth rule: at step k a new node arrives and attaches to a uniformly random existing node.
- After $n-1$ steps the network G is a tree with n nodes.



Random recursive trees

- Initial structure S is a single node.
- Growth rule: at step k a new node arrives and attaches to a uniformly random existing node.
- After $n-1$ steps the network G is a tree with n nodes.



Hidden exchangeability

- Let's revisit the Pólya urn simulation. link

Hidden exchangeability

- Let's revisit the Pólya urn simulation. link
- We can see an embedded random recursive tree.

Hidden exchangeability

- Let's revisit the Pólya urn simulation. link
- We can see an embedded random recursive tree.
- Pólya urn connection $\Rightarrow$ tree edges that arrive early will split cause macroscopic splits of the population.

Hidden exchangeability

- Let's revisit the Pólya urn simulation. link
- We can see an embedded random recursive tree.
- Pólya urn connection $\Rightarrow$ tree edges that arrive early will split cause macroscopic splits of the population.
- Edges that arrive late will split off tiny subtrees from the bulk.

Hidden exchangeability

- Let's revisit the Pólya urn simulation. link
- We can see an embedded random recursive tree.
- Pólya urn connection $\Rightarrow$ tree edges that arrive early will split cause macroscopic splits of the population.
- Edges that arrive late will split off tiny subtrees from the bulk.
- Using exchangeability, can then optimally analyze the performance of the maximum likelihood estimator for root reconstruction (Addario-Berry, Fontaine, Khanfir, Langevin, Têtu 2024).

Thank you