Random BSTs

Outline for Today

- ***Markov's Inequality on Moments***
  - Excellent bounds… if you can calculate them.
- ***Should We Even Balance Trees?***
  - It's a lot of work – is it worth it?
- ***Heights of Randomized BSTs***
  - A beautiful analysis.

Markov’s Inequality on Moments

Markov's Inequality: If X is a nonnegative random variable and $a > 0$, then

$$\Pr\left[X \geq a\right] \leq \frac{\mathrm{E}\left[X\right]}{a}.$$

Suppose $U \sim \text{Uniform}(0, 1)$.
Then $\Pr[U \geq 0.999] = 0.001$.

What bound does
Markov's inequality give?

Answer at
https://cs166.stanford.edu/pollev

Markov's Inequality

- Markov's inequality applies to *all* nonnegative random variables. Thus its bounds are sometimes weak.
- ***Clever Trick:*** Use Markov's inequality to bound a variable *other* than the one you're interested in.
  - This is where Chebyshev's inequality and Chernoff bounds come from.

$$
\begin{aligned}
\Pr\left[X \ge a\right] &= \Pr\Big[\overbrace{\underbrace{X^2}_{X^2 \text{ is nonnegative.}} \ge a^2}^{f : x \mapsto x^2 \text{ is monotone increasing.}}\Big] \\
&\le \frac{\mathrm{E}\left[X^2\right]}{a^2}
\end{aligned}
$$

$$\Pr\left[X \geq a\right] \leq \frac{\mathrm{E}\left[X^2\right]}{a^2}$$

Suppose $U \sim \text{Uniform}(0, 1)$.
Then $\Pr[U \geq 0.999] = 0.001$.

What bound does this version of Markov's inequality give?

Answer at
https://cs166.stanford.edu/pollev

$f : x \mapsto x^t$ is monotone increasing.

$$\Pr[X \geq a] = \Pr[\overbrace{\underbrace{X^t}\geq a^t}] \quad \textit{(for any } t > 0\textit{)}$$

$$\leq \frac{\mathrm{E}[X^t]}{a^t}$$

X^t is nonnegative because X is nonnegative.

$$\Pr[X \geq a] \leq \min_{t>0} \frac{\mathrm{E}[X^t]}{a^t}$$

Suppose $U \sim$ Uniform(0, 1).
Then $\Pr[U \geq 0.999] = 0.001$.

What does this version of Markov's inequality give us?

$$
\begin{aligned}
\Pr\left[U \geq 0.999\right] &\leq \min_{t>0} \frac{\mathrm{E}\left[U^t\right]}{\left(\frac{999}{1000}\right)^t} \\
&= \min_{t>0} \left(\frac{1000}{999}\right)^t \cdot \frac{1}{t+1} \\
&\approx 0.0027
\end{aligned}
$$

Not bad given that we only needed to know $\mathrm{E}[U^t]$!

When Is This Useful?

- This version of Markov's inequality can often provide much tighter bounds than the original version.
- However, you have to be able to compute $\mathrm{E}[X^t]$ for various t.
  - (Or provide an upper bound on it.)
- This makes it tricky to use in many cases, but when it works, it *really* works.

Random Binary Search Trees

Why do we need to balance trees?

Demo Time!

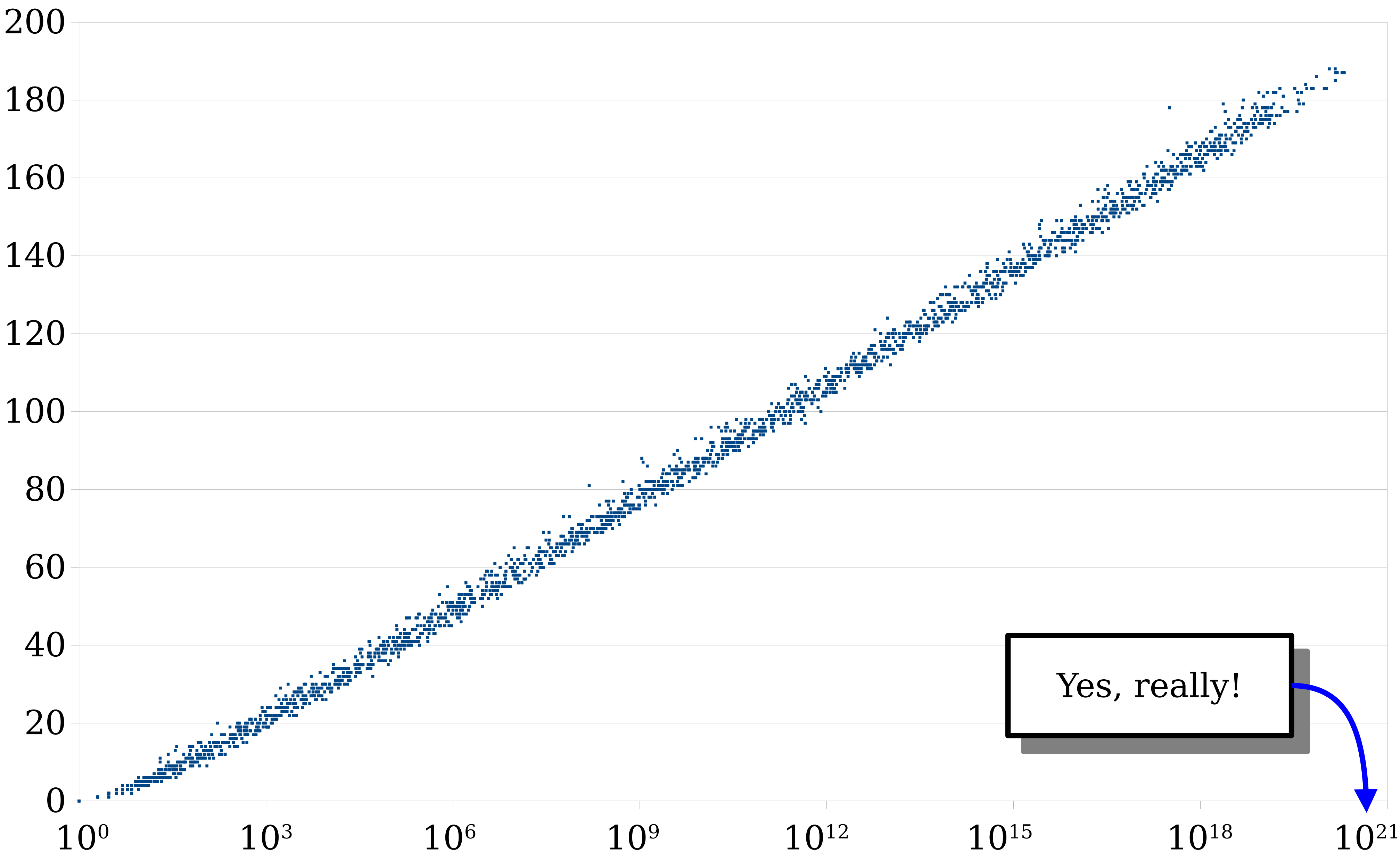
Observations

- Raw BSTs are a Really Bad Idea if there are any trends in your data.
  - (Or if they're maliciously crafted!)
- If the elements inserted into a BST are added in a "random order," then the resulting tree looks pretty balanced.
- Is that a coincidence?

Some Formalisms

- Suppose we have a collection of n elements. We insert those elements into a BST in random order.
  - Equivalently: We build a BST where the elements to insert are i.i.d. random variables with “sufficiently large” support that we don’t need to worry about duplicates.
  - Equivalently: We assign an i.i.d. “weight” to each key and insert keys in increasing weight order.
- Let 🌲$_n$ be a random variable representing the height of the resulting tree. What can we say about 🌲$_n$?
  - Is it low on expectation?
  - Is it concentrated tightly around that expectation?
- ***Meta note:*** We’re now looking at *random data* fed into a *deterministic data structure,* rather than *deterministic data* fed into a *randomized data structure*.

First, Some Data



200
180
160
140
120
100
80
60
40
20
0
10^0
10^3
10^6
10^9
10^{12}
10^{15}
10^{18}
10^{21}
Yes, really!

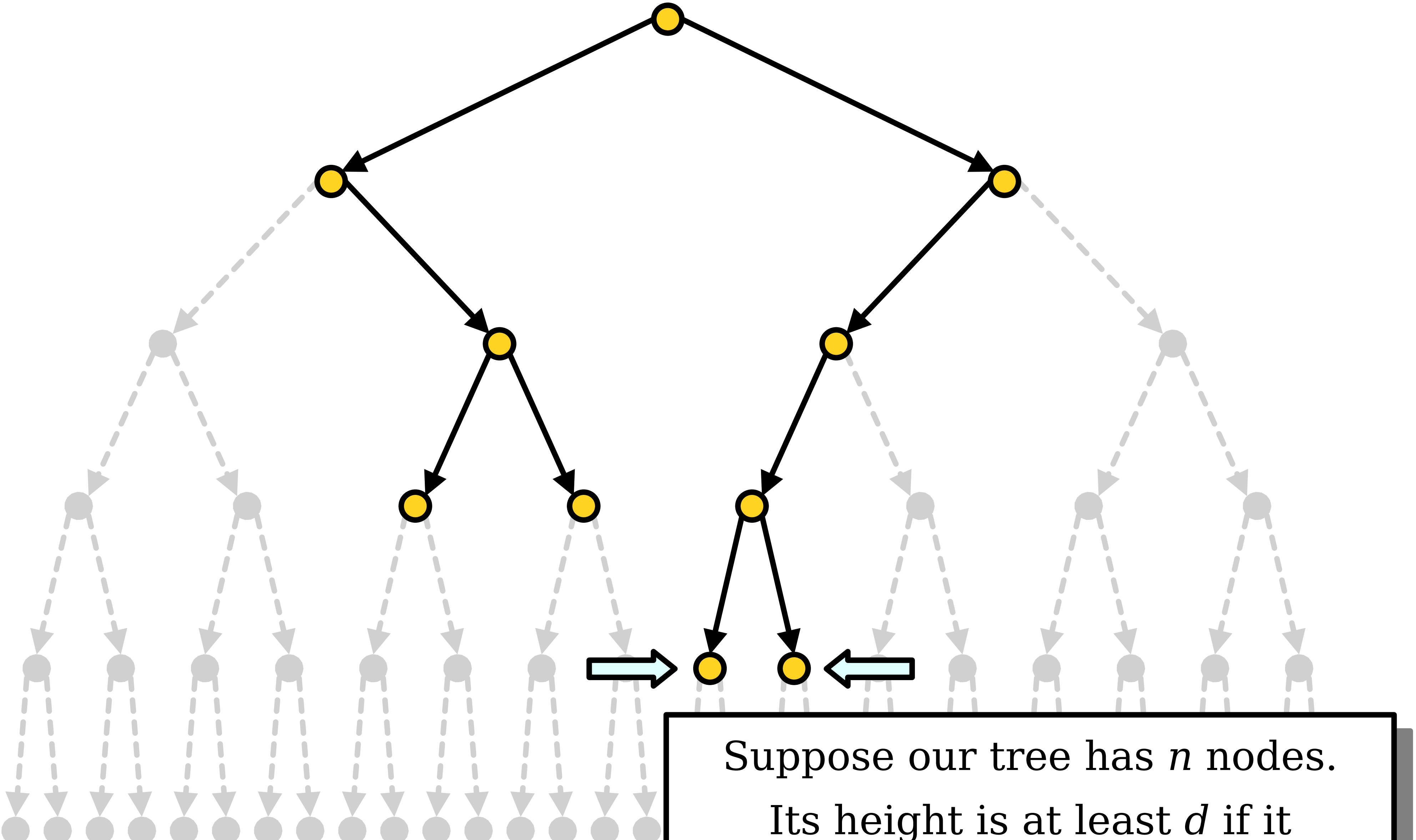
Theorem: There is a constant $\gamma \approx 2.988206...$ where

$$\mathrm{E}\left[\text{🌲}_n\right] = \gamma \lg n - \mathrm{O}(\log \log n)$$

Additionally, for every $c > \gamma$, we have that

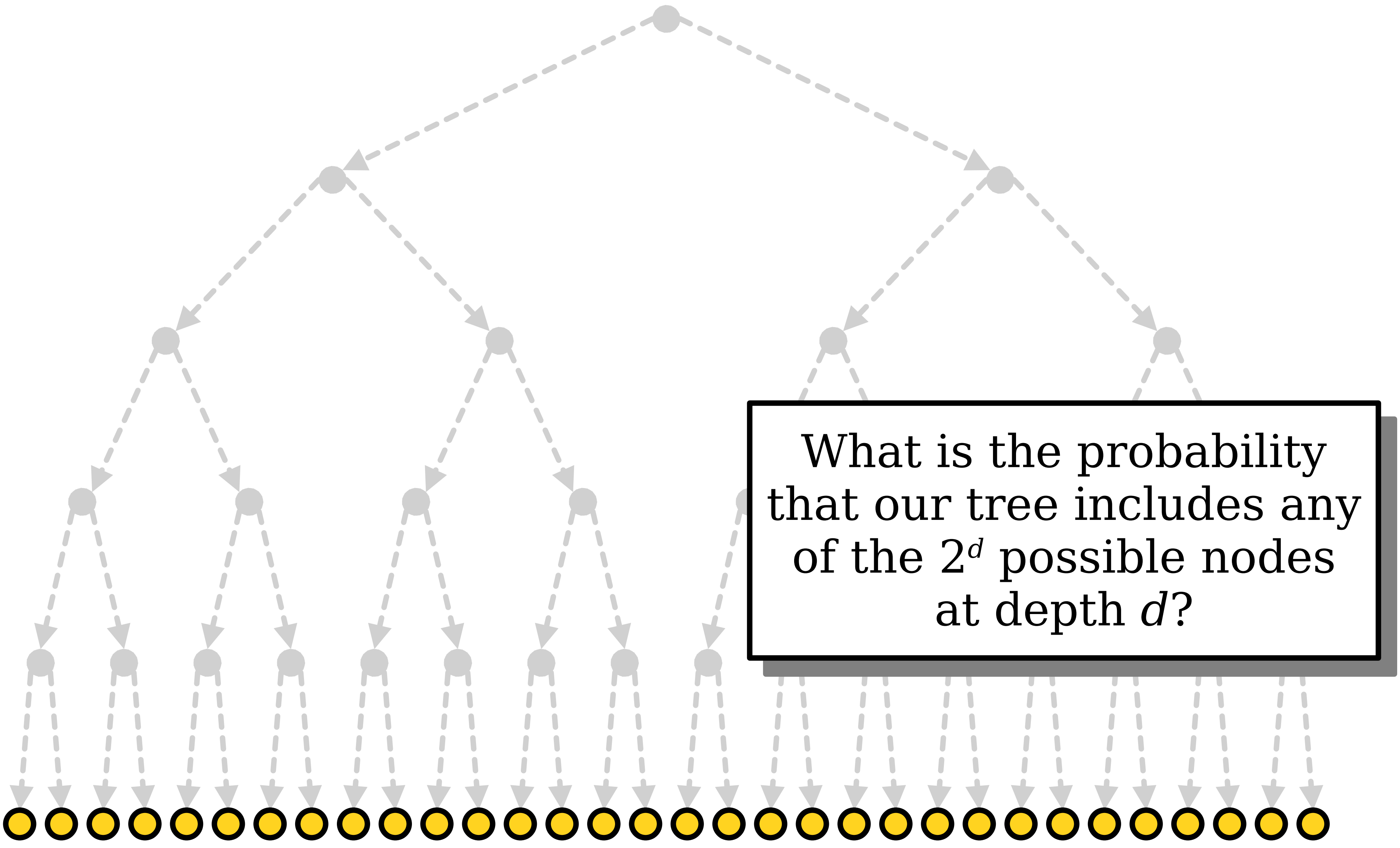
$$\lim_{n \to \infty} \Pr\left[\text{🌲}_n \geq c \lg n\right] = 0.$$

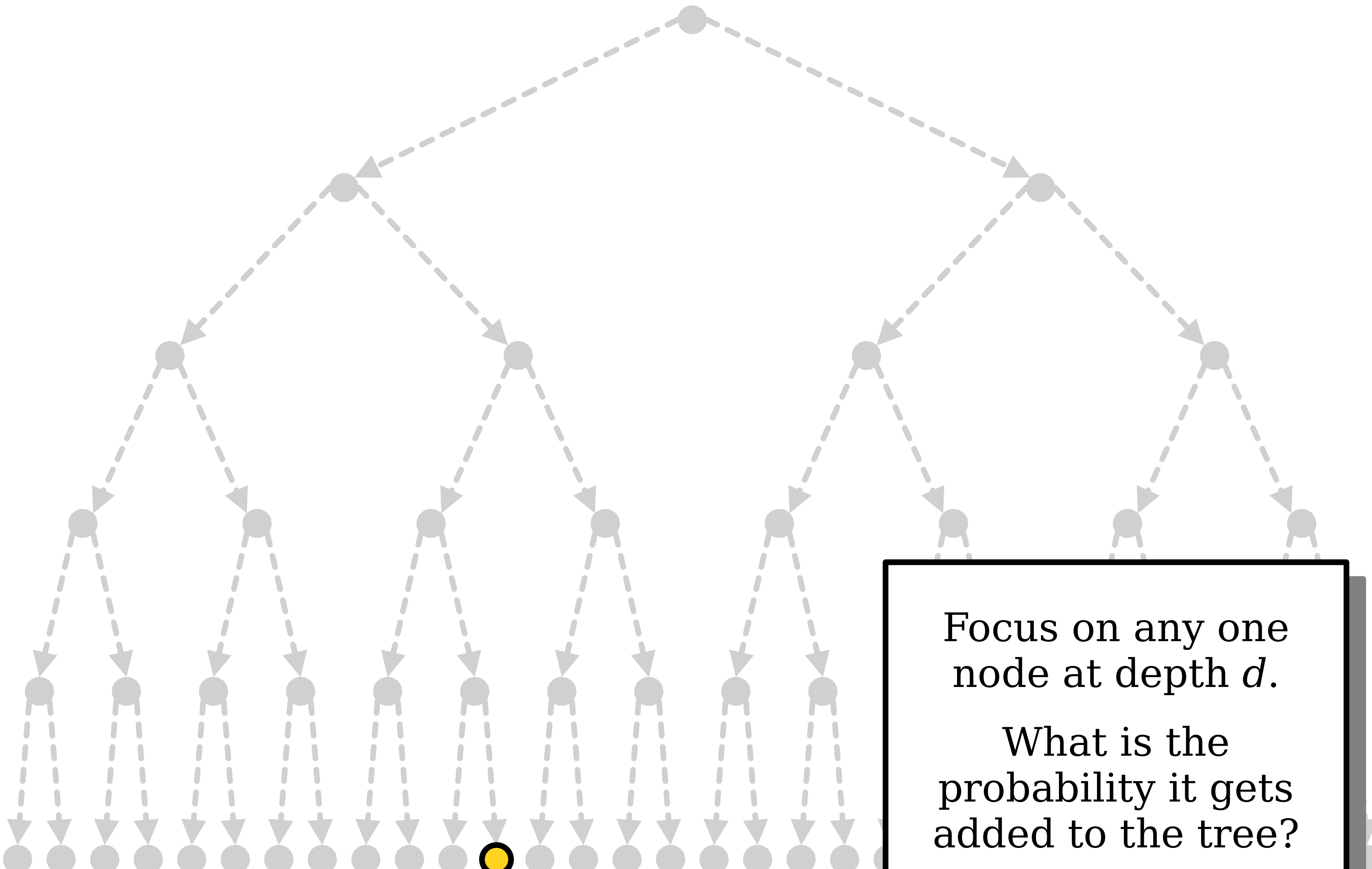
How do we reason about the height of a randomly-built binary search tree?



Suppose our tree has n nodes.

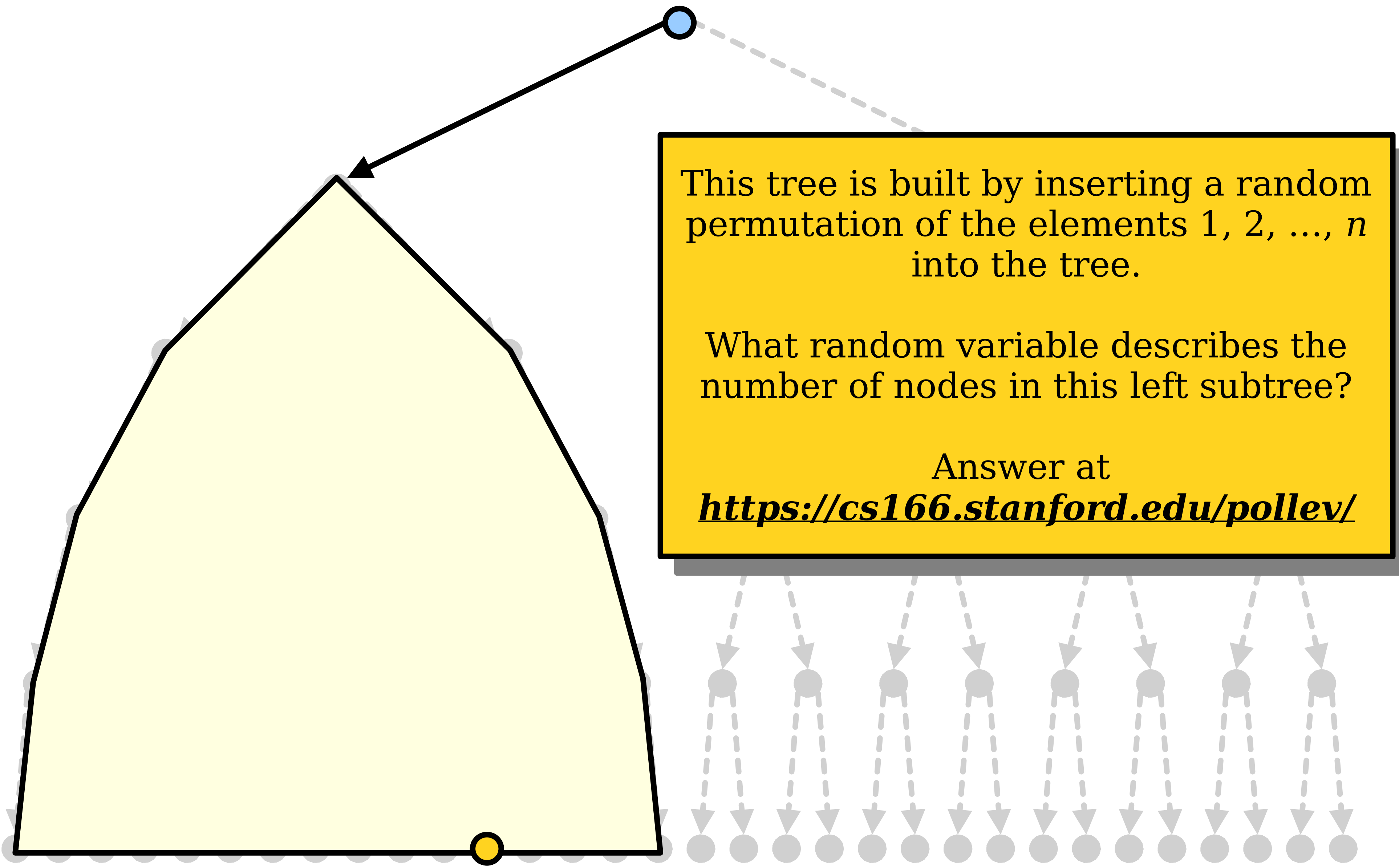
Its height is at least d if it includes at least one node at depth d.


What is the probability
that our tree includes any
of the 2^d possible nodes
at depth d?



Focus on any one node at depth *d*.

What is the probability it gets added to the tree?

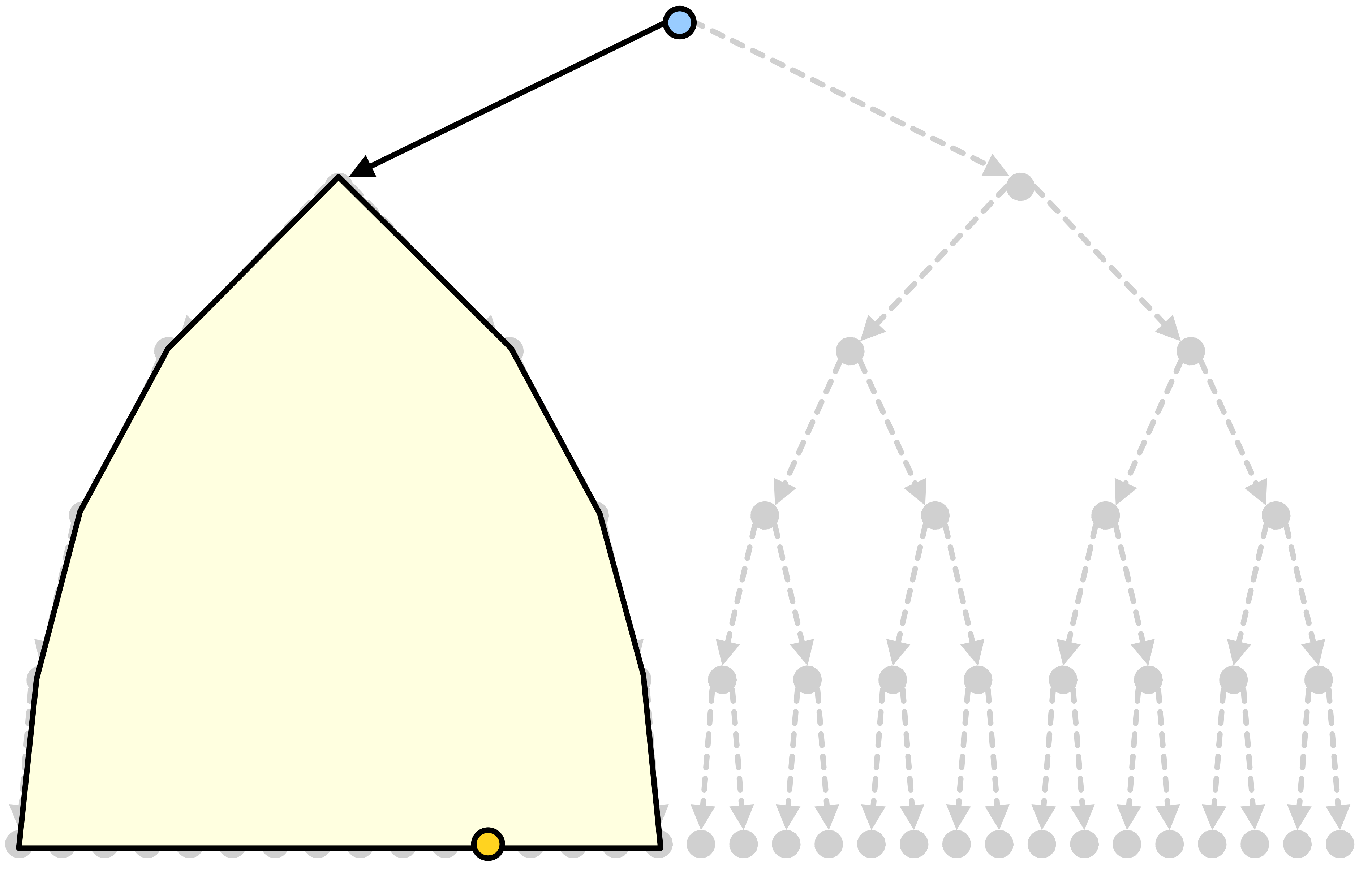


This tree is built by inserting a random permutation of the elements 1, 2, …, n into the tree.

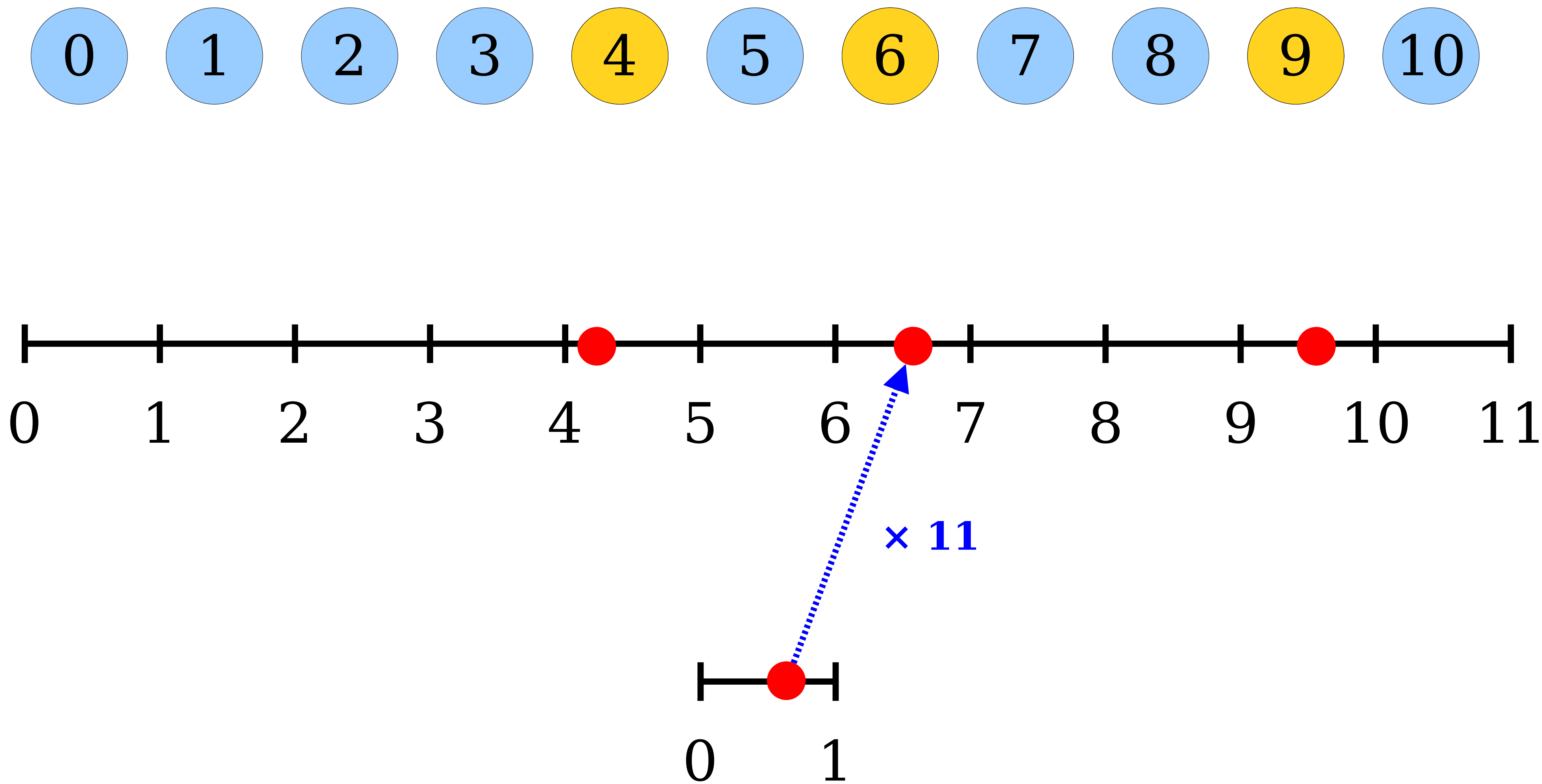
What random variable describes the number of nodes in this left subtree?

Answer at
https://cs166.stanford.edu/pollev/

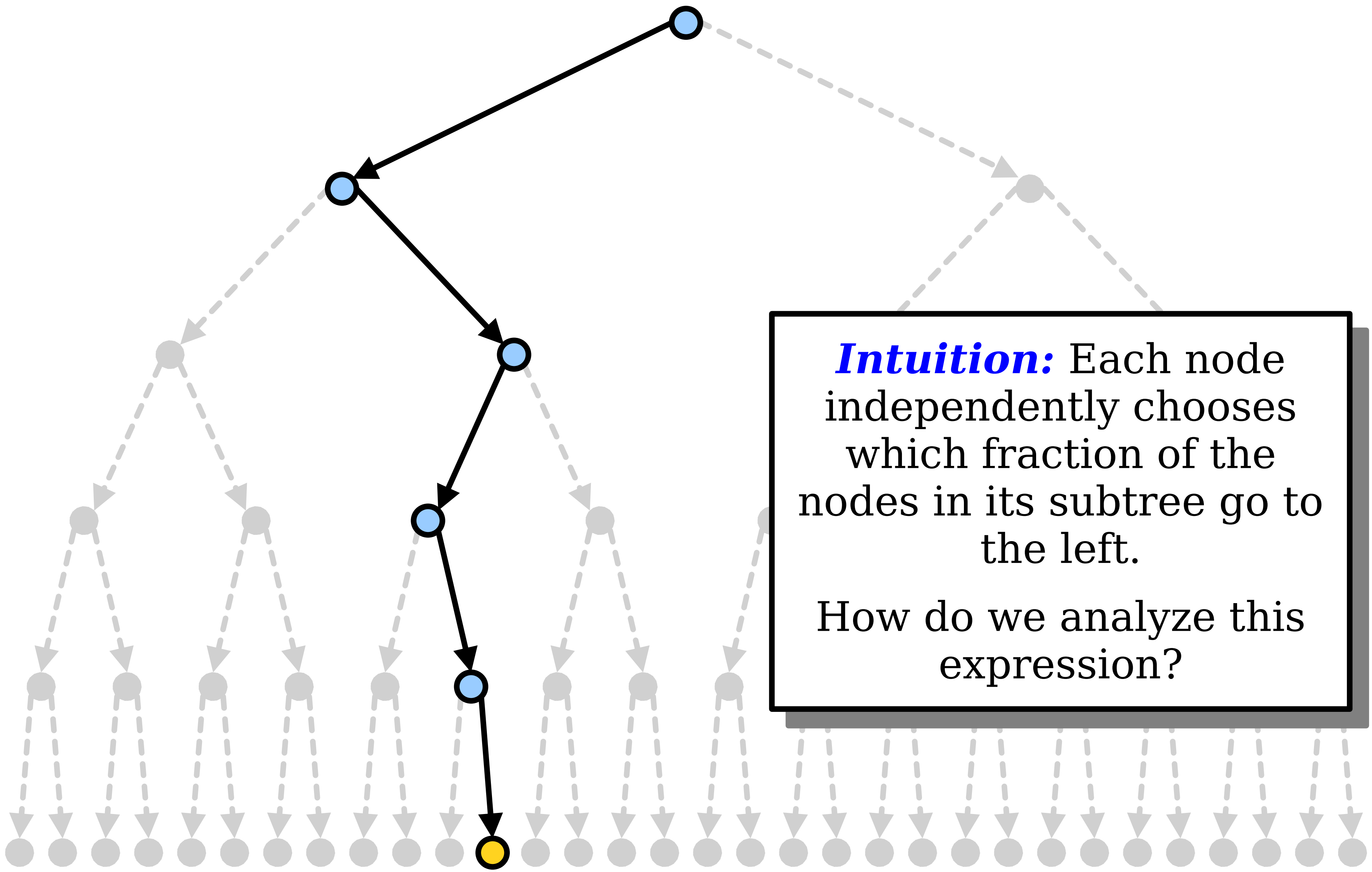
Nodes in this subtree: n



Nodes in this subtree: DiscreteUniform$\{0, n - 1\}$



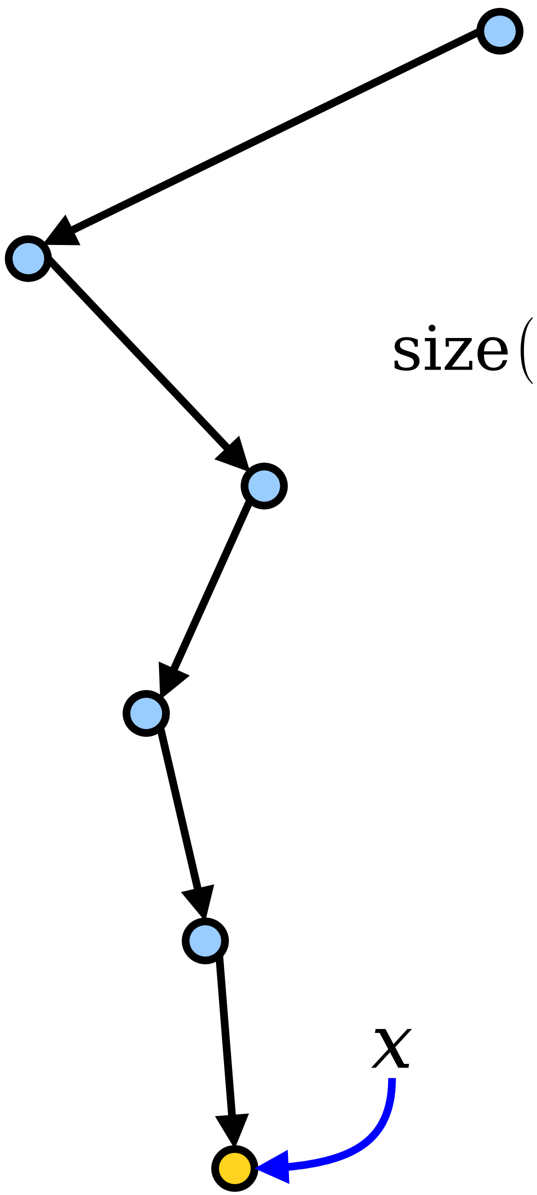
$$k \sim \lfloor \mathrm{U}(0, 1) \cdot n \rfloor$$



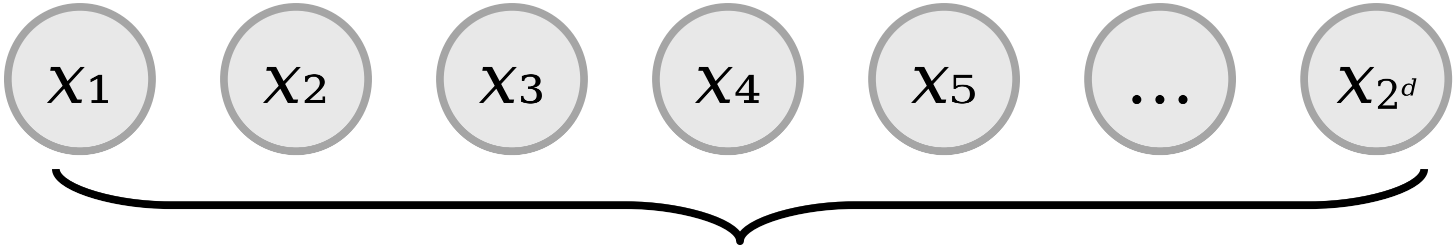
Intuition: Each node independently chooses which fraction of the nodes in its subtree go to the left.

How do we analyze this expression?

Nodes: $\lfloor U(0,1) \cdot \lfloor U(0,1) \cdot \lfloor U(0,1) \cdot \lfloor U(0,1) \cdot \lfloor U(0,1) \cdot n \rfloor \rfloor \rfloor \rfloor \rfloor$



$$
\begin{aligned}
\mathrm{size}(x) &\sim \lfloor \mathrm{Uniform}(0,1) \cdot \ldots \cdot \lfloor \mathrm{Uniform}(0,1) \cdot n \rfloor \ldots \rfloor \\
&\le \mathrm{Uniform}(0,1) \cdot \ldots \cdot \mathrm{Uniform}(0,1) \cdot n \\
&= n \cdot \prod_{i=1}^{\mathrm{depth}(x)} \mathrm{Uniform}(0,1)
\end{aligned}
$$



2^d Possible Nodes

$$
\begin{aligned}
\Pr\left[\text{🌲}_n \geq d\right] &= \Pr\left[\text{tree includes a node at depth } d\right] \\
&= \Pr\left[\bigcup_{i=1}^{2^d} x_i \text{ included in tree}\right] \\
&\leq \sum_{i=1}^{2^d} \Pr\left[x_i \text{ included in tree}\right]
\end{aligned}
$$

The ***union bound***:

$\Pr[\cup \mathcal{E}_i] \leq \sum \Pr[\mathcal{E}_i]$.

$$\underbrace{x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad \ldots \quad x_{2^d}}_{2^d \text{ Possible Nodes}}$$

$$
\begin{aligned}
\Pr\left[🌲_n \;\geq\; d\right] \;&\leq\; \sum_{i=1}^{2^d} \Pr\left[x_i \text{ included in tree}\right] \\
&=\; \sum_{i=1}^{2^d} \Pr\left[\text{subtree at } x_i \text{ has at least one node}\right] \\
&\leq\; \sum_{i=1}^{2^d} \Pr\left[n \cdot \prod_{j=1}^{d} \text{Uniform}\left(0\ ,\ 1\right) \geq 1\right] \\
&=\; 2^d \cdot \Pr\left[n \cdot \prod_{j=1}^{d} \text{Uniform}\left(0\ ,\ 1\right) \geq 1\right]
\end{aligned}
$$

$$\Pr\left[🌲_n \geq d\right] \leq 2^d \cdot \Pr\left[n \cdot \prod_{i=1}^{d} \mathrm{Uniform}(0,1) \geq 1\right]$$

$$\leq 2^d \cdot \min_{t>0} \mathrm{E}\left[\left(n \cdot \prod_{i=1}^{d} \mathrm{Uniform}(0,1)\right)^t\right]$$

Markov's inequality:

$$\Pr[X \geq a] \leq \min_{t>0} \frac{\mathrm{E}[X^t]}{a^t}$$

$$
\begin{aligned}
\Pr\left[🌲_n \geq d\right] &\leq 2^d \cdot \Pr\left[n \cdot \prod_{i=1}^{d} \mathrm{Uniform}\left(0\ ,\ 1\right) \geq 1\right] \\
&\leq 2^d \cdot \min_{t>0} \mathrm{E}\left[\left(n \cdot \prod_{i=1}^{d} \mathrm{Uniform}\left(0\ ,\ 1\right)\right)^t\right] \\
&= 2^d \cdot \min_{t>0} \cdot \mathrm{E}\left[n^t \cdot \prod_{i=1}^{d} \mathrm{Uniform}\left(0\ ,\ 1\right)^t\right]
\end{aligned}
$$

$$
\begin{aligned}
\Pr\left[\text{🌲}_n \geq d\right] &\leq 2^d \cdot \Pr\left[n \cdot \prod_{i=1}^{d} \mathrm{Uniform}(0,1) \geq 1\right] \\
&\leq 2^d \cdot \min_{t>0} \mathrm{E}\left[\left(n \cdot \prod_{i=1}^{d} \mathrm{Uniform}(0,1)\right)^t\right] \\
&= 2^d \cdot \min_{t>0} \cdot \mathrm{E}\left[n^t \cdot \prod_{i=1}^{d} \mathrm{Uniform}(0,1)^t\right] \\
&= 2^d \cdot \min_{t>0} n^t \cdot \mathrm{E}\left[\prod_{i=1}^{d} \mathrm{Uniform}(0,1)^t\right]
\end{aligned}
$$

$$\begin{aligned}
\Pr\left[🌲_n \;\geq\; d\right] &\leq 2^d \cdot \Pr\left[n \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\;,\;1) \;\geq\; 1\right] \\
&\leq 2^d \cdot \min_{t>0} \mathrm{E}\left[\left(n \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\;,\;1)\right)^t\right] \\
&= 2^d \cdot \min_{t>0} \cdot \mathrm{E}\left[n^t \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\;,\;1)^t\right] \\
&= 2^d \cdot \min_{t>0} n^t \cdot \mathrm{E}\left[\prod_{i=1}^{d} \mathrm{Uniform}\,(0\;,\;1)^t\right] \\
&= 2^d \cdot \min_{t>0} n^t \cdot \prod_{i=1}^{d} \mathrm{E}\left[\mathrm{Uniform}\,(0\;,\;1)^t\right]
\end{aligned}$$

If X and Y are uncorrelated, then $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$.

$$
\begin{aligned}
\Pr\left[\text{🌲}_n \;\geq\; d\right] &\leq 2^d \cdot \Pr\left[n \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\,,\,1) \;\geq\; 1\right] \\
&\leq 2^d \cdot \min_{t>0}\, \mathrm{E}\left[\left(n \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\,,\,1)\right)^t\right] \\
&= 2^d \cdot \min_{t>0} \cdot \mathrm{E}\left[n^t \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\,,\,1)^t\right] \\
&= 2^d \cdot \min_{t>0}\, n^t \cdot \mathrm{E}\left[\prod_{i=1}^{d} \mathrm{Uniform}\,(0\,,\,1)^t\right] \\
&= 2^d \cdot \min_{t>0}\, n^t \cdot \prod_{i=1}^{d} \mathrm{E}\left[\mathrm{Uniform}\,(0\,,\,1)^t\right] \\
&= 2^d \cdot \min_{t>0}\, n^t \cdot \prod_{i=1}^{d} \frac{1}{t+1}
\end{aligned}
$$

$$
\begin{aligned}
\Pr\left[🌲_n \;\geq\; d\right] \;&\leq\; 2^d \cdot \Pr\left[n \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\;,\;1) \;\geq\; 1\right] \\
&\leq\; 2^d \cdot \min_{t>0} \mathrm{E}\left[\left(n \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\;,\;1)\right)^t\right] \\
&=\; 2^d \cdot \min_{t>0} \cdot \mathrm{E}\left[n^t \cdot \prod_{i=1}^{d} \mathrm{Uniform}\,(0\;,\;1)^t\right] \\
&=\; 2^d \cdot \min_{t>0} n^t \cdot \mathrm{E}\left[\prod_{i=1}^{d} \mathrm{Uniform}\,(0\;,\;1)^t\right] \\
&=\; 2^d \cdot \min_{t>0} n^t \cdot \prod_{i=1}^{d} \mathrm{E}\left[\mathrm{Uniform}\,(0\;,\;1)^t\right] \\
&=\; 2^d \cdot \min_{t>0} n^t \cdot \prod_{i=1}^{d} \frac{1}{t+1} \\
&=\; \min_{t>0} n^t \cdot \left(\frac{2}{t+1}\right)^d
\end{aligned}
$$

$$\Pr\left[\,\text{🌲}_n \;\geq\; d\,\right] \;\leq\; \min_{t>0}\, n^t \cdot \left(\frac{2}{t+1}\right)^d$$

$$=\; n^{\frac{d}{\ln n}-1}\left(\frac{2\ln n}{d}\right)^d$$

Simple calculus exercise:
show that this expression
is minimized when

$$t+1 \;=\; \frac{d}{\ln n}\,.$$

$$\Pr\left[\text{🌲}_n \geq d\right] \leq n^{\frac{d}{\ln n}-1}\left(\frac{2\ln n}{d}\right)^d$$

We know from intuition and experiment that the expected depth should be something logarithmic.

$$\Pr\left[\text{🌲}_n \geq \boldsymbol{c \ln n}\right] \leq n^{\frac{\boldsymbol{c \ln n}}{\ln n} - 1}\left(\frac{2\ln n}{\boldsymbol{c \ln n}}\right)^{\boldsymbol{c \ln n}}$$

We know from intuition and experiment that the expected depth should be something logarithmic.

$$\Pr\left[🌲_n \;\geq\; c \ln n\right] \;\leq\; n^{\frac{c \ln n}{\ln n} - 1} \left(\frac{2 \ln n}{c \ln n}\right)^{c \ln n}$$

$$= \boxed{n^{c-1} \left(\frac{2}{c}\right)^{c \ln n}}$$

$$= \left(e^{\ln n}\right)^{c-1} \left(\frac{2}{c}\right)^{c \ln n}$$

$$= \left(e^{c-1}\right)^{\ln n} \left(\left(\frac{2}{c}\right)^{c}\right)^{\ln n}$$

$$= \left(e^{c-1} \left(\frac{2}{c}\right)^{c}\right)^{\ln n}$$

$$= \boxed{\left(\frac{1}{e} \left(\frac{2\,e}{c}\right)^{c}\right)^{\ln n}}$$

Vigorously whack with our Algebra Hammer!

$$\Pr\left[\text{🌲}_n \geq c \ln n\right] \leq \boxed{\left(\frac{1}{e}\left(\frac{2e}{c}\right)^c\right)}^{\ln n}$$

If this term is less than one, this probability drops to zero as n gets larger and larger.

📈 ***How does this blue term behave?***

Let $f(x) = \frac{1}{e}\left(\frac{2e}{x}\right)^x$ and let γ' be the unique solution to $f(\gamma') = 1$, subject to $\gamma' > 1$.

$$\gamma' \approx 4.31107040700100503 50\ldots$$

Then for any $c > \gamma'$, we have

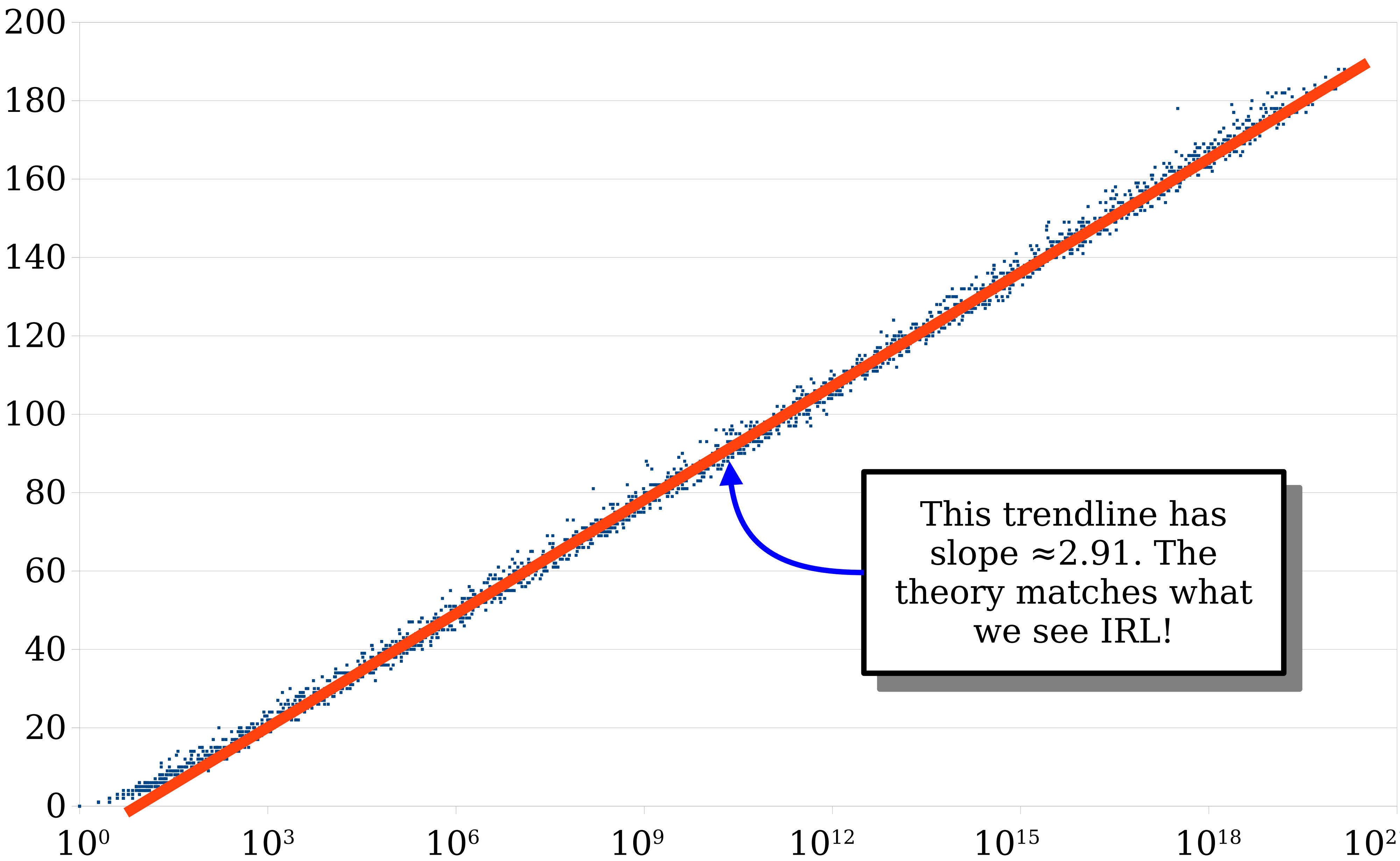
$$\lim_{n \to \infty} \Pr\left[\text{🌲}_n \;\geq\; c \ln n\right] \;=\; 0.$$

Let $\gamma = \gamma' \ln 2$.

$$\gamma \approx 2.98820629780816 25\ldots$$

Then for any $c > \gamma$, we have

$$\lim_{n \to \infty} \Pr\left[\text{🌲}_n \;\geq\; c \lg n\right] \;=\; 0.$$


200
180
160
140
120
100
80
60
40
20
0
10^0
10^3
10^6
10^9
10^{12}
10^{15}
10^{18}
10^{21}
This trendline has slope ≈2.91. The theory matches what we see IRL!

Practical Consequences

- You can make nonrandom data look random!
  - The ***treap*** data structure associates each key with a random value, then chooses its shape as if keys were inserted in sorted order by keys.
  - The ***zip-zip tree*** (2022) uses the same basic idea as a treap but with needs fewer random bits.
  - The ***skiplist*** is widely used as an alternative to BSTs and works by assigning each key a geometrically-random variable determining the structure shape.
- These are all worth a read.

Further Theory Results

- Later work (Reed, 2000) proved that there are constants γ and δ where

  $$\mathrm{E}[\text{🌲}_n] = \gamma \lg n - \delta \lg \lg n + \mathrm{O}(1)$$

  and that

  $$\mathrm{Var}[\text{🌲}_n] = \mathrm{O}(1),$$

  meaning random tree heights are indeed tightly concentrated around their expectation.
- Robson (1978, published 1995) devised the algorithm I used earlier to sample heights of random n-node BSTs in expected time $\mathrm{O}(\log^4 n)$.

Next Time

- ***Amortized Analysis***
  - A little accounting trickery never hurt anyone, right?
- ***The Potential Method***
  - Carbon credits, Theoryland edition.
- ***Three Motivating Applications***
  - Data structures that are faster than they look.