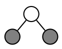
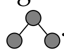
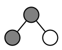
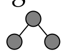
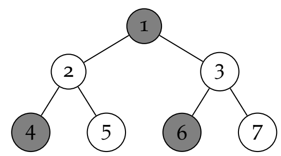
Marking Trees

° 2016-10-07, rev. 8c220f7

Bob maintains a complete binary tree of height $h > 1$ with $N = 2^h - 1$ nodes called $1, \ldots, N$. Nodes are in one of two states: marked or unmarked; initially, all nodes are unmarked. Each round Alice sends an integer $i \in \{1, \ldots, N\}$ according to a random process specified below. Bob marks node i in his tree. Bob then uses the following marking rules repeatedly:

1. An internal node gets marked when both its children are marked, so  becomes .
2. A non-root node gets marked if its parent and sibling are marked, so  becomes  and similarly for a left child.

Bob applies these rules as often as possible, often leading to a cascade of markings. When no more rules apply, this round ends. For instance, if Alice sends "5" in the figure to the right then the entire tree gets marked in this round.



The whole process ends when every node in Bob's tree is marked. We study 3 random processes:

1. Each round, Alice sends the name of a node independently and uniformly at random from $\{1, \ldots, N\}$.
2. Each round, Alice sends the name of a node uniformly at random *from those she has not sent before*. Another way of saying this is that Alice starts by picking a random permutation π on $\{1, \ldots, N\}$ and then sends $\pi(i)$ in round i.
3. Each round, Alice sends the name of a node uniformly at random *from those that are not yet marked*. (Alice can see Bob's tree.)

Hints: For R_2 and R_3, you may be tempted to pick random numbers uniformly from $\{1, \ldots, N\}$ each round and and throw away those already sent. That takes too much time. Instead, for R_2 start with a random permutation using the "Knuth shuffle". Look it up.

To understand what is going on, maybe it helps you to see which nodes Alice sent. (In particular, what is the last node she sends before the process stops?)

Lab Report: Marking Trees

by Alice Cooper and Bob Marley[1]

Results

For $i \in \{1, 2, 3\}$, the number of rounds R_i spent until the tree is completely marked in process i is given in the following table. The table shows the result of $[\cdots]$ repeated trails.[2]

N	R_1	R_2	R_3
3	2.6 ± 0.9		
7			
15			
31			
63			
127			
255			
511			
1023	$3.2 \pm 0.5 \times 10^3$		
$\vdots$			
524 287	$3.2 \pm 0.2 \times 10^6$		
1 048 575			

Analysis

Our experimental data indicates that $\mathbf{E}[R_1]$ is [...], while $\mathbf{E}[R_2]$ $[\cdots]$, and $\mathbf{E}[R_3]$ $[\cdots]$.[3]

Theoretically, the behaviour of R_1 can be explained as follows: $[\cdots]$ [4]

[1] Complete the report by filling in your names and the parts marked [...]. Remove the sidenotes in your final hand-in.

[2] Report your empirical data. Give each value as the mean plus/minus one standard deviation. Use whatever best practices for reporting data you may have learned; here's a crash course that suffices for our purposes: (i) Calculate mean and standard deviation ($m = 2.5074$, $s = 0.889341021813$) from a number of repeated experiments. (ii) Round s to one significant digit ($s = 0.9$). (iii) Round m to the decimal place corresponding to the first significant digit in s ($m = 2.5$, $s = 0.9$). (iv) Report $m \pm s$ (2.5 ± 0.9). (v) Use scientific notation.

If you've taken too many statistics classes, feel free to go to town with graphs and confidence intervals and so forth.

[3] For each of the tree processes, try to express the observed behaviour of R_i using standard terminology from the analys of algorithms. For instance, use expressions such as "$E[R_1]$ is logarithmic in N" or "$E[R_2]$ is somewhere between $\Omega(N^{1/2})$ and $\Omega(N^{3/2})$".

[4] This is the difficult part. You need to write a few lines that explain the random process underlying R_1 and derive an expression for $\mathbf{E}[R_1]$. (Hopefully it's the same as your empirical analysis!) Once you recognize what's going on, this should be easy; it involves no complicated calculations.

Hint: If you're stuck at R_1, do the following experiment as a warmup. Process 0 is like process 1, except that Bob doesn't use his marking rules: the only nodes that get marked are those sent by Alice. Implement this (it's easy) and analyse the behaviour both theoretically and practically.

Optional: Explain the behaviours of R_2 and R_3 as well. The behaviour of R_2 is quite a bit harder; while R_3 is just cute.

Perspective

This lab establishes minimal skills in simulation of random processes, introduces the Knuth shuffle for those who haven't seen it, and some basic probability theory about occupancy problems. In process 2, Alice's messages are not independent, which can lead to temping errors in the analysis. Process 3 is just surprising (and maybe fun to implement), but not much about randomness is to be learned from it.

The exercise is built on an assigment of Michael Mitzenmacher.[5]

The front page image shows *Acinonyx jubatus* marking a tree. Photo by Joachim Huber, under the Creative Commons Attribution-Share Alike 2.0 Generic license.[6]

[5] Michael Mitzenmacher, An Experimental Assignment on Random Processes, SIGACT News, 27 December 2000. See also section 5.8 in M. Mitzenmacher and E. Upfal, Probability and Computing – Randomized Algorithms and Probabilistic Analysis. Cambridge University Press, 2005.

[6] commons.wikimedia.org/wiki/File:-Acinonyx_jubatus_-Southern_Namibia-8.jpg