Coalescing Particles on an Interval

BY MICHAEL LARSEN AND RUSSELL LYONS

Abstract. At time 0, we begin with a particle at each integer in $[0, n]$. At each positive integer time, one of the particles remaining in $[1, n]$ is chosen at random and moved one to the left, coalescing with any particle that might already be there. How long does it take until all particles coalesce (at 0)?

Key words. Coalescence, random walk, asymptotics, parallel processing.

Schabanel (1996) presents a study of self-organizing balanced binary search trees. The simplest case is that of a linear tree (a chained list); here, rotations are not allowed. The root is at 0 and the leaf is at n. Each node of the tree has a rank, which is its distance from the leaf. However, the ranks are initially unknown. Each node initially estimates its rank as 0. At each time, we choose at random one of the nodes that has not updated the estimate of its parent since the most recent time it has itself been updated; this random node then updates the estimate of its parent by giving it the estimate of one more than the current estimate of its own rank. Eventually, all nodes have the correct estimate of their own rank, which is indicated by the agreement of each node's estimate with that of its parent.

A simple way to study this process was given by Schabanel (1996): Begin with a particle at each integer in $[0, n]$. The presence of a particle at a node indicates that it has not updated its parent since the last time that it has itself been updated. Thus, at each positive integer time, one of the particles remaining in $[1, n]$ is chosen at random and moved one to the left, coalescing with any particle that might already be there. How long does it take until all particles coalesce (at 0)? This is the time at which all nodes have the correct estimate of their own rank.

Let T_n be the time until all particles coalesce. As usual, we say that $f(n)$ is asymptotic to $g(n)$ and write $f(n) \sim g(n)$ if $f(n)/g(n) \to 1$ as $n \to \infty$. Schabanel (1996) estimated $\mathbf{E}[T_n]$ to be about $0.75n^{3/2}$ through simulation. This is indeed the correct order:

Theorem. *With the above notation,*

$$\mathbf{E}[T_n] = \frac{2n(2n+1)}{3}\binom{2n}{n}\frac{1}{2^{2n}} \sim \frac{4n^{3/2}}{3\sqrt{\pi}} = 0.752^{+}n^{3/2} \tag{1}$$

and

$$\mathrm{Var}(T_n) \le \frac{(8+o(1))n^{5/2}}{15\sqrt{\pi}} . \tag{2}$$

REMARK. We conjecture that the variance of T_n is $\sim Cn^{5/2}$ for some constant $C \le 8/(15\sqrt{\pi})$. Presumably, T_n obeys a central limit theorem.

Proof. Label a particle according to its starting position. Let τ_k be the number of moves that the particle starting at k makes until it coalesces with the particle starting at $k-1$. Then

$$T_n = \sum_{k=1}^{n} \tau_k .$$

More generally, define $\tau(u,v)$ to be the number of moves of particle u until it coalesces with particle v for $u \ge v$. Define the generating function

$$G(\lambda,x,y) := \sum_{i,j=0}^{\infty} c(\lambda,i,j)x^i y^j ,$$

where

$$c(\lambda,i,j) := \mathbf{E}\big[e^{\lambda\tau(j+i,j)}\big] .$$

Note that $c(\lambda,i,j)$ is the same for all $n \ge j+i$, so that, during our calculations of the generating function, we may treat n as infinity. Also, note that $c(\lambda,i,j) \le e^{\lambda(j+i)}$, whence $G(\lambda,x,y)$ converges for $|x|,|y| < e^{-\lambda}$. Clearly, $c(\lambda,0,j) = 1$ for all $j \ge 0$ and $c(\lambda,i,0) = e^{\lambda i}$ for all $i \ge 0$. Also, consideration of which of the particles at $j+i$ and j moves before the other shows that

$$c(\lambda,i,j) = \big(e^{\lambda}c(\lambda,i-1,j) + c(\lambda,i+1,j-1)\big)/2$$

for $i,j \ge 1$. Therefore,

$$(2x - e^{\lambda}x^2 - y)G(\lambda,x,y) = \frac{2x-y}{1-y} + \frac{e^{\lambda}x^2}{1-e^{\lambda}x} - xyH(\lambda,y) , \tag{3}$$

where

$$H(\lambda,y) := \sum_{j\ge 0} c(\lambda,1,j)y^j .$$

Since the left-hand side of (3) vanishes where $2x - e^{\lambda}x^2 - y = 0$, i.e., where

$$x = e^{-\lambda}\left(1 - \sqrt{1-e^{\lambda}y}\right) ,$$

so does the right-hand side. This yields the equation

$$H(\lambda, y) = \frac{1-\sqrt{1-e^{\lambda}y}}{y}\left(\frac{1}{1-y}+\frac{1}{\sqrt{1-e^{\lambda}y}}\right).$$

In particular, we find that

$$\sum_{j=0}^{\infty} \mathbf{E}[\tau_j]y^j = \frac{\partial H}{\partial \lambda}(0,y) = (1-y)^{-3/2}$$

and

$$\sum_{j=0}^{\infty} \mathbf{E}[\tau_j^2]y^j = \frac{\partial^2 H}{\partial \lambda^2}(0,y) = (1-y)^{-5/2}.$$

Hence,

$$\mathbf{E}[\tau_k] = (-1)^k\binom{-3/2}{k} = \frac{2k+1}{2^{2k}}\binom{2k}{k}. \tag{4}$$

Also, $\mathbf{E}[T_n]$ is the coefficient of y^{n-1} in $(\partial H/\partial\lambda)(0,y)/(1-y) = (1-y)^{-5/2}$. This gives (1). Likewise, $\sum_{k=1}^{n} \mathbf{E}[\tau_k^2]$ is the coefficient of y^{n-1} in $(\partial^2 H/\partial\lambda^2)(0,y)/(1-y) = (1-y)^{-7/2}$, i.e.,

$$\sum_{k=1}^{n} \mathbf{E}[\tau_k^2] = (-1)^{n-1}\binom{-7/2}{n-1} = \frac{(2n+3)(2n+2)(2n+1)}{15}\binom{2n}{n+1}\frac{1}{2^{2n}} \sim \frac{8n^{5/2}}{15\sqrt{\pi}}.$$

In conjunction with (4), this will give (2) as soon as we show that the random variables τ_k are negatively correlated.

To complete the proof, we need to show that for $a > b$, we have $\mathbf{E}[\tau_a\tau_b] \le \mathbf{E}[\tau_a]\mathbf{E}[\tau_b]$. We shall show, more generally, that for $x \ge w \ge y \ge z \ge 0$,

$$\mathbf{E}[\tau(x,w)\tau(y,z)] \le \mathbf{E}\tau(x,w)\mathbf{E}\tau(y,z). \tag{5}$$

To prove (5), we may ignore all particles other than x, w, y, and z. Let $\mathbf{E}^{(k)}$ be expectation with respect to the probability measure that lets the two pairs of particles ((x,w) and (y,z)) evolve independently after the first k moves in total. Note: we consider two pairs of particles after time k, regardless of which ones have coalesced. For example, even if $k=0$ and $w=y$, we consider separate and independent pairs (x,w) and (y,z). Since $\mathbf{E}^{(\infty)} = \mathbf{E}$, the equation (5) will follow from

$$\mathbf{E}\tau(x,w)\mathbf{E}\tau(y,z) = \mathbf{E}^{(0)}\tau(x,w)\tau(y,z) \ge \mathbf{E}^{(1)}\tau(x,w)\tau(y,z) \ge \mathbf{E}^{(2)}\tau(x,w)\tau(y,z) \ge \cdots. \tag{6}$$

To prove (6), let

$$F_k(x,w,y,z) := \mathrm{Cov}^{(k)}\Big(\tau(x,w),\tau(y,z)\Big) := \mathbf{E}^{(k)}\tau(x,w)\tau(y,z) - \mathbf{E}\tau(x,w)\mathbf{E}\tau(y,z)\,.$$

Then we want to show that $F_k \ge F_{k+1}$ for all $k \ge 0$. We shall first show that $F_0 \ge F_1$.

Now we have $F_0(x,w,y,z) = F_1(x,w,y,z)$ if $x = w$, $w > y$, $y = z$, or $z = 0$. So assume that $x > w = y > z > 0$. Considering the first move, we see that

$$\mathbf{E}\tau(x,y) = (1/2)\big(1 + \mathbf{E}\tau(x-1,y) + \mathbf{E}\tau(x,y-1)\big)$$

and likewise

$$\mathbf{E}\tau(y,z) = (1/2)\big(1 + \mathbf{E}\tau(y-1,z) + \mathbf{E}\tau(y,z-1)\big)\,.$$

We also have

$$\begin{aligned}\mathbf{E}^{(1)}[\tau(x,y)\tau(y,z)] = (1/3)\Big\{&\big(1+\mathbf{E}\tau(x-1,y)\big)\mathbf{E}\tau(y,z) + \\ &+ \mathbf{E}\tau(x,y-1)\big(1+\mathbf{E}\tau(y-1,z)\big) + \mathbf{E}\tau(x,y)\mathbf{E}\tau(y,z-1)\Big\}\,.\end{aligned}$$

Substitution, expansion, factoring, and resubstitution leads to

$$\begin{aligned}&\mathbf{E}\tau(x,y)\mathbf{E}\tau(y,z) - \mathbf{E}^{(1)}[\tau(x,y)\tau(y,z)] \\ &\quad = \frac{1}{12}\Big\{\mathbf{E}\tau(x,y-1) - 1 - \mathbf{E}\tau(x-1,y)\Big\}\Big\{\mathbf{E}\tau(y,z-1) - 1 - \mathbf{E}\tau(y-1,z)\Big\} \\ &\quad = \frac{1}{3}\Big\{\mathbf{E}\tau(x,y-1) - \mathbf{E}\tau(x,y)\Big\}\Big\{\mathbf{E}\tau(y,z-1) - \mathbf{E}\tau(y,z)\Big\}\,,\end{aligned}$$

which is clearly positive. Thus, $0 = F_0 \ge F_1$.

We shall now deduce that $F_k \ge F_{k+1}$ for $k \ge 1$. Let $\mathcal{F}_k$ be the σ-field generated by the first k moves. Then for any j, we have

$$\begin{aligned}\mathrm{Cov}^{(j)}\Big(\tau(x,w),\tau(y,z)\Big) = \mathrm{Cov}^{(j)}\Big(&\mathbf{E}^{(j)}\big[\tau(x,w)\mid\mathcal{F}_k\big],\ \mathbf{E}^{(j)}\big[\tau(y,z)\mid\mathcal{F}_k\big]\Big) + \\ &+ \mathbf{E}^{(j)}\mathrm{Cov}^{(j)}\Big(\tau(x,w),\tau(y,z)\ \Big|\ \mathcal{F}_k\Big)\,.\end{aligned} \tag{7}$$

Now $\mathbf{E}^{(j)}\big[\tau(x,w)\mid\mathcal{F}_k\big]$ is the same for all $j \ge k$. Also, given any $\mathcal{F}_k$-measurable random variables X, Y, we have that $\mathrm{Cov}^{(j)}(X,Y)$ is the same for all $j \ge k$. Therefore, the first term on the right-hand side of (7) is the same for all $j \ge k$. In addition,

$$\mathrm{Cov}^{(k)}\Big(\tau(x,w),\tau(y,z)\ \Big|\ \mathcal{F}_k\Big) = 0$$

since the particles move independently after k moves together; while

$$\mathrm{Cov}^{(k+1)}\Big(\tau(x,w),\tau(y,z)\ \Big|\ \mathcal{F}_k\Big) \le 0$$

since $F_1 \le 0$. Putting these together, we obtain $F_k \ge F_{k+1}$, as desired. ∎

Acknowledgement. We are grateful to Claire Kenyon for bringing this problem to our attention and to Harry Kesten for the representation of T_n as $\sum \tau_k$. We thank the referee for clarifying our proof of negative correlation. We are indebted to Robin Pemantle for permission to include his ideas on the generating function that give the exact expression for $\mathbf{E}T_n$, rather than our original proof, which gave only the asymptotics.

REFERENCE

SCHABANEL, N. (1996). *Equilibrage AVL distribué d'arbres binaires de recherche.* DEA thesis, Ecole Normale de Lyon.

Michael Larsen, Department of Mathematics, University of Missouri-Columbia, Columbia, MO 65211 `larsen@math.missouri.edu`

Russell Lyons, Department of Mathematics, Indiana University, Bloomington, IN 47405-5701 `rdlyons@indiana.edu` `http://php.indiana.edu/~rdlyons/`