Generating functions

Let X be an rv distributed on $\mathbb{N}_{\geq 0}$ with distribution $\mu = (\mu_0, \mu_1, \ldots)$.

Definition 7. *The* probability generating function (pgf) *of X (or of μ) is $E(z^X) = \sum_{i\geq 0} \mu_i z^i$, the ordinary generating function of the probability distribution.*

Sometimes this will be only a formal tool, but sometimes we will be interested in it as a function of $z \in \mathbb{C}$. Below we'll use the notation

$$g(z) = g_X(z) = E(z^X).$$

Here are some properties.

1. g is differentiable in the open disk $|z| < 1$ and continuous in the closed disk $z \leq 1$.
2. $g(0) = \mu_0$ and $g(1) = 1$; g is monotone nondecreasing for $z \in [0,1]$, and strictly increasing unless $\mu_0 = 1$.

Note the pgf is quite different from the moment generating function (mgf) [1]

Lemma 9. *If X and Y are independent rvs on $\mathbb{N}_{\geq 0}$ then $g_{X+Y}(z) = g_X(z) g_Y(z)$.*

(Exercise.)

Before going on let us recall the so-called "tower property" of conditional expectations from **??**. If X and Y are rvs then $E(X|Y)$ is also an rv, and

$$E(X) = E(E(X|Y)).$$

Now let $X_1, \ldots$ be iid rvs on $\mathbb{N}_{\geq 0}$, and N another rv on $\mathbb{N}_{\geq 0}$ which is independent of all the X_i. Let $S = \sum_{i=1}^{N} X_i$ (and set $S = 0$ if $N = 0$).

Theorem 10. $g_S(z) = g_N(g_X(z))$.

Proof. Plugging in the definition, $g_S(z) = E(z^S)$, and applying the tower property, this equals

$$\begin{aligned} &= E(E(z^S|N)) \\ &= \sum_{n\geq 0} E(z^{\sum_1^N X_i}) \Pr(N = n) \end{aligned}$$

(Notice this works even if $N = 0$ given our definition of S in that case.) Applying independence of the X_i's,

$$\begin{aligned} &= \sum_{n\geq 0} (E(z^X))^n \Pr(N = n) \\ &= \sum_{n\geq 0} \Pr(N = n)(g_X(z))^n \\ &= g_N(g_X(z)) \end{aligned}$$

□

[1]

Definition 8. *The moment generating function of X (or of μ) is the exponential generating function of the moments, namely $E(e^{zX})$.*

To verify the terminology, see: $E(e^{zX}) = \sum_{i\geq 0} \mu_i e^{zi} = \sum_{j\geq 0} \frac{z^j}{j!} \sum_{i\geq 0} \mu_i i^j = \sum_{j\geq 0} \frac{z^j}{j!} E(X^j)$.

Back to branching processes

Proof. of Theorem 6. Define $g(z)$ to be the generating function for μ, i.e., $g(z) = \sum_{i\geq 0} \mu_i z^i$.

Before the formal proof let's give the idea; this paragraph however is not rigorous. Let X be the number of children of the root. The tree is finite in any of the following disjoint events: $X = 0$; $X = 1$ and the single subtree is finite; $X = 2$ and both subtrees are finite; etc. Suppose $p = \Pr(T \text{ is finite})$. Since what happens in distinct subtrees is independent, p satisfies $p = g(p)$, i.e., it is a fixed point of g. The conditions on solutions of this equation match the statement of the theorem.

Now to the formal proof. Let Z_n be the number of descendants of the root at level n (the root being level 0, so $Z_0 = 1$). Note that a simple property of the probability generating function is that $\Pr(Z_n = 0) = g_{Z_n}(0)$.

"Extinction is forever:"

$$[\![Z_n = 0]\!] \subseteq [\![Z_{n+1} = 0]\!]$$

So letting $p_n = \Pr(Z_n = 0)$, p_n is a monotone non-decreasing sequence, therefore tending to a limit p, which is $\Pr(\bigcup [\![Z_n = 0]\!]) = \Pr(\text{Tree } T \text{ is finite})$.

By Theorem 10 and induction on n, $g_{Z_n} = \underbrace{g \circ \ldots \circ g}_{n}$. In particular $p_n = g_{Z_n}(0) = \underbrace{g \circ \ldots \circ g}_{n}(0)$; the last expression can be rewritten $g(\underbrace{g \circ \ldots \circ g}_{n-1}(0))$ so $p_n = g(p_{n-1})$.

By continuity of g, it follows that $g(p) = p$, i.e., p is a fixed point of g.

Lemma 11. *p is the least nonnegative fixed point of g.*

Proof. If $g(0) = 0$ then $\mu_0 = 0$, i.e., the tree has no leaves, so certainly $p = 0$.

Otherwise, suppose z_0 is a nonnegative fixed point of g, so $z_0 > 0$. Recall that g is nondecreasing, so $p_1 = g(0) \leq g(z_0) = z_0$. Likewise by induction, $p_n \leq z_0$: $p_n = g(p_{n-1}) \leq g(z_0) = z_0$. So all p_n are $\leq z_0$, and since p is their limit point, $p \leq z_0$. See Fig. 2.3. □

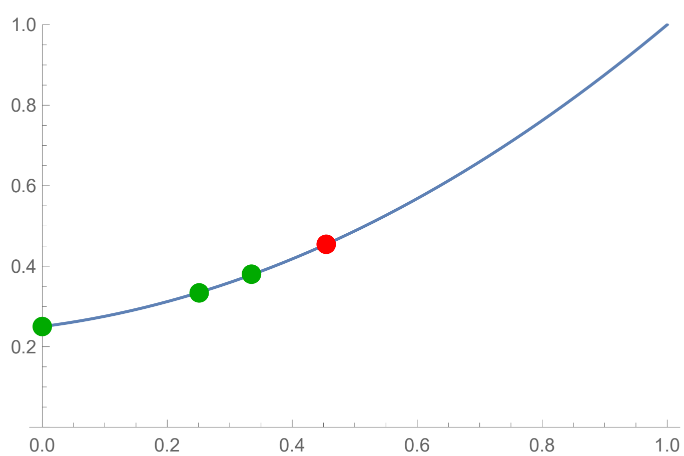

Figure 2.3: Example $g(z) = 1/4 + z/5 + 11z^2/20$. Marked $(0, p_0), (p_0, p_1), (p_1, p_2)$, and (p, p) (here $p = 5/11$).

So we have established that $[T \text{ is a.s. finite}] \iff [g - \text{Id does not have a root less than } 1]$.

Note that

(a) $(g - \text{Id})(0) = \mu_0$;

(b) $(g - \text{Id})''(z) = 2\mu_2 + 6\mu_3 p + \ldots$ so $(g - \text{Id})''(z) \geq 0$ for all $z \geq 0$.

(c) $(g - \text{Id})(1) = 0$

(d) $(g - \text{Id})'(1) = \bar{\mu} - 1$.

From (a,b,c): [$g - \text{Id}$ does not have a root less than 1] $\iff$ ([$(g - \text{Id})'(1) < 0$] or [$(g - \text{Id})'(1) = 0$ and $\mu_0 + \mu_1 < 1$])

Now applying (d), this is equivalent to ([$\bar{\mu} < 1$] or [$\bar{\mu} = 1$ and $\mu_1 < 1$]), as desired.

See Fig. 2.4 for examples. The top (red) curve has $\bar{\mu} < 1$; the next (green) has $\bar{\mu} = 1$ but $\mu_0 + \mu_1 < 1$; the flat (yellow) curve has $\bar{\mu} = 1$ and $\mu_1 = 1$; and the blue curve has $\bar{\mu} > 1$. □

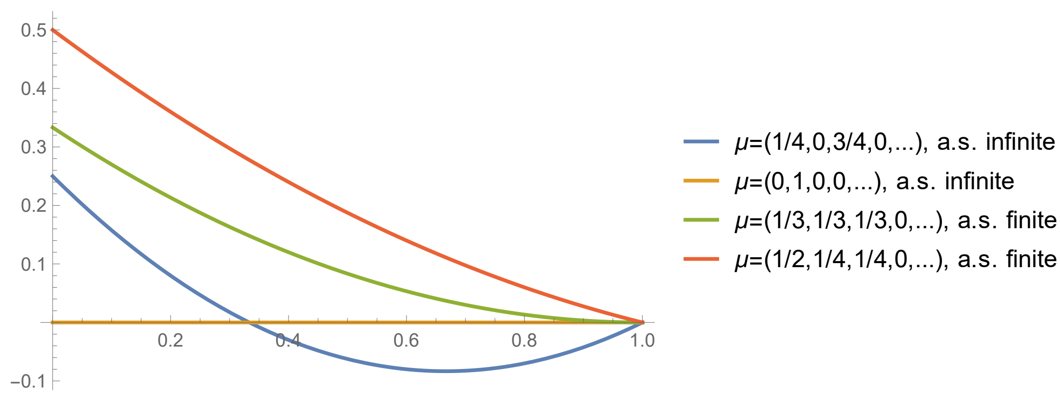

Figure 2.4: $g - \text{Id}$ for various μ

What do branching processes tell us about algorithms

The above discussion sheds light on two algorithms we've seen. The termination of the Moser-Tardos algorithm for the local lemma (recall Sec. **??**) rests essentially on the fact that in that case, $\bar{\mu} < 1$. The MAJ3 query complexity falls in the regime where $\bar{\mu} > 1$, specifically $\bar{\mu} = 8/3$; and this gave us savings over the naïve bound of 3.