On Distributions Computable by Random Walks on Graphs

Guy Kindler[*] Dan Romik[†]

Key words. Finite-state generator, automata, random walks on graphs, random number generation.

AMS Subject Classifications. 65C10, 68Q05, 68Q70.

Abstract. We answer a question raised by Donald E. Knuth and Andrew C. Yao, concerning the class of polynomials on $[0, 1]$ that can be realized as the distribution function of a random variable, whose binary expansion is the output of a finite state automaton driven by unbiased coin tosses. The polynomial distribution functions which can be obtained in this way are precisely those with rational coefficients, whose derivative has no irrational roots on $[0, 1]$.

We also show, strengthening a result of Knuth and Yao, that all smooth distribution functions which can be obtained by such automata are polynomials.

1. Introduction

In a 1976 paper, Donald Knuth and Andrew C. Yao laid the foundations for a complexity theory of probability distribution functions. They defined a computability class of distribution functions that can be "computed" by a random walk on an edge-labelled graph (this can also be thought of as a finite-state automaton driven by a sequence of random bits). They called such a graph a *finite-state generator*, or f.s.g.

Formally, an f.s.g. is a finite directed graph whose vertices are called *states*, with one designated state called the *initial state*. Some of the edges in the graph are labelled with *output strings*, which are finite binary strings. The *output* of the f.s.g. is the random sequence of bits $\alpha_1\alpha_2\alpha_3...$ obtained by performing a simple random walk on its states, starting from the initial state, and writing down sequentially the output strings that are encountered along the way. We identify the output with the real-valued random variable $0 \le X \le 1$ whose binary expansion is the output sequence $\alpha_1\alpha_2\alpha_3...$, namely

$$X = \sum_{n=1}^{\infty} \frac{\alpha_n}{2^n}$$

[*] The Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, Israel. email: `puzne@math.huji.ac.il`

[†] Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel. email: `romik@wisdom.weizmann.ac.il`

A distribution function $F(x)$ supported on $[0,1]$ (that is, $F(0-) = 0$ and $F(1) = 1$) is called *computable* by an f.s.g., or just computable, if it can be realized as the distribution function of a random variable X generated by an f.s.g.

A natural question is to identify all computable distribution functions. Clearly there is a countable number of such functions, so the class of computable distribution functions is rather small. However, since the set of such distributions contains many Cantor-like distributions, and other singular distributions which do not have a simple description, one soon realizes that this question is (probably) too general to possess a meaningful answer.

On the other hand, if the discussion is limited to "nice" distributions, e.g. piecewise smooth distribution functions, then a beautiful algebraic connection is revealed. Knuth and Yao showed that if F is a computable distribution function, and F is real-analytic in an interval $(a,b) \subset [0,1]$ then it must be a polynomial with rational coefficients there. (Theorem 2 below shows that it is enough to require that F be smooth in (a,b).) They constructed a family of polynomial distribution functions which are computable, but left open the question ([3], question (v) on page 427) of precisely which polynomials are distribution functions that can be computed by an f.s.g. The question was raised again by Yao [5], who gave some necessary conditions.

The purpose of this paper is to show that Yao's necessary conditions are sufficient. Our main result is

Theorem 1. A polynomial $Q(x)$ which is monotone increasing on $[0,1]$ and satisfies $Q(0) = 0, Q(1) = 1$, can be realized as the distribution function of a random variable that is generated by an f.s.g., if and only if
1. $Q(x)$ has rational coefficients;
2. $Q'(x)$ has no irrational roots in $[0,1]$.

We prove two additional results. The next theorem further substantiates Knuth and Yao's claim that polynomials form the main class of interesting computable distribution functions, by showing that if a computable distribution function is smooth, then it is a polynomial. This strengthens Theorem 7.4 of [3], which shows the same for *analytic* computable distribution functions.

Theorem 2. Let F be a computable distribution function. If F is infinitely differentiable on an interval $(a,b) \subset [0,1]$, then F is a polynomial there.

The last theorem investigates some structural properties of f.s.g.'s that compute non-smooth distributions. Recall that any distribution function F can be decomposed into a mixture

$$F = \lambda F_{\text{ac}} + (1-\lambda) F_{\text{sing}}\,, \qquad 0 \le \lambda \le 1 \tag{1}$$

of an absolutely continuous distribution function F_{ac} and a singular distribution function F_{sing} (for the purpose of this paper we include the atomic part of F in F_{sing} – see also the comment in Section 5). λ is determined uniquely, and if $0 < \lambda < 1$, namely if F is not purely singular or absolutely continuous, then F_{ac} and F_{sing} are also determined uniquely (otherwise, one of them is trivially not).

Theorem 3. Let $F(x)$ be a computable distribution function, let $F = \lambda F_{\text{ac}} + (1-\lambda)F_{\text{sing}}$ be the decomposition of F as in (1), and assume that $0 < \lambda < 1$. Then λ is rational, and F_{ac} and F_{sing} are both computable.

In the proof of Theorem 3 it is shown that essentially, the contributions to the absolutely continuous and singular parts, respectively, come from different parts of the f.s.g. which do not interact.

Remarks. The above definition of an f.s.g. is a slight variation on those of [3, 5], but is easily seen to be equivalent, in the sense that the class of computable distribution functions is the same. In [3, 5] it was required that the outdegree of each vertex in the graph be 2 (this restriction is natural when an f.s.g. is interpreted as a coin-tossing automaton). In Section 3 below, we use another equivalent variation on the f.s.g. model.

Our paper was inspired by the recent work of Mossel and Peres [4], which deals with questions somewhat similar to ours. Mossel and Peres characterize the class of functions $f : (0,1) \to (0,1)$ for which there exists a finite state automaton whose input is a sequence of random bits with bias p and whose output is a single random bit with bias $f(p)$. Those functions are precisely the rational functions of p with rational coefficients.

Structure of the paper. In the next section we prove Theorem 1. The "only if" part was already proved in [3] and [5]. For the "if" part, we rely essentially on Knuth and Yao's construction involving the order statistics of uniform random variables. It is amusing that order statistics should play a distinguished role in this problem, and that in fact by taking scalings and rational mixtures of polynomials constructed using order statistics one obtains the most general class of constructible polynomials.

In section 3 we prove Theorem 3. In section 4 we prove Theorem 2. In section 5 we give an example of a computable distribution function which is absolutely continuous but whose density is everywhere locally unbounded, and discuss related open problems.

2. Proof of Theorem 1

It will be convenient, in the proof of Theorem 1, to deal with density functions rather than cumulative distribution functions. Let $\mathcal{D}$ be the set of piecewise polynomial density functions on $[0,1]$. Let $\mathcal{C}$ be those elements $q(x) \in \mathcal{D}$ such that the corresponding cumulative distribution function $Q(x) = \int_0^x q(t)dt$ is computable. The elements of $\mathcal{C}$ are called computable (piecewise polynomial) densities.

The following theorem summarizes Knuth and Yao's constructions of computable densities:

Theorem 4. **[3]** (i) If $0 \le a < b \le 1$ are rational, then the uniform density on $[a,b]$ is computable.
(ii) If $0 \le a < b \le 1$ are rational, then the density

$$f(x) = \frac{(n+1)!}{k!(n-k)!(b-a)^{n+1}} \ (x-a)^k (b-x)^{n-k} \mathbf{1}_{[a,b]}(x)$$

of the $(k+1)$'th order statistic of $n+1$ independent random variables distributed uniformly on $[a,b]$, is computable.
(iii) If $f_1, f_2, ..., f_n$ are computable densities, then any rational mixture of the form $f = \sum_{i=1}^n a_i f_i$ where $0 < a_i \in \mathbb{Q}$, $\sum_i a_i = 1$, is also computable.

Let $q \in \mathcal{D}$ be a polynomial density function such that $Q(x) = \int_0^x q(t)dt$ satisfies the conditions of Theorem 1. In terms of q, this simply means that q has rational coefficients, and no irrational roots in $[0,1]$. Our aim is to show that q is computable. Let $0 = r_0 < r_1 < r_2 < ... < r_{k-1} < r_k = 1$ be the roots of q in $[0,1]$, together with 0 and 1 if they are not roots. In view of Theorem 4(iii), it is enough to show that each of the densities

$$q_i(x) = \frac{1}{\int_{r_i}^{r_{i+1}} q(t)dt}\, q(x)\mathbf{1}_{[r_i,r_{i+1}]}(x), \qquad i = 0,1,2,...,k-1$$

(the density q conditioned on the interval $[r_i, r_{i+1}]$) is computable. This is because q is then a mixture of the q_i with rational coefficients.

Now fix i, $0 \le i \le k-1$. q_i is a density that is 0 outside the interval $[r_i, r_{i+1}]$. Inside this interval q_i has the form

$$q_i(x) = c(x-r_i)^j (r_{i+1}-x)^l h(x), \tag{2}$$

where $c \in \mathbb{Q} \cap (0,\infty)$, $j,l \ge 0$, and $h(x)$ is a polynomial with rational coefficients that is *strictly positive* on $[r_i, r_{i+1}]$, and integrates to 1 there. Our claim now relies on the following

Proposition 1. $h(x)$ can be expressed as a rational mixture (a convex combination with rational coefficients) of polynomials which have the form

$$c(x-t_1)^{v_1}(x-t_2)^{v_2}...(x-t_{m-1})^{v_{m-1}}(-x+t_m)^{v_m} \tag{3}$$

for some *rational* $r_i \le t_1 < t_2 < ... < t_m \le r_{i+1}$, and which integrate to 1 on $[r_i, r_{i+1}]$ – the constant c takes care of this, and is therefore necessarily rational. The powers $v_1, v_2, ..., v_m$ above must be even, with the exception that if $t_1 = r_i$ then v_1 can be odd, and if $t_m = r_{i+1}$, v_m can be odd (this is why the last term in (3) is written differently than the other terms).

Proposition 1 implies our claim that q_i is computable. To see this, let f be a polynomial density on $[r_i, r_{i+1}]$ which has the form (3) (note that not only h, but also q_i is a mixture of such polynomials, by (2)). We prove that f is computable by showing that its restriction to each subinterval $[t_j, t_{j+1}]$ (normalized to have integral 1) is a computable density. On $[t_j, t_{j+1}]$, write f as

$$f(x) = c\,[(x-t_j)+(t_j-t_1)]^{v_1}[(x-t_j)+(t_j-t_2)]^{v_2}...(x-t_j)^{v_j}\cdot$$
$$(t_{j+1}-x)^{v_{j+1}}[(t_{j+1}-x)+(t_{j+2}-t_{j+1})]^{v_{j+2}}...[(t_{j+1}-x)+(t_m-t_{j+1})]^{v_m}$$

Now expand out the products, observing that $t_j - t_1$, $t_j - t_2$, ..., $t_j - t_{j-1}$, $t_{j+2} - t_{j+1}$, ..., $t_m - t_{j+1}$ are all positive rational numbers. This gives a representation of f as a rational mixture of polynomials proportional to $(x-t_j)^\alpha (t_{j+1}-x)^\beta$, hence by Theorem 4(ii),(iii), the restriction of f to $[t_j, t_{j+1}]$ is computable.

Our goal is now to prove Proposition 1. We start by discussing how a non-negative polynomial density on an interval can be represented as a convex combination of polynomial densities which are *not necessarily rational*:

Lemma 1. Let $C_n[a,b]$ be the closed convex set of non-negative polynomials of degree at most n on an interval $[a,b]$ that integrate to 1 there. Then $C_n[a,b]$ is a compact set, and its extreme points are precisely the polynomials in $C_n[a,b]$ of degree exactly n which have the form (3) for some $a \leq t_1 < t_2 < ... < t_m \leq b$ and positive even $v_1, v_2, ..., v_m$ (again, with the exception that if $t_1 = a$, v_1 can be odd, and if $t_m = b$, v_m can be odd).

As was indicated to us by a referee, a proof of Lemma 1 appears in the 1953 paper [1] by Karlin and Shapley (Theorem 9.2, page 28). We include the proof here for completeness.

Proof. Recall that a bounded closed set within a finite-dimensional normed space is compact. The space of n-degree polynomials is finite dimensional, and it can be equipped with the norm defined by $||f|| = \int_a^b |f(x)|dx$. The set $C_n[a,b]$ is bounded with respect to this norm (all of its elements have norm 1), and it is obviously closed, hence it is compact.

Now let $f \in C_n[a,b]$ be a polynomial of degree n with all n roots (counting multiplicities) in the interval $[a,b]$ (the evenness of the multiplicities is automatic from the non-negativity requirement), and suppose $f = \alpha g + (1-\alpha)h$, where $g, h \in C_n[a,b]$, and $0 < \alpha < 1$. From positivity we have that wherever f vanishes, g and h must also vanish with at least the same order, so they share the same n roots as f and are therefore equal to it, since they are both of degree at most n and integrate to 1. Thus f is an extreme point of $C_n[a,b]$.

Conversely, if $f \in C_n[a,b]$ does *not* have n roots in the interval $[a,b]$, then it can be represented as

$$f(x) = c(x-t_1)^{v_1}(x-t_2)^{v_2}...(-x+t_m)^{v_m} \cdot g(x) =: w(x) \cdot g(x),$$

where $a \leq t_1 < t_2 < ... < t_m \leq b$, the sum of the multiplicities $\deg w = \sum_i v_i$ is strictly less than n, the constant $c > 0$ is chosen so that $g \in C_n[a,b]$, and g has no roots in $[a,b]$. Now, either of two cases must hold: if g is a constant, then $\deg f = \deg w < n$, and then

$$f(x) = \left(\int_a^b \frac{t-a}{b-a} f(t)dt\right)\left(\left(\int_a^b \frac{t-a}{b-a} f(t)dt\right)^{-1} \frac{x-a}{b-a} f(x)\right)$$
$$+ \left(\int_a^b \frac{b-t}{b-a} f(t)dt\right)\left(\left(\int_a^b \frac{b-t}{b-a} f(t)dt\right)^{-1} \frac{b-x}{b-a} f(x)\right)$$

represents f as a convex combination of two unequal polynomials in $C_n[a,b]$. Otherwise, $\deg g \geq 1$, in which case, letting $\epsilon = \min_{x \in [a,b]} g(x)$, the equation

$$f(x) = \left(\frac{\int_a^b w(t)(g(t)-\epsilon)dt}{2}\right) \cdot \frac{w(x)(g(x)-\epsilon)}{\int_a^b w(t)(g(t)-\epsilon)dt}$$
$$+ \left(\frac{\int_a^b w(t)(g(t)+\epsilon)dt}{2}\right) \cdot \frac{w(x)(g(x)+\epsilon)}{\int_a^b w(t)(g(t)+\epsilon)dt}$$

represents f as a convex combination of two polynomials in $C_n[a, b]$ which (because $\deg g \geq 1$) are not equal. Therefore f is not an extreme point of $C_n[a, b]$. ■

Proof of Proposition 1. First, note that it is enough to show that $h(x)$ can be expressed as a mixture of polynomials of the form (3), without insisting on a *rational* mixture: this is since for a linear system of equations with rational coefficients, the set of rational solutions is dense in the set of real solutions.

Now, the idea of the proof is to first use Lemma 1 to represent $h(x)$ as a convex combination of polynomials of the form (3), with $r_i \leq t_1 < t_2 < ... < t_m \leq r_{i+1}$ not necessarily rational. The t_i's are then slightly perturbed to make them rational.

Proposition 1 follows from the three lemmas below as follows. First, note that since $h(x)$ has no roots, it is actually an interior point of $C_n[r_i, r_{i+1}]$, where $n = \deg h$ (we consider $C_n[r_i, r_{i+1}]$ as a subset of the affine vector space of polynomials of degree at most n that integrate to 1 on $[r_i, r_{i+1}]$). By Lemma 2, this implies that $h(x)$ is also in the interior of the convex hull of some finite set P of polynomials of the form (3). According to Lemma 3 the polynomials in P may be perturbed slightly while maintaining $h(x)$ in the interior of their convex hull. Finally, Lemma 4 implies that these perturbations can be chosen so that the roots of the polynomials become rational.

Lemma 2. For a set B, denote by B° the interior of B. Let K be a compact convex body in a finite-dimensional vector space V, and let $n = \dim(V)$. Then for every interior point $x \in K^\circ$ there exists extreme points $y_1, \ldots, y_m$ of K, such that $x \in \mathrm{Conv}^\circ(y_1, ..., y_m)$. The number of points, m, is at most $2n$.

Lemma 3. Let $x, y_1, \ldots, y_n$ be points in a finite-dimensional vector space V. Suppose that $x \in \mathrm{Conv}^\circ(y_1, .., y_n)$. Then there exists a neighborhood $\mathcal{U}$ of $0 \in V$ with the following property. If $z_1, \ldots, z_n \in V$ satisfy $z_i - y_i \in \mathcal{U}$ for all i, then $x \in \mathrm{Conv}^\circ(z_1, .., z_n)$.

Lemma 4. The set of extreme points in $C_n[r_i, r_{i+1}]$ all of whose roots are rational is dense in the set of extreme points of $C_n[r_i, r_{i+1}]$ (with the obvious topology).

Lemmas 3 and 4 are obvious, hence we only prove Lemma 2. Note that the bound $2n$ on the number of required extreme points in Lemma 2 is tight, as can be seen by taking $x = 0$ and $K = \mathrm{Conv}(\pm e_1, \ldots, \pm e_n)$.

Proof of Lemma 2. Assume that $K^\circ \neq \emptyset$, so that there will be something to prove. Without loss of generality, assume that $x = 0$. We choose a basis $y_1, \ldots, y_n$ for V whose elements are extreme points of K, as follows.

Take y_1 to be any extreme point of K ($y_1 \neq 0$). Having chosen $y_1, \ldots, y_i$ for $i < n$, we set $H_i = \mathrm{span}(y_1, \ldots, y_i)$. Since K contains a neighborhood of 0, it cannot be contained in H_i. Therefore there exists an extreme point y_{i+1} of K, satisfying $y_{i+1} \notin H_i$ (for example, there exists an extreme point maximizing the convex function $\mathrm{dist}(\cdot, H_i)$, where dist is computed according to some norm on V. Recall that a convex function defined on a closed convex body always attains its maximum on some extreme point). This process obviously yields a basis for V.

Take z to be the intersection of the boundary of K with the ray $\{t\cdot(-y_1-y_2-...-y_n) : t > 0\}$. Obviously, the convex hull of $y_1, ..., y_n, z$ contains a neighborhood of 0. Now let H_z be an affine hyperplane supporting K at z. The intersection of K with H_z is a convex body in a vector space of dimension $\leq n-1$, and therefore by Carathéodory's theorem (see [2]) z is a convex combination of at most n extreme points $y_{n+1}, \dots, y_m$ in it. Since these are also extreme points of K, and since obviously $\mathrm{Conv}(y_1, \dots, y_n, z) \subseteq \mathrm{Conv}(y_1, \dots, y_m)$, the proof is complete. ■

3. Proof of Theorem 3

In the next two sections, we modify slightly our model of finite state generators to an equivalent model. In the modified model, the outgoing edges are labelled with transition probabilities, which are arbitrary rational numbers in $(0,1]$ (and which sum to 1 for any given state). The random walk which is performed is then a weighted random walk with these transition probabilities. We also require every edge to be labelled with a single output bit.

The equivalence of the two models is simple, and was noted in [3], p. 421-422.

Let S be the set of states of such a modified f.s.g. An alternative description of the f.s.g. is in terms of the matrix of transition probabilities, which we denote by

$$A = (p_{s\to s'})_{s,s'\in S}$$

A is a Markov transition matrix with rational entries, and is decomposed as the sum of two sub-stochastic matrices with rational entries

$$A = A_0 + A_1$$

where A_0 has non-zero entries for those edges whose output label is "0", and A_1 has non-zero entries for those edges with output label "1". Specifying the f.s.g. is equivalent to specifying the matrices A_0, A_1 and the initial state s_0.

Let $F = \lambda F_{\mathrm{ac}} + (1-\lambda)F_{\mathrm{sing}}$ be as in Theorem 3, and suppose that S is the set of states of a given f.s.g. that computes F, with initial state $s_0 \in S$. For any state $s \in S$, let F^s be the distribution function generated by the same f.s.g. with the initial state replaced by s. Thus, $F = F^{s_0}$. Thinking of the F^s as measures on $[0,1]$, we denote for any Borel subset $B \subset [0,1]$

$$F(B) = \int_B dF(x)$$

A state $s \in S$ is said to be of absolutely continuous (a.c.) type, if F^s is an absolutely continuous measure. Call s of singular type (or just singular) if F^s is a singular measure. Call s *pure* if it is either absolutely continuous or singular.

Lemma 5. 1. If $s \in S$ is pure, and $s' \in S$ is a state such that there exists a path in the graph of the f.s.g. leading from s to s', then s' is pure and of the same type as s.
2. If the graph of the f.s.g. is strongly connected (namely there is a path from any state to any other state), then all the states are pure (and are therefore of the same type by part 1).

Proof. Let $\mu = (F^s)_{s\in S}$ be the vector-valued measure whose coordinates are the measures F^s. The definition of the f.s.g. and the measures F^s can be translated into the following system of equations satisfied by μ: For any Borel subset $B \subset [0,1]$ and any state $s \in S$,

$$F^s(B) = \sum_{s \xrightarrow{0} s'} p_{s\to s'} F^{s'}(2B \cap [0,1]) + \sum_{s \xrightarrow{1} s'} p_{s\to s'} F^{s'}((2B-1)\cap[0,1]),$$

with $s \xrightarrow{\alpha} s'$ meaning that s has an outgoing edge to s', labelled by the output bit α. In matrix notation, this can be written as

$$\mu(B) = A_0\mu(2B\cap[0,1]) + A_1\mu((2B-1)\cap[0,1]) \tag{4}$$

where μ is thought of as a column vector.

Now let s be an a.c. state, and let s' be a state such that $s \xrightarrow{\alpha} s'$, with α being either 0 or 1. Then for any Borel set $B \subset [0,1]$ which has Lebesgue measure 0, we have

$$0 = F^s((B+\alpha)/2) \geq p_{s\to s'} F^{s'}(B) \tag{5}$$

Therefore $F^{s'}$ is also a.c. Similarly, if s is singular, then, taking $C \subset [0,1]$ a set of Lebesgue measure 0 such that $F^s(C) = 1$, and $B = [0,1] \setminus (2C - \alpha)$, again (5) holds. This proves that s' is singular.

For part 2 of the Lemma, observe first that (4) uniquely determines a vector $\mu = (F^s)_{s\in S}$ of probability measures on $[0,1]$ – this is equivalent to saying that the output of the f.s.g. is a well-defined random variable. Now, for any state $s \in S$, let $F^s = \lambda(s)F^s_{\text{ac}} + (1-\lambda(s))F^s_{\text{sing}}$ be the decomposition of F^s into a mixture of an a.c. probability measure and a singular probability measure. We claim that, when the graph of the f.s.g. is strongly connected, the coefficients $\lambda(s)$ in these decompositions are all equal. This is because, by (4), $\lambda(s)$ is a harmonic function on this (finite) graph and is therefore constant (take as the subset B in (4) the union of the supports of all the measures F^s_{sing}).

So if $0 < \lambda = \lambda(s) < 1$ then we have shown that

$$\mu = \lambda\mu_{\text{ac}} + (1-\lambda)\mu_{\text{sing}}$$

where μ_{ac} and μ_{sing} are vector-valued measures each coordinate of which is a probability measure. But then, both μ_{ac} and μ_{sing} are easily seen to be solutions of (4), and therefore we have found two different (in fact, mutually singular) solutions to (4), in contradiction to the fact that (4) has exactly one solution. Therefore λ must be 0 or 1, and all the states are pure. ■

Corollary. $\lambda = \lambda(s_0)$ is rational, and $F^{s_0}_{\text{ac}}$, $F^{s_0}_{\text{sing}}$ are computable.

Proof. The states of the f.s.g. decompose into strongly connected components. Call a strongly connected component *terminal*, if it has no edges going out to other strongly connected components. Clearly, with probability one the random walk on the states must end up in a terminal component. Looking at a terminal component as a sub-f.s.g., Lemma 5 implies that its states must be pure, since the measures F_s for the sub-f.s.g. are the same as for the original one. Call a strongly connected component with pure states either a.c. or singular, according to the type of its states.

The above discussion leads to an identification of the mixture coefficient $\lambda(s_0)$: it is simply the probability that the random walk eventually ends up in one of the a.c. terminal components. This probability is clearly rational, as it can be represented as the solution of a (well-posed) system of linear equations with rational coefficients. From the discussion it is also easy to see how to build an f.s.g. that computes F_{ac}: simply delete any edges going into singular components, and renormalize the transition probabilities so that the sum of the probabilities of outgoing edges for any state is 1. (In other words, the new f.s.g. is the old f.s.g. conditioned never to go into a singular component.) A similar construction replacing the words "singular" and "a.c." computes F_{sing}. ■

4. Proof of Theorem 2.

Let F be a distribution function, computable by a given f.s.g. with state set S and initial state s_0, which is infinitely differentiable on an interval $(a,b) \subset [0,1]$. Let $x \in (a,b)$ be a dyadic number, i.e. of the form $x = k/2^m$ for some integers $m \geq 1, 0 \leq k < 2^m$. For every $n \geq m$, we shall apply (4) n times repeatedly starting with the set

$$B = \left[x, x + \frac{1}{2^n}\right]$$

Some notation will help: If the binary expansion of x is $x = 0.\alpha_1\alpha_2...\alpha_n$ (the last $n-m$ digits are 0), and for $\alpha \in \{0,1\}$ we denote by T_α the set operation

$$T_\alpha(C) = 2C - \alpha, \qquad C \subset [0,1]$$

then applying (4) successively gives the vector equation string

$$\begin{aligned} \mu(B) &= A_{\alpha_1}\mu(T_{\alpha_1}(B)) = A_{\alpha_1}A_{\alpha_2}\mu(T_{\alpha_2} \circ T_{\alpha_1}(B)) = ... \\ &= A_{\alpha_1}A_{\alpha_2}...A_{\alpha_{n-1}}A_{\alpha_n}\mu(T_{\alpha_n} \circ ... \circ T_{\alpha_1}(B)) \\ &= (A_{\alpha_1}A_{\alpha_2}...A_{\alpha_{m-1}}A_{\alpha_m})(A_{\alpha_{m+1}}....A_{\alpha_n})\mu([0,1]) \\ &= (A_{\alpha_1}A_{\alpha_2}...A_{\alpha_{m-1}}A_{\alpha_m})A_0^{n-m}\mu([0,1]) =: A_x A_0^{n-m}\mu([0,1]) = A_x A_0^{n-m}\mathbf{1}. \end{aligned}$$

Here, $\mathbf{1}$ is the vector of all ones $(1)_{s\in S}$, and A_x is, as above, the matrix with rational entries obtained by multiplying A_0's and A_1's corresponding to the m bits in the binary expansion of x. Taking the s_0-th coordinate in the above equation we obtain

$$F(B) = F\left(x + \frac{1}{2^n}\right) - F(x) = \mathbf{1}_{s_0}^\top A_x A_0^{n-m}\mathbf{1} \tag{6}$$

where $\mathbf{1}_{s_0}$ is the state vector all of whose coordinates are 0 except the s_0-th coordinate, which is 1. Now observe that, since F is infinitely differentiable at x, then for any j the left-hand side of (6) has the asymptotic expansion as $n \to \infty$

$$F\left(x + \frac{1}{2^n}\right) - F(x) = F'(x)\cdot\frac{1}{2^n} + \frac{F''(x)}{2}\cdot\frac{1}{2^{2n}} + ... + \frac{F^{(j)}(x)}{j!}\cdot\frac{1}{2^{jn}} + O\left(\frac{1}{2^{(j+1)n}}\right)$$

For the right-hand side, on the other hand, we can write down a complete expansion in terms of the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_l$ of the matrix A_0: clearly it must be of the form

$$\sum_{i=1}^{l} c_i \lambda_i^n p_{\lambda_i}(n)$$

for some constants c_i and polynomials $p_{\lambda_i}(t)$ derived from x, the matrices A_x, A_0 and the vectors $\mathbf{1}_{s_0}, \mathbf{1}$ (the polynomials p_{λ_i} appear when A_0 is not diagonalizable).

Equating the two expansions as $n \to \infty$, we conclude the following.

Lemma 6. At any dyadic $x \in (a, b)$, F can have at most $|S|$ nonzero derivatives.

The proof of Theorem 2 will be complete once we prove the following simple lemma:

Lemma 7. Let F be an infinitely differentiable function on an interval (a, b), and let $\mathcal{D} \subseteq (a, b)$ be a dense subset, such that in every point $x \in \mathcal{D}$, F has at most l nonzero derivatives. Then F is a polynomial on (a, b) of degree at most l.

Proof. Suppose for the sake of contradiction that F is not a polynomial of degree at most l. Then there exists a point $x \in (a, b)$ where its $(l + 1)$'th derivative is nonzero. By continuity, there exists a subsegment $(a_{l+1}, b_{l+1}) \subseteq (a, b)$ where the $(l + 1)$'th derivative of F is nonzero.

The l'th derivative is strictly monotone on (a_{l+1}, b_{l+1}), and hence it crosses zero at most once. Hence there is a subsegment $(a_l, b_l) \subseteq (a_{l+1}, b_{l+1})$ where both the l'th derivative and the $(l+1)$'th derivative are nonzero. Continuing by induction one obtains an interval $(a_1, b_1) \subseteq (a, b)$ where all derivatives up to order $(l + 1)$ are non-zero. This is a contradiction to the assumption that F has at most l nonzero derivatives in every point of $\mathcal{D}$ (since $\mathcal{D} \cap (a_1, b_1) \neq \emptyset$). ■

5. Open problems

Several natural questions arise from the paper:

1. Our proof of Theorem 1, which is presented in a somewhat abstract form, can easily be translated into an algorithm for constructing an f.s.g. that computes a given polynomial distribution function F. The resulting algorithm, however, seems to generate extremely large f.s.g.'s, as a function of the degree of the given polynomial and the denominators of its coefficients.

   It is interesting to determine the complexity class of finding the smallest f.s.g. that computes a given polynomial. Another interesting question is to give a sharp bound on the number of states required to compute a polynomial of given parameters.

2. One may consider the same questions that are discussed here, in the case of pushdown automata. Partial results in this direction are given in [5].

3. It may be of interest to investigate the computable distribution functions among the absolutely continuous (and not necessarily smooth) distributions. This class contains some peculiar specimens, such as the distribution computed by the f.s.g. in Figure 1 below. This distribution is absolutely continuous, yet its density function is nowhere locally bounded.

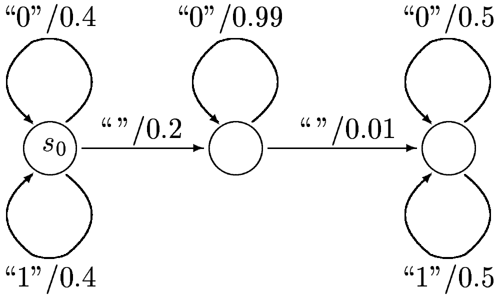

Figure 1: An f.s.g. generating a nowhere bounded density

4. A sufficient condition for the distribution function F computed by a given f.s.g. to be a.c., is that any terminal component of the graph (considered as a sub-f.s.g.), outputs a uniform distribution on $[0, 1]$ starting from any of its states. Is this condition necessary?

5. Characterize all the *atomic* computable distributions.

References

[1] S. Karlin, L. S. Shapley, Geometry of Moment Spaces, Memoirs of the Amer. Math. Soc. 12, 1953.

[2] P. J. Kelly, M. L. Weiss, Geometry and Convexity, Wiley, 1979.

[3] D. E. Knuth, A. C. Yao, The complexity of nonuniform random number generation. Algorithms and Complexity: New Directions and Recent Results, ed. J. F. Traub, Addison-Wesley, 1976.

[4] E. Mossel, Y. Peres, New coins from old: computing with unknown bias. To appear.

[5] A. C. Yao, Context-free grammars and random number generation. Combinatorial Algorithms on Words, ed. A. Apostolico and Z. Galil, NATO ASI SERIES, Springer-Verlag 1985.