Notas de
Algoritmos, Azar y Autómatas

Manuel Panichelli

September 5, 2021

Chapter 1

Introducción

1.1 Azar

Azar es **imposibilidad de predecir**, **falta de patrones**, imposibilidad de abreviar, comprimir.

Vamos a categorizar el azar según diferentes modelos de cómputo

- Autómatas finitos
- Autómatas de pila
- Máquinas de turing

Def. 1. Una secuencia es **azarosa** (para los autómatas de la clase C) cuando, esencialmente, la única forma de describirla (mediante un autómata de la clase C) es nombrando explícitamente cada uno de sus símbolos.

Esto quiere decir que no tiene patrones (porque sino podríamos nombrar menos) y que no se puede comprimir. *Esencialmente* porque se pueden hacer pequeñas conversiones. Por ejemplo, las cadenas de $\{a^n b^n \mid n \in \mathbb{N}\}$ son azarosas para AF pero no para AP (porque es un lenguaje libre de contexto pero no regular).

Hay distintos *grados de azar*:

1. **Azar puro**: Impredecibilidad / incompresibilidad para máquinas de turing
2. **Azar básico**: Impredicibilidad / incompresibilidad para autómatas finitos.

1. Una secuencia es **random** si, esencialmente, sus *segmentos iniciales* solo se pueden describir explícitamente por una Turing Machine (no pueden ser comprimidos por una TM)

2. Una secuencia es **normal** si, esencialmente, sus segmentos iniciales solo se pueden describir explicitamente por un autómata finito.

Cosas que no copié

1. Kolmogorov / program size complexity
2. Definicion de azar de Chaitin basado en kolmogorov
3. Martin Löf random

1.2 Numeros normales

Def. Una **base** es un entero ≥ 2. Para un $x \in \mathbb{R}$ en el intervalo unitario[1], su **expansión** en base b es una **secuencia** $a_1 a_2 a_3 \ldots$ de enteros de $0, 1, \ldots, b-1$ tales que

$$x = 0.a_1 a_2 a_3 \ldots,$$

donde $x = \sum_{k \geq 1} \frac{a_k}{b_k}$ y x no termina con una cola de $b-1$ (esto lo hacemos para tener una representación única de todos los numeros racionales)

Cuando se de por sentada la base b denotamos los primeros n digitos de la expansión de x con $x[1 \ldots n]$

Def. 2 (Números normales, Borel 1909)**.** Un número real x es,

- **Simplemente normal a base** b si en la expansión de x en base b, cada digito ocurre con una frecuencia de $1/b$ en el límite.

  (En el límite todos los símbolos tienen la misma frecuencia)

- **Normal a base** b si para cada entero positivo k, cada bloque de k digitos (arrancando de cualquier posición) ocurre en la expansión de x en base b con una frecuencia en el límite de $1/b^k$

- **Absolutamente normal** si es normal para todas las bases.

Ejemplos:

- 0.01 002 0003 00004 000005 0000006 00000007 000000008 ... no es simplemente normal a base 10 (el 0 tiene más frecuencia que el resto)

- 0.0123456789 0.0123456789 0.0123456789 0.0123456789 ... es simpelemente normal a base 10, pero no es simplemente normal a base 100.

  Pasar de base 10 a base 100 es tomar combinaciones de dos dígitos en base 10 de forma contigua

- El ternario de cantor no es simplemente normal a base 3 (las expansiones no tienen el dígito 1)

[1] El intervalo unitario es el intervalo cerrado $[0, 1]$

- Los numeros racionales no son normales a ninguna base

  Si agarro un número racional, por ej 3.14

$$3.14 \rightsquigarrow 3.140000000\ldots$$

  en base 10 tiene un período que se repite

- La constante de Liouville $\sum_{n\geq 1} 10^{-n!}$ no es normal a base 10

Teorema 1 (Borel 1909)**.** Casi todos los números reales son absolutamente normales.

Son las constantes matemáticas usuales como π, e o $\sqrt{2}$ absolutamente normales? O al menos simplemente normales a alguna base? Es una pregunta abierta.

Teorema 2 (Champernowne, 1933)**.** Todos los numeros naturales en base 10 concatenados es normal a base 10.

$$0.123456789101112131415161718192021\ldots$$

No se sabe si es normal a bases que no son potencias de 10

Teorema 3 (Cassels 1959; Schmidth 1961)**.** Casi todos los números del ternario de Cantor son normales a base 2.

Teorema 4 (Bailey y Borwein 2012)**.** El número de Stoneham $\alpha_{2,3} = \sum_{k\geq 1} \frac{1}{3^k 2^{3^k}}$ es normal a base 2 pero no simplemente normal a base 6.

1.2.1 Normalidad y autómatas finitos

Def. 3. Una secuencia $x = a_1 a_2 a_3 \ldots$ es **compresible** por un trasductor finito T si y solo si en la corrida en T $q_0 \xrightarrow{a_1|v_1} q_1 \xrightarrow{a_2|v_2} q_2 \xrightarrow{a_3|v_3} q_3 \ldots$ satisface que

$$\liminf_{n\to\infty} \frac{|v_1 v_2 \ldots v_n|}{n} < 1.$$

Recordar que los a son símbolols y los v cadenas, posiblemente vacías.

Teorema 5. Una secuencia es **normal** si y solo si es **incompresible por todo one-to-one transducer**.

Teorema (Becher, Casrton, Heiber 2013)**.** Los transductores finitos uno a uno no deterministicos con contadores no pueden comprimir secuencias normales.

Teorema.

Los trasductores de pila no determinísticos pueden comprimir secuencias normales.

0123456789 9876543210 00 01 02 03 ... 98 99 99 98 97 ... 03 02 01 00 000 001 002 ...

Va pusheando y cuando detecta el cambio empieza a desapilar. Parecido al APD que reconoce $w\#w^r$

Chapter 2

3 secuencias normales

2.1 Notación

- Un *alfabeto* es un conjunto finito de símbolos. Por ej A
- A^ω es el conjunto de todas las palabras infinitas
- A^* (la clausura de Kleene) es el conjunto de todas las palabras finitas
- $A^{\leq k}$ es el conjunto de todas las palabras de longitud hasta k
- A^k es el conjunto de palabras de longitud exactamente k.
- Si w es una cadena $|w|$ es su longitud.
- Las posiciones de las cadenas se numeran desde 1
- $w[i]$ es el simbolo iésimo de w y $w[i \dots j]$ es el substring de i a j.
- La cadena vacía es λ

Def. 4. El número de ocurrencias alineadas y no alineadas de una cadena es

$$|w|_u = |\{i : w[i \dots i + |u| - 1] = u\}|,$$
$$||w||_u = |\{i : w[i \dots i + |u| - 1] = u \text{ y } i \equiv 1 \bmod |u|\}|$$

Por ejemplo, $|aaaaa|_{aa} = 4$ y $||aaaaa||_{aa} = 2$.

Cuando u es un símbolo las definiciones coinciden. Y la de alineadas son posiciones que son múltiplos de $|u|$

(La de alineadas tiene $\equiv 1$ en vez de $\equiv 0$ ya que las posiciones se numeran de 1)

Prop. Las ocurrencias alineadas de una palabra de longitud r sobre un alfabeto A coinciden con las ocurrencias del símbolo correspondiente sobre el alfabeto A^r.

Proof. Sean un alfabeto A, una longitud r y un alfabeto B con $|A|^r$ símbolos (la cantidad de símbolos que tiene el alfabeto A^r). A^r (el conjunto de palabras de longitud r sobre el alfabeto A) y B son isomorfos, existe

$$\pi : A^r \to B$$

que se induce del orden lexicográfico en cada conjunto (se puede hacer un matching 1 a 1). Por lo tanto, para cada $w \in A^*$ tal que $|w|$ es múltiplo de r,

$$|\pi(w)| = |w|/r.$$

(Una palabra de longitud múltiplo de r es una cadena de r símbolos de A^r, luego la longitud de la palabra en B que tiene símbolos unitarios digamos es esa).

Luego,

$$\forall u \in A^r \ (||w||_u = |\pi(w)|_{\pi(u)}).$$

$\square$

Por ejemplo, sean $A = \{0, 1\}$, $r = 3$, y B tal que $|A^r| = |B|$,

$$B = \{\underset{000}{0}, \underset{001}{1}, \underset{010}{2}, \underset{011}{3}, \underset{100}{4}, \underset{101}{5}, \underset{110}{6}, \underset{111}{7}\}$$

Luego la cadena,

100 100 111 000

4470

La cantidad de ocurrencias de 100 coinciden con las de 4.

Def. 5 (Normalidad no alineada, Borel)**.** Un número real x es **normal a base *b*** si para cada bloque u,

$$\underset{n\to\infty}{lim} \frac{|x[1 \dots n]|_u}{n} = \frac{1}{b^{|u|}}.$$

En el límite,como $b^{|u|}$ son todos los bloques posibles de longitud $|u|$, $1/b^{|u|}$ seria que cada uno tiene la misma frecuencia.

Teorema 6 (Piatetski-Shapiro 1957)**.** Sea x un número real, $b \geq 2$ un entero y $A = \{0, \dots, b-1\}$. Las siguientes son equivalentes

1. x es normal a base b
2. Existe una constante C tal que para infinitas longitudes ℓ y para todo $w \in A^\ell$

$$\lim_{n\to\infty} \text{sum} \frac{|x[1 \dots n]|_w}{n} < C \cdot b^{-\ell}.$$

3. Existe una constante C tal que para infinitas longitudes ℓ y para todo $w \in A^\ell$

$$\lim_{n\to\infty} \mathrm{sum} \frac{||x[1 \ldots n\ell]||_w}{n} < C \cdot b^{-\ell}.$$

Para el ejercicio hay que hacer la 3ra.

2.2 Tres secuencias normales

- Secuencias de Brujin infinitas
- A la Champernowne (binario)

  01 00 01 10 11 000 001 010 011 100 101 110 111 0000...

- Una secuencia normal tal que la subsecuencia en las posiciones pares es idéntica a toda la secuencia.

2.3 De Brujin

Def. 6 (De Brujin 1946)**.** Definiciones de De Brujin,

- Un **collar de De Brujin** de orden n sobre un alfabeto A es una secuencia cíclica de longitud $|A|^n$ tal que cada palabra de longitud n ocurre en ella exactamente una vez.

  Ejemplos: 01; 0011; 00011101 (el 100 está en la pos 8 por ej.).

- Una **palabra de De Brujin** (no cíclica) de orden n sobre el alfabeto A es una palabra de longitud $|A|^n + n - 1$ (se le agrega todo lo que uno podría hacer con un ciclo) tal que cada palabra de longitud n ocurre en ella exactamente una vez.

  Ejemplos: 01; 00110; 0001110100.

- Una **palabra infinita de De Brujin** $w = a_1a_2\ldots$ en un alfabeto de al menos tres símbolos es una palabra infinita tal que,

  $$\forall n.\ a_1 \ldots a_{|A|^n+n-1}$$

  es una palabra de De Brujin de orden n.

  Ejemplo: 012, una palabra de De Brujin de orden 1, se puede extender a la siguiente de orden 2: 0122002110.

  Si el alfabeto tiene dos símbolos, una palabra infinita de De Brujin $w = a_1a_2\ldots$ es aquella que para cada n impar, $a_1 \ldots a_{|A|^n+n-1}$ es una palabra de De Brujin de orden n.

Def. 7. Un **grafo de De Brujin** $G_A(n)$ es un digrafo cuyos vértices son palabras de longitud n sobre el alfabeto A y sus ejes los pares (au, ub) para alguna palabra u de longitud $n-1$ y posiblemente dos símbolos diferentes a, b.

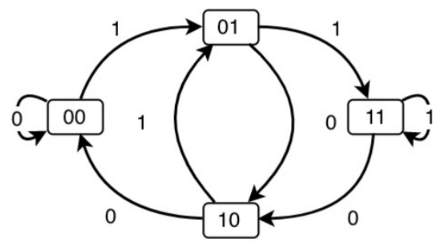

Figure 2.1: Ejemplo de grafo de De Brujin de orden 2 para $A = \{0, 1\}$

- Tiene $|A|^n$ vertices y $|A|^{n+1}$ arcos
- Es *fuertemente conexo* (existe un camino dirigido entre todo par de vértices)[1]
- Es *regular*, $\forall v.d_{in}(v) = d_{out}(v)$ (los loops suman uno a la entrada y salida)
- Es Euleriano (por teorema de Euler, solo hace falta que sea regular y fuertemente conexo).

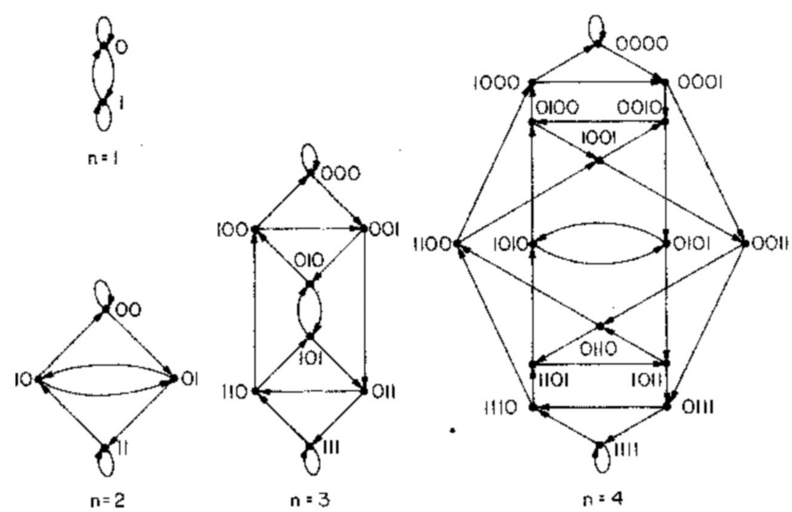

Fig. 3. The de Bruijn graphs of order $n = 1$, 2, 3, and 4.

Figure 2.2: Grafos de De Brujin de ordenes 1, 2, 3 y 4 sobre $A = \{0, 1\}$

[1]Conexo a secas en digrafos es que el grafo subyacente (sacándole direcicones) sea conexo

Def. El **grafo de línea** de un grafo G, es otro grafo que tiene como vértilos ejes de G y como ejes los caminos de longitud 2.

Prop. 1. Toda secuencia de De Brujin de orden $n+1$ sobre un alfabeto de $|A|$ simbolos se puede construir como un ciclo Euleriano en $G_A(n)$.

Prop. 2 (Becher, Heiber 2011)**.** Dado un alfabeto A con al menos tres símbolos, toda secuencia de De Brujin de orden n se puede extender a una de orden $n+1$

Proof. Dado un alfabeto A, suponiendo que E es un ciclo Euleriano de $G_A(n)$. Como $G_A(n+1)$ es el grafo de línea de $G_A(n)$, E es un ciclo Hamiltoniano en $G_A(n+1)$.

Está la demo en las clases, no la terminé de ver.

Todo ciclo euleriano va a ser hamiltoniano en el grafo de línea, porque los vertices son los ejes □

Para computar una palabra infinita de De Brujin puedo para cada $n \geq 1$ extender un ciclo Hamiltoniano en un grafo de De Brujin de orden n a uno Euleriano en el mismo grafo. Esto se hace en tiempo exponencial de n, y no se conoce ningun algoritmo eficiente.

Teorema 7 (Ugalde 2000)**.** Las palabras infinitas de De Brujin son normales.

Si el alfabeto A tiene dos símbolos, se puede considerar el alfabeto A' de 4 símbolos que se obtiene con el morfismo que mapea bloques de dos simbolos en A a un simbolo en A' y probar normalidad ahí.