

LLEGANDO AL LÍMITE DE LA NORMALIDAD

Pablo Ariel Heiber — Charla de Borrachos

Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

26 de octubre de 2012

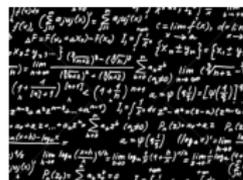


¿Son las siguientes secuencias aleatorias?

- 000000000000000000000000000000000000...
- 010101010101010101010101010101010101...
- 01001000100001000001000000100000001000...
- 000001000000000000000000010001000000001100...
- 0101011101010011000100001010101010110110...



¿Por qué?



¿QUÉ ES LA NORMALIDAD?

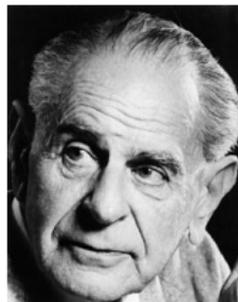


La frecuencia de un resultado dado entre n equiprobables en una secuencia es $1/n$

Ley (débil) de los grandes números (Bernoulli, 1713, luego Chebyshev, Markov, Borel, y siguen las firmas...)

En una secuencia aleatoria todas las subcadenas de una longitud dada tienen la misma frecuencia

Popper en “Logik der Forschung” (“La lógica de la investigación científica”), 1934



La cantidad de información (entropía) se maximiza cuando todos los símbolos son equiprobables

Shannon en “A mathematical theory of communication” (“Una teoría matemática de la comunicación”), 1948

¿QUÉ ES LA NORMALIDAD?

- Concepto muy débil de azar
- Estadísticamente balanceado
- Los dígitos cumplen la ley débil de los grandes números ...
- ... y los grupos de k dígitos también, para cualquier $k \in \mathbb{N}$



Borel

x ES NORMAL (BOREL, 1909)

\forall palabra de ℓ dígitos w :

$$\lim_{n \rightarrow \infty} \frac{\text{cant. ocurrencias de } w \text{ en } x[1..n]}{n} = \frac{1}{2^\ell}$$

Sí y no

- Es la más débil de muchas definiciones
- Otras definiciones:
 - Secuencias no distinguibles con predicados *computables*
No pertenecen a conjuntos chicos y bla-computables
 - Secuencias impredecibles *computablemente*
Una bla-compu no puede adivinar dígitos con alta probabilidad
 - Secuencias incompresibles *computablemente*
Una bla-compu no puede comprimir con alta probabilidad
- En general las otras definiciones implican normalidad

ESTADÍSTICA VS COMPUTABILIDAD

Son formas distintas de establecer condiciones de azar, pero
¿Son irreconciliables?

ESTADÍSTICA VS COMPUTABILIDAD

Son formas distintas de establecer condiciones de azar, pero
¿Son irreconciliables?

123456789101
112131415161
718192021222
324252627282
930313233343

La concatenación de los enteros positivos escritos en base 10 es Normal

Champernowne en “The Construction of Decimals Normal in the Scale of Ten” (“La construcción de decimales normales en base 10”), 1933

Es fácil escribir programas para:

- Escupir tantos dígitos como se quiera.
- Dado n , computar el n -ésimo dígito en tiempo y memoria óptima $\mathcal{O}(\log n)$.



Si las frecuencias no están balanceadas se puede comprimir / obtener radio de compresión < 1

Teorema de la codificación de fuentes (Shannon 1948)

Se puede “implementar” el teorema de Shannon con Autómatas Finitos sin Pérdida

Huffman en “A Method for the Construction of Minimum-Redundancy Codes” (“Un método para la construcción de códigos de mínima redundancia”), 1952



\neg normal \Rightarrow compresible con autómatas finitos
incompresible con autómatas finitos \Rightarrow normal



Un autómata finito no puede predecir dígitos de números normales con alta probabilidad

Schnorr y Stimm en “Endliche Automaten und Zufallsfolgen” (“Autómatas finitos y secuencias aleatorias”), 1971

normal \Leftrightarrow incompresible \Leftrightarrow impredecible (con aut. finitos)

- Schnorr y Stimm, 1971
- Dai, Lathrop, Lutz y Mayordomo en “Finite-state dimension” (“Dimensión de autómatas finitos”), 2004
- Bourke, Hitchcock, y Vinodchandran en “Entropy rates and finite-state dimension” (“Razones de entropía y dimensión de autómatas finitos”), 2005

Autómatas finitos y secuencias aleatorias (trad.)

Becher, Heiber, López-Bristot y Rosner, 2012

Números normales y autómatas finitos

Becher, Heiber, 2012, demostración directa y elemental



incompresible
por DFA



Chomsky

compresible por
log TM

AUTÓMATAS FINITOS

- Una pasada: Se puede leer el input solo de izquierda a derecha
- No determinismo: Múltiples opciones en un momento dado. Requiere validación
- Memoria: Estados codifican finita memoria
- Tiempo-real: Luego de leer algo de input se hacen solo $\mathcal{O}(1)$ operaciones

MODELOS DE MEMORIA INFINITA

- Contadores / pila o cola con alfabeto de un símbolo
- Pila con alfabeto de más de un símbolo
- Cola con alfabeto de más de un símbolo
- Acceso aleatorio (arreglo)

	Una pasada	Tiempo real	Determi- nístico	Memoria finita	
DFA	✓	✓	✓	✓	
2DFA	✗	✓	✓	✓	
NDFA	✓	✓	✗	✓	
k cDFA	✓	✓	✓	✗	(k contadores)
k cNDFA	✓	✓	✗	✗	(k contadores)
NPDA	✓	✓	✗	✗	(pila)
TM	✓	✗	✓	✗	(cola/arreglo)

¿Y DÓNDE ESTÁ EL FAMOSO LÍMITE?

	Una pasada	Tiempo real	Determinístico	Memoria finita	
DFA	✓	✓	✓	✓	
2DFA	✗	✓	✓	✓	
NDFA	✓	✓	✗	✓	
k cDFA	✓	✓	✓	✗	(k contadores)
k cNDFA	✓	✓	✗	✗	(k contadores)
NPDA	✓	✓	✗	✗	(pila)
TM	✓	✗	✓	✗	(cola/arreglo)

¿Y DÓNDE ESTÁ EL FAMOSO LÍMITE?

	Una pasada	Tiempo real	Determi- nístico	Memoria finita	
DFA	✓	✓	✓	✓	
2DFA	✗	✓	✓	✓	
NDFA	✓	✓	✗	✓	
k cDFA	✓	✓	✓	✗	(k contadores)
k cNDFA	✓	✓	✗	✗	(k contadores)
NPDA	✓	✓	✗	✗	(pila)
TM	✓	✗	✓	✗	(cola/arreglo)

¿QUE ANDA FALTANDO?

	Una pasada	Tiempo real	Determinístico	Memoria finita	
<i>kcDFA</i>	✓	✓	✓	✗	
<i>kcN DFA</i>	✓	✓	✗	✗	
<i>2kcDFA</i>	✗	✓	✗	✗	
<i>2kcN DFA</i>	✗	✓	✗	✗	
PDA	✓	✓	✓	✗	(pila)
NPDA	✓	✓	✗	✗	(pila)

¿QUE ANDA FALTANDO? TIEMPO REAL

	Una pasada	Tiempo real	Determi- nístico	Memoria finita	
DFA	✓	✗	✓	✓	
2DFA	✗	✗	✓	✓	
NDFA	✓	✗	✗	✓	
1cDFA	✓	✗	✓	✗	
(> 2)cDFA	✓	✗	✓	✗	
NPDA	✓	✗	✗	✗	(pila)

Lo que más afecta la normalidad es el incremento de memoria.



¿PREGUNTAS?

