

Análisis Probabilístico de Algoritmos

Pablo Rotondo

LIGM, Université Gustave Eiffel

ECI, Buenos Aires, 28 de Julio a 1 de Agosto, 2025.

Modalidad del curso

- Curso dividido en 3 grandes módulos temáticos
- Cada clase estará dividida en dos partes:
15 min intro/exos + 1h15 + 15 min de pausa + 1h15
- Examen escrito al final de la última clase. Duración 1h

¿De qué trata este curso?

Analysis of Algorithms (AofA) is a field at the boundary of computer science and mathematics. The goal is to obtain a precise understanding of the *asymptotic, average-case characteristics of algorithms and data structures*.
[...]

The area of Analysis of Algorithms is frequently traced to 27 July 1963, when **Donald E. Knuth** wrote “Notes on Open Addressing”.

Del sitio de la comunidad **AofA**

<https://www.math.aau.at/AofA/>



Wikipedia, CC BY-SA 3.0.

1. Introducción al análisis probabilístico de algoritmos:
 - Motivación, ejemplos clásicos (sorting, hashing, ...)
 - Modelos modernos (branch prediction).

2. Introducción a la Combinatoria analítica:
 - Funciones generatrices ordinarias y exponenciales.
 - Singularidades, extracción de coeficientes y Teorema de Transferencia.
 - Aplicaciones algorítmicas.

3. Aplicaciones a la generación aleatoria de estructuras discretas¹:
 - Método recursivo.
 - Boltzmann samplers.

¹Si tiempo.

Outline

1. Introducción al análisis de algoritmos
 - Algoritmos de sorting
 - Tablas de Hash
2. Aplicaciones a la predicción de saltos
3. Combinatoria Analítica: métodos simbólicos
4. Combinatoria Analítica: Teoría Analítica

Introducción: análisis de algoritmos

Estudiar teóricamente la performance de un algoritmo:

- independientemente del lenguaje de programación,
- independientemente del hardware.

Introducción: análisis de algoritmos

Estudiar teóricamente la performance de un algoritmo:

- independientemente del lenguaje de programación,
- independientemente del hardware.

⇒ contar operaciones concretas efectuadas.

Introducción: análisis de algoritmos

Estudiar teóricamente la performance de un algoritmo:

- independientemente del lenguaje de programación,
- independientemente del hardware.

⇒ contar operaciones concretas efectuadas.

En los estudios más clásicos:

- Se considera solo el peor caso.
- Solo en orden de magnitud cuando el tamaño del input $n \rightarrow \infty$.
Por ejemplo $O(n^2)$, $O(n \log n)$, etc.

Introducción: análisis de algoritmos

Estudiar teóricamente la performance de un algoritmo:

- independientemente del lenguaje de programación,
- independientemente del hardware.

⇒ contar operaciones concretas efectuadas.

En los estudios más clásicos:

- Se considera solo el peor caso.
- Solo en orden de magnitud cuando el tamaño del input $n \rightarrow \infty$.
Por ejemplo $O(n^2)$, $O(n \log n)$, etc.

Ejemplo

Consideremos el problema de ordenar un array de n elementos distintos.

Si contamos comparaciones:

1. Mergesort $\Theta(n \log n)$ en peor caso, Bubble sort y Quicksort $\Theta(n^2)$,
2. pero Quicksort se comporta en $O(n \log n)$ en media (valor esperado !)

Nociones básicas de probabilidad

La **media** de una variable aleatoria discreta X es

$$\mathbb{E}[X] = \sum_{k \in \mathbb{Z}} k \cdot \Pr(X = k),$$

cuando la suma converge absolutamente, es decir $\mathbb{E}[|X|] < \infty$.

Recordamos las siguientes propiedades básicas:

- Desigualdad triangular: $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
- La media es lineal $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
- Para una función indicatriz $\mathbf{1}_A$ tenemos $\mathbb{E}[\mathbf{1}_A] = \Pr(A)$.

Nociones básicas de probabilidad

La **media** de una variable aleatoria discreta X es

$$\mathbb{E}[X] = \sum_{k \in \mathbb{Z}} k \cdot \Pr(X = k),$$

cuando la suma converge absolutamente, es decir $\mathbb{E}[|X|] < \infty$.

Recordamos las siguientes propiedades básicas:

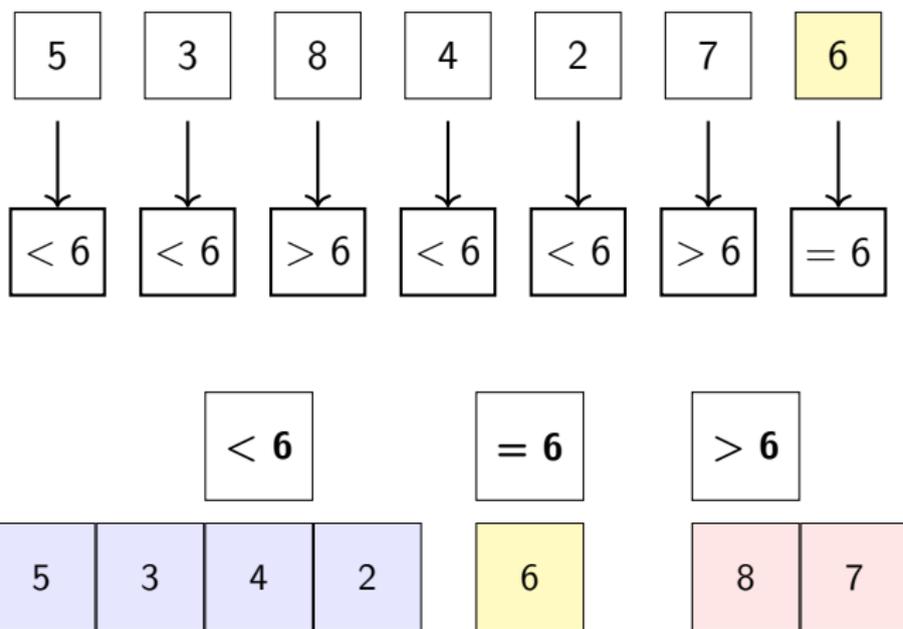
- Desigualdad triangular: $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
- La media es lineal $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
- Para una función indicatriz $\mathbf{1}_A$ tenemos $\mathbb{E}[\mathbf{1}_A] = \Pr(A)$.

Fórmula de la probabilidad total

Sean eventos S_1, S_2, \dots disjuntos con $\cup_i S_i = \Omega$ (todo) :

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \Pr(S_k) \times \mathbb{E}[X|S_k].$$

Quicksort



Quicksort particiona según un pivot y luego continua recursivamente.

Quicksort

Por simplicidad² consideramos que el pivot es elegido determinísticamente:

```
def partition(arr, low, high):
    pivot = arr[high]
    i = low

    for j in range(low, high):
        if arr[j] < pivot:
            arr[i], arr[j] = arr[j], arr[i]
            i += 1

    arr[i], arr[high] = arr[high], arr[i]
    return i
```

Peor caso: o todos mayores, o todos menores que el pivot:

- Cantidad de comparaciones = $1 + 2 + \dots + (n - 1) = \frac{n(n-1)}{2} = \Theta(n^2)$.
- Puede suceder si el array está ya ordenado !

²Mejor sería un pivot aleatorio, o permutar la entrada para evitar ataques.

Quicksort: modelo aleatorio

Veamos ahora qué sucede si el array es una permutación aleatoria

- cada permutación π de $(1, 2, \dots, n)$ tiene probabilidad $p(\pi) = 1/n!$
- equivalente a elegir n números aleatorios del intervalo $[0, 1]$
 \implies argumento de simetría !

Quicksort: modelo aleatorio

Veamos ahora qué sucede si el array es una permutación aleatoria

- cada permutación π de $(1, 2, \dots, n)$ tiene probabilidad $p(\pi) = 1/n!$
- equivalente a elegir n números aleatorios del intervalo $[0, 1]$
 \implies argumento de simetría !

Nos interesa la cantidad de comparaciones $C_n(\pi)$ necesarias para ordenar:

- En media $E_n = \mathbb{E}[C_n] = \sum_{\pi \in \mathcal{S}_n} C_n(\pi) \times p(\pi) = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} C_n(\pi)$,
- En distribución $\Pr(C_n > \lambda) = \sum_{\pi \in \mathcal{S}_n: C_n(\pi) > \lambda} p(\pi)$.

Quicksort: comportamiento en media

Para la media E_n de la cantidad de comparaciones C_n tenemos:

Proposición

Quicksort satisface $E_n = 2(n+1)H_n - 4n$, donde $H_n = 1 + \frac{1}{2} + \dots + \frac{1}{n}$ son las sumas armónicas.

Entonces $E_n \sim 2n \log n$ donde \log es el logaritmo natural (neperiano).³

³Las sumas armónicas satisfacen $H_n \sim \int_1^n \frac{dx}{x} = \log n$.

Análisis en media de Quicksort

Con probabilidad $1/n$ el rango de $\pi(n)$ es j , entonces

$$\begin{aligned} E_n &= \sum_{j=1}^n \mathbb{E}[C_n | \text{pivot} = j] \cdot \Pr(\text{pivot} = j), && \text{(prob. total)} \\ &= \frac{1}{n} \times \sum_{j=1}^n \mathbb{E}[C_n | \text{pivot} = j], \\ &= \frac{1}{n} \times \sum_{j=1}^n \mathbb{E}[C_{j-1} + \tilde{C}_{n-j} + n - 1], && (C_{j-1} \text{ y } \tilde{C}_{n-j} \text{ indep.}) \\ &= \frac{1}{n} \times \sum_{j=0}^{n-1} (E_j + E_{n-1-j}) + n - 1. && \text{(linealidad esperanza)} \end{aligned}$$

En la tercera línea \tilde{C}_{n-j} es el costo de ordenar el array de la parte alta, que contiene $j + 1, \dots, n$ en el orden inicial. Su distribución es la misma que C_{n-j} .

Estudio en media

La **media** es una buena medida cuando pensamos ejecutar **muchas veces** un algoritmo.

Teorema (Ley de los grandes números)

Si X_1, X_2, \dots son independientes e idénticamente distribuidas, con $\mathbb{E}[|X_1|] < \infty$, entonces con probabilidad 1 :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}[X_1].$$

Estudio en media

La **media** es una buena medida cuando pensamos ejecutar **muchas veces** un algoritmo.

Teorema (Ley de los grandes números)

Si X_1, X_2, \dots son independientes e idénticamente distribuidas, con $\mathbb{E}[|X_1|] < \infty$, entonces con probabilidad 1 :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}[X_1].$$

- ¿y si lo queremos ejecutar solamente una vez?
- ¿la **media** refleja la complejidad de **una sola ejecución**?

Estudio en media

La **media** es una buena medida cuando pensamos ejecutar **muchas veces** un algoritmo.

Teorema (Ley de los grandes números)

Si X_1, X_2, \dots son independientes e idénticamente distribuidas, con $\mathbb{E}[|X_1|] < \infty$, entonces con probabilidad 1 :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}[X_1].$$

- ¿y si lo queremos ejecutar solamente una vez?
- ¿la **media** refleja la complejidad de **una sola ejecución**? En general: **no**.

Concentración en probabilidad

Decimos que una secuencia de variables aleatorias X_n satisface $X_n \sim f(n)$ en probabilidad sii, para cada $\varepsilon > 0$ fijo,

$$\Pr(X_n \in [(1 - \varepsilon)f(n), (1 + \varepsilon)f(n)]) \rightarrow 1.$$

⁴De hecho se sabe mucho más al respecto, ver el artículo: C. McDiarmid y R. Hayward. 1992. *Strong concentration for Quicksort*. SODA '92.

Concentración en probabilidad

Decimos que una secuencia de variables aleatorias X_n satisface $X_n \sim f(n)$ en probabilidad si, para cada $\varepsilon > 0$ fijo,

$$\Pr(X_n \in [(1 - \varepsilon)f(n), (1 + \varepsilon)f(n)]) \rightarrow 1.$$

Probaremos más tarde que la cantidad de comparaciones C_n en quicksort⁴ satisface $C_n \sim 2n \log n$ en probabilidad.

Proposición

La cantidad de comparaciones satisface $C_n \sim 2n \log n$ en probabilidad.

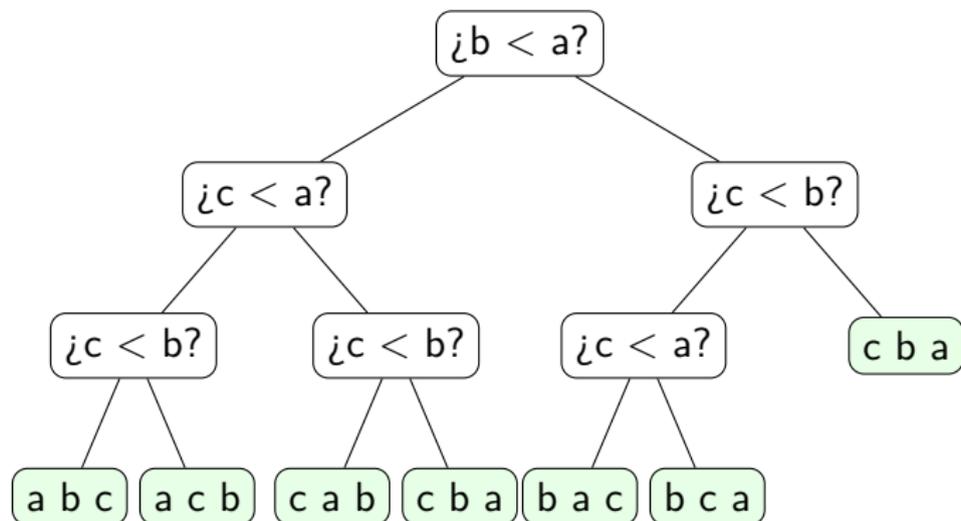
Las funciones generatrices nos ahorrarán muchos cálculos.

⁴De hecho se sabe mucho más al respecto, ver el artículo: C. McDiarmid y R. Hayward. 1992. *Strong concentration for Quicksort*. SODA '92.

Optimalidad en media

Algoritmo basado en comparaciones se representa como árbol binario:

- nodos internos corresponden a comparaciones; rama izquierda False, rama derecha True.
- hojas corresponden a los posibles output del algoritmo.



Optimalidad en media y entropía

Hojas son permutaciones $\Rightarrow n!$ hojas.

- Altura del árbol [peor caso] es al menos $\log_2(n!)$,
- $\log_2(n!) \sim n \log_2 n$

Optimalidad en media y entropía

Hojas son permutaciones $\Rightarrow n!$ hojas.

- Altura del árbol [peor caso] es al menos $\log_2(n!)$,
- $\log_2(n!) \sim n \log_2 n$

Pero esto es también cierto para la media. Sea ℓ_π la profundidad de la hoja π , notar que $C_n(\pi) = \ell_\pi$ es la cantidad de comparaciones,

Teorema (Profundidad media de un árbol binario)

Para cualquier distribución $\mathbf{p} = (p(\pi))_\pi$ sobre las hojas

$$\mathbb{E}[\ell] = \sum \ell_\pi p(\pi) \geq H_2(\mathbf{p}),$$

donde $H_2(\mathbf{p}) = -\sum_\pi p(\pi) \log_2 p(\pi)$ es la entropía binaria.

Optimalidad en media y entropía

Hojas son permutaciones $\Rightarrow n!$ hojas.

- Altura del árbol [peor caso] es al menos $\log_2(n!)$,
- $\log_2(n!) \sim n \log_2 n$

Pero esto es también cierto para la media. Sea ℓ_π la profundidad de la hoja π , notar que $C_n(\pi) = \ell_\pi$ es la cantidad de comparaciones,

Teorema (Profundidad media de un árbol binario)

Para cualquier distribución $\mathbf{p} = (p(\pi))_\pi$ sobre las hojas

$$\mathbb{E}[\ell] = \sum \ell_\pi p(\pi) \geq H_2(\mathbf{p}),$$

donde $H_2(\mathbf{p}) = -\sum_\pi p(\pi) \log_2 p(\pi)$ es la entropía binaria.

En nuestro caso $p(\pi) = 1/n!$ para cada permutación, y $\mathbb{E}[C_n] \geq \log_2 n!$.

Prueba: entropía es cota inferior

Lema

Para un árbol binario completo $\sum_h \text{hojas} 2^{-\ell_h} = 1$

Lema

Para todo $x > 0$, $\log x \leq x - 1$. La igualdad se verifica sii $x = 1$.

Prueba: entropía es cota inferior

Lema

Para un árbol binario completo $\sum_h \text{hojas} 2^{-\ell_h} = 1$

Lema

Para todo $x > 0$, $\log x \leq x - 1$. La igualdad se verifica sii $x = 1$.

Prueba del Teorema. Combinamos los lemas con las propiedades del logaritmo, interpretando $-\ell_h = \log 2^{-\ell_h}$.

QuickSort: modelo del input

- En nuestro modelo de quicksort el input π es una **permutación uniforme** :

$$\Pr(\pi = (a_1, \dots, a_n)) = (n!)^{-1}.$$

QuickSort: modelo del input

- En nuestro modelo de quicksort el input π es una **permutación uniforme** :

$$\Pr(\pi = (a_1, \dots, a_n)) = (n!)^{-1}.$$

- Corresponde a considerar n números (flotantes) de $[0, 1]$.

QuickSort: modelo del input

- En nuestro modelo de quicksort el input π es una **permutación uniforme** :

$$\Pr(\pi = (a_1, \dots, a_n)) = (n!)^{-1}.$$

- Corresponde a considerar n números (flotantes) de $[0, 1]$.
- Razonable sin conocimiento a priori del input.

QuickSort: modelo del input

- En nuestro modelo de quicksort el input π es una **permutación uniforme** :

$$\Pr(\pi = (a_1, \dots, a_n)) = (n!)^{-1}.$$

- Corresponde a considerar n números (flotantes) de $[0, 1]$.
- Razonable sin conocimiento a priori del input.

Otros algoritmos (Powersort, Timsort, ...) suponen que el input puede estar **parcialmente ordenado** en pedazos:

- input dividido en *runs* crecientes/decrecientes de longitud a_1, \dots, a_r .

$$\underbrace{[1, 5, 7]}_{a_1=3}, \underbrace{[2, 4, 9]}_{a_2=3}, \underbrace{[6, 4, 4]}_{a_3=3}, \underbrace{[12, 4]}_{a_4=2};$$

QuickSort: modelo del input

- En nuestro modelo de quicksort el input π es una **permutación uniforme** :

$$\Pr(\pi = (a_1, \dots, a_n)) = (n!)^{-1}.$$

- Corresponde a considerar n números (flotantes) de $[0, 1]$.
- Razonable sin conocimiento a priori del input.

Otros algoritmos (Powersort, Timsort, ...) suponen que el input puede estar **parcialmente ordenado** en pedazos:

- input dividido en *runs* crecientes/decrecientes de longitud a_1, \dots, a_r .

$$\left[\underbrace{1, 5, 7}_{a_1=3}, \underbrace{2, 4, 9}_{a_2=3}, \underbrace{6, 4, 4}_{a_3=3}, \underbrace{12, 4}_{a_4=2} \right]; \text{ o quizás } \left[\underbrace{1, 5, 7}_{a_1=3}, \underbrace{2, 4, 9}_{a_2=3}, \underbrace{6, 4}_{a_3=2}, \underbrace{4, 12}_{a_4=2}, \underbrace{4}_{a_5=1} \right].$$

QuickSort: modelo del input

- En nuestro modelo de quicksort el input π es una **permutación uniforme** :

$$\Pr(\pi = (a_1, \dots, a_n)) = (n!)^{-1}.$$

- Corresponde a considerar n números (flotantes) de $[0, 1]$.
- Razonable sin conocimiento a priori del input.

Otros algoritmos (Powersort, Timsort, ...) suponen que el input puede estar **parcialmente ordenado** en pedazos:

- input dividido en *runs* crecientes/decrecientes de longitud a_1, \dots, a_r .

$$\left[\underbrace{1, 5, 7}_{a_1=3}, \underbrace{2, 4, 9}_{a_2=3}, \underbrace{6, 4, 4}_{a_3=3}, \underbrace{12, 4}_{a_4=2} \right]; \text{ o quizás } \left[\underbrace{1, 5, 7}_{a_1=3}, \underbrace{2, 4, 9}_{a_2=3}, \underbrace{6, 4}_{a_3=2}, \underbrace{4, 12}_{a_4=2}, \underbrace{4}_{a_5=1} \right].$$

- merge(sort) inteligente aprovecha los runs existentes!

QuickSort: modelo del input

- En nuestro modelo de quicksort el input π es una **permutación uniforme** :

$$\Pr(\pi = (a_1, \dots, a_n)) = (n!)^{-1}.$$

- Corresponde a considerar n números (flotantes) de $[0, 1]$.
- Razonable sin conocimiento a priori del input.

Otros algoritmos (Powersort, Timsort, ...) suponen que el input puede estar **parcialmente ordenado** en pedazos:

- input dividido en *runs* crecientes/decrecientes de longitud a_1, \dots, a_r .

$$\underbrace{[1, 5, 7]}_{a_1=3}, \underbrace{[2, 4, 9]}_{a_2=3}, \underbrace{[6, 4, 4]}_{a_3=3}, \underbrace{[12, 4]}_{a_4=2}; \text{ o quizás } \underbrace{[1, 5, 7]}_{a_1=3}, \underbrace{[2, 4, 9]}_{a_2=3}, \underbrace{[6, 4]}_{a_3=2}, \underbrace{[4, 12]}_{a_4=2}, \underbrace{[4]}_{a_5=1}.$$

- merge(sort) inteligente aprovecha los runs existentes!

La elección del modelo probabilista es un paso clave.

Fusión de dos runs

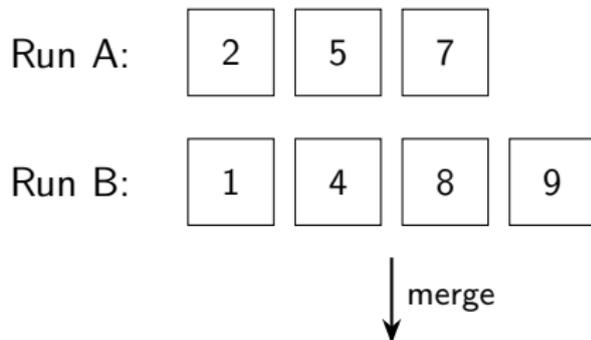
Run A:

2	5	7
---	---	---

Run B:

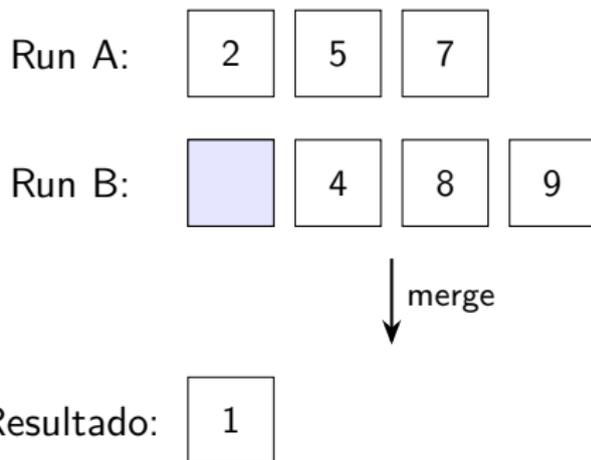
1	4	8	9
---	---	---	---

Fusión de dos runs

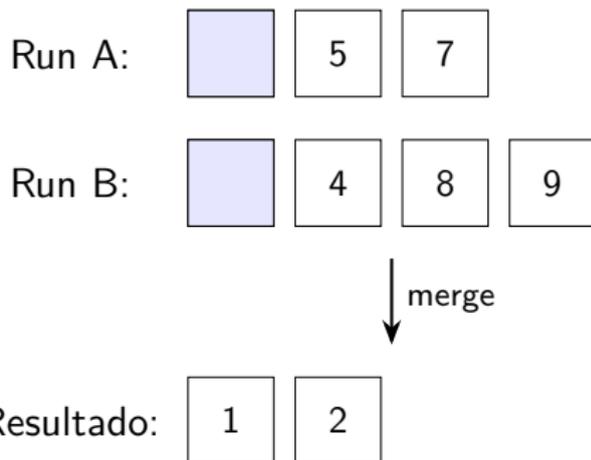


Resultado:

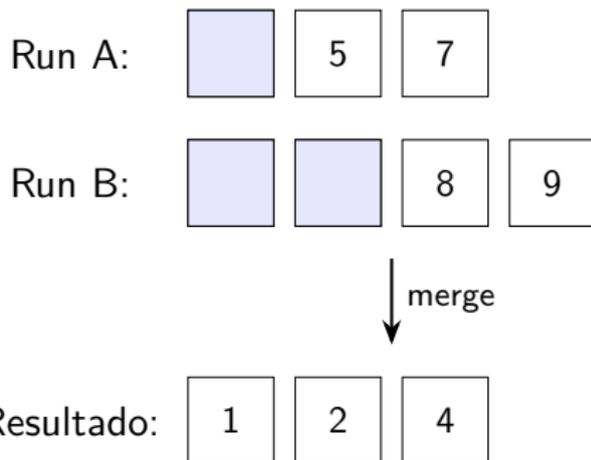
Fusión de dos runs



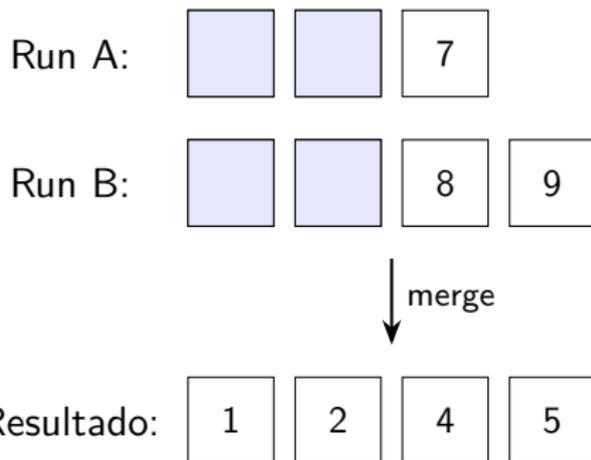
Fusión de dos runs



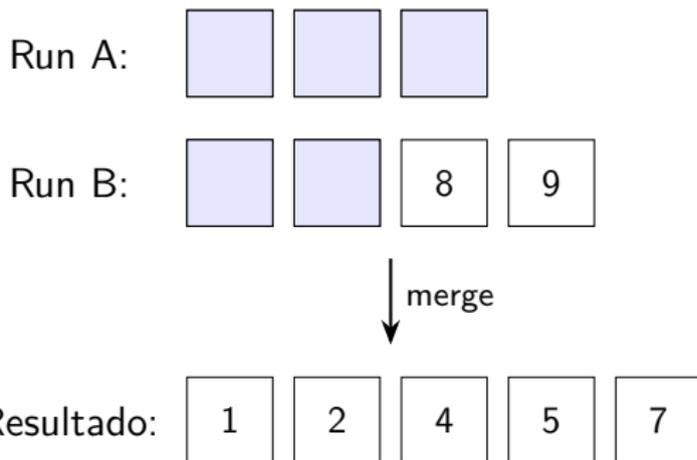
Fusión de dos runs



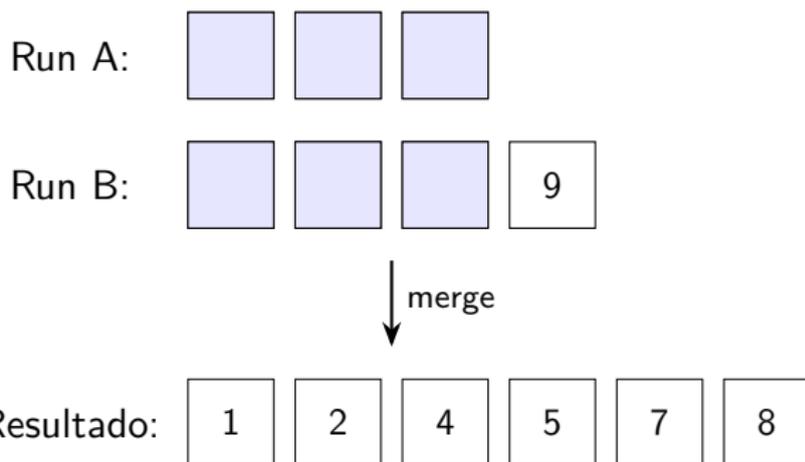
Fusión de dos runs



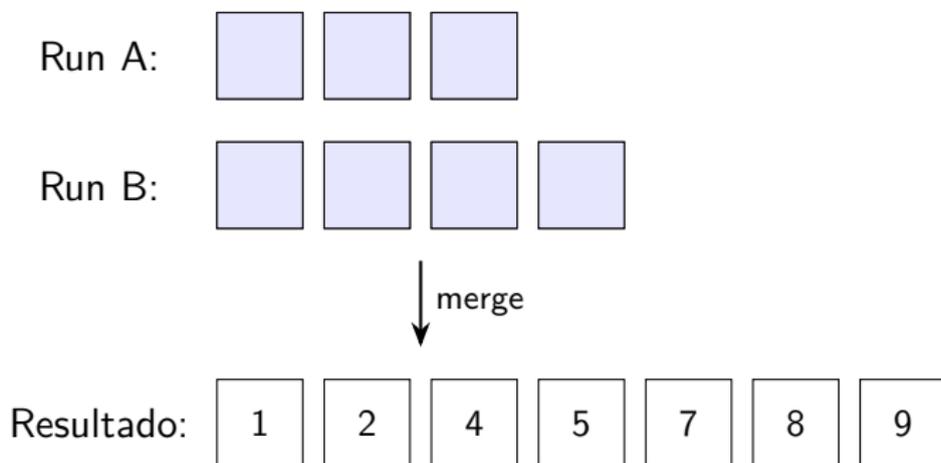
Fusión de dos runs



Fusión de dos runs



Fusión de dos runs



Suposición: la fusión (merge) de dos runs (corridas), de longitud a_1 y a_2 , cuesta $a_1 + a_2$.

Teorema

El costo C de cualquier algoritmo basado en la fusión de runs^a satisface

$$C(\pi) \geq n \cdot \mathcal{H}(\pi),$$

donde $\mathcal{H} = H_2(a_1/n, \dots, a_r/n) = -\sum \frac{a_i}{n} \log_2 \frac{a_i}{n}$ es la entropía de run de π .

^aSin contar la detección de runs.

Entropía de runs

Teorema

El costo C de cualquier algoritmo basado en la fusión de runs^a satisface

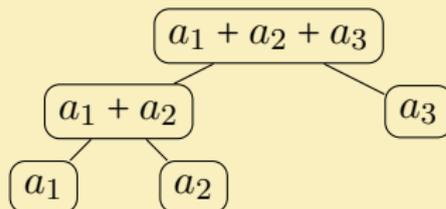
$$C(\pi) \geq n \cdot \mathcal{H}(\pi),$$

donde $\mathcal{H} = H_2(a_1/n, \dots, a_r/n) = -\sum \frac{a_i}{n} \log_2 \frac{a_i}{n}$ es la entropía de run de π .

^aSin contar la detección de runs.

Demostración.

- Estrategia de fusión corresponde a árbol binario \Rightarrow costo $C = \sum a_i l_i$.



$$C = 2 \cdot a_1 + 2 \cdot a_2 + 1 \cdot a_3$$

Entropía de runs

Teorema

El costo C de cualquier algoritmo basado en la fusión de runs^a satisface

$$C(\pi) \geq n \cdot \mathcal{H}(\pi),$$

donde $\mathcal{H} = H_2(a_1/n, \dots, a_r/n) = -\sum \frac{a_i}{n} \log_2 \frac{a_i}{n}$ es la entropía de run de π .

^aSin contar la detección de runs.

Demostración.

- Estrategia de fusión corresponde a árbol binario \Rightarrow costo $C = \sum a_i l_i$.
- Renormalizando obtenemos el resultado. □

Entropía de runs

Teorema

El costo C de cualquier algoritmo basado en la fusión de runs satisface

$$C(\pi) \geq n \cdot \mathcal{H}(\pi),$$

donde $\mathcal{H} = H_2(a_1/n, \dots, a_r/n) = -\sum \frac{a_i}{n} \log_2 \frac{a_i}{n}$ es la entropía de run de π .

\mathcal{H} puede ser mucho menor que $\log_2 n$.

Proposición

Tenemos $\mathcal{H} \leq \log_2 r$ donde r es la cantidad de runs.

Entropía de runs

Teorema

El costo C de cualquier algoritmo basado en la fusión de runs satisface

$$C(\pi) \geq n \cdot \mathcal{H}(\pi),$$

donde $\mathcal{H} = H_2(a_1/n, \dots, a_r/n) = -\sum \frac{a_i}{n} \log_2 \frac{a_i}{n}$ es la entropía de run de π .

\mathcal{H} puede ser mucho menor que $\log_2 n$.

Proposición

Tenemos $\mathcal{H} \leq \log_2 r$ donde r es la cantidad de runs.

\Rightarrow Existen varios algoritmos en tiempo $\Theta(n\mathcal{H} + n)$.

Entropía de runs

No se pierde mucho trabajando solo con fusiones.

Teorema (Barbay, Navarro, '13)

Sea $\mathcal{C} = \mathcal{C}(a_1, \dots, a_r)$ la clase de las permutaciones con runs de largo a_1, a_2, \dots, a_r , con $a_i \geq 2$ para $i = 1, \dots, r - 1$.

Para todo algoritmo \mathcal{A} basado en la comparación de pares de elementos, existe un elemento $\pi \in \mathcal{C}$ que requiere al menos $n\mathcal{H} - 3n$ comparaciones.

Entropía de runs

No se pierde mucho trabajando solo con fusiones.

Teorema (Barbay, Navarro, '13)

Sea $\mathcal{C} = \mathcal{C}(a_1, \dots, a_r)$ la clase de las permutaciones con runs de largo a_1, a_2, \dots, a_r , con $a_i \geq 2$ para $i = 1, \dots, r - 1$.

Para todo algoritmo \mathcal{A} basado en la comparación de pares de elementos, existe un elemento $\pi \in \mathcal{C}$ que requiere al menos $n\mathcal{H} - 3n$ comparaciones.

Borrador de prueba.

Siempre existe π que requiere al menos $\log_2 |\mathcal{C}|$ operaciones.

Se necesita una cota [no trivial^a], en este caso $2^{r-1} |\mathcal{C}| \geq \binom{n}{a_1, \dots, a_r}$. □

^aVer referencias, en particular <https://arxiv.org/pdf/1805.08612>

Tim Peters⁵ diseñó en 2002 un nuevo algoritmo para Python:

This describes an adaptive, stable, natural mergesort, modestly called timsort (hey, I earned it <wink>). It has supernatural performance on many kinds of partially ordered arrays (less than $\lg(N!)$ comparisons needed, and as few as $N-1$), yet as fast as Python's previous highly tuned samplesort hybrid on random arrays.

In a nutshell, the main routine marches over the array once, left to right, alternately identifying the next run, then merging it into the previous runs "intelligently". Everything else is complication for speed, and some hard-won measure of memory efficiency.

⁵<https://svn.python.org/projects/python/trunk/Objects/listsort.txt>

TimSort principio e historia

- Leer **runs** de izquierda a derecha, agregándolas a una **pila (stack)**.
- La pila $\rightarrow R_1, R_2, \dots$ debe satisfacer un *invariante*:
si el invariante no se cumple, desencadena secuencia de fusiones.
- Merges se realizan entre runs adyacentes (localidad/cache).

TimSort principio e historia

- Leer **runs** de izquierda a derecha, agregándolas a una **pila (stack)**.
- La pila $\rightarrow R_1, R_2, \dots$ debe satisfacer un *invariante*:
si el invariante no se cumple, desencadena secuencia de fusiones.
- Merges se realizan entre runs adyacentes (localidad/cache).

Invariante inspirado por **Fibonacci**:

$$r_{i+2} > r_i + r_{i+1}, \quad r_{i+1} > r_i,$$

donde $r_i = |R_i|$ son las longitudes.

- Varias condiciones de merge \Rightarrow originalmente con bugs ! 🐛
- Algoritmo era usado en Python [ahora PowerSort], usado en Java.
- Ha inspirado muchos algoritmos nuevos, basados en runs.

TimSort reglas e historia

Sea la pila $\rightarrow R_1, R_2, \dots$ y las longitudes $\rightarrow r_1, r_2, \dots$

while *runs fuera de pila* $\neq \emptyset$ **do**

tomar primer run (izq. a der.) y agregar a la pila ;

while *True* **do**

if $r_1 > r_3$ **then**

| merge R_2 and R_3

else if $r_1 \geq r_2$ **then**

| merge R_1 and R_2 ;

else if $r_1 + r_2 \geq r_3$ **then**

| merge R_1 and R_2 ;

else if $r_2 + r_3 \geq r_4$ **then**

| merge R_1 and R_2 ;

else

| **break**

TimSort reglas e historia

Sea la pila $\rightarrow R_1, R_2, \dots$ y las longitudes $\rightarrow r_1, r_2, \dots$

while *runs fuera de pila* $\neq \emptyset$ **do**

tomar primer run (izq. a der.) y agregar a la pila ;

while *True* **do**

if $r_1 > r_3$ **then**

| merge R_2 and R_3

else if $r_1 \geq r_2$ **then**

| merge R_1 and R_2 ;

else if $r_1 + r_2 \geq r_3$ **then**

| merge R_1 and R_2 ;

else if $r_2 + r_3 \geq r_4$ **then**

| merge R_1 and R_2 ;

else

| **break**

– La condición en rojo no estaba en la versión original. Bug descubierto por [de Gouw et al \(2015\)](#) cuando trataban de probar formalmente la corrección. 🦋

– Existía un segundo bug: el tamaño máximo de la pila en Java. Descubierto por [Auger et al \(2018\)](#) al realizar el análisis preciso del algoritmo. 🦋

Optimalidad en media y entropía

Teorema (Auger, Jugé, Nicaud, Pivoteau '18)

En el peor caso TimSort es $1,5 n\mathcal{H} + O(n)$.

Optimalidad en media y entropía

Teorema (Auger, Jugé, Nicaud, Pivoteau '18)

En el peor caso TimSort es $1,5n\mathcal{H} + O(n)$.

TimSort no es óptimo

Teorema (Wild, Munro '18)

En el peor caso PowerSort es $n\mathcal{H} + O(n)$.

Optimalidad en media y entropía

Teorema (Auger, Jugé, Nicaud, Pivoteau '18)

En el peor caso TimSort es $1,5n\mathcal{H} + O(n)$.

TimSort no es óptimo

Teorema (Wild, Munro '18)

En el peor caso PowerSort es $n\mathcal{H} + O(n)$.

Una permutación aleatoria (típica) tiene muchos runs cortos!

$n = 20$: [11, 18, 1, 5, 2, 14, 20, 3, 8, 15, 6, 4, 16, 17, 13, 10, 19, 9, 7, 12].

$$\mathcal{H} = 3,1\dots, \quad \log_2 20 = 4,3\dots$$

Probabilidad de run de largo $\geq k$

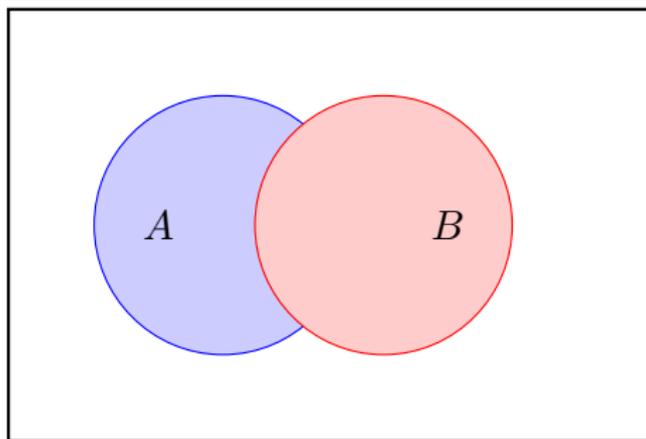
Sea S_i : run de longitud $\geq k$ comienza en i . Notar que $\Pr(S_i) \leq 2/k!$

Probabilidad de run de largo $\geq k$

Sea S_i : run de longitud $\geq k$ comienza en i . Notar que $\Pr(S_i) \leq 2/k!$

Técnica: Union bound

$$\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$$



Probabilidad de run de largo $\geq k$

Sea S_i : run de longitud $\geq k$ comienza en i . Notar que $\Pr(S_i) \leq 2/k!$

Técnica: Union bound

$$\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$$

Por el union bound tenemos:

$$P(n, k) := \Pr(\exists \text{run de longitud } \geq k) = \Pr\left(\bigcup_i S_i\right) \leq \sum_i \Pr(S_i) \leq 2n/k!.$$

Proposición

$$P = P(n, k) \leq 2 \exp(\log n - k \log k + k).$$

Entropía de corridas de permutación aleatoria

Utilizando

$$P = P(n, k) \leq 2 \exp(\log n - k \log k + k),$$

obtenemos que para $k \geq 2 \frac{\log n}{\log \log n}$, $P(n, k) \rightarrow 0$. Las runs son cortas !

Proposición

Con alta probabilidad (es decir $p \rightarrow 1$) todas las runs A_1, \dots, A_r de una permutación aleatoria uniforme satisfacen $A_i \leq 2 \frac{\log n}{\log \log n}$.

Entropía de corridas de permutación aleatoria

Utilizando

$$P = P(n, k) \leq 2 \exp(\log n - k \log k + k),$$

obtenemos que para $k \geq 2 \frac{\log n}{\log \log n}$, $P(n, k) \rightarrow 0$. Las runs son cortas !

Proposición

Con alta probabilidad (es decir $p \rightarrow 1$) todas las runs A_1, \dots, A_r de una permutación aleatoria uniforme satisfacen $A_i \leq 2 \frac{\log n}{\log \log n}$.

Corolario

Con alta probabilidad, para una permutación aleatoria uniforme^a,
 $\mathcal{H} \geq \sum \frac{A_i}{n} \log_2 \left(\frac{n}{2(\log n)/\log \log n} \right) = \log_2 n + O(\log \log n)$, $\mathcal{H} \leq \log_2 n$.

^aLa constante del término O en realidad se puede calcular explícitamente y no depende de la secuencia de conjuntos elegidos, cuya probabilidad tiende a 1.

Entropía de corridas de permutación aleatoria

Utilizando

$$P = P(n, k) \leq 2 \exp(\log n - k \log k + k),$$

obtenemos que para $k \geq 2 \frac{\log n}{\log \log n}$, $P(n, k) \rightarrow 0$. Las runs son cortas !

Proposición

Con alta probabilidad (es decir $p \rightarrow 1$) todas las runs A_1, \dots, A_r de una permutación aleatoria uniforme satisfacen $A_i \leq 2 \frac{\log n}{\log \log n}$.

Corolario

Con alta probabilidad, para una permutación aleatoria uniforme^a,
 $\mathcal{H} \geq \sum \frac{A_i}{n} \log_2 \left(\frac{n}{2(\log n)/\log \log n} \right) = \log_2 n + O(\log \log n)$, $\mathcal{H} \leq \log_2 n$.

^aLa constante del término O en realidad se puede calcular explícitamente y no depende de la secuencia de conjuntos elegidos, cuya probabilidad tiende a 1.

⇒ Modelo de permutaciones uniformes \neq modelo de runs largas

Con alta probabilidad y en media

Probamos que, con probabilidad $p \rightarrow 1$,

$$\mathcal{H} = \log_2 n + O(\log \log n),$$

es decir, que esto se cumple para $\pi \in A_n \subseteq S_n$ con $\Pr(A_n) \rightarrow 1$.

Pregunta

¿Qué podemos decir sobre la esperanza $\mathbb{E}[\mathcal{H}]$?

Con alta probabilidad y en media

Probamos que, con probabilidad $p \rightarrow 1$,

$$\mathcal{H} = \log_2 n + O(\log \log n),$$

es decir, que esto se cumple para $\pi \in A_n \subseteq S_n$ con $\Pr(A_n) \rightarrow 1$.

Pregunta

¿Qué podemos decir sobre la esperanza $\mathbb{E}[\mathcal{H}]$?

$$\begin{aligned}\mathbb{E}[\mathcal{H}] &= \Pr(A_n) \times \mathbb{E}[\mathcal{H} | A_n] + \Pr(A_n^c) \times \mathbb{E}[\mathcal{H} | A_n^c], \\ &\geq \Pr(A_n) \times \mathbb{E}[\mathcal{H} | A_n], \\ &= \Pr(A_n) \times (\log_2 n + O(\log \log n)) .\end{aligned}$$

Con alta probabilidad y en media

Probamos que, con probabilidad $p \rightarrow 1$,

$$\mathcal{H} = \log_2 n + O(\log \log n),$$

es decir, que esto se cumple para $\pi \in A_n \subseteq S_n$ con $\Pr(A_n) \rightarrow 1$.

Pregunta

¿Qué podemos decir sobre la esperanza $\mathbb{E}[\mathcal{H}]$?

$$\begin{aligned}\mathbb{E}[\mathcal{H}] &= \Pr(A_n) \times \mathbb{E}[\mathcal{H} | A_n] + \Pr(A_n^c) \times \mathbb{E}[\mathcal{H} | A_n^c], \\ &\geq \Pr(A_n) \times \mathbb{E}[\mathcal{H} | A_n], \\ &= \Pr(A_n) \times (\log_2 n + O(\log \log n)).\end{aligned}$$

Para la cota superior tenemos suerte: $\mathcal{H} \leq \log_2(n)$ siempre.

Conclusión: $\mathbb{E}[\mathcal{H}] \sim \log_2 n$ también.

Problema: número de runs

```
def runs(arr): # arr = permutacion
    res = []
    i, n = 0, len(arr)
    while i < n:
        j = i + 1
        if j < n and arr[i] <= arr[j]:
            # creciente
            while j < n and arr[j - 1] <= arr[j]:
                j += 1
        elif j < n and arr[i] > arr[j]:
            # decreciente
            while j < n and arr[j - 1] > arr[j] :
                j += 1
        else:
            # elemento aislado
            j = i + 1
        res.append(j - i)
        i = j
    return res
```

Problema: número de runs

Problema

La cantidad esperada de runs es $\mathbb{E}[r] \sim cn$ para una cierta $c > 0$.

Veamos la permutación como una secuencia X_1, X_2, \dots de números iid de $[0, 1]$.

- (a) Probar $runs(X_1, \dots, X_{i+j}) \leq runs(X_1, \dots, X_i) + runs(X_{i+1}, \dots, X_{i+j})$.
- (b) Probar que $e_k := \mathbb{E}[runs(X_1, \dots, X_k)]$ satisface $e_{i+j} \leq e_i + e_j$ para todo $i, j \geq 0$. Concluir que $e_k/k \rightarrow c$ para cierta $c \geq 0$.
- (c) Mostrar que la constante es positiva $c > 0$.

⁵Pista (b). Lema de Fekete...

⁵Pista (c). ¿Qué podemos decir si $X_i < X_{i+1}$ y $X_{i+1} > X_{i+2}$?

Entropía de corrida: modelo aleatorio

Distribución de Zipf

Dado $\alpha > 1$, consideramos

$$\Pr(\ell = k) \propto k^{-\alpha}.$$

Cuando $\alpha \leq 2$, la longitud esperada de ℓ es infinita.

Entropía de corrida: modelo aleatorio

Distribución de Zipf

Dado $\alpha > 1$, consideramos

$$\Pr(\ell = k) \propto k^{-\alpha}.$$

Cuando $\alpha \leq 2$, la longitud esperada de ℓ es infinita.

- Valores más irregulares. Ejemplo con $\alpha = 3/2$

1	1	2	1	6	8	9	14	3	5
953	1	6	32	2	24	1	1	3	1
21	1	26	2	1	1	9	2	49	4
1	1	1	1	2	48	68	4	189	2

- Usada para modelar frecuencias de palabras en lenguaje natural.

Entropía de corrida: modelo aleatorio

Distribución de Zipf

Dado $\alpha > 1$, consideramos

$$\Pr(\ell = k) \propto k^{-\alpha}.$$

Cuando $\alpha \leq 2$, la longitud esperada de ℓ es infinita.

- Valores más irregulares. Ejemplo con $\alpha = 3/2$

1	1	2	1	6	8	9	14	3	5
953	1	6	32	2	24	1	1	3	1
21	1	26	2	1	1	9	2	49	4
1	1	1	1	2	48	68	4	189	2

- Usada para modelar frecuencias de palabras en lenguaje natural.

¿Modelo más razonable? ¿Producir permutación con longitudes dadas?

Para aprender más

-  *Nicolas Auger, Vincent Jugé, Cyril Nicaud, y Carine Pivoteau,*
On the Worst-Case Complexity of TimSort
<https://arxiv.org/pdf/1805.08612>
-  *Jérémy Barbay y Gonzalo Navarro,*
On compressing permutations and adaptive sorting.
<http://dx.doi.org/10.1016/j.tcs.2013.10.019>
-  *Nearly-Optimal Mergesorts: Fast, Practical Sorting Methods That Optimally Adapt to Existing Runs,*
<https://doi.org/10.4230/LIPIcs.ESA.2018.63>

Tablas de Hash

Motivación

Implementar un array asociativo m :

- universo \mathcal{U} de claves $k \in \mathcal{U}$ grande,
- asociar a cada clave k un valor $m[k]$,
- insertar, buscar, borrar...

Tablas de Hash

Motivación

Implementar un array asociativo m :

- universo \mathcal{U} de claves $k \in \mathcal{U}$ grande,
- asociar a cada clave k un valor $m[k]$,
- insertar, buscar, borrar...

Las *tablas de Hash*:

- **Idea** : utilizar un array A pequeño, de tamaño $K \ll |\mathcal{U}|$
 - considerar una función $h : \mathcal{U} \rightarrow \mathbb{Z}$ *pseudo-aleatoria*,
 - insertar k en $A[i]$ donde $i = h(k) \bmod K$. [modelo : i es uniforme]
- **Problema** : colisiones, dos keys k_1 y k_2 con $h(k_1) = h(k_2)$.

La paradoja del cumpleaños

Las colisiones están relacionadas con la famosa *paradoja del cumpleaños*:

Paradoja del cumpleaños

¿Cuál es el **número mínimo de personas** requerido para que la probabilidad de que dos o más personas tengan el mismo cumpleaños^a sea mayor que 1/2?

^asolo el día del año, no el año

La paradoja del cumpleaños

Las colisiones están relacionadas con la famosa *paradoja del cumpleaños*:

Paradoja del cumpleaños

¿Cuál es el **número mínimo de personas** requerido para que la probabilidad de que dos o más personas tengan el mismo cumpleaños^a sea mayor que $1/2$?

^asolo el día del año, no el año

Más en general, ¿cuántas personas para la primera “colisión”?

La paradoja del cumpleaños

Las colisiones están relacionadas con la famosa *paradoja del cumpleaños*:

Paradoja del cumpleaños

¿Cuál es el **número mínimo de personas** requerido para que la probabilidad de que dos o más personas tengan el mismo cumpleaños^a sea mayor que $1/2$?

^asolo el día del año, no el año

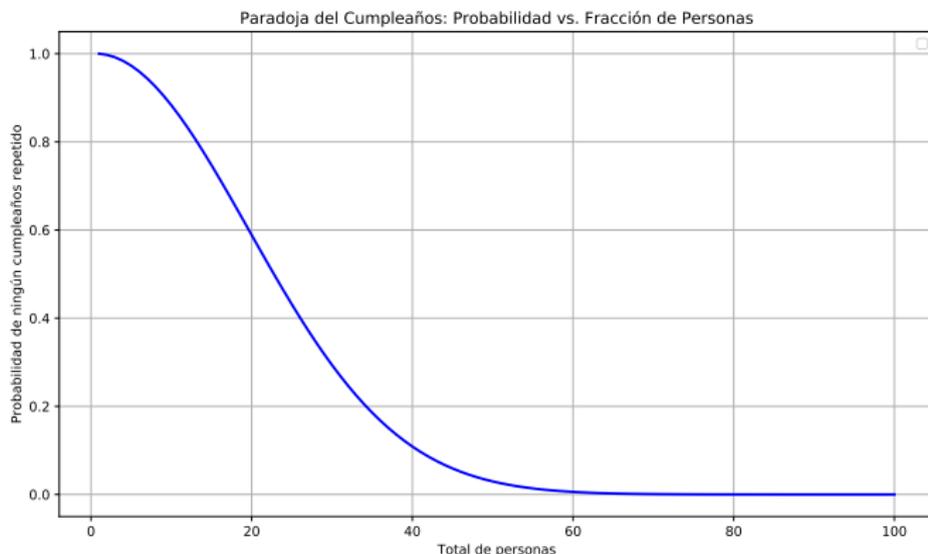
Más en general, ¿cuántas personas para la primera “colisión”?

- Supongamos que tenemos K posibles valores ($K = 365$),
- y consideramos n elementos (las personas),
- ¿cuál es la probabilidad de que hayan dos elementos iguales?

Modelo: cada valor tiene probabilidad $1/K$, elementos independientes.

La paradoja del cumpleaños

$$p_n = \Pr(n \text{ valores distintos}) = \prod_{i=1}^{n-1} \left(1 - \frac{i}{K}\right).$$



La probabilidad de al menos un cumpleaños repetido es $q_n = 1 - p_n$.

La paradoja del cumpleaños

Estimemos la probabilidad de n valores distintos: $p_n = \prod_{i=1}^{n-1} \left(1 - \frac{i}{K}\right)$,

La paradoja del cumpleaños

Estimemos la probabilidad de n valores distintos: $p_n = \prod_{i=1}^{n-1} \left(1 - \frac{i}{K}\right)$,

Proposición

Considerando $n \sim \sqrt{2\theta K}$ con $K \rightarrow \infty$, $p_n \sim e^{-\theta}$.

La paradoja del cumpleaños

Estimemos la probabilidad de n valores distintos: $p_n = \prod_{i=1}^{n-1} \left(1 - \frac{i}{K}\right)$,

Proposición

Considerando $n \sim \sqrt{2\theta K}$ con $K \rightarrow \infty$, $p_n \sim e^{-\theta}$.

Demostración.

Para la prueba usamos las desigualdades $1 + x \leq e^x$, válida para $x \in \mathbb{R}$, y $1 + x \geq e^{x-x^2/2}$, válida para $x \in [0, 1]$. □

La paradoja del cumpleaños

Estimemos la probabilidad de n valores distintos: $p_n = \prod_{i=1}^{n-1} \left(1 - \frac{i}{K}\right)$,

Proposición

Considerando $n \sim \sqrt{2\theta K}$ con $K \rightarrow \infty$, $p_n \sim e^{-\theta}$.

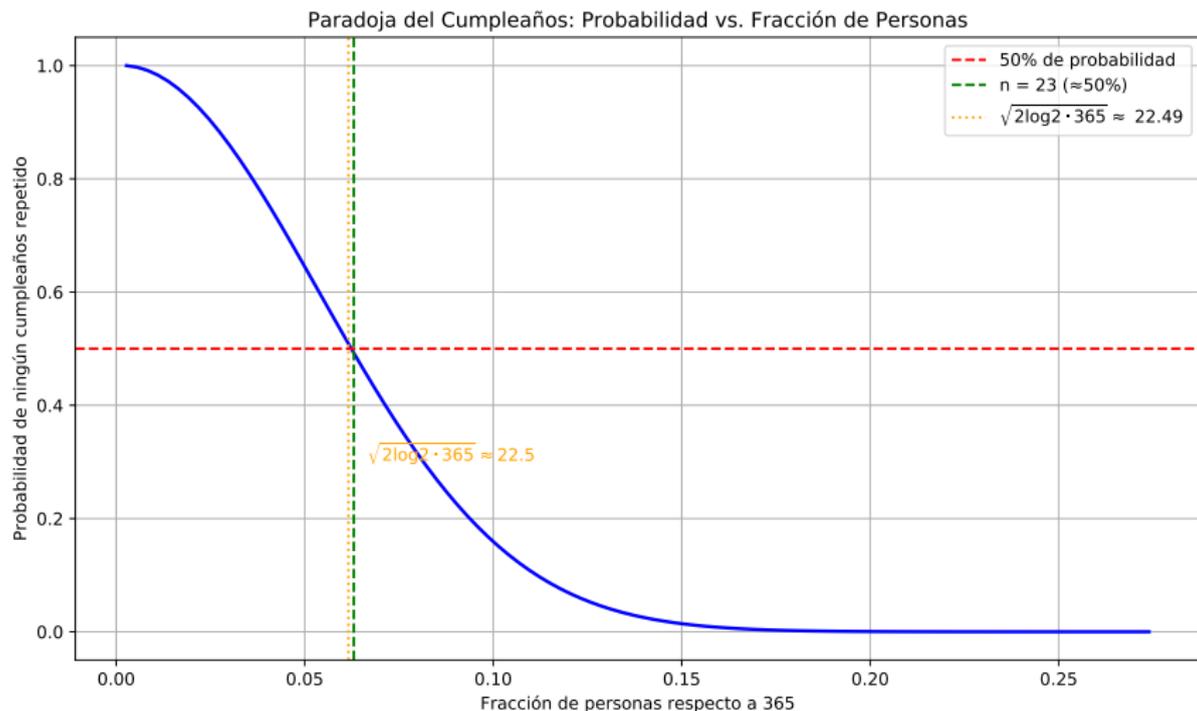
Demostración.

Para la prueba usamos las desigualdades $1 + x \leq e^x$, válida para $x \in \mathbb{R}$, y $1 + x \geq e^{x-x^2/2}$, válida para $x \in [0, 1]$. □

Primera colisión ocurre (con gran proba.) cuando n es de orden \sqrt{K} .

La paradoja del cumpleaños

Ilustración de la aproximación: $n \sim \sqrt{2\theta K}$, $p_n \sim e^{-\theta}$ con $\theta = \log 2$



Tablas de Hash

Tablas de Hash requieren un mecanismo de resolución de colisiones:

- **Política de resolución de colisiones :**

1. [External Hashing] Cada célula $A[i]$ contiene una lista encadenada.
2. [Internal Hashing / Open addressing] Si la célula está ya ocupada, buscar otra en el mismo array.

- **Política de rehashing :**

- si tasa de ocupación⁶ del array A es alta, nuevo array de tamaño mayor,
- necesario re-insertar todo. [paso lento!]

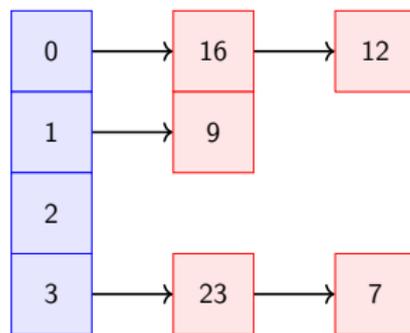
⁶“Load factor” en inglés.

Tablas de Hash

Tablas de Hash requieren un mecanismo de resolución de colisiones:

- **Política de resolución de colisiones :**

1. [External Hashing] Cada célula $A[i]$ contiene una lista encadenada.



2. [Internal Hashing / Open addressing] Si la célula está ya ocupada, buscar otra en el mismo array.

- **Política de rehashing :**

- si tasa de ocupación⁶ del array A es alta, nuevo array de tamaño mayor,
- necesario re-insertar todo. [paso lento!]

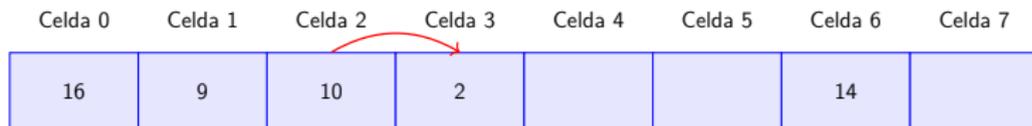
⁶“Load factor” en inglés.

Tablas de Hash

Tablas de Hash requieren un mecanismo de resolución de colisiones:

■ Política de resolución de colisiones :

1. [External Hashing] Cada célula $A[i]$ contiene una lista encadenada.
2. [Internal Hashing / Open addressing] Si la célula está ya ocupada, buscar otra en el mismo array.



[la flecha roja es indicativa]

■ Política de rehashing :

- si tasa de ocupación⁶ del array A es alta, nuevo array de tamaño mayor,
- necesario re-insertar todo. [paso lento!]

⁶“Load factor” en inglés.

Rehashing

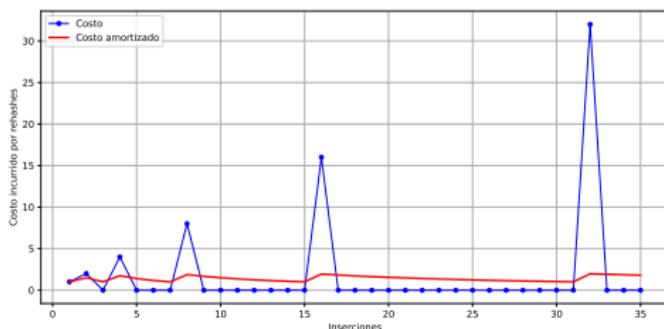
Cuando el load factor $\alpha = n/K$ excede un valor dado γ (p.e., $\gamma = 0,85$), considerar un array nuevo con capacidad⁷ $K' = 2K$.

Proposición

El costo amortizado por inserción es constante.

Concepto: costo amortizado

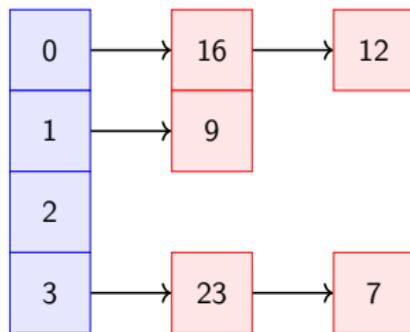
En lugar de considerar el costo de una sola operación c_t , nos interesa el costo medio de la secuencia total de operaciones $\frac{1}{T} \sum_{t=1}^T c_t$.



⁷Aquí no consideramos que el tamaño puede reducirse.

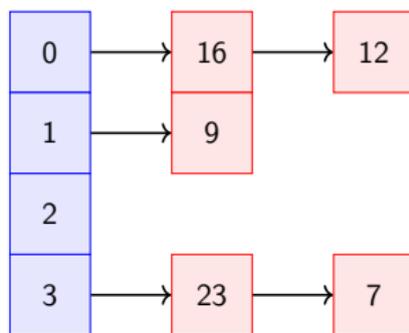
External hashing

Cada celda $A[i]$ contiene una lista encadenada



External hashing

Cada celda $A[i]$ contiene una lista encadenada

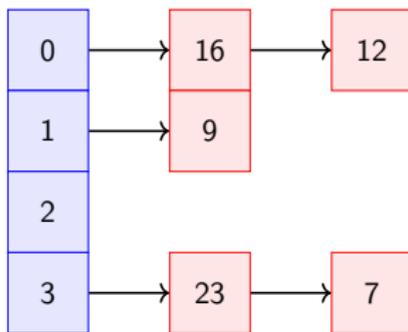


Observación

Cada celda contiene en media $\alpha = n/K$ elementos.

External hashing

Cada celda $A[i]$ contiene una lista encadenada



Observación

Cada celda contiene en media $\alpha = n/K$ elementos.

Proposición (Lookup)

Con alta probabilidad ($K \rightarrow \infty$), ninguna lista tiene longitud mayor que $2 \frac{\log K}{\log \log K}$.

External hashing

Consideremos solo inserciones. Sea $\gamma > 0$ la tasa de ocupación máxima, $n/K \leq \gamma$.

Proposición (Lookup)

Con alta probabilidad ($K \rightarrow \infty$), ninguna lista tiene longitud mayor que $2 \frac{\log K}{\log \log K}$.

External hashing

Consideremos solo inserciones. Sea $\gamma > 0$ la tasa de ocupación máxima, $n/K \leq \gamma$.

Proposición (Lookup)

Con alta probabilidad ($K \rightarrow \infty$), ninguna lista tiene longitud mayor que $2 \frac{\log K}{\log \log K}$.

Sketch de demostración.

Sea X_1, \dots, X_n la secuencia de celda elegidas para las n inserciones.

1. Consideramos celda C_0 , y sea $C_0(n)$ la lista luego de n inserciones.
2. Probamos $\Pr(|C_0(n)| \geq m) \leq \binom{n}{m} K^{-m}$ por el union-bound.
3. Si consideramos cualquier celda, por el union-bound

$$P_m := \Pr(\exists j : |C_j(n)| \geq m) \leq K \times \binom{n}{m} K^{-m}.$$

Ahora, tomando $m = 2 \frac{\log K}{\log \log K}$ probamos que la cota tiende a 0. □

Internal hashing / Open addressing

Internal hashing: Si la celda está ya ocupada, buscar otra en el mismo array.

- Internal hashing / Open addressing es más común en la actualidad.
- Muchas estrategias para decidir la secuencia (probe sequence).

Internal hashing / Open addressing

Internal hashing: Si la celda está ya ocupada, buscar otra en el mismo array.

- Internal hashing / Open addressing es más común en la actualidad.
- Muchas estrategias para decidir la secuencia (probe sequence).

Probing sequence / secuencia de búsqueda

Para buscar/insertar un elemento x :

- Comenzar por $i_0 = h(x) \text{ mód } K$.
- Si posición ocupada por otra clave, seguir para i_1, i_2, \dots etc.

Internal hashing / Open addressing

Internal hashing: Si la celda está ya ocupada, buscar otra en el mismo array.

- Internal hashing / Open addressing es más común en la actualidad.
- Muchas estrategias para decidir la secuencia (probe sequence).

Probing sequence / secuencia de búsqueda

Para buscar/insertar un elemento x :

- Comenzar por $i_0 = h(x) \text{ mód } K$.
- Si posición ocupada por otra clave, seguir para i_1, i_2, \dots etc.

Módulo K ,

- **Linear probing:** $i_1 = i_0 + 1, i_2 = i_1 + 1, \dots$
- **Quadratic probing:** $i_1 = i_0 + 1, i_2 = i_1 + 2, \dots, i_j = i_{j-1} + j, \dots$
- **Double hashing:** $\Delta(x) = h_2(x), i_1 = i_0 + \Delta, i_2 = i_1 + \Delta, \dots$

Secuencia de búsqueda: modelo

El comportamiento de *linear* y *quadratic probing* es complejo:

- **Linear probing** presenta el llamado *primary clustering*, pero aprovecha localidad (memoria cache).



Secuencia de búsqueda: modelo

El comportamiento de *linear* y *quadratic probing* es complejo:

- **Linear probing** presenta el llamado *primary clustering*, pero aprovecha localidad (memoria cache).



- **Quadratic probing** se comporta inicialmente en modo similar a linear probing, pero luego los saltos aumentan en tamaño.



Secuencia de búsqueda: modelo

El comportamiento de *linear* y *quadratic probing* es complejo:

- **Linear probing** presenta el llamado *primary clustering*, pero aprovecha localidad (memoria cache).



- **Quadratic probing** se comporta inicialmente en modo similar a linear probing, pero luego los saltos aumentan en tamaño.



Modelos simplificados para el análisis:

- **Random probing**: secuencia de búsqueda de números aleatorios uniformes (incluso repetidos).
- **Uniform probing**: secuencia de búsqueda es una permutación $\pi \in \mathcal{S}_K$ aleatoria.

Secuencia de búsqueda: modelo

Parámetros de interés [análisis sin supresiones]

1. Búsqueda exitosa: buscar un elemento presente.
2. Búsqueda no exitosa: buscar un elemento no presente.

Secuencia de búsqueda: modelo

Parámetros de interés [análisis sin supresiones]

1. Búsqueda exitosa: buscar un elemento presente.
2. Búsqueda no exitosa: buscar un elemento no presente.

First Come First Serve (FCFS)

Los elementos se insertan donde termina su búsqueda no exitosa.

Es decir, los elementos ya insertados no se desplazan

En tiempo: Inserción n -ésima = búsqueda no exitosa con $n - 1$ elementos

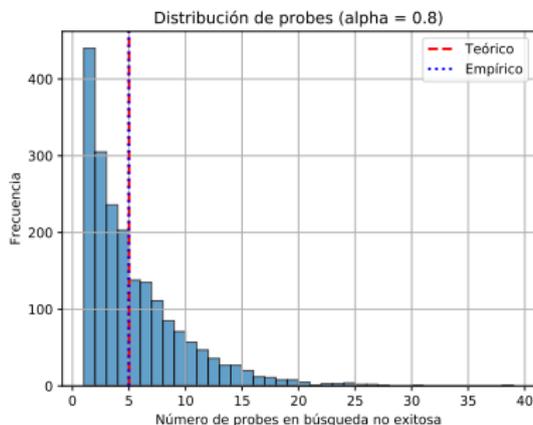
Random probing: búsqueda no exitosa

Buscar un elemento no presente corresponde a una inserción.

Teorema

El costo medio de una búsqueda no exitosa, cuando hay n elementos, es

$$U_n = \frac{1}{1 - \alpha}, \quad \alpha = \frac{n}{K}.$$



No hay concentración: ley \sim geométrica.

Random probing: búsqueda exitosa

Teorema

El costo medio de una búsqueda exitosa es

$$S_n = \frac{1}{\alpha} \log \left(\frac{1}{1-\alpha} \right) + O(n^{-1}).$$

Random probing: búsqueda exitosa

Teorema

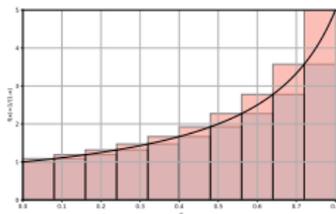
El costo medio de una búsqueda exitosa es

$$S_n = \frac{1}{\alpha} \log \left(\frac{1}{1-\alpha} \right) + O(n^{-1}).$$

Técnica: aproximar sumas con integrales

Si f es positiva, monótona y acotada en $[a, b]$:

$$\sum_{j=a \cdot N}^{b \cdot N - 1} f\left(\frac{j}{N}\right) \cdot \frac{1}{N} = \int_a^b f(x) dx + O(N^{-1}).$$



Uniform hashing

En Uniform Hashing las secuencias de búsqueda son permutaciones de \mathcal{S}_K

- Eliminar la posibilidad de elementos repetidos **no cambia sustancialmente el resultado**.
- Esto es **esperado**: si la secuencia de búsqueda es $\ll \sqrt{K}$ no esperamos repetidos (**paradoja del cumpleaños**).

Uniform hashing

En Uniform Hashing las secuencias de búsqueda son permutaciones de \mathcal{S}_K

- Eliminar la posibilidad de elementos repetidos **no cambia sustancialmente el resultado**.
- Esto es **esperado**: si la secuencia de búsqueda es $\ll \sqrt{K}$ no esperamos repetidos (**paradoja del cumpleaños**).

Teorema (Búsqueda en Uniform hashing, Peterson '57)

El costo medio de una búsqueda con uniform hashing es

$$U_n = \frac{K+1}{K-n+1} \sim \frac{1}{1-\alpha}, \quad S_n \sim \frac{1}{\alpha} \log\left(\frac{1}{1-\alpha}\right).$$

Linear probing

Linear probing es más complejo

Teorema (Búsqueda en Linear probing, Knuth '63)

$$\text{No exitosa} \sim \frac{1}{2} \left(1 + \frac{1}{(1-\alpha)^2} \right), \quad \text{Exitosa} \sim \frac{1}{2} \left(1 + \frac{1}{1-\alpha} \right).$$

Linear probing

Linear probing es más complejo

Teorema (Búsqueda en Linear probing, Knuth '63)

$$\text{No exitosa} \sim \frac{1}{2} \left(1 + \frac{1}{(1-\alpha)^2} \right), \quad \text{Exitosa} \sim \frac{1}{2} \left(1 + \frac{1}{1-\alpha} \right).$$

Se trata del artículo histórico de Donald. E. Knuth que da origen al área:

NOTES ON "OPEN" ADDRESSING

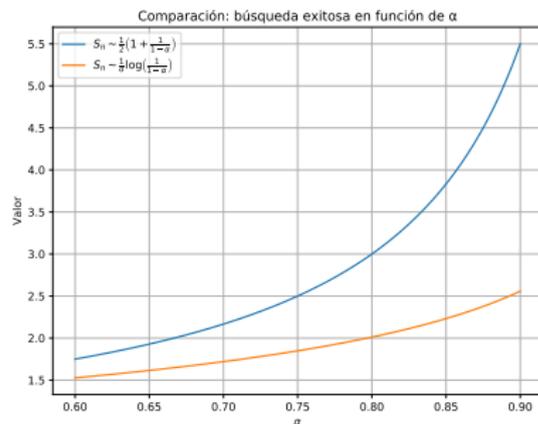
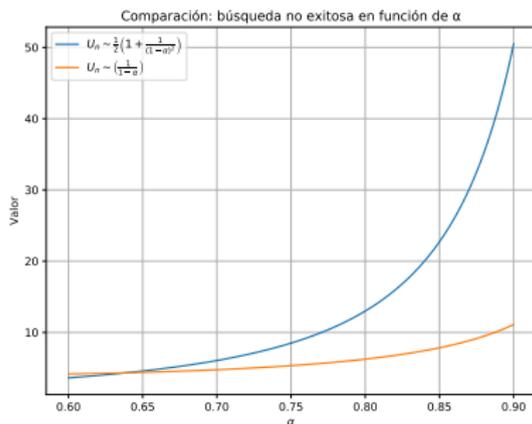
[MY FIRST ANALYSIS OF AN ALGORITHM, ORIGINALLY

DONE DURING SUMMER 1962 IN MADISON.]

DON KNUTH 7/22/63

1. Introduction and Definitions. Open addressing is a widely-used technique for keeping "symbol tables." The method was first used in 1954 by Samuel, Amdahl, and Boehme in an assembly program for the IBM 701. An extensive discussion of the method was given by Peterson in 1957 [1], and frequent references have been made to it ever since (e.g. Schay and Spruth [2], Iverson [3]). However, the timing characteristics have apparently never been exactly established, and indeed the author has heard reports of several reputable mathematicians who failed to find the solutions after some trial. Therefore it is the purpose of this note to indicate one way by which the solution can be obtained.

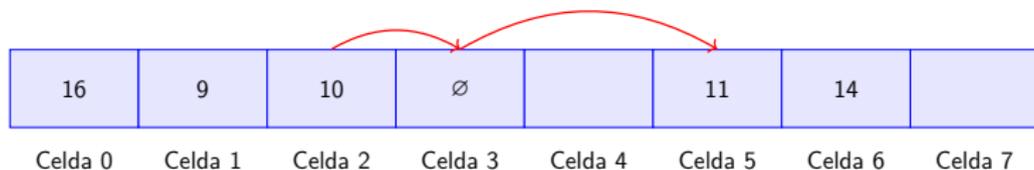
Comparación de tiempos de búsqueda según α



Y si hay supresiones

Para borrar:

- Introducir *tombstones* (marcas especiales) para indicar que la celda alguna vez fue ocupada,



[las flechas rojas son solo indicativas]

- Las tombstones ocupan una celda, y se cuenta para los rehashings.
- Se puede insertar un elemento en un tombstone.

Para aprender más

-  *Donald E. Knuth*,
The Art of Computer Programming, Vol. 3: Sorting and Searching.
-  *Donald E. Knuth*
Notes on Open Addressing.
<https://jeffe.cs.illinois.edu/teaching/datastructures/2011/notes/knuth-OALP.pdf>
-  *Conrado Martínez, Cyril Nicaud y Pablo Rotondo*
Mathematical models to analyze Lua hybrid tables.
Preprint <https://arxiv.org/abs/2208.13602>

Outline

1. Introducción al análisis de algoritmos
2. Aplicaciones a la predicción de saltos
 - MinMax: records en permutaciones
 - Exponenciación sesgada
3. Combinatoria Analítica: métodos simbólicos
4. Combinatoria Analítica: Teoría Analítica

MinMax: un ejemplo paradójico

Sean los algoritmos siguientes para encontrar simultáneamente el mínimo y el máximo de un array T de largo n .

```
min = max = T[0];
for(i = 1; i < n; i++) {
    if (T[i] < min)
        min = T[i];
    if (T[i] > max)
        max = T[i];
}
```

MinMax “ingenuo”

$2n - 2$ comparaciones

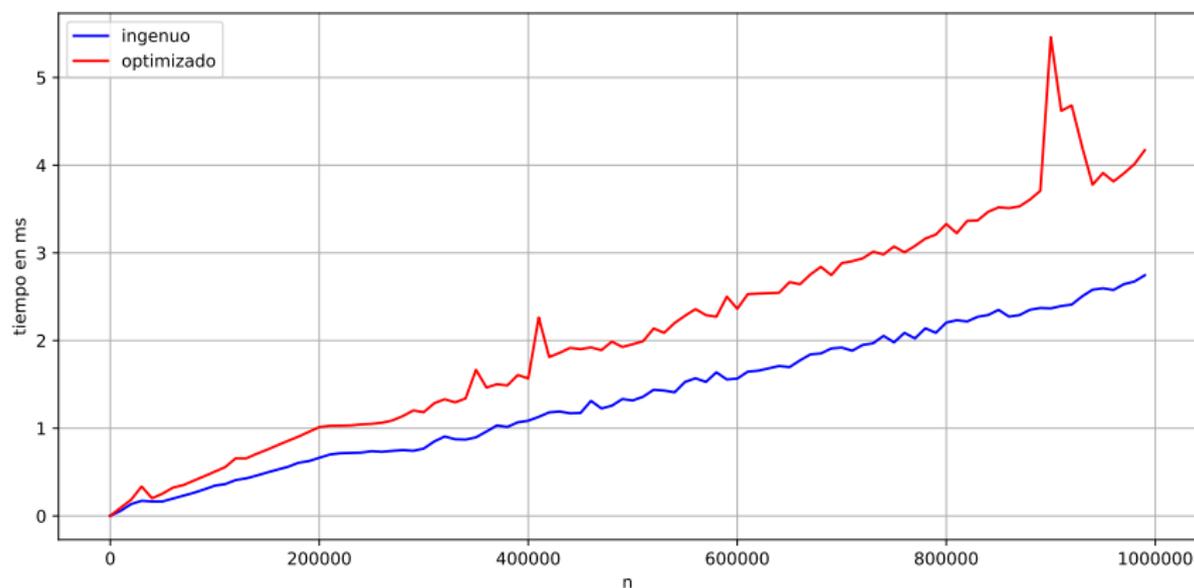
```
min = max = T[n-1];
for(i = 0; i < n - 1; i += 2) {
    if (T[i] < T[i+1]) {
        if (T[i] < min)
            min = T[i];
        if (T[i+1] > max)
            max = T[i+1];
    } else {
        if (T[i+1] < min)
            min = T[i+1];
        if (T[i] > max)
            max = T[i];
    }
}
```

MinMax “optimizado”

$\sim \frac{3}{2}n$ comparaciones

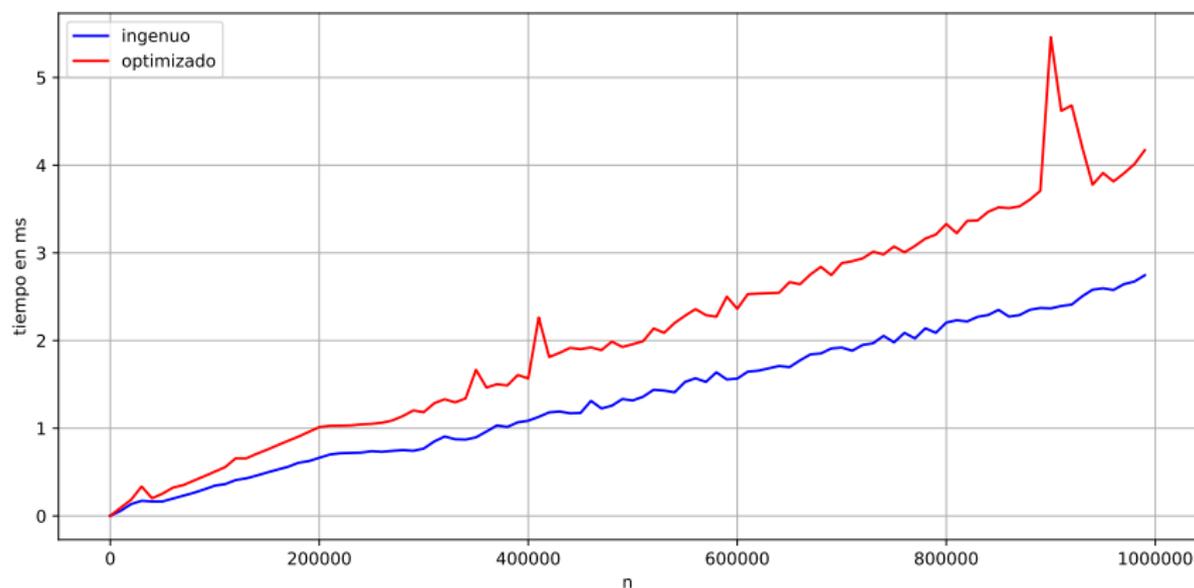
MinMax: resultados prácticos para los algoritmos

Considerando T como una permutación aleatoria:



MinMax: resultados prácticos para los algoritmos

Considerando T como una permutación aleatoria:



¿Por qué? ¿modelo?

Optimizaciones de “bajo nivel”

La arquitectura de la computadora incluye varias optimizaciones:

- La jerarquía de memoria (*memoria cache*),
- Operaciones *SIMD* (Single Instruction, Multiple Data),
- El [pipeline](#) del procesador.

Optimizaciones de “bajo nivel”

La arquitectura de la computadora incluye varias optimizaciones:

- La jerarquía de memoria (*memoria cache*),
- Operaciones *SIMD* (Single Instruction, Multiple Data),
- El **pipeline** del procesador.

En nuestro caso no hay SIMD, y acceso a memoria es esencialmente el mismo en los dos algoritmos:

⇒ nos vamos a concentrar en el pipeline.

Optimizaciones de “bajo nivel”

La arquitectura de la computadora incluye varias optimizaciones:

- La jerarquía de memoria (*memoria cache*),
- Operaciones *SIMD* (Single Instruction, Multiple Data),
- El **pipeline** del procesador.

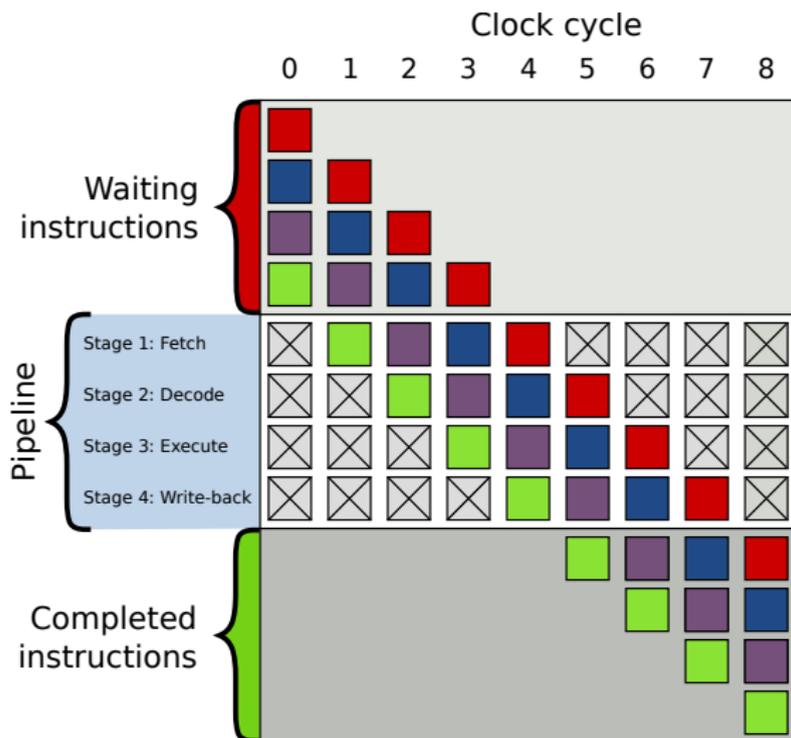
En nuestro caso no hay SIMD, y acceso a memoria es esencialmente el mismo en los dos algoritmos:

⇒ nos vamos a concentrar en el pipeline.

El pipeline del procesador:

- ejecutar una instrucción requiere *fetch*, *decode*, *execute*, *write*:
traer instrucción de memoria, decodificar, ejecutar, escribir
- en un **ciclo de reloj** se pueden realizar en **paralelo** para **varias instrucciones** sucesivas, en distintas etapas del pipeline.

El pipeline del procesador



⁷Fuente: Wikipedia, por en:User:Cburnett, **CC BY-SA 3.0**,
<https://commons.wikimedia.org/w/index.php?curid=1499754>

El pipeline del procesador

Problema. un if provoca un dilema:

¿qué rama (branch) de ejecución tomar (fetch)?

⇒ error de predicción provoca pérdida del pipeline (paralelismo)

El pipeline del procesador

Problema. un if provoca un dilema:

¿qué rama (branch) de ejecución tomar (fetch)?

⇒ error de predicción provoca pérdida del pipeline (paralelismo)

Branch prediction. diseñar esquemas para **predecir** el resultado de un if

- Locales (cada if separado), globales, mixtos, ...

El pipeline del procesador

Problema. un if provoca un dilema:

¿qué rama (branch) de ejecución tomar (fetch)?

⇒ error de predicción provoca pérdida del pipeline (paralelismo)

Branch prediction. diseñar esquemas para **predecir** el resultado de un if

- Locales (cada if separado), globales, mixtos, ...
- Memoria: ¿cuánta historia recuerda un predictor?

El pipeline del procesador

Problema. un if provoca un dilema:

¿qué rama (branch) de ejecución tomar (fetch)?

⇒ error de predicción provoca pérdida del pipeline (paralelismo)

Branch prediction. diseñar esquemas para **predecir** el resultado de un if

- Locales (cada if separado), globales, mixtos, ...
- Memoria: ¿cuánta historia recuerda un predictor?

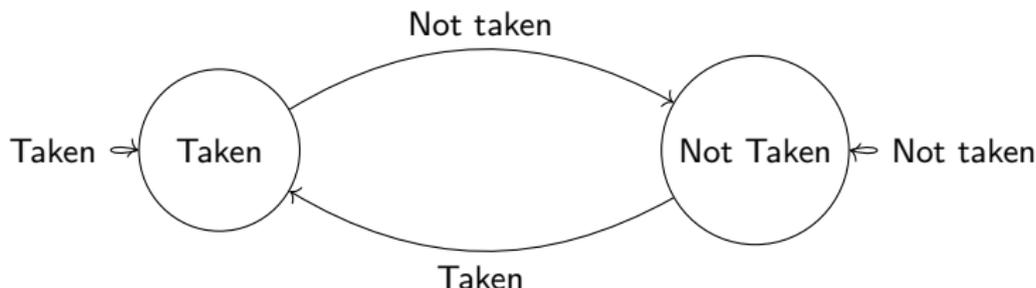
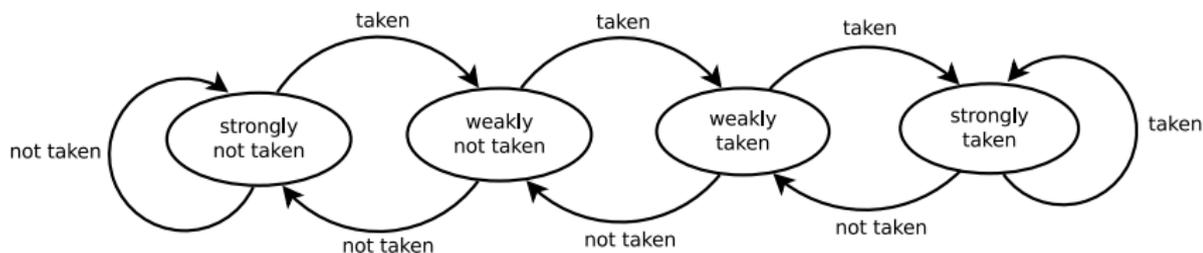


Figura: Predictor 1 Bit

Esquemas de predicción de branching

Por simplicidad consideraremos los siguientes predictores **locales**:

- Predictor de 1 bit [pagina precedente],
- Predictor de 2 bits saturado



- Predictor de 3 bits saturado ...

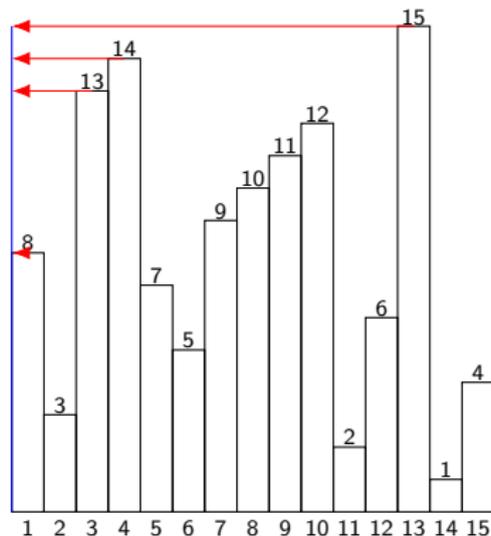
⁷Fuente: Wikipedia, Afog derivative work: ENORMATOR (talk), **CC BY-SA 3.0**, File:Branch_prediction_2bit_saturating_counter-dia.svg

Records en permutaciones

Errores de predicción en los dos MinMax relacionados con los **records**

Definición

Una posición k en una permutación π es un record máximo (mínimo) si $\pi_i < \pi_k$ (resp. $\pi_i > \pi_k$) para todo $i < k$.



Records en permutaciones

Errores de predicción en los dos MinMax relacionados con los **records**

Definición

Una posición k en una permutación π es un record máximo (mínimo) sii $\pi_i < \pi_k$ (resp. $\pi_i > \pi_k$) para $i < k$.

Records en permutaciones

Errores de predicción en los dos MinMax relacionados con los **records**

Definición

Una posición k en una permutación π es un record máximo (mínimo) sii $\pi_i < \pi_k$ (resp. $\pi_i > \pi_k$) para $i < k$.

- a) En la línea (3), la condición es verdadera sii i es un record de mínimo.
- b) En la línea (5), la condición es verdadera sii i es un record de máximo.
- ```
1 | min = max = T[0];
2 | for(i = 1; i < n; i++) {
3 | if (T[i] < min)
4 | min = T[i];
5 | if (T[i] > max)
6 | max = T[i];
7 | }
```

# Records en permutaciones

Errores de predicción en los dos MinMax relacionados con los **records**

## Definición

Una posición  $k$  en una permutación  $\pi$  es un record máximo (mínimo) sii  $\pi_i < \pi_k$  (resp.  $\pi_i > \pi_k$ ) para  $i < k$ .

- a) En la línea (3), la condición es verdadera sii  $i$  es un record de mínimo.
- b) En la línea (5), la condición es verdadera sii  $i$  es un record de máximo.
- ```
1 | min = max = T[0];
2 | for(i = 1; i < n; i++) {
3 |     if (T[i] < min)
4 |         min = T[i];
5 |     if (T[i] > max)
6 |         max = T[i];
7 | }
```

Observación

Por simetría basta estudiar records de máximo.

La cantidad esperada de records

Sea $R_n(\pi)$ la cantidad de records en $\pi \in \mathcal{S}_n$, y $e_n := \mathbb{E}[R_n]$.

Proposición

La cantidad esperada de records es $e_n = H_n = \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{n}$.

La cantidad esperada de records

Sea $R_n(\pi)$ la cantidad de records en $\pi \in \mathcal{S}_n$, y $e_n := \mathbb{E}[R_n]$.

Proposición

La cantidad esperada de records es $e_n = H_n = \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{n}$.

Demostración.

Sea $E_j = \{\pi \in \mathcal{S}_n : j \text{ es un record de } \pi\}$.

Observar que:

1. Tenemos $R_n = \sum_{j=1}^n \mathbf{1}_{E_j}$, así $e_n = \sum_{j=1}^n \Pr(E_j)$,

La cantidad esperada de records

Sea $R_n(\pi)$ la cantidad de records en $\pi \in \mathcal{S}_n$, y $e_n := \mathbb{E}[R_n]$.

Proposición

La cantidad esperada de records es $e_n = H_n = \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{n}$.

Demostración.

Sea $E_j = \{\pi \in \mathcal{S}_n : j \text{ es un record de } \pi\}$.

Observar que:

1. Tenemos $R_n = \sum_{j=1}^n \mathbf{1}_{E_j}$, así $e_n = \sum_{j=1}^n \Pr(E_j)$,
2. Los eventos E_j satisfacen $\Pr(E_j) = \frac{1}{j}$. □.

Branch misses: MinMax ingenuo

Proposición[Auger,Nicaud,Pivoteau'16]

La cantidad esperada de errores de predicción en el MinMax ingenuo para una permutación aleatoria es asintóticamente:

$4 \log n$ (predictor 1-bit), $2 \log n$ (predictor 2-bit,3-bit,...)

Branch misses: MinMax optimizado

Proposición [Auger,Nicaud,Pivoteau'16]

La cantidad esperada de errores de predicción en el MinMax optimizado es asintóticamente:

$$\frac{1}{4}n + O(\log n),$$

para los predictores de 1, 2, 3, ... bits.

a) Condición de línea (3), es decir $T[i] < T[i+1]$, se cumple con **probabilidad 1/2** para un i dado.

b) El evento $T[i] < T[i+1]$ es **independiente de la historia**

$$T[0], \dots, T[i-2], T[i-1].$$

c) Los otros ifs contribuyen $O(\log n)$.

```
1 min = max = T[n-1];
2 for(i = 0; i < n - 1; i
  += 2) {
3     if (T[i] < T[i+1]) {
4         if (T[i] < min)
5             min = T[i];
6         if (T[i+1] > max)
7             max = T[i+1];
8     } else {
9         if (T[i+1] < min)
```

...

Branch misses en una sola ejecución

Probamos $\mathbb{E}[R_n] \sim \log n$, pero ¿y si ejecutamos el algoritmo una sola vez?

⁷Para MinMax ingenuo, sabemos que la cantidad de errores de predicción es $O(R_n)$.

Branch misses en una sola ejecución

Probamos $\mathbb{E}[R_n] \sim \log n$, pero ¿y si ejecutamos el algoritmo una sola vez?

Proposición

Se cumple que $R_n \sim \log n$ en probabilidad

⁷Para MinMax ingenuo, sabemos que la cantidad de errores de predicción es $O(R_n)$.

Branch misses en una sola ejecución

Probamos $\mathbb{E}[R_n] \sim \log n$, pero ¿y si ejecutamos el algoritmo una sola vez?

Proposición

Se cumple que $R_n \sim \log n$ en probabilidad

Recordamos. Una secuencia de variables aleatorias X_n satisface $X_n \sim f(n)$ en probabilidad sii, para cada $\varepsilon > 0$ fijo,

$$\Pr(X_n \in [(1 - \varepsilon)f(n), (1 + \varepsilon)f(n)]) \rightarrow 1.$$

⁷Para MinMax ingenuo, sabemos que la cantidad de errores de predicción es $O(R_n)$.

Branch misses en una sola ejecución

Probamos $\mathbb{E}[R_n] \sim \log n$, pero ¿y si ejecutamos el algoritmo una sola vez?

Proposición

Se cumple que $R_n \sim \log n$ en probabilidad

Recordamos. Una secuencia de variables aleatorias X_n satisface $X_n \sim f(n)$ en probabilidad sii, para cada $\varepsilon > 0$ fijo,

$$\Pr(X_n \in [(1 - \varepsilon)f(n), (1 + \varepsilon)f(n)]) \rightarrow 1.$$

\Rightarrow Típicamente R_n está “cerca” de $\log n$.

⁷Para MinMax ingenuo, sabemos que la cantidad de errores de predicción es $O(R_n)$.

Desigualdad de Chebyshev

Proposición (Concentración)

Supongamos que $\mathbb{E}[X_n^2] \sim \mathbb{E}[X_n]^2$, y $\mathbb{E}[X_n] \rightarrow \infty$, cuando $n \rightarrow \infty$.

Entonces $X_n \sim \mathbb{E}[X_n]$ en probabilidad.

Lema (Chebyshev)

Sea X una variable aleatoria, entonces

$$\Pr(|X - \mathbb{E}[X]| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}.$$

Concentración de la cantidad de records

Demostración.

Probamos que $d_n := \mathbb{E}[R_n^2] \sim \mathbb{E}[R_n]^2 = e_n^2$.

Sea $E_j = \{\pi \in S_n : j \text{ es un record de } \pi\}$. Observar que:

1. Los eventos E_j satisfacen $\Pr(E_j) = \frac{1}{j}$.

Concentración de la cantidad de records

Demostración.

Probamos que $d_n := \mathbb{E}[R_n^2] \sim \mathbb{E}[R_n]^2 = e_n^2$.

Sea $E_j = \{\pi \in S_n : j \text{ es un record de } \pi\}$. Observar que:

1. Los eventos E_j satisfacen $\Pr(E_j) = \frac{1}{j}$.
2. Los eventos E_j y E_k son independientes para $j \neq k$,
$$\Pr(E_j \cap E_k) = \frac{1}{j \cdot k} = \Pr(E_j) \cdot \Pr(E_k).$$

Concentración de la cantidad de records

Demostración.

Probamos que $d_n := \mathbb{E}[R_n^2] \sim \mathbb{E}[R_n]^2 = e_n^2$.

Sea $E_j = \{\pi \in S_n : j \text{ es un record de } \pi\}$. Observar que:

1. Los eventos E_j satisfacen $\Pr(E_j) = \frac{1}{j}$.
2. Los eventos E_j y E_k son independientes para $j \neq k$,

$$\Pr(E_j \cap E_k) = \frac{1}{j \cdot k} = \Pr(E_j) \cdot \Pr(E_k).$$

\Rightarrow las indicatrices $X_j = \mathbf{1}_{E_j}$ son independientes: para $j \neq k$

$$\mathbb{E}[X_j X_k] = \mathbb{E}[X_j] \mathbb{E}[X_k] = \frac{1}{j \cdot k}.$$

Concentración de la cantidad de records

Demostración.

Probamos que $d_n := \mathbb{E}[R_n^2] \sim \mathbb{E}[R_n]^2 = e_n^2$.

Sea $E_j = \{\pi \in S_n : j \text{ es un record de } \pi\}$. Observar que:

1. Los eventos E_j satisfacen $\Pr(E_j) = \frac{1}{j}$.
2. Los eventos E_j y E_k son independientes para $j \neq k$,

$$\Pr(E_j \cap E_k) = \frac{1}{j \cdot k} = \Pr(E_j) \cdot \Pr(E_k).$$

\Rightarrow las indicatrices $X_j = \mathbf{1}_{E_j}$ son independientes: para $j \neq k$

$$\mathbb{E}[X_j X_k] = \mathbb{E}[X_j] \mathbb{E}[X_k] = \frac{1}{j \cdot k}.$$

Usando $R_n = \sum_{j=1}^n X_j$ obtenemos

$$\mathbb{E}[R_n^2] = H_n + 2 \sum_{j=2}^n \frac{H_{j-1}}{j}.$$

Concentración de la cantidad de records

Demostración.

Probamos que $d_n := \mathbb{E}[R_n^2] \sim \mathbb{E}[R_n]^2 = e_n^2$.

Sea $E_j = \{\pi \in S_n : j \text{ es un record de } \pi\}$. Observar que:

1. Los eventos E_j satisfacen $\Pr(E_j) = \frac{1}{j}$.
2. Los eventos E_j y E_k son independientes para $j \neq k$,

$$\Pr(E_j \cap E_k) = \frac{1}{j \cdot k} = \Pr(E_j) \cdot \Pr(E_k).$$

\Rightarrow las indicatrices $X_j = \mathbf{1}_{E_j}$ son independientes: para $j \neq k$

$$\mathbb{E}[X_j X_k] = \mathbb{E}[X_j] \mathbb{E}[X_k] = \frac{1}{j \cdot k}.$$

Usando $R_n = \sum_{j=1}^n X_j$ obtenemos

$$\mathbb{E}[R_n^2] = H_n + 2 \sum_{j=2}^n \frac{H_{j-1}}{j}.$$

Considerando $(\log n)^2 \sim H_n^2 = \sum_{j=1}^n \frac{1}{j^2} + 2 \sum_{j=2}^n \frac{H_{j-1}}{j}$



Concentración del MinMax optimizado

Sea $A_i = \{ \text{branch miss en } T[i] < T[i + 1] \}$. Observamos que:

- $\Pr(A_i) = 1/2$,
- A_i es independiente de A_j para $|i - j| > 1$.

Concentración del MinMax optimizado

Sea $A_i = \{ \text{branch miss en } T[i] < T[i + 1] \}$. Observamos que:

- $\Pr(A_i) = 1/2$,
- A_i es independiente de A_j para $|i - j| > 1$.

Tenemos $m \approx \frac{n}{2}$ variables aleatorias Bernoulli $\frac{1}{2} - \frac{1}{2}$ independientes:

$$C_m := \sum_{i=0}^{m-1} \mathbf{1}_{A_{2i}}$$

Concentración del MinMax optimizado

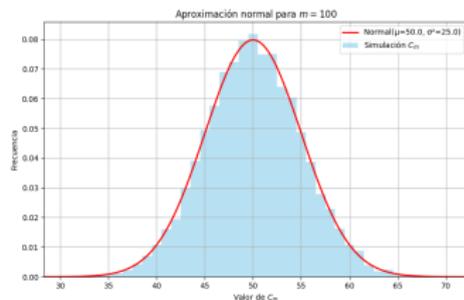
Sea $A_i = \{ \text{branch miss en } T[i] < T[i + 1] \}$. Observamos que:

- $\Pr(A_i) = 1/2$,
- A_i es independiente de A_j para $|i - j| > 1$.

Tenemos $m \approx \frac{n}{2}$ variables aleatorias Bernoulli $\frac{1}{2} - \frac{1}{2}$ independientes:

$$C_m := \sum_{i=0}^{m-1} \mathbf{1}_{A_{2i}}$$

Se tiene el **Teorema Central del Límite**: $(C_m - m/2)/\sqrt{m/4} \rightarrow N(0, 1)$ en ley.



Sesgar algoritmos para acelerarlos

Se necesita un compromiso:

- Un **if** con una condición que es True con proba. 50% (e independiente del pasado) es un problema para el predictor.
- Un **if** con una condición que no es 50 – 50 e independiente del pasado presenta redundancias.

Sesgar algoritmos para acelerarlos

Se necesita un compromiso:

- Un **if** con una condición que es True con proba. 50% (e independiente del pasado) es un problema para el predictor.
- Un **if** con una condición que no es 50 – 50 e independiente del pasado presenta redundancias.

Veamos otro ejemplo: la exponenciación...

Exponenciación sesgada

```
r = 1;
while (n > 0) {
    // n es impar
    if (n & 1)
        r = r * x;
    n /= 2;
    x = x * x;
}
```

Potencia clásica

```
r = 1;
while (n > 0) {
    t = x * x;
    // n1 n0 != 0 0
    if (n & 3) {
        if (n & 1)
            r = r * x;
        if (n & 2)
            r = r * t;
    }
    n /= 4;
    x = t * t;
}
```

Potencia sesgada

Exponenciación sesgada

```
r = 1;
while (n > 0) {
    // n es impar
    if (n & 1)
        r = r * x;
    n /= 2;
    x = x * x;
}
```

Potencia clásica

```
r = 1;
while (n > 0) {
    t = x * x;
    // n1 n0 != 0 0
    if (n & 3) {
        if (n & 1)
            r = r * x;
        if (n & 2)
            r = r * t;
    }
    n /= 4;
    x = t * t;
}
```

Potencia sesgada

- En la potencia clásica, a priori cada bit de n es $1/2 - 1/2$ independiente.
- En la potencia sesgada, el primer **if** aumenta la probabilidad de los otros dos!

Exponenciación sesgada

```
r = 1;
while (n > 0) {
    // n es impar
    if (n & 1)
        r = r * x;
    n /= 2;
    x = x * x;
}
```

Potencia clásica

```
r = 1;
while (n > 0) {
    t = x * x;
    // n1 n0 != 0 0
    if (n & 3) {
        if (n & 1)
            r = r * x;
        if (n & 2)
            r = r * t;
    }
    n /= 4;
    x = t * t;
}
```

Potencia sesgada

- En la potencia clásica, a priori cada bit de n es $1/2 - 1/2$ independiente.
- En la potencia sesgada, el primer **if** aumenta la probabilidad de los otros dos!
- Igual cantidad de multiplicaciones, pero más ifs ! ¿Quién ganará?

Análisis de la exponenciación sesgada

Modelo. Consideramos $k > 0$ y $n \in [0, 2^{2k} - 1]$ aleatorio:

$$n = n_{2k-1}n_{2k-2} \dots n_1n_0,$$

con cada n_i independiente y $n_i \sim \text{Ber}(1/2)$.

Consideramos predictores de 1-bit y 2-bits.

Análisis de la exponenciación sesgada

Modelo. Consideramos $k > 0$ y $n \in [0, 2^{2k} - 1]$ aleatorio:

$$n = n_{2k-1}n_{2k-2} \dots n_1n_0,$$

con cada n_i independiente y $n_i \sim \text{Ber}(1/2)$.

Consideramos predictores de 1-bit y 2-bits.

Plan para el análisis. Modelamos estado de predictor como una *Cadena de Markov*, nos interesa contar transiciones asociadas a “branch-miss”

Análisis de la exponenciación sesgada

Modelo. Consideramos $k > 0$ y $n \in [0, 2^{2k} - 1]$ aleatorio:

$$n = n_{2k-1}n_{2k-2} \dots n_1n_0,$$

con cada n_i independiente y $n_i \sim \text{Ber}(1/2)$.

Consideramos predictores de 1-bit y 2-bits.

Plan para el análisis. Modelamos estado de predictor como una *Cadena de Markov*, nos interesa contar transiciones asociadas a “branch-miss”

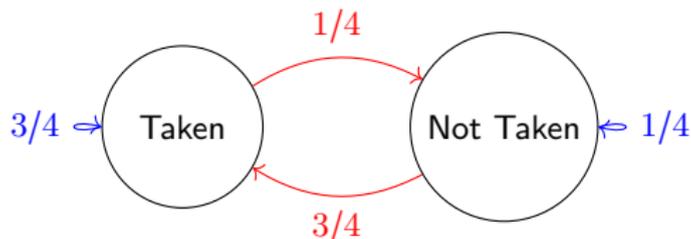


Figura: Predictor 1 Bit para if exterior

Modelo: cadenas de Markov

- Leemos pares [independientes] (n_{2i+1}, n_{2i}) , $i = 0, 1, 2, \dots, k - 1$.
- Seguimos el estado del predictor de cada if.

Resultado : cadenas de Markov.

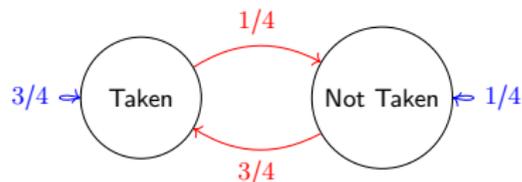


Figura: Predictor 1 Bit para if exterior

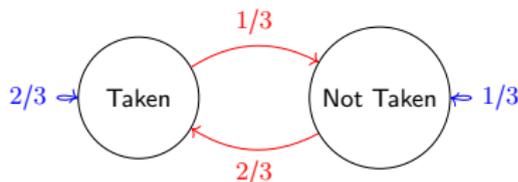


Figura: Predictor 1 Bit para ifs interiores

- Para el **if exterior**, tenemos $n \& 3 \neq 0$ con $\Pr(\text{Taken}) = \frac{3}{4}$.
- Para los **ifs interiores**, dado que pasamos el if exterior, tenemos $n \& 1$ y $n \& 2$ con $\Pr(\text{Taken}) = \frac{2}{3}$.
- Branch misses en rojo en las figuras.

Cadena de Markov y distribución estacionaria

Un proceso X_0, X_1, \dots con valores en $\{s_1, \dots, s_K\}$, el conjunto de estados, es una Cadena de Markov sii existe una matriz $P \in \mathcal{M}_{K \times K}([0, 1])$, fija, que define las probabilidades de transición

$$[P]_{i,j} = \Pr(X_{n+1} = s_j | X_n = s_i),$$

para todo $n \geq 0$.

Cadena de Markov y distribución estacionaria

Un proceso X_0, X_1, \dots con valores en $\{s_1, \dots, s_K\}$, el conjunto de estados, es una Cadena de Markov sii existe una matriz $P \in \mathcal{M}_{K \times K}([0, 1])$, fija, que define las probabilidades de transición

$$[P]_{i,j} = \Pr(X_{n+1} = s_j | X_n = s_i),$$

para todo $n \geq 0$.

Lema

Sea $\mu^{(n)} = (\mu_1^{(n)}, \dots, \mu_K^{(n)})$ la distribución de X_n , i.e., $\Pr(X_n = s_i) = \mu_i^{(n)}$. Entonces tenemos la recurrencia matricial $\mu^{(n+1)} = \mu^{(n)} P$.

Cadena de Markov y distribución estacionaria

Un proceso X_0, X_1, \dots con valores en $\{s_1, \dots, s_K\}$, el conjunto de estados, es una Cadena de Markov sii existe una matriz $P \in \mathcal{M}_{K \times K}([0, 1])$, fija, que define las probabilidades de transición

$$[P]_{i,j} = \Pr(X_{n+1} = s_j | X_n = s_i),$$

para todo $n \geq 0$.

Lema

Sea $\mu^{(n)} = (\mu_1^{(n)}, \dots, \mu_K^{(n)})$ la distribución de X_n , i.e., $\Pr(X_n = s_i) = \mu_i^{(n)}$. Entonces tenemos la recurrencia matricial $\mu^{(n+1)} = \mu^{(n)} P$.

Definición

Un vector $\pi = (\pi_1, \dots, \pi_K)$ con $\pi_i \geq 0$ y $\sum \pi_i = 1$ es una distribución estacionaria para P sii $\pi = \pi P$.

Teorema Ergódico para Cadenas de Markov

Para asegurar que la distribución converge a una estacionaria $\mu^{(n)} \rightarrow \pi$, necesitamos algunas condiciones técnicas relacionadas con el digrafo de P .

Teorema Ergódico para Cadenas de Markov

Para asegurar que la distribución converge a una estacionaria $\mu^{(n)} \rightarrow \pi$, necesitamos algunas condiciones técnicas relacionadas con el digrafo de P .

Definición

- Una Cadena de Markov es **irreducible** sii existe un camino con probabilidad positiva entre cada par de estados.
- Una Cadena de Markov es **aperiódica** sii el máximo común divisor de la longitud de todos los ciclos es 1.

Teorema Ergódico para Cadenas de Markov

Para asegurar que la distribución converge a una estacionaria $\mu^{(n)} \rightarrow \pi$, necesitamos algunas condiciones técnicas relacionadas con el digrafo de P .

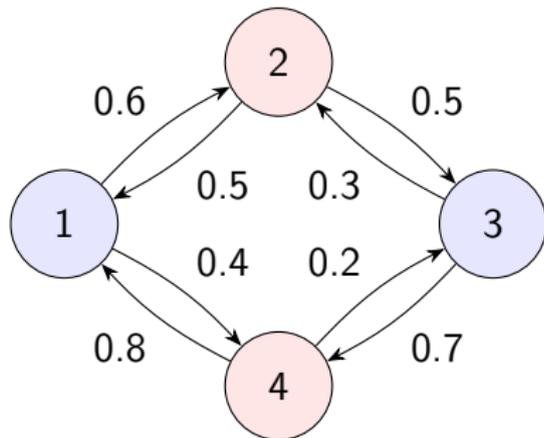
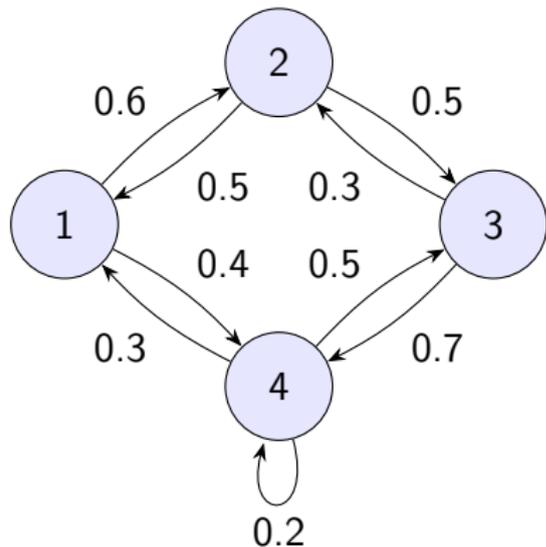
Definición

- Una Cadena de Markov es **irreducible** sii existe un camino con probabilidad positiva entre cada par de estados.
- Una Cadena de Markov es **aperiódica** sii el máximo común divisor de la longitud de todos los ciclos es 1.

La condición de irreductibilidad permite afirmar que no tenemos nodos “transitorios”, que no visitaremos más a partir de un cierto momento.

Cadenas periódicas y aperiódicas

Una cadena aperiódica (izq.) y una periódica (der.):



- La aperiodicidad se cumple inmediatamente cuando tenemos *loops*.
- Una cadena periódica presenta periodicidades en $\mu^{(n)}$.

Teorema Ergódico para Cadenas de Markov

Teorema

Sea (X_0, X_1, \dots) una Cadena de Markov **irreducible** y **aperiódica** con matriz de transición P y distribución inicial arbitraria $\mu^{(0)}$.

Existe una única distribución estacionaria π tal que $\mu^{(n)} \rightarrow \pi$. Más aún, π no depende de la elección de la distribución inicial $\mu^{(0)}$.

En este caso π es el único vector propio de $\lambda = 1$ para P , con $\sum \pi_i = 1$.

Teorema Ergódico para Cadenas de Markov

Teorema

Sea (X_0, X_1, \dots) una Cadena de Markov **irreducible** y **aperiódica** con matriz de transición P y distribución inicial arbitraria $\mu^{(0)}$.

Existe una única distribución estacionaria π tal que $\mu^{(n)} \rightarrow \pi$. Más aún, π no depende de la elección de la distribución inicial $\mu^{(0)}$.

En este caso π es el único vector propio de $\lambda = 1$ para P , con $\sum \pi_i = 1$.

Para contar las transiciones

Proposición

En las hipótesis del teorema, para cada transición $v = (s_i, s_j)$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{1}_{(X_k, X_{k+1})=v} = \pi_i P_{i,j}.$$

Teorema Ergódico aplicado: 1 bit

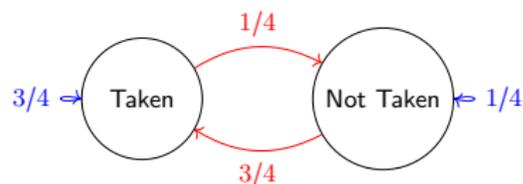


Figura: Predictor 1 Bit para if exterior

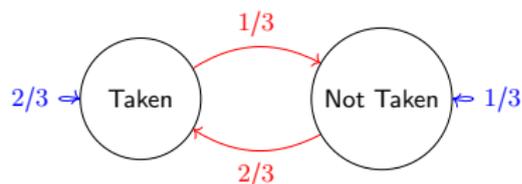
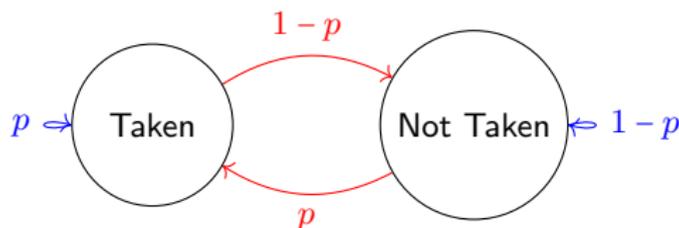


Figura: Predictor 1 Bit para ifs interiores

Sea entonces



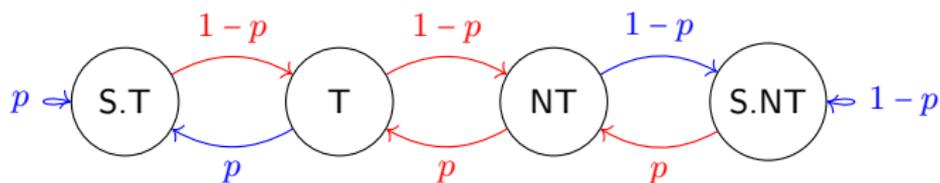
Basta estudiar $\pi = \pi(p)$, donde $1 = T$, $2 = NT$. En este caso

$$\pi = (p, 1-p).$$

Y para las transiciones en rojo tenemos la frecuencia $\alpha_1(p) := 2p(1-p)$.

Teorema Ergódico aplicado: 2 bits

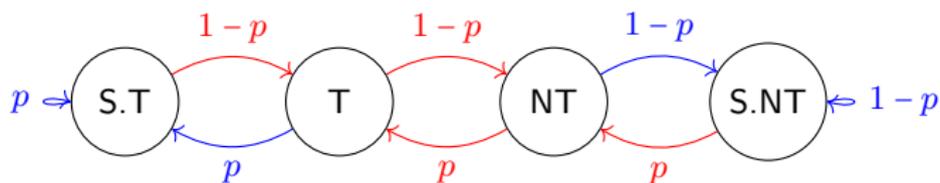
Para dos bits tenemos



Hay que calcular $\pi = \pi(p)$, donde $1 = S.T$, $2 = T$, $3 = NT$, $4 = S.NT$.

Teorema Ergódico aplicado: 2 bits

Para dos bits tenemos



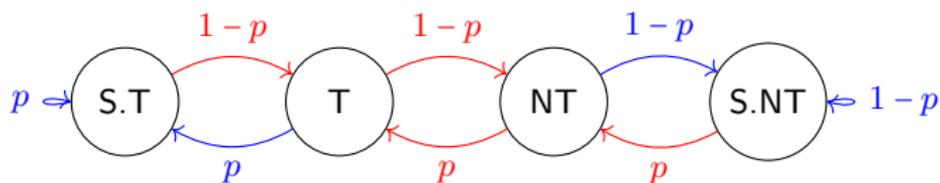
Hay que calcular $\pi = \pi(p)$, donde $1 = S.T$, $2 = T$, $3 = NT$, $4 = S.NT$.

En este caso [cálculo en pizarrón]

$$\pi_i = C \cdot \left(\frac{1-p}{p}\right)^{i-1}, \quad C = \frac{1 - \left(\frac{1-p}{p}\right)}{1 - \left(\frac{1-p}{p}\right)^4} = \frac{p^3}{1 - 2p(1-p)}.$$

Teorema Ergódico aplicado: 2 bits

Para dos bits tenemos



Hay que calcular $\pi = \pi(p)$, donde $1 = S.T$, $2 = T$, $3 = NT$, $4 = S.NT$.

En este caso [cálculo en pizarrón]

$$\pi_i = C \cdot \left(\frac{1-p}{p}\right)^{i-1}, \quad C = \frac{1 - \left(\frac{1-p}{p}\right)}{1 - \left(\frac{1-p}{p}\right)^4} = \frac{p^3}{1 - 2p(1-p)}.$$

El conjunto de transiciones marcadas en rojo tiene frecuencia:

$$\alpha_2(p) = \pi_1(1-p) + \pi_2(1-p) + \pi_3p + \pi_4p = \frac{p(1-p)}{1 - 2p(1-p)}.$$

Errores de predicción en la exponenciación sesgada

Proposición (Simplificada)

Sea $N = 4^k$ y consideremos $n \in \{0, \dots, N - 1\}$ uniforme.

La cantidad media de errores de predicción, cuando $k \rightarrow \infty$, con el predictor para i -bits es asintóticamente

$$k \times (\alpha_i(3/4) + \frac{3}{4} \cdot 2 \cdot \alpha_i(2/3)), \quad k = \frac{1}{2} \log_2 N.$$

El factor $\frac{3}{4}$ es la probabilidad de efectuar los ifs internos.

$\frac{7}{48} \approx 0,52$, $\frac{9}{20} = 0,45$, $\frac{1095}{2788} \approx 0,39$.

Errores de predicción en la exponenciación sesgada

Proposición (Simplificada)

Sea $N = 4^k$ y consideremos $n \in \{0, \dots, N-1\}$ uniforme.

La cantidad media de errores de predicción, cuando $k \rightarrow \infty$, con el predictor para i -bits es asintóticamente

$$k \times (\alpha_i(3/4) + \frac{3}{4} \cdot 2 \cdot \alpha_i(2/3)), \quad k = \frac{1}{2} \log_2 N.$$

El factor $\frac{3}{4}$ es la probabilidad de efectuar los ifs internos.

Proposición (Auger, Nicaud, Pivoteau'2016)

Sea $N \rightarrow \infty$ arbitrario y consideremos $n \in \{0, \dots, N-1\}$ uniforme.

La cantidad esperada de errores de predicción en la exponenciación sesgada para los predictores saturados de 1, 2 y 3 bits es:

$$M_{1 \text{ bit}}(N) \sim \log_2(N) \times \frac{25}{48}, \quad M_{2 \text{ bit}}(N) \sim \log_2(N) \times \frac{9}{20}, \quad M_{3 \text{ bit}}(N) \sim \log_2(N) \times \frac{1095}{2788}.$$

$${}^7 \frac{25}{48} \approx 0,52, \quad \frac{9}{20} = 0,45, \quad \frac{1095}{2788} \approx 0,39.$$

Para aprender más



Olle Häggström,

Finite Markov Chains and Algorithmic Applications.

London Mathematical Society Student Texts 52.



Nicolas Auger, Cyril Nicaud, y Carine Pivoteau,

Good Predictions Are Worth a Few Comparisons.

<https://www-igm.univ-mlv.fr/~nicaud/articles/stacs16.pdf>



Cyril Nicaud, Carine Pivoteau y Stéphane Vialette

Branch Prediction Analysis of Morris-Pratt and Knuth-Morris-Pratt Algorithms.

<https://arxiv.org/abs/2503.13694>



Conrado Martínez, Markus E. Nebel y Sebastian Wild

Analysis of branch misses in quicksort.

<https://dl.acm.org/doi/10.5555/2790216.2790227>

Outline

1. Introducción al análisis de algoritmos
2. Aplicaciones a la predicción de saltos
3. **Combinatoria Analítica: métodos simbólicos**
 - Funciones generatrices
 - Clases combinatorias no etiquetadas
 - Clases combinatorias etiquetadas
 - Funciones generatrices multivariadas
4. Combinatoria Analítica: Teoría Analítica

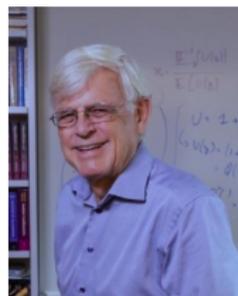
La Combinatoria Analítica

Analytic Combinatorics aims at predicting precisely the properties of *large structured combinatorial configurations*, through an approach based extensively on *analytic methods*. *Generating functions* are the central objects of study of the theory.

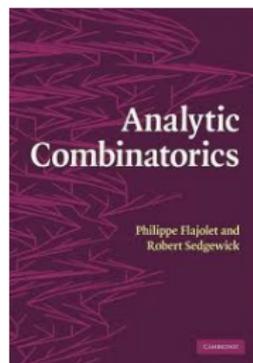
– Philippe Flajolet (1948–2011), Robert Sedgewick (1946–)



Philippe Flajolet, © Inria / Foto C. Tourniaire



Robert Sedgewick, Wikipedia, CC BY-SA 4.0



- Orígenes se remontan a Euler y la Teoría Analítica de Números,
- La Combinatoria Analítica fue desarrollada en gran medida por **Philippe Flajolet** durante los años 80 y 90,
- La Combinatoria Analítica se encuentra resumida en la *Magnum Opus* de P. Flajolet y R. Sedgewick: **Analytic Combinatorics**, <https://algo.inria.fr/flajolet/Publications/books.html>.

Principio de la Combinatoria Analítica

¿Por qué estudiar Combinatoria Analítica?

- Método sistemático para estudiar de algoritmos y estructuras discretas.
- Permite obtener comportamientos típicos para objetos de gran tamaño.
- Generación aleatoria sistemática.

Principio de la Combinatoria Analítica

¿Por qué estudiar Combinatoria Analítica?

- Método sistemático para estudiar de algoritmos y estructuras discretas.
- Permite obtener comportamientos típicos para objetos de gran tamaño.
- Generación aleatoria sistemática.

If you can specify it, you can analyze it !
– P. Flajolet y R. Sedgewick.

Principio de la Combinatoria Analítica

¿Por qué estudiar Combinatoria Analítica?

- Método sistemático para estudiar de algoritmos y estructuras discretas.
- Permite obtener comportamientos típicos para objetos de gran tamaño.
- Generación aleatoria sistemática.

If you can specify it, you can analyze it !

– P. Flajolet y R. Sedgewick.

Método general de la Combinatoria Analítica

Un análisis se divide en dos pasos

1. **Paso simbólico:** a partir de una especificación [recursiva, iterativa, ...] del problema, se obtiene una ecuación para la *función generatriz* asociada.
2. **Paso analítico:** usando un *Teorema de Transferencia*, las propiedades analíticas de la función generatriz se transforman en asintóticos.

Funciones generatrices

A generating function is a clothesline on which we hang up a sequence of numbers for display.

Una función generatriz es como un tendedero en el que colgamos una secuencia de números para verla.

– Herbert S. Wilf (1931–2012)



Wikipedia, CC BY-SA 3.0.

$$\{a_n\}_{n=0}^{\infty} \longleftrightarrow A(z) = \sum_{n=0}^{\infty} a_n z^n$$

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z}, \quad \sum_{n=0}^{\infty} n z^n = \frac{z}{(1-z)^2}, \quad \sum_{n=0}^{\infty} H_n z^n = \frac{1}{1-z} \log\left(\frac{1}{1-z}\right).$$

Funciones generatrices

Dada una sucesión de números $\{a_n\}_{n=0}^{\infty}$, le asociamos su *función generatriz ordinaria* $A(z)$, OGF en inglés

$$\{a_n\}_{n=0}^{\infty} \longleftrightarrow A(z) = \sum_{n=0}^{\infty} a_n z^n .$$

Escribimos también $[z^n]A(z) = a_n$.

Funciones generatrices

Dada una sucesión de números $\{a_n\}_{n=0}^{\infty}$, le asociamos su *función generatriz ordinaria* $A(z)$, OGF en inglés

$$\{a_n\}_{n=0}^{\infty} \longleftrightarrow A(z) = \sum_{n=0}^{\infty} a_n z^n .$$

Escribimos también $[z^n]A(z) = a_n$.

⚠ Por ahora tratamos las series como **objetos formales**:

- Se puede *operar con los términos* (como con polinomios),
- Permite determinar los coeficientes,
- Pero *no se puede evaluar* la serie!

Funciones generatrices como objeto formal

$$\{a_n\}_{n=0}^{\infty} \longleftrightarrow A(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Objeto formal: forma un anillo conmutativo. Si $B(z) = \sum_{n=0}^{\infty} b_n z^n$,

$$A(z) \pm B(z) := \sum_{n=0}^{\infty} (a_n \pm b_n) z^n, \quad A(z) \cdot B(z) := \sum_{n=0}^{\infty} c_n z^n,$$

donde $c_n = \sum_{k=0}^n a_k b_{n-k}$ es el *producto de Cauchy*.

Funciones generatrices como objeto formal

$$\{a_n\}_{n=0}^{\infty} \longleftrightarrow A(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Objeto formal: forma un anillo conmutativo. Si $B(z) = \sum_{n=0}^{\infty} b_n z^n$,

$$A(z) \pm B(z) := \sum_{n=0}^{\infty} (a_n \pm b_n) z^n, \quad A(z) \cdot B(z) := \sum_{n=0}^{\infty} c_n z^n,$$

donde $c_n = \sum_{k=0}^n a_k b_{n-k}$ es el *producto de Cauchy*.

Ejemplo. función generatriz de las sumas parciales

$$\left\{ \sum_{k=0}^n a_k \right\}_{n=0}^{\infty} \longleftrightarrow \frac{1}{1-z} A(z)$$

Funciones generatrices como objeto formal

$$\{a_n\}_{n=0}^{\infty} \longleftrightarrow A(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Objeto formal: forma un anillo conmutativo. Si $B(z) = \sum_{n=0}^{\infty} b_n z^n$,

$$A(z) \pm B(z) := \sum_{n=0}^{\infty} (a_n \pm b_n) z^n, \quad A(z) \cdot B(z) := \sum_{n=0}^{\infty} c_n z^n,$$

donde $c_n = \sum_{k=0}^n a_k b_{n-k}$ es el *producto de Cauchy*.

Ejemplo. función generatriz de las sumas parciales

$$\left\{ \sum_{k=0}^n a_k \right\}_{n=0}^{\infty} \longleftrightarrow \frac{1}{1-z} A(z)$$

$$\{1\}_{n=0}^{\infty} \leftrightarrow \frac{1}{1-z}, \quad \{n+1\}_{n=0}^{\infty} \leftrightarrow \frac{1}{(1-z)^2}, \quad \left\{ \sum_{k=0}^n (k+1) \right\}_{n=0}^{\infty} \leftrightarrow \frac{1}{(1-z)^3}, \quad \dots$$

Funciones generatrices como objeto formal 2

Proposición

Una función generatriz $A(z) = \sum_{n=0}^{\infty} a_n z^n$ tiene un inverso multiplicativo, es decir $\frac{1}{A(z)} = \sum_{n=0}^{\infty} b_n z^n$, sii $a_0 \neq 0$.

Funciones generatrices como objeto formal 2

Proposición

Una función generatriz $A(z) = \sum_{n=0}^{\infty} a_n z^n$ tiene un inverso multiplicativo, es decir $\frac{1}{A(z)} = \sum_{n=0}^{\infty} b_n z^n$, sii $a_0 \neq 0$.

Un ejemplo fundamental es

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z} \iff (1-z) \cdot \sum_{n=0}^{\infty} z^n = 1.$$

Ejemplo: los números de Fibonacci

Los números de Fibonacci están definidos por $f_0 = 0$, $f_1 = 1$ y,

$$f_{n+2} = f_{n+1} + f_n$$

para $n \geq 0$.

Ejemplo: los números de Fibonacci

Los números de Fibonacci están definidos por $f_0 = 0$, $f_1 = 1$ y,

$$f_{n+2} = f_{n+1} + f_n$$

para $n \geq 0$.

Sea $F(z) = \sum f_n z^n$. Tenemos

Función generatriz de Fibonacci

$$F(z) = \frac{z}{1 - z - z^2}.$$

Ejemplo: los números de Fibonacci 2

Vamos a encontrar la fórmula para los números de Fibonacci: $F(z) = \frac{z}{1-z-z^2}$.

⁸Es decir $r_1 = \phi^{-1}$ y $r_2 = -\phi$ donde $\phi = (1 + \sqrt{5})/2$ satisface $\phi^2 = \phi + 1$.

Ejemplo: los números de Fibonacci 2

Vamos a encontrar la fórmula para los números de Fibonacci: $F(z) = \frac{z}{1-z-z^2}$.

– Por **fracciones simples**, existen C_1, C_2 tales que

$$\frac{z}{1-z-z^2} = \frac{C_1}{r_1-z} + \frac{C_2}{r_2-z},$$

donde r_1 y r_2 son las raíces⁸ de $1-z-z^2=0$.

⁸Es decir $r_1 = \phi^{-1}$ y $r_2 = -\phi$ donde $\phi = (1 + \sqrt{5})/2$ satisface $\phi^2 = \phi + 1$.

Ejemplo: los números de Fibonacci 2

Vamos a encontrar la fórmula para los números de Fibonacci: $F(z) = \frac{z}{1-z-z^2}$.

– Por **fracciones simples**, existen C_1, C_2 tales que

$$\frac{z}{1-z-z^2} = \frac{C_1}{r_1-z} + \frac{C_2}{r_2-z},$$

donde r_1 y r_2 son las raíces⁸ de $1-z-z^2=0$.

– Observamos que

$$\frac{C_j}{r_j-z} = \frac{C_j/r_j}{1-z/r_j} = (C_j/r_j) \sum_{n=0}^{\infty} r_j^{-n} z^n,$$

de donde obtenemos $f_n = C_1 r_1^{-n-1} + C_2 r_2^{-n-1}$ para todo $n \geq 0$.

⁸Es decir $r_1 = \phi^{-1}$ y $r_2 = -\phi$ donde $\phi = (1 + \sqrt{5})/2$ satisface $\phi^2 = \phi + 1$.

Ejemplo: los números de Fibonacci 2

Vamos a encontrar la fórmula para los números de Fibonacci: $F(z) = \frac{z}{1-z-z^2}$.

– Por **fracciones simples**, existen C_1, C_2 tales que

$$\frac{z}{1-z-z^2} = \frac{C_1}{r_1-z} + \frac{C_2}{r_2-z},$$

donde r_1 y r_2 son las raíces⁸ de $1-z-z^2=0$.

– Observamos que

$$\frac{C_j}{r_j - x_j} = \frac{C_j/r_j}{1 - z/r_j} = (C_j/r_j) \sum_{n=0}^{\infty} r_j^{-n} z^n,$$

de donde obtenemos $f_n = C_1 r_1^{-n-1} + C_2 r_2^{-n-1}$ para todo $n \geq 0$.

– **Calculando constantes**,

$$f_n = \frac{1}{\sqrt{5}} \phi^{n+1} + C_2 (-\phi)^{-n-1} \sim \frac{1}{\sqrt{5}} \phi^{n+1}.$$

⁸Es decir $r_1 = \phi^{-1}$ y $r_2 = -\phi$ donde $\phi = (1 + \sqrt{5})/2$ satisface $\phi^2 = \phi + 1$.

Ejemplo: fracciones simples

Ejemplo. Encontrar los coeficientes de

$$A(z) = \frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)}.$$

⁹Expandiendo el sistema de ecuaciones, por ejemplo.

Ejemplo: fracciones simples

Ejemplo. Encontrar los coeficientes de

$$A(z) = \frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)}.$$

Por *fracciones simples* existen coeficientes a, b, c tales que

$$\frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)} = \frac{a}{(1 - z)^2} + \frac{b}{1 - z} + \frac{c}{1 - 2z}.$$

⁹Expandiendo el sistema de ecuaciones, por ejemplo.

Ejemplo: fracciones simples

Ejemplo. Encontrar los coeficientes de

$$A(z) = \frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)}.$$

Por *fracciones simples* existen coeficientes a, b, c tales que

$$\frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)} = \frac{a}{(1 - z)^2} + \frac{b}{1 - z} + \frac{c}{1 - 2z}.$$

Podemos calcular⁹ $a = -1, b = 0, c = 2$.

⁹Expandiendo el sistema de ecuaciones, por ejemplo.

Ejemplo: fracciones simples

Ejemplo. Encontrar los coeficientes de

$$A(z) = \frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)}.$$

Por *fracciones simples* existen coeficientes a, b, c tales que

$$\frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)} = \frac{a}{(1 - z)^2} + \frac{b}{1 - z} + \frac{c}{1 - 2z}.$$

Podemos calcular⁹ $a = -1, b = 0, c = 2$. Entonces

$$[z^n]A(z) = -[z^n]\frac{1}{(1 - z)^2} + 2[z^n]\frac{1}{1 - 2z}$$

⁹Expandiendo el sistema de ecuaciones, por ejemplo.

Ejemplo: fracciones simples

Ejemplo. Encontrar los coeficientes de

$$A(z) = \frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)}.$$

Por *fracciones simples* existen coeficientes a, b, c tales que

$$\frac{1 - 2z + 2z^2}{(1 - z)^2(1 - 2z)} = \frac{a}{(1 - z)^2} + \frac{b}{1 - z} + \frac{c}{1 - 2z}.$$

Podemos calcular⁹ $a = -1, b = 0, c = 2$. Entonces

$$[z^n]A(z) = -[z^n]\frac{1}{(1 - z)^2} + 2[z^n]\frac{1}{1 - 2z}$$

y deducimos $[z^n]A(z) = 2^{n+1} - (n + 1)$ usando

$$\frac{1}{(1 - z)^2} = \sum_{n=0}^{\infty} (n + 1)z^n, \quad \frac{1}{1 - 2z} = \sum_{n=0}^{\infty} 2^n z^n.$$

⁹Expandiendo el sistema de ecuaciones, por ejemplo.

Funciones generatrices como objeto formal 3

Composición

Dadas $A(z) = \sum_{n=0}^{\infty} a_n$ y $B(z) = \sum_{n=0}^{\infty} b_n z^n$, si $b_0 = 0$ podemos definir

$$A(B(z)) = \sum a_n (B(z))^n = \sum a_n z^n (b_1 + b_2 z^1 + \dots)^n .$$

Funciones generatrices como objeto formal 3

Composición

Dadas $A(z) = \sum_{n=0}^{\infty} a_n z^n$ y $B(z) = \sum_{n=0}^{\infty} b_n z^n$, si $b_0 = 0$ podemos definir

$$A(B(z)) = \sum a_n (B(z))^n = \sum a_n z^n (b_1 + b_2 z^1 + \dots)^n .$$

Ejemplo:

$$\sum_{n=0}^{\infty} \left(\frac{z}{1-z} \right)^n = \frac{1}{1 - \frac{z}{1-z}} = \frac{1-z}{1-2z} = 1 + \sum_{n=1}^{\infty} 2^{n-1} z^n .$$

Funciones generatrices como objeto formal 4

Derivación

Dada $A(z) = \sum_{n=0}^{\infty} a_n z^n$, definimos

$$A'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1} .$$

La derivación cumple las reglas clásicas del producto y el cociente.

Funciones generatrices como objeto formal 4

Derivación

Dada $A(z) = \sum_{n=0}^{\infty} a_n z^n$, definimos

$$A'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

La derivación cumple las reglas clásicas del producto y el cociente.

Ejemplo: Si $A(z) \leftrightarrow \{a_n\}$,

$$z \partial_z A(z) = z A'(z) \leftrightarrow \{n a_n\}.$$

Funciones generatrices como objeto formal 4

Derivación

Dada $A(z) = \sum_{n=0}^{\infty} a_n z^n$, definimos

$$A'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

La derivación cumple las reglas clásicas del producto y el cociente.

Ejemplo: Si $A(z) \leftrightarrow \{a_n\}$,

$$z \partial_z A(z) = z A'(z) \leftrightarrow \{n a_n\}.$$

Ejemplo. Probar que

$$\sum_{n=0}^{\infty} \binom{n}{k} z^n = \frac{z^k}{(1-z)^{k+1}}.$$

Funciones generatrices como objeto formal 5

Integración

Dada $A(z) = \sum_{n=0}^{\infty} a_n z^n$, definimos

$$\int_0^x A(z) dz = \sum_{n=0}^{\infty} \frac{a_n}{n+1} z^{n+1}.$$

La derivación cumple las reglas clásicas con respecto a la derivación.

Funciones generatrices como objeto formal 5

Integración

Dada $A(z) = \sum_{n=0}^{\infty} a_n z^n$, definimos

$$\int_0^x A(z) dz = \sum_{n=0}^{\infty} \frac{a_n}{n+1} z^{n+1}.$$

La derivación cumple las reglas clásicas con respecto a la derivación.

Ejemplos.

$$\sum_{n=1}^{\infty} \frac{1}{n} z^n = \log\left(\frac{1}{1-z}\right), \quad \sum_{n=1}^{\infty} H_n x^n = \frac{1}{1-x} \log\left(\frac{1}{1-x}\right).$$

Tabla de funciones generatrices

Sucesión a_n	OGF $F(z) = \sum a_n z^n$
1	$\frac{1}{1-z}$
n	$\frac{z}{(1-z)^2}$
$\frac{1}{n}, n \geq 1$	$-\log\left(\frac{1}{1-z}\right)$
$H_n, n \geq 1$	$-\frac{1}{1-z} \log\left(\frac{1}{1-z}\right)$
$\binom{n}{m}, m \in \mathbb{Z}_{\geq 0}$	$\frac{z^m}{(1-z)^{m+1}}$
$\binom{\alpha}{n}, \alpha \in \mathbb{R}$	$(1+z)^\alpha$
Fibonacci f_n	$\frac{z}{1-z-z^2}$
$\frac{1}{n!}$	e^z

Clases combinatorias no etiquetadas

Definición

Una *clase combinatoria* es un par $(\mathcal{A}, |\cdot|_{\mathcal{A}})$, formado por un conjunto \mathcal{A} numerable de objetos y una función de talla $|\cdot|_{\mathcal{A}}$ tal que

- $|a|_{\mathcal{A}} \in \mathbb{Z}_{\geq 0}$ para cada $a \in \mathcal{A}$,
- para cada $n \in \mathbb{Z}_{\geq 0}$, el conjunto de $a \in \mathcal{A}$ tales que $|a|_{\mathcal{A}} = n$ es finito.

Clases combinatorias no etiquetadas

Definición

Una *clase combinatoria* es un par $(\mathcal{A}, |\cdot|_{\mathcal{A}})$, formado por un conjunto \mathcal{A} numerable de objetos y una función de talla $|\cdot|_{\mathcal{A}}$ tal que

- $|a|_{\mathcal{A}} \in \mathbb{Z}_{\geq 0}$ para cada $a \in \mathcal{A}$,
- para cada $n \in \mathbb{Z}_{\geq 0}$, el conjunto de $a \in \mathcal{A}$ tales que $|a|_{\mathcal{A}} = n$ es finito.

Para una clase combinatoria podemos definir

$$A(z) = \sum_{a \in \mathcal{A}} z^{|a|}.$$

Clases combinatorias no etiquetadas

Definición

Una *clase combinatoria* es un par $(\mathcal{A}, |\cdot|_{\mathcal{A}})$, formado por un conjunto \mathcal{A} numerable de objetos y una función de talla $|\cdot|_{\mathcal{A}}$ tal que

- $|a|_{\mathcal{A}} \in \mathbb{Z}_{\geq 0}$ para cada $a \in \mathcal{A}$,
- para cada $n \in \mathbb{Z}_{\geq 0}$, el conjunto de $a \in \mathcal{A}$ tales que $|a|_{\mathcal{A}} = n$ es finito.

Para una clase combinatoria podemos definir

$$A(z) = \sum_{a \in \mathcal{A}} z^{|a|}.$$

Observación

$A(z) = \sum_{n=0}^{\infty} a_n z^n$ donde a_n es la cantidad de elementos de tamaño n ,

$$\mathcal{A}_n := \{a \in \mathcal{A} : |a| = n\}.$$

Operaciones: clases combinatorias no etiquetadas

Unión $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$

Dadas clases combinatorias \mathcal{B} y \mathcal{C} con $\mathcal{B} \cap \mathcal{C} = \emptyset$, $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$ es una clase combinatoria con función de talla:

$$|a|_{\mathcal{A}} = \begin{cases} |a|_{\mathcal{B}} & \text{si } a \in \mathcal{B}, \\ |a|_{\mathcal{C}} & \text{si } a \in \mathcal{C}. \end{cases}$$

Observar que las funciones generatrices satisfacen $A(z) = B(z) + C(z)$.

Operaciones: clases combinatorias no etiquetadas

Unión $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$

Dadas clases combinatorias \mathcal{B} y \mathcal{C} con $\mathcal{B} \cap \mathcal{C} = \emptyset$, $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$ es una clase combinatoria con función de talla:

$$|a|_{\mathcal{A}} = \begin{cases} |a|_{\mathcal{B}} & \text{si } a \in \mathcal{B}, \\ |a|_{\mathcal{C}} & \text{si } a \in \mathcal{C}. \end{cases}$$

Observar que las funciones generatrices satisfacen $A(z) = B(z) + C(z)$.

Producto $\mathcal{A} = \mathcal{B} \times \mathcal{C}$

Dadas clases combinatorias \mathcal{B} y \mathcal{C} , el producto cartesiano $\mathcal{A} = \mathcal{B} \times \mathcal{C}$ es una clase combinatoria con función de talla:

$$|(b, c)|_{\mathcal{A}} = |b|_{\mathcal{B}} + |c|_{\mathcal{C}}.$$

Observar que las funciones generatrices satisfacen $A(z) = B(z) \cdot C(z)$.

Ejemplo: Strings bien parentizadas

- La clase de strings bien parentizadas

$$\mathcal{S} = \{\varepsilon, (), ()(), (()), \dots\}$$

se puede caracterizar por

$$\mathcal{S} = \varepsilon + (\mathcal{S})\mathcal{S}.$$

¹⁰Veremos luego los fundamentos analíticos que permiten obtener esta solución.

Ejemplo: Strings bien parentizadas

- La **clase de strings bien parentizadas**

$$\mathcal{S} = \{\varepsilon, (), ()(), (()), \dots\}$$

se puede caracterizar por

$$\mathcal{S} = \varepsilon + (\mathcal{S})\mathcal{S}.$$

- Para la **talla**, contemos la cantidad de paréntesis abiertos “(”.

¹⁰Veremos luego los fundamentos analíticos que permiten obtener esta solución.

Ejemplo: Strings bien parentizadas

- La **clase de strings bien parentizadas**

$$\mathcal{S} = \{\varepsilon, (), ()(), (()), \dots\}$$

se puede caracterizar por

$$\mathcal{S} = \varepsilon + (\mathcal{S})\mathcal{S}.$$

- Para la **talla**, contemos la cantidad de paréntesis abiertos “(”. Entonces

$$S(z) = 1 + z(S(z))^2.$$

¹⁰Veremos luego los fundamentos analíticos que permiten obtener esta solución.

Ejemplo: Strings bien parentizadas

- La **clase de strings bien parentizadas**

$$\mathcal{S} = \{\varepsilon, (), ()(), (()), \dots\}$$

- se puede caracterizar por

$$\mathcal{S} = \varepsilon + (\mathcal{S})\mathcal{S}.$$

- Para la **talla**, contemos la cantidad de paréntesis abiertos “(”. Entonces

$$S(z) = 1 + z(S(z))^2.$$

- Esta ecuación se puede **resolver**¹⁰, obteniendo

$$S(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

¹⁰Veremos luego los fundamentos analíticos que permiten obtener esta solución.

Construcciones recursivas: árboles

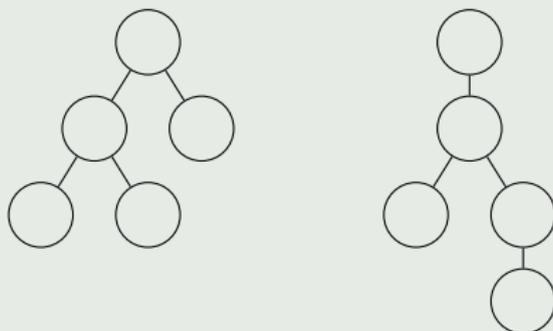
Convención

- Denotamos por \mathcal{E} un elemento vacío, con $|\mathcal{E}| = 0$.
- Denotamos por \mathcal{Z} un “átomo”, con $|\mathcal{Z}| = 1$.

Ejemplo

Los árboles unarios-binarios \mathcal{A} están especificados por

$$\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}.$$



$$A(z) = z + zA(z) + z(A(z))^2.$$

Construcciones básicas: secuencia

Secuencia $\mathcal{A} = \text{Seq}(\mathcal{B})$

Dada una clase combinatoria \mathcal{B} , sin elementos de talla 0, $\mathcal{B}_0 = \emptyset$, definimos

$$\begin{aligned}\mathcal{A} &= \{(a_1, \dots, a_k) : k \geq 0, a_i \in \mathcal{A}\} \\ &= \{\epsilon\} + \mathcal{A} + \mathcal{A} \times \mathcal{A} + \dots,\end{aligned}$$

con la función de talla asociada a los productos.

Se cumple:

$$A(z) = \frac{1}{1 - B(z)}.$$

Construcciones básicas: secuencia

Secuencia $\mathcal{A} = \text{Seq}(\mathcal{B})$

Dada una clase combinatoria \mathcal{B} , sin elementos de talla 0, $\mathcal{B}_0 = \emptyset$, definimos

$$\begin{aligned}\mathcal{A} &= \{(a_1, \dots, a_k) : k \geq 0, a_i \in \mathcal{A}\} \\ &= \{\epsilon\} + \mathcal{A} + \mathcal{A} \times \mathcal{A} + \dots,\end{aligned}$$

con la función de talla asociada a los productos.

Se cumple:

$$A(z) = \frac{1}{1 - B(z)}.$$

Ejemplo. Los enteros $\mathbb{Z}_{\geq 0}$ como una clase combinatoria: $\text{Seq}(\{\bullet\})$

Ejemplo: particiones enteras

Definición: particiones enteras

Una partición de $n \in \mathbb{Z}_{\geq 1}$ es una secuencia $a_1 \leq a_2 \leq \dots \leq a_k$ de enteros positivos tales que $a_1 + a_2 + \dots + a_k = n$.

Es decir, formas de sumar n si no consideramos orden:

$$7 = 7; 7 = 6 + 1; 7 = 5 + 2; 7 = 5 + 1 + 1; 7 = \dots$$

Ejemplo: particiones enteras

Definición: particiones enteras

Una partición de $n \in \mathbb{Z}_{\geq 1}$ es una secuencia $a_1 \leq a_2 \leq \dots \leq a_k$ de enteros positivos tales que $a_1 + a_2 + \dots + a_k = n$.

Es decir, formas de sumar n si no consideramos orden:

$$7 = 7; 7 = 6 + 1; 7 = 5 + 2; 7 = 5 + 1 + 1; 7 = \dots$$

Equivalente: una partición es elegir *cuantos unos, cuantos dos, cuantos tres*, etc. :

$$\mathcal{P} = \text{Seq}(\mathbf{1}) \times \text{Seq}(\mathbf{2}) \times \text{Seq}(\mathbf{3}) \times \dots$$

con la talla $|\mathbf{k}|_{\mathcal{P}} = k$ para cada k .

Ejemplo: particiones enteras

Definición: particiones enteras

Una partición de $n \in \mathbb{Z}_{\geq 1}$ es una secuencia $a_1 \leq a_2 \leq \dots \leq a_k$ de enteros positivos tales que $a_1 + a_2 + \dots + a_k = n$.

Es decir, formas de sumar n si no consideramos orden:

$$7 = 7; 7 = 6 + 1; 7 = 5 + 2; 7 = 5 + 1 + 1; 7 = \dots$$

Equivalente: una partición es elegir *cuantos unos, cuantos dos, cuantos tres*, etc. :

$$\mathcal{P} = \text{Seq}(\mathbf{1}) \times \text{Seq}(\mathbf{2}) \times \text{Seq}(\mathbf{3}) \times \dots$$

con la talla $|\mathbf{k}|_{\mathcal{P}} = k$ para cada k .

$$P(z) = \prod_{n=1}^{\infty} \frac{1}{1 - z^n}.$$

Multiset no etiquetado

La construcción que acabamos de ver es el *Multiset*

MSet

Dada una clase combinatoria sin etiquetas \mathcal{A} , con $\mathcal{A}_0 = \emptyset$, su multiset es

$$\text{MSet}(\mathcal{A}) = \prod_{a \in \mathcal{A}} \text{Seq}(a).$$

Proposición

La función generatriz ordinaria de $\text{MSet}(\mathcal{A})$ es

$$M(z) = \prod_{n=1}^{\infty} \left(\frac{1}{1-z^n} \right)^{a_n} = \exp \left(\sum_{n=1}^{\infty} \frac{1}{n} A(z^n) \right).$$

Multiset no etiquetado

La construcción que acabamos de ver es el *Multiset*

MSet

Dada una clase combinatoria sin etiquetas \mathcal{A} , con $\mathcal{A}_0 = \emptyset$, su multiset es

$$\text{MSet}(\mathcal{A}) = \prod_{a \in \mathcal{A}} \text{Seq}(a).$$

Proposición

La función generatriz ordinaria de $\text{MSet}(\mathcal{A})$ es

$$M(z) = \prod_{n=1}^{\infty} \left(\frac{1}{1-z^n} \right)^{a_n} = \exp \left(\sum_{n=1}^{\infty} \frac{1}{n} A(z^n) \right).$$

Para las particiones enteras

$$P(z) = \exp \left(\sum_{n=1}^{\infty} \frac{1}{n} \frac{z^n}{1-z^n} \right).$$

Powerset no etiquetado

Pregunta

¿Y si los sumandos tuvieran que ser distintos?

Powerset no etiquetado

Pregunta

¿Y si los sumandos tuvieran que ser distintos?

PowerSet

Dada una clase combinatoria sin etiquetas \mathcal{A} , con $\mathcal{A}_0 = \emptyset$, su power-set es

$$\text{PSet}(\mathcal{A}) = \prod_{a \in \mathcal{A}} (\mathcal{E} + a).$$

Notación. \mathcal{E} para el elemento vacío, con $|\mathcal{E}| = 0$.

Powerset no etiquetado

Pregunta

¿Y si los sumandos tuvieran que ser distintos?

PowerSet

Dada una clase combinatoria sin etiquetas \mathcal{A} , con $\mathcal{A}_0 = \emptyset$, su power-set es

$$\text{PSet}(\mathcal{A}) = \prod_{a \in \mathcal{A}} (\mathcal{E} + a).$$

Notación. \mathcal{E} para el elemento vacío, con $|\mathcal{E}| = 0$.

Proposición

La función generatriz ordinaria de $\text{PSet}(\mathcal{A})$ es

$$\prod_{n=1}^{\infty} (1 + z^n)^{a_n} = \exp \left(\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} A(z^n) \right).$$

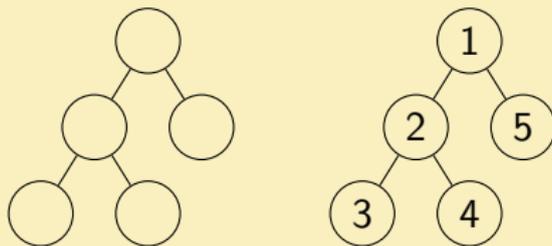
Construcciones para objetos no etiquetados

Construcción	OGF
\mathcal{E}	1
\mathcal{Z}	z
$\mathcal{A} + \mathcal{B}$	$A(z) + B(z)$
$\mathcal{A} \times \mathcal{B}$	$A(z) \times B(z)$
$\text{Seq}(\mathcal{A})$	$\frac{1}{1-A(z)}$
$\text{MSet}(\mathcal{A})$	$\exp\left(\sum_{n=1}^{\infty} \frac{1}{n} A(z^n)\right)$
$\text{PSet}(\mathcal{A})$	$\exp\left(\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} A(z^n)\right)$

Objetos etiquetados

Clases etiquetadas

En un objeto etiquetado de talla n , cada uno de los n "átomos" tiene un número distinto asociado de $[n]$.



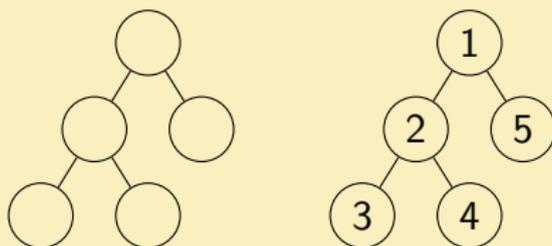
Objetos etiquetados caracterizados por:

figura de base con átomos vacíos + etiquetas para los átomos.

Objetos etiquetados

Clases etiquetadas

En un objeto etiquetado de talla n , cada uno de los n "átomos" tiene un número distinto asociado de $[n]$.



Objetos etiquetados caracterizados por:

figura de base con átomos vacíos + etiquetas para los átomos.

Esto introduce cambios:

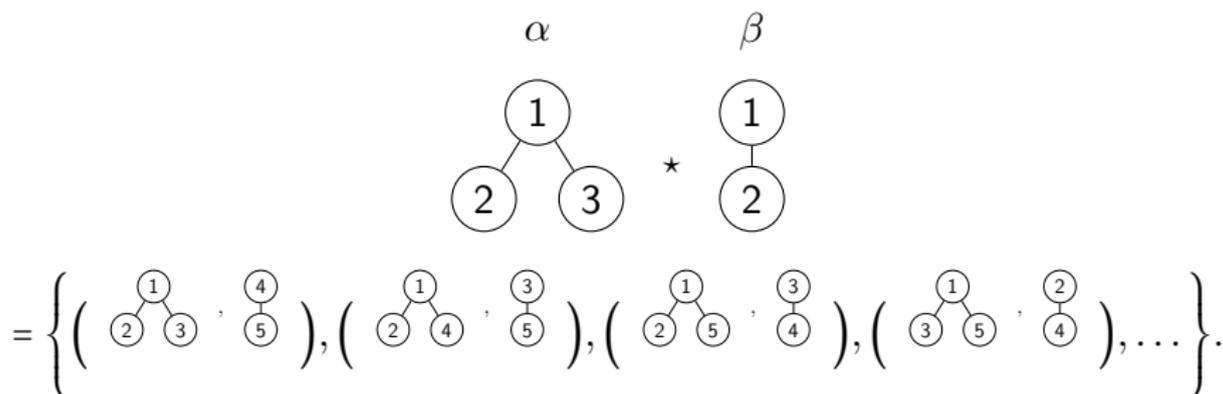
- La regla del producto cambia: hay que particionar las etiquetas.
- Funciones generatrices exponenciales (EGF): $A(z) = \sum_{n=0}^{\infty} \frac{a_n}{n!} z^n$

Clases combinatorias etiquetadas

Producto etiquetado: objetos etiquetados

Sean α y β dos objetos etiquetados, con $|\alpha| = n$ y $|\beta| = m$.

El producto $\alpha \star \beta$ es el conjunto de **todos los pares** con los átomos **etiquetados correctamente** de 1 a $n + m$.

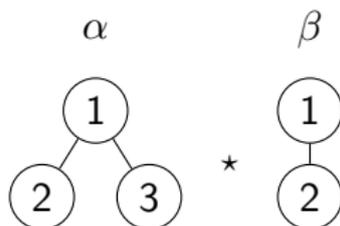


Clases combinatorias etiquetadas

Producto etiquetado: objetos etiquetados

Sean α y β dos objetos etiquetados, con $|\alpha| = n$ y $|\beta| = m$.

El producto $\alpha \star \beta$ es el conjunto de **todos los pares** con los átomos **etiquetados correctamente** de 1 a $n + m$.



$$= \left\{ \left(\begin{array}{c} 1 \\ \diagdown \quad \diagup \\ 2 \quad 3 \end{array}, \begin{array}{c} 4 \\ \circ \\ 5 \end{array} \right), \left(\begin{array}{c} 1 \\ \diagdown \quad \diagup \\ 2 \quad 4 \end{array}, \begin{array}{c} 3 \\ \circ \\ 5 \end{array} \right), \left(\begin{array}{c} 1 \\ \diagdown \quad \diagup \\ 2 \quad 5 \end{array}, \begin{array}{c} 3 \\ \circ \\ 4 \end{array} \right), \left(\begin{array}{c} 1 \\ \diagdown \quad \diagup \\ 3 \quad 5 \end{array}, \begin{array}{c} 2 \\ \circ \\ 4 \end{array} \right), \dots \right\}$$

En total $\binom{3+2}{3} = 10$ elementos !

Tabla de construcciones etiquetadas

Construcción	Función Generatriz Exponencial
\mathcal{E}	1
\mathcal{Z}	z
$\mathcal{A} + \mathcal{B}$	$A(z) + B(z)$
$\mathcal{A} * \mathcal{B}$	$A(z) \times B(z)$
$\text{Seq}(\mathcal{A})$	$\frac{1}{1-A(z)}$
$\text{Set}_k(\mathcal{A})$	$\frac{(A(z))^k}{k!}$
$\text{Set}(\mathcal{A})$	$\exp(A(z))$

⚠ No trataremos las clases etiquetadas, pero se puede consultar las referencias.

Funciones generatrices multivariadas

Para estudiar un parámetro, como la cantidad de ciclos, introducimos nuevas variables formales (quizás más de una):

$$A(z, u) = \sum_{n,k} a_{n,k} z^n u^k, \quad B(z, u) = \sum_{n,k} b_{n,k} \frac{z^n}{n!} u^k.$$

Funciones generatrices multivariadas

Para estudiar un parámetro, como la cantidad de ciclos, introducimos nuevas variables formales (quizás más de una):

$$A(z, u) = \sum_{n,k} a_{n,k} z^n u^k, \quad B(z, u) = \sum_{n,k} b_{n,k} \frac{z^n}{n!} u^k.$$

El valor medio del parámetro k está relacionado con las derivadas en $u = 1$.

Proposición

$$\frac{[z^n] \partial_u A(z, 1)}{[z^n] A(z, 1)} = \frac{\sum_k k \cdot a_{n,k}}{a_n}, \quad \frac{[z^n] \partial_u B(z, 1)}{[z^n] B(z, 1)} = \frac{\sum_k k \cdot b_{n,k}}{b_n}.$$

Clases combinatorias multivariadas

Objetos parametrizados, no sólo por talla.

Supongamos un parámetro $\chi: \mathcal{A} \rightarrow \mathbb{Z}_{\geq 0}$:

$$\mathcal{A}_{n,k} = \{a \in \mathcal{A} : |a| = n, \chi(a) = k\}.$$

En el caso no etiquetado definimos

$$A(z, u) = \sum_{a \in \mathcal{A}} z^{|a|} u^{\chi(a)} = \sum_{n,k \geq 0} a_{n,k} z^n u^k.$$

¹¹Más precisamente $\chi((\alpha, \beta)) = \chi(\alpha) + \chi(\beta)$.

Clases combinatorias multivariadas

Objetos parametrizados, no sólo por talla.

Supongamos un parámetro $\chi: \mathcal{A} \rightarrow \mathbb{Z}_{\geq 0}$:

$$\mathcal{A}_{n,k} = \{a \in \mathcal{A} : |a| = n, \chi(a) = k\}.$$

En el caso no etiquetado definimos

$$A(z, u) = \sum_{a \in \mathcal{A}} z^{|a|} u^{\chi(a)} = \sum_{n,k \geq 0} a_{n,k} z^n u^k.$$

Nuestras construcciones funcionan verbatim:

- Si parámetros χ son aditivos¹¹, construcciones de diccionario funcionan,
- Se introduce \mathcal{U} , un nuevo átomo con función generatriz $= u$,
- El átomo marca \mathcal{U} el parámetro de interés.

¹¹Más precisamente $\chi((\alpha, \beta)) = \chi(\alpha) + \chi(\beta)$.

Ejemplo: composiciones de enteros

Las composiciones son como las particiones enteras, pero consideramos también los órdenes posibles: una secuencia de enteros

$$\mathcal{C} = \text{Seq}(\{\mathbf{1}, \mathbf{2}, \mathbf{3}, \dots\}) \simeq \text{Seq}(\text{Seq}_{\geq 1}(\mathcal{Z})).$$

Marcando la cantidad de sumandos tenemos

$$\mathcal{C} = \text{Seq}(\mathcal{U} \times \text{Seq}_{\geq 1}(\mathcal{Z})) \implies C(z, u) = \frac{1}{1 - \frac{uz}{1-z}}.$$

Proposición

La cantidad media de sumandos en una composición de n es $\sim n/2$.

Funciones generatrices de probabilidad

Con las funciones de conteo podemos tratar la distribución uniforme:

Si tomamos un objeto uniforme A de talla n , $\Pr_n(\chi(A) = k) = a_{n,k}/a_n$.

Funciones generatrices de probabilidad

Con las funciones de conteo podemos tratar la distribución uniforme:

Si tomamos un objeto uniforme A de talla n , $\Pr_n(\chi(A) = k) = a_{n,k}/a_n$.

¿Y si la distribución **no** fuera **uniforme**?

- Consideramos las *funciones generatrices de probabilidad* **PGF**:

$$P^{(A)}(z, u) := \sum_{n, k \geq 0} \mathbf{P}_n^{(A)}(\chi = k) \cdot z^n u^k, \quad P^{(A)}(z, 1) = \sum_{n=0}^{\infty} 1 \cdot z^n = \frac{1}{1-z}.$$

Funciones generatrices de probabilidad

Con las funciones de conteo podemos tratar la distribución uniforme:

Si tomamos un objeto uniforme A de talla n , $\Pr_n(\chi(A) = k) = a_{n,k}/a_n$.

¿Y si la distribución **no** fuera **uniforme**?

- Consideramos las *funciones generatrices de probabilidad* **PGF**:

$$P^{(A)}(z, u) := \sum_{n, k \geq 0} \mathbf{P}_n^{(A)}(\chi = k) \cdot z^n u^k, \quad P^{(A)}(z, 1) = \sum_{n=0}^{\infty} 1 \cdot z^n = \frac{1}{1-z}.$$

- No hay diccionarios a priori, depende del caso concreto [funciona, por ejemplo, si los objetos son secuencias independientes]

Funciones generatrices de probabilidad

Con las funciones de conteo podemos tratar la distribución uniforme:

Si tomamos un objeto uniforme A de talla n , $\Pr_n(\chi(A) = k) = a_{n,k}/a_n$.

¿Y si la distribución **no** fuera **uniforme**?

- Consideramos las *funciones generatrices de probabilidad* **PGF**:

$$P^{(A)}(z, u) := \sum_{n, k \geq 0} \mathbf{P}_n^{(A)}(\chi = k) \cdot z^n u^k, \quad P^{(A)}(z, 1) = \sum_{n=0}^{\infty} 1 \cdot z^n = \frac{1}{1-z}.$$

- No hay diccionarios a priori, depende del caso concreto [funciona, por ejemplo, si los objetos son secuencias independientes]
- Fórmulas para momentos se cumplen:

$$(u\partial_u)^1 P(z, u) \Big|_{u=1} = \sum_{n=0}^{\infty} \mathbb{E}_n[\chi] z^n, \quad (u\partial_u)^2 P(z, u) \Big|_{u=1} = \sum_{n=0}^{\infty} \mathbb{E}_n[\chi^2] z^n, \dots$$

Para aprender más

-  *Philippe Flajolet y Robert Sedgewick*
Analytic Combinatorics.
<https://algo.inria.fr/flajolet/Publications/books.html>
-  *Herbert S. Wilf*
Generatingfunctionology.
<https://www2.math.upenn.edu/~wilf/DownldGF.html>

1. Introducción al análisis de algoritmos
2. Aplicaciones a la predicción de saltos
3. Combinatoria Analítica: métodos simbólicos
4. **Combinatoria Analítica: Teoría Analítica**
 - Radio de convergencia y coeficientes
 - Funciones analíticas
 - Asintóticos de funciones racionales
 - Asintóticos generales: Teorema de Transferencia
 - Aplicaciones

*Generating functions are a **bridge** between **discrete mathematics**, on the one hand, and **continuous analysis** (particularly complex variable theory) on the other. [...]*

To omit those [analytical] parts of the subject [...] is like listening to a stereo broadcast of, say, Beethoven's Ninth Symphony, using only the left audio channel.

– *Herbert S. Wilf*, prefacio de *generatingfunctionology*, 1989.

Introducción

Cuando la serie $f(z) = \sum_n a_n z^n$ converge, sus propiedades analíticas (como función) están fuertemente relacionadas con los coeficientes.

Si tenemos

$$[z^n]f(z) \sim \theta(n)R^{-n}$$

1. **Primer principio:** el radio de convergencia R de la serie de potencias determina el crecimiento exponencial.
2. **Segundo principio:** el factor sub-exponencial $\theta(n)$ está relacionado con el tipo de singularidades de $f(z)$.

Radio de convergencia

Una serie $\sum_{n \geq 0} c_n$ converge cuando $\lim_{N \rightarrow \infty} \sum_{n=0}^N c_n$ existe:

- Convergencia absoluta: cuando $\sum |c_n|$ converge,
- Convergencia condicional: cuando converge pero $\sum |c_n| = \infty$.

Radio de convergencia

Una serie $\sum_{n \geq 0} c_n$ converge cuando $\lim_{N \rightarrow \infty} \sum_{n=0}^N c_n$ existe:

- Convergencia absoluta: cuando $\sum |c_n|$ converge,
- Convergencia condicional: cuando converge pero $\sum |c_n| = \infty$.

Sea $f(z) = \sum_{n \geq 0} a_n z^n$ serie de potencias

Radio de convergencia

El radio de convergencia de $f(z)$ es $R \geq 0$ (quizás $R = \infty$) tal que la serie converge absolutamente para $|z| < R$ y diverge para $|z| > R$.

Coeficientes y radio de convergencia

El **primer principio** está dado por el siguiente teorema:

Teorema (Cauchy-Hadamard)

El radio de convergencia $0 \leq R \leq \infty$ de una serie de potencias $f(z) = \sum_{n \geq 0} a_n z^n$ está determinado por

$$R = \frac{1}{\limsup_{n \rightarrow \infty} |a_n|^{1/n}}.$$

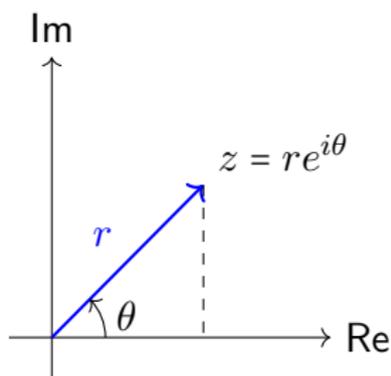
En otras palabras, cuando $0 < R < \infty$, para cualquier $\epsilon > 0$:

- tenemos $|a_n|^{1/n} < \frac{1}{R} + \epsilon$ para todo n suficientemente grande,
- tenemos $|a_n|^{1/n} > \frac{1}{R} - \epsilon$ para una infinidad de n .

Los números complejos

Un *complejo* se escribe $z = x + iy$, $x, y \in \mathbb{R}$, donde i es raíz de $x^2 + 1 = 0$.

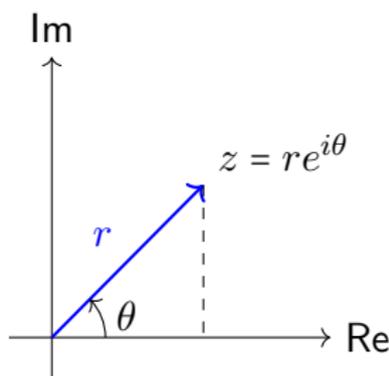
- Escribimos $x = \Re z$, $y = \Im z$, la parte real e imaginaria de z .
- Representación en **polares**: $z = re^{i\theta}$ donde $r = |z|$ es el módulo y $\theta = \arg(z) \in (-\pi, \pi]$ es el **argumento** o fase de z .



Los números complejos

Un *complejo* se escribe $z = x + iy$, $x, y \in \mathbb{R}$, donde i es raíz de $x^2 + 1 = 0$.

- Escribimos $x = \Re z$, $y = \Im z$, la parte real e imaginaria de z .
- Representación en **polares**: $z = re^{i\theta}$ donde $r = |z|$ es el módulo y $\theta = \arg(z) \in (-\pi, \pi]$ es el **argumento** o fase de z .



Ahora introduciremos nociones básicas de análisis complejo.

Diferenciación en \mathbb{C}

Consideramos **dominios** \mathcal{D} : abiertos, conexos por caminos.

Definición

Sea $f: \mathcal{D} \rightarrow \mathbb{C}$. Decimos que f es derivable en $z_0 \in \mathcal{D}$ si

$$\lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

existe, y denotamos el límite $f'(z_0)$.

Una función f derivable en cada punto de \mathcal{D} se dice **holomorfa** en \mathcal{D} .

Diferenciación en \mathbb{C}

Consideramos **dominios** \mathcal{D} : abiertos, conexos por caminos.

Definición

Sea $f: \mathcal{D} \rightarrow \mathbb{C}$. Decimos que f es derivable en $z_0 \in \mathcal{D}$ si

$$\lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

existe, y denotamos el límite $f'(z_0)$.

Una función f derivable en cada punto de \mathcal{D} se dice **holomorfa** en \mathcal{D} .

La derivación compleja es más fuerte que la real:

- Las reglas del producto, división y cadena se cumplen,
- Además tenemos las ecuaciones de **Cauchy-Riemann**

$$\partial_x g = \partial_y h, \quad \partial_x h = -\partial_y g,$$

donde $g(x, y) := \Re(f(x + iy))$ y $h(x, y) := \Im(f(x + iy))$.

Definición

Una función $f: \mathcal{D} \rightarrow \mathbb{C}$ se dice analítica en $z_0 \in \mathcal{D}$ si existe una bola $B_r(z_0) \subset \mathcal{D}$ sobre la cual

$$f(z) = \sum_{n \geq 0} c_n (z - z_0)^n,$$

con la serie convergente para $z \in B_r(z_0)$. Si f es analítica en cada punto de \mathcal{D} se dice analítica en \mathcal{D} .

Funciones analíticas

Definición

Una función $f: \mathcal{D} \rightarrow \mathbb{C}$ se dice analítica en $z_0 \in \mathcal{D}$ sii existe una bola $B_r(z_0) \subset \mathcal{D}$ sobre la cual

$$f(z) = \sum_{n \geq 0} c_n (z - z_0)^n,$$

con la serie convergente para $z \in B_r(z_0)$. Si f es analítica en cada punto de \mathcal{D} se dice analítica en \mathcal{D} .

Teorema

Una función es holomorfa en \mathcal{D} si y solamente si es analítica.

Funciones analíticas: derivación es integración

Teorema (Fórmula de Cauchy)

Si $f(z)$ es analítica en $z = z_0$, con $f(z) = \sum_{n \geq 0} c_n (z - z_0)^n$,

$$c_n = \frac{f^{(n)}(z_0)}{n!} = \frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{(z - z_0)^{n+1}} dz,$$

para cualquier círculo γ suficientemente pequeño centrado en z_0 .

Funciones analíticas: derivación es integración

Teorema (Fórmula de Cauchy)

Si $f(z)$ es analítica en $z = z_0$, con $f(z) = \sum_{n \geq 0} c_n (z - z_0)^n$,

$$c_n = \frac{f^{(n)}(z_0)}{n!} = \frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{(z - z_0)^{n+1}} dz,$$

para cualquier círculo γ suficientemente pequeño centrado en z_0 .

- Una función derivable una vez, lo es infinitas veces.
- Los círculos γ se pueden substituir por cualquier otra curva simple, cuyo interior esté contenido el dominio de analiticidad.
- Fórmula para coeficientes permite aproximar con método de Laplace.

Funciones analíticas: derivación es integración

Teorema (Fórmula de Cauchy)

Si $f(z)$ es analítica en $z = z_0$, con $f(z) = \sum_{n \geq 0} c_n (z - z_0)^n$,

$$c_n = \frac{f^{(n)}(z_0)}{n!} = \frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{(z - z_0)^{n+1}} dz,$$

para cualquier círculo γ suficientemente pequeño centrado en z_0 .

- Una función derivable una vez, lo es infinitas veces.
- Los círculos γ se pueden substituir por cualquier otra curva simple, cuyo interior esté contenido el dominio de analiticidad.
- Fórmula para coeficientes permite aproximar con método de Laplace.

⚠ No es la perspectiva que adoptaremos (por simplicidad), pero es importante para teoremas más avanzados (Teorema de Transferencia, Saddle-Point,...).

Singularidades

Una *singularidad* es un punto en el cual $f(z)$ no es analítica/holomorfa.

Tipo	Condición	Ejemplo
Polo (n)	$(z - a)^n f(z) \rightarrow \text{finito}$	$\frac{1}{(z-a)^2}$
Esencial	No existe límite	$e^{1/z}$
Rama	Multivaluada	$\sqrt{z}, \log z$

El logaritmo

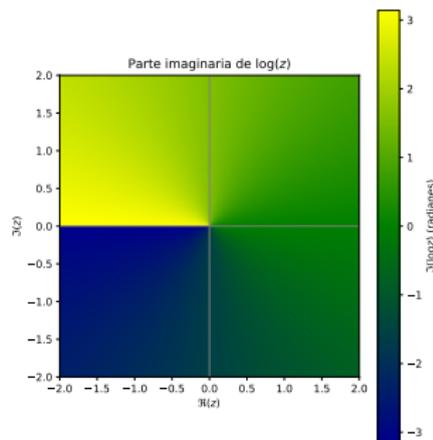
Como $e^{2\pi ik} = 1$ para todo $k \in \mathbb{Z}$, hay infinitas definiciones posibles para $\log z$.

Consideramos

$$\log z = \log |z| + \arg(z) \times i,$$

donde $\arg(z)$ está comprendido en $(-\pi, \pi]$.

Figura. El logaritmo es analítico en $\mathbb{C} \setminus (-\infty, 0]$ y **no** se puede extender.



Potencias no enteras

Potencias no enteras

Definimos $g(z) = z^\alpha$ como $g(z) := \exp(\alpha \times \log z)$ para $\mathbb{C} \setminus (-\infty, 0]$,

- cuando $\alpha \in \mathbb{Z}$ esto coincide con la definición clásica,
- para el resto obtenemos una definición sobre $\mathbb{C} \setminus (-\infty, 0]$.

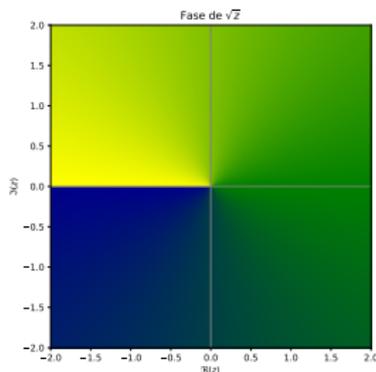
Potencias no enteras

Potencias no enteras

Definimos $g(z) = z^\alpha$ como $g(z) := \exp(\alpha \times \log z)$ para $\mathbb{C} \setminus (-\infty, 0]$,

- cuando $\alpha \in \mathbb{Z}$ esto coincide con la definición clásica,
- para el resto obtenemos una definición sobre $\mathbb{C} \setminus (-\infty, 0]$.

Ejemplo. $z \mapsto \sqrt{z}$ es analítica en $\mathbb{C} \setminus (-\infty, 0]$, pero no se puede extender más.



Teorema

Consideramos un dominio \mathcal{D} [abierto y conexo por caminos].

Si existe $(z_n)_n$ en \mathcal{D} con $f(z_n) = 0$ y $z_n \rightarrow z_\infty \in \mathcal{D}$, entonces $f(z) \equiv 0$.

Extensión analítica

Teorema

Consideramos un dominio \mathcal{D} [abierto y conexo por caminos].
Si existe $(z_n)_n$ en \mathcal{D} con $f(z_n) = 0$ y $z_n \rightarrow z_\infty \in \mathcal{D}$, entonces $f(z) \equiv 0$.

Un corolario importantísimo de este resultado es el siguiente:

Corolario: unicidad de la extensión analítica

Consideremos dominios $\mathcal{D} \subset \hat{\mathcal{D}}$. Sea $f: \mathcal{D} \rightarrow \mathbb{C}$ y supongamos que $\hat{f}: \hat{\mathcal{D}} \rightarrow \mathbb{C}$ es analítica y satisface $\hat{f}|_{\mathcal{D}} \equiv f$.

Entonces \hat{f} es la única extensión analítica de f a $\hat{\mathcal{D}}$.

Extensión analítica

- La serie $f(z) = \sum_{n=0}^{\infty} z^n$ define una función analítica sobre $|z| < 1$.

Sabemos que $f(z) = \frac{1}{1-z}$ vale sobre $|z| < 1$: $\tilde{f}(z) = \frac{1}{1-z}$ es su extensión analítica a $\mathbb{C} \setminus \{1\}$.

Extensión analítica

- La serie $f(z) = \sum_{n=0}^{\infty} z^n$ define una función analítica sobre $|z| < 1$.

Sabemos que $f(z) = \frac{1}{1-z}$ vale sobre $|z| < 1$: $\tilde{f}(z) = \frac{1}{1-z}$ es su extensión analítica a $\mathbb{C} \setminus \{1\}$.

- Existe una única función analítica en $\mathbb{C} \setminus (-\infty, 0]$ que extiende $t \mapsto \log t$, definida en $t \in \mathbb{R}_{>0}$. [usar el teorema precedente]

Extensión analítica

- La serie $f(z) = \sum_{n=0}^{\infty} z^n$ define una función analítica sobre $|z| < 1$.
Sabemos que $f(z) = \frac{1}{1-z}$ vale sobre $|z| < 1$: $\tilde{f}(z) = \frac{1}{1-z}$ es su extensión analítica a $\mathbb{C} \setminus \{1\}$.
- Existe una única función analítica en $\mathbb{C} \setminus (-\infty, 0]$ que extiende $t \mapsto \log t$, definida en $t \in \mathbb{R}_{>0}$. [usar el teorema precedente]

Teorema (Singularidades removibles)

Si $f(z)$ es analítica en $\mathcal{D} \setminus \{z_0\}$ y es acotada cuando $z \rightarrow z_0$, entonces se puede extender analíticamente a z_0 .

Ejemplo: $\sin(z)/z$ tiene una singularidad removible en $z = 0$.

Determinar el radio de convergencia

Esencialmente: las funciones generatrices convergen “hasta que encuentran una singularidad”.

Proposición

Sea $f(z) = \sum_{n \geq 0} a_n z^n$ con radio de convergencia $R < \infty$, entonces

1. no tiene ninguna singularidad en el disco abierto $|z| < R$.
2. tiene al menos una singularidad en el círculo $|z| = R$,

Determinar el radio de convergencia

Esencialmente: las funciones generatrices convergen “hasta que encuentran una singularidad”.

Proposición

Sea $f(z) = \sum_{n \geq 0} a_n z^n$ con radio de convergencia $R < \infty$, entonces

1. no tiene ninguna singularidad en el disco abierto $|z| < R$.
2. tiene al menos una singularidad en el círculo $|z| = R$,

Determinar el radio de convergencia

Esencialmente: las funciones generatrices convergen “hasta que encuentran una singularidad”.

Proposición

Sea $f(z) = \sum_{n \geq 0} a_n z^n$ con radio de convergencia $R < \infty$, entonces

1. no tiene ninguna singularidad en el disco abierto $|z| < R$.
2. tiene al menos una singularidad en el círculo $|z| = R$,

Una singularidad z_0 con $|z_0| = R$ se dice **dominante**.

Determinar el radio de convergencia

Esencialmente: las funciones generatrices convergen “hasta que encuentran una singularidad”.

Proposición

Sea $f(z) = \sum_{n \geq 0} a_n z^n$ con radio de convergencia $R < \infty$, entonces

1. no tiene ninguna singularidad en el disco abierto $|z| < R$.
2. tiene al menos una singularidad en el círculo $|z| = R$,

Una singularidad z_0 con $|z_0| = R$ se dice **dominante**.

Teorema (Pringsheim)

Supongamos que $f(z) = \sum a_n z^n$ es analítica en $z = 0$ (es decir, $R > 0$) y $a_n \geq 0$ para todo n .

Si $R < \infty$, entonces $z = R$ es una singularidad.

Funciones racionales

Una función $f(z)$ es racional si $f(z) = \frac{p(z)}{q(z)}$ con $p(z)$ y $q(z)$ polinomios.

En este caso el procedimiento es directo:

- Podemos suponer sin pérdida de generalidad que $\gcd(p(z), q(z)) = 1$.
- Entonces las singularidades corresponden a los ceros de $q(z)$.
- Aplicamos fracciones simples y el siguiente lema

Lema

$$[z^n] \frac{1}{(1 - z/z_0)^{m+1}} = \binom{n+m}{m} z_0^{-n} \sim \frac{n^m}{m!} z_0^{-n}.$$

Funciones racionales

Lema

$$[z^n] \frac{1}{(1 - z/z_0)^{m+1}} = \binom{n+m}{m} z_0^{-n} \sim \frac{n^m}{m!} z_0^{-n}.$$

Ejemplo.

$$f(z) = \frac{1}{(1-2z)^2(1-z)} = \frac{2}{(1-2z)^2} - \frac{2}{1-2z} + \frac{1}{1-z}$$
$$\Rightarrow [z^n]f(z) \sim 2n \times 2^n$$

Funciones racionales

Lema

$$[z^n] \frac{1}{(1 - z/z_0)^{m+1}} = \binom{n+m}{m} z_0^{-n} \sim \frac{n^m}{m!} z_0^{-n}.$$

Ejemplo.

$$f(z) = \frac{1}{(1-2z)^2(1-z)} = \frac{2}{(1-2z)^2} - \frac{2}{1-2z} + \frac{1}{1-z}$$
$$\Rightarrow [z^n]f(z) \sim 2n \times 2^n$$

¿y si no conocemos todas las singularidades?

Funciones racionales

Lema

$$[z^n] \frac{1}{(1 - z/z_0)^{m+1}} = \binom{n+m}{m} z_0^{-n} \sim \frac{n^m}{m!} z_0^{-n}.$$

Ejemplo.

$$f(z) = \frac{1}{(1-2z)^2(1-z)} = \frac{2}{(1-2z)^2} - \frac{2}{1-2z} + \frac{1}{1-z}$$
$$\Rightarrow [z^n]f(z) \sim 2n \times 2^n$$

¿y si no conocemos todas las singularidades?

⇒ estudiamos las **singularidades dominantes**

Funciones racionales

Lema

$$[z^n] \frac{1}{(1 - z/z_0)^{m+1}} = \binom{n+m}{m} z_0^{-n} \sim \frac{n^m}{m!} z_0^{-n}.$$

Ejemplo.

$$f(z) = \frac{1}{(1-2z)^2(1-z)} = \frac{2}{(1-2z)^2} - \frac{2}{1-2z} + \frac{1}{1-z}$$
$$\Rightarrow [z^n]f(z) \sim 2n \times 2^n$$

¿y si no conocemos todas las singularidades?

\Rightarrow estudiamos las **singularidades dominantes**

Singularidades dominantes con *orden m máximo* determinan asintóticos.

$$f(z) \sim \frac{2}{(1-2z)^2}, (z \rightarrow 1/2) \implies [z^n]f(z) \sim 2n \times 2^n.$$

Singularidades dominantes 2

Si hay varias singularidades dominantes en $|z| = R$, se pueden producir fenómenos de oscilación.

Singularidades dominantes 2

Si hay varias singularidades dominantes en $|z| = R$, se pueden producir fenómenos de oscilación.

Ejemplo

$$\frac{1}{1-z^3} = \sum_{n=0}^{\infty} z^{3n} \leftrightarrow \{1, 0, 0, 1, 0, 0, 1, 0, 0, \dots\},$$

en este caso porque las raíces son $z_0 = 1$, $z_1 = e^{2\pi i/3}$, $z_2 = e^{4\pi i/3}$.

Singularidades dominantes 2

Si hay varias singularidades dominantes en $|z| = R$, se pueden producir fenómenos de oscilación.

Ejemplo

$$\frac{1}{1-z^3} = \sum_{n=0}^{\infty} z^{3n} \leftrightarrow \{1, 0, 0, 1, 0, 0, 1, 0, 0, \dots\},$$

en este caso porque las raíces son $z_0 = 1$, $z_1 = e^{2\pi i/3}$, $z_2 = e^{4\pi i/3}$.

$$\frac{1}{1-z^3} = \frac{1/3}{1-z} + \frac{1/3}{1-z/e^{2\pi i/3}} + \frac{1/3}{1-z/e^{4\pi i/3}}.$$

Singularidades dominantes 2

Si hay varias singularidades dominantes en $|z| = R$, se pueden producir fenómenos de oscilación.

Ejemplo

$$\frac{1}{1-z^3} = \sum_{n=0}^{\infty} z^{3n} \leftrightarrow \{1, 0, 0, 1, 0, 0, 1, 0, 0, \dots\},$$

en este caso porque las raíces son $z_0 = 1$, $z_1 = e^{2\pi i/3}$, $z_2 = e^{4\pi i/3}$.

$$\frac{1}{1-z^3} = \frac{1/3}{1-z} + \frac{1/3}{1-z/e^{2\pi i/3}} + \frac{1/3}{1-z/e^{4\pi i/3}}.$$

⇒ importante considerar **todas** las singularidades dominantes.

Ejemplo: el cambio con monedas

Pregunta

Sean a y b enteros positivos coprimos (las monedas). ¿De cuántas maneras se puede representar n como $n = xa + yb$ con enteros $x, y \geq 0$?

Ejemplo: el cambio con monedas

Pregunta

Sean a y b enteros positivos coprimos (las monedas). ¿De cuántas maneras se puede representar n como $n = xa + yb$ con enteros $x, y \geq 0$?

Sea a_n la cantidad de representaciones, notamos que

$$A(z) = \sum a_n z^n = \frac{1}{1 - z^a} \times \frac{1}{1 - z^b}.$$

Ejemplo: el cambio con monedas

Pregunta

Sean a y b enteros positivos coprimos (las monedas). ¿De cuántas maneras se puede representar n como $n = xa + yb$ con enteros $x, y \geq 0$?

Sea a_n la cantidad de representaciones, notamos que

$$A(z) = \sum a_n z^n = \frac{1}{1 - z^a} \times \frac{1}{1 - z^b}.$$

Notamos que todas las raíces de $1 - z^a$ y $1 - z^b$ son raíces de la unidad.

Ejemplo: el cambio con monedas

Pregunta

Sean a y b enteros positivos coprimos (las monedas). ¿De cuántas maneras se puede representar n como $n = xa + yb$ con enteros $x, y \geq 0$?

Sea a_n la cantidad de representaciones, notamos que

$$A(z) = \sum a_n z^n = \frac{1}{1-z^a} \times \frac{1}{1-z^b}.$$

Notamos que todas las raíces de $1 = z^a$ y $1 = z^b$ son raíces de la unidad. Como a y b son coprimos, todas son *raíces simples salvo* $z = 1$ que es *doble*:

$$A(z) = \frac{A}{(1-z)^2} + \frac{B}{1-z} + \sum_{u:u^b=1, u \neq 1} \frac{c_u}{1-z/u} + \sum_{u:u^a=1, u \neq 1} \frac{c_u}{1-z/u}.$$

Ejemplo: el cambio con monedas

Pregunta

Sean a y b enteros positivos coprimos (las monedas). ¿De cuántas maneras se puede representar n como $n = xa + yb$ con enteros $x, y \geq 0$?

Sea a_n la cantidad de representaciones, notamos que

$$A(z) = \sum a_n z^n = \frac{1}{1-z^a} \times \frac{1}{1-z^b}.$$

Notamos que todas las raíces de $1 = z^a$ y $1 = z^b$ son raíces de la unidad. Como a y b son coprimos, todas son *raíces simples salvo* $z = 1$ que es *doble*:

$$A(z) = \frac{A}{(1-z)^2} + \frac{B}{1-z} + \sum_{u:u^b=1, u \neq 1} \frac{c_u}{1-z/u} + \sum_{u:u^a=1, u \neq 1} \frac{c_u}{1-z/u}.$$

Se sigue que $a_n \sim n \times A$. Para calcular A observamos que

$$A = \lim_{z \rightarrow 1} (1-z)^2 A(z) = \frac{1}{ab}.$$

Singularidades generales

Conocemos algunas **singularidades de otros tipos**:

$$\log\left(\frac{1}{1-z}\right) = \sum_{n=1}^{\infty} \frac{z^n}{n}, \quad \frac{1}{1-z} \log\left(\frac{1}{1-z}\right) = \sum_{n=1}^{\infty} H_n z^n,$$

... integrando $\log \log\left(\frac{1}{1-z}\right) = \sum_{n=1}^{\infty} \frac{1}{n+1} H_n z^{n+1}$.

Singularidades generales

Conocemos algunas **singularidades de otros tipos**:

$$\log\left(\frac{1}{1-z}\right) = \sum_{n=1}^{\infty} \frac{z^n}{n}, \quad \frac{1}{1-z} \log\left(\frac{1}{1-z}\right) = \sum_{n=1}^{\infty} H_n z^n,$$

... integrando $\log \log\left(\frac{1}{1-z}\right) = \sum_{n=1}^{\infty} \frac{1}{n+1} H_n z^{n+1}$.

Pero:

- ¿Y si $f(z) = (1-z)^{-\alpha}$ con $\alpha \in \mathbb{R} \setminus \mathbb{Z}_{\geq 1}$?
- ¿Y si $f(z) \sim \frac{1}{1-z} \log\left(\frac{1}{1-z}\right)$ cuando $z \rightarrow 1$ en lugar de tener igualdad?

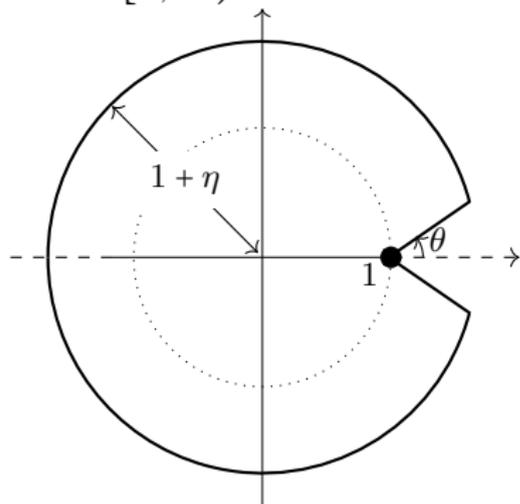
Teorema de Transferencia

Consideramos dominios más generales que $\mathbb{C} \setminus [1, \infty)$:

Dominio “Camembert” o “Pacman” 🍷

$$\Delta(\theta, \eta) = \{z : |z| \leq 1 + \eta, \quad \arg(z-1) \geq \theta\}.$$

Supongamos $f(z)$ analítica en $\Delta(\theta, \eta)$,
excepto quizás $z = 1$.



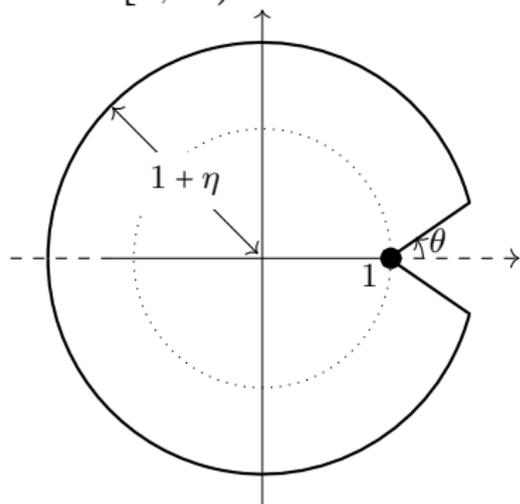
Teorema de Transferencia

Consideramos dominios más generales que $\mathbb{C} \setminus [1, \infty)$:

Dominio “Camembert” o “Pacman” 🍷

$$\Delta(\theta, \eta) = \{z : |z| \leq 1 + \eta, \quad \arg(z-1) \geq \theta\}.$$

Supongamos $f(z)$ analítica en $\Delta(\theta, \eta)$,
excepto quizás $z = 1$.



Teorema (Flajolet, Odlyzko)

Si $f(z) \sim \frac{1}{(1-z)^\alpha} \left(\log\left(\frac{1}{1-z}\right)\right)^\beta$ cuando $z \rightarrow 1$, $\alpha \neq 0, -1, -2, \dots$, entonces

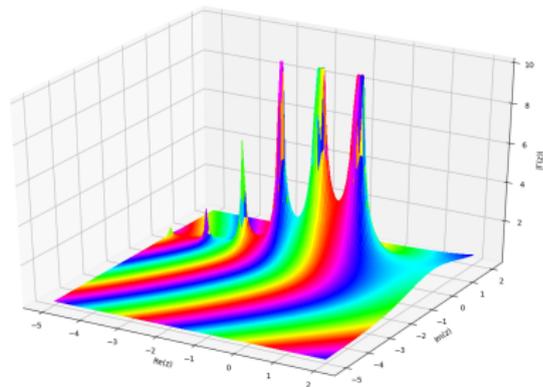
$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} (\log n)^\beta.$$

La función Gamma

Para $\Re(z) > 0$ se define por

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt.$$

- Satisface $\Gamma(z+1) = z\Gamma(z)$: extiende factoriales $\Gamma(n+1) = n!$, $n \in \mathbb{Z}_{\geq 0}$.
- Otros valores importantes $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ y $\Gamma(-\frac{1}{2}) = -2\sqrt{\pi}$
- Se puede extender analíticamente a $\mathbb{C} \setminus \{-1, -2, \dots\}$. Polos simples en enteros negativos $\Gamma(z) \sim \frac{(-1)^n/n!}{z+n}$, $z \rightarrow -n$



Teorema de Transferencia II

El Teorema de Transferencia es *muy flexible*:

- Claramente la **singularidad puede ser $\rho \neq 1$** : se aplica el resultado a $g(z) = f(z/\rho)$ en ese caso.
- Se aplica cambiando $(\log \frac{1}{1-z})^\beta$ por cualquier producto finito $(\log \frac{1}{1-z})^{\beta_1} (\log \log \frac{1}{1-z})^{\beta_2} (\log \log \log \frac{1}{1-z})^{\beta_3} \dots$
- Se aplica cambiando equivalentes \sim por cotas O y o .
- Se puede generalizar a **cantidad finita de singularidades**¹² en $|z| = 1$.

¹²Atención a las cancelaciones al sumar equivalentes.

Ejemplo: Strings bien parentizadas

Recordamos que la función generatriz de las **strings bien parentizadas** es

$$S(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

Ejemplo: Strings bien parentizadas

Recordamos que la función generatriz de las **strings bien parentizadas** es

$$S(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

Usando el Teorema de Transferencia:

$$[z^n]S(z) = -\frac{1}{2}[z^{n+1}]\sqrt{1 - 4z} \sim -\frac{1}{2} \frac{n^{-1/2-1}}{\Gamma(-1/2)} 4^{n+1} = \frac{1}{\sqrt{\pi}} \cdot \frac{4^n}{n^{3/2}}$$

Quicksort

Recordatorio

La cantidad esperada de comparaciones $E_n = \mathbb{E}[C_n]$ satisface $E_n = 2 \cdot (n + 1)H_n - 4n$.

Quicksort

Recordatorio

La cantidad esperada de comparaciones $E_n = \mathbb{E}[C_n]$ satisface $E_n = 2 \cdot (n + 1)H_n - 4n$.

Vamos a probar esto ahora con funciones generatrices. Recordamos que

$$E_n = \frac{1}{n} \sum_{j=0}^{n-1} (E_j + E_{n-j} + n - 1) .$$

Recordatorio

La cantidad esperada de comparaciones $E_n = \mathbb{E}[C_n]$ satisface $E_n = 2 \cdot (n + 1)H_n - 4n$.

Vamos a probar esto ahora con funciones generatrices. Recordamos que

$$E_n = \frac{1}{n} \sum_{j=0}^{n-1} (E_j + E_{n-j} + n - 1) .$$

Sea $E(z) = \sum_n E_n z^n$, la recurrencia se traduce en

$$zE'(z) = \frac{2z}{1-z}E(z) + \frac{2z^2}{(1-z)^3} .$$

Recordatorio

La cantidad esperada de comparaciones $E_n = \mathbb{E}[C_n]$ satisface $E_n = 2 \cdot (n + 1)H_n - 4n$.

Vamos a probar esto ahora con funciones generatrices. Recordamos que

$$E_n = \frac{1}{n} \sum_{j=0}^{n-1} (E_j + E_{n-j} + n - 1).$$

Sea $E(z) = \sum_n E_n z^n$, la recurrencia se traduce en

$$zE'(z) = \frac{2z}{1-z}E(z) + \frac{2z^2}{(1-z)^3}.$$

Resolviendo,

$$E(z) = \frac{2}{(1-z)^2} \log\left(\frac{1}{1-z}\right) - \frac{2z}{(1-z)^2},$$

y vemos inmediatamente que $E_n = 2n \log n + O(n)$.

Un paréntesis sobre las EDO

En general, la ecuación $u'(z) = a(z)u(z) + b(z)$ tiene una única solución

Teorema

Si $a(z)$ y $b(z)$ son analíticas en $z = 0$, en un entorno de $z = 0$

$$u(z) = \exp\left(\int_0^z a(v)dv\right) \cdot \left(u(0) + \int_0^z \exp\left(-\int_0^u a(v)dv\right) \cdot b(u)du\right).$$

Un paréntesis sobre las EDO

En general, la ecuación $u'(z) = a(z)u(z) + b(z)$ tiene una única solución

Teorema

Si $a(z)$ y $b(z)$ son analíticas en $z = 0$, en un entorno de $z = 0$

$$u(z) = \exp\left(\int_0^z a(v)dv\right) \cdot \left(u(0) + \int_0^z \exp\left(-\int_0^u a(v)dv\right) \cdot b(u)du\right).$$

Demostración.

Definimos $v(z) := \exp\left(-\int_0^z a(v)dv\right) \cdot u(z)$ y derivamos. □

Un paréntesis sobre las EDO

En general, la ecuación $u'(z) = a(z)u(z) + b(z)$ tiene una única solución

Teorema

Si $a(z)$ y $b(z)$ son analíticas en $z = 0$, en un entorno de $z = 0$

$$u(z) = \exp\left(\int_0^z a(v)dv\right) \cdot \left(u(0) + \int_0^z \exp\left(-\int_0^u a(v)dv\right) \cdot b(u)du\right).$$

Demostración.

Definimos $v(z) := \exp\left(-\int_0^z a(v)dv\right) \cdot u(z)$ y derivamos. □

Las singularidades de una EDO lineal están relacionadas con las singularidades de los coeficientes:

Ejercicio

Sea $u'(z) = a(z)u(z)$. Si $a(z)$ es analítica en $\mathbb{C} \setminus [\rho, \infty)$, tiene un polo simple $a(z) \sim \frac{\alpha}{z-\rho}$ en $z = \rho$, probar que $u(z) \sim (z - \rho)^{-\alpha} \times h(z)$ con h analítica en $z = \rho$.

Quicksort II

Ahora vamos a probar que $C_n \sim E_n$ en probabilidad:

probamos que $D_n = \mathbb{E}[C_n^2]$ satisface $D_n \sim E_n^2$.

Quicksort II

Ahora vamos a probar que $C_n \sim E_n$ en probabilidad:

probamos que $D_n = \mathbb{E}[C_n^2]$ satisface $D_n \sim E_n^2$.

Por el mismo argumento, con C_j y \tilde{C}_{n-j} independientes,

$$D_n = \frac{1}{n} \times \sum_{j=0}^{n-1} \mathbb{E}[(C_j + \tilde{C}_{n-1-j} + n - 1)^2].$$

Quicksort II

Ahora vamos a probar que $C_n \sim E_n$ en probabilidad:

probamos que $D_n = \mathbb{E}[C_n^2]$ satisface $D_n \sim E_n^2$.

Por el mismo argumento, con C_j y \tilde{C}_{n-j} independientes,

$$D_n = \frac{1}{n} \times \sum_{j=0}^{n-1} \mathbb{E}[(C_j + \tilde{C}_{n-1-j} + n - 1)^2].$$

Expandiendo [detalles]

$$D_n = \frac{2}{n} \times \sum_{j=0}^{n-1} D_j + \frac{2}{n} \sum_{j=0}^{n-1} E_j E_{n-1-j} - (n-1)^2 + 2(n-1)E_n.$$

Quicksort II

Ahora vamos a probar que $C_n \sim E_n$ en probabilidad:

probamos que $D_n = \mathbb{E}[C_n^2]$ satisface $D_n \sim E_n^2$.

Por el mismo argumento, con C_j y \tilde{C}_{n-j} independientes,

$$D_n = \frac{1}{n} \times \sum_{j=0}^{n-1} \mathbb{E}[(C_j + \tilde{C}_{n-1-j} + n - 1)^2].$$

Expandiendo [detalles]

$$D_n = \frac{2}{n} \times \sum_{j=0}^{n-1} D_j + \frac{2}{n} \sum_{j=0}^{n-1} E_j E_{n-1-j} - (n-1)^2 + 2(n-1)E_n.$$

Paso simbólico

La función generatriz $D(z) = \sum_n D_n z^n$ satisface

$$zD'(z) = \frac{2z}{1-z} D(z) + 2(E(z))^2 - \frac{6z^3}{(1-z)^4} - \frac{2z^2}{(1-z)^3} + 2z^2 E''(z).$$

Quicksort III

Basta seguir los términos¹³ dominantes:

Técnica: integración singular

Si $A(z) = o(B(z))$ y $|\int_0^z B(z)dz| \rightarrow \infty$ $z \rightarrow 1$, $\int A(z)dz = o(\int B(z)dz)$.

¹³En este caso los términos son de forma $\frac{c}{(1-z)^m} (\log \frac{1}{1-z})^k$ con $m, k \geq 0$ enteros.

Quicksort III

Basta seguir los términos¹³ dominantes:

Técnica: integración singular

Si $A(z) = o(B(z))$ y $|\int_0^z B(z)dz| \rightarrow \infty$ $z \rightarrow 1$, $\int A(z)dz = o(\int B(z)dz)$.

– Si $K(z) = (1-z)^2 D(z)$ entonces

$$K'(z) = (1-z)^2 \left(D'(z) - \frac{2z}{1-z} D(z) \right) \sim (1-z)^2 \frac{8}{(1-z)^4} \left(\log \left(\frac{1}{1-z} \right) \right)^2$$

¹³En este caso los términos son de forma $\frac{c}{(1-z)^m} \left(\log \frac{1}{1-z} \right)^k$ con $m, k \geq 0$ enteros.

Quicksort III

Basta seguir los términos¹³ dominantes:

Técnica: integración singular

Si $A(z) = o(B(z))$ y $|\int_0^z B(z)dz| \rightarrow \infty$ $z \rightarrow 1$, $\int A(z)dz = o(\int B(z)dz)$.

– Si $K(z) = (1-z)^2 D(z)$ entonces

$$K'(z) = (1-z)^2 \left(D'(z) - \frac{2z}{1-z} D(z) \right) \sim (1-z)^2 \frac{8}{(1-z)^4} \left(\log \left(\frac{1}{1-z} \right) \right)^2$$

– Por integración singular

$$(1-z)^2 D(z) = K(z) \sim \int_0^z \frac{8}{(1-z)^2} \left(\log \left(\frac{1}{1-z} \right) \right)^2 dz \sim \frac{8}{1-z} \left(\log \left(\frac{1}{1-z} \right) \right)^2.$$

¹³En este caso los términos son de forma $\frac{c}{(1-z)^m} \left(\log \frac{1}{1-z} \right)^k$ con $m, k \geq 0$ enteros.

Quicksort III

Basta seguir los términos¹³ dominantes:

Técnica: integración singular

Si $A(z) = o(B(z))$ y $|\int_0^z B(z)dz| \rightarrow \infty$ $z \rightarrow 1$, $\int A(z)dz = o(\int B(z)dz)$.

– Si $K(z) = (1-z)^2 D(z)$ entonces

$$K'(z) = (1-z)^2 \left(D'(z) - \frac{2z}{1-z} D(z) \right) \sim (1-z)^2 \frac{8}{(1-z)^4} \left(\log \left(\frac{1}{1-z} \right) \right)^2$$

– Por integración singular

$$(1-z)^2 D(z) = K(z) \sim \int_0^z \frac{8}{(1-z)^2} \left(\log \left(\frac{1}{1-z} \right) \right)^2 dz \sim \frac{8}{1-z} \left(\log \left(\frac{1}{1-z} \right) \right)^2.$$

– Se deduce $D(z) \sim \frac{8}{(1-z)^3} \left(\log \left(\frac{1}{1-z} \right) \right)^2$ y $D_n \sim 4n^2 (\log n)^2$.

¹³En este caso los términos son de forma $\frac{c}{(1-z)^m} \left(\log \frac{1}{1-z} \right)^k$ con $m, k \geq 0$ enteros.

Árboles Binarios de Búsqueda

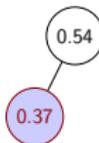
Un árbol binario de búsqueda (ABB) de tamaño n se construye insertando n números aleatorios de $[0, 1]$ a un árbol inicialmente vacío:

- nodos no cambian de posición, elementos se insertan en hojas nuevas,
- elementos en rama izquierda son menores que la raíz, y en derecha mayores.

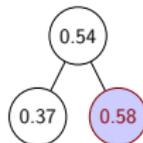
(1) Insertar 0.54



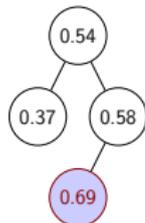
(2) Insertar 0.37



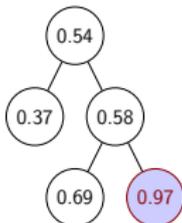
(3) Insertar 0.58



(4) Insertar 0.69

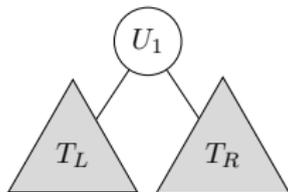


(5) Insertar 0.97



Árboles Binarios de Búsqueda II

Dados U_1, \dots, U_n iid de $[0, 1]$, construimos un ABB de n nodos,

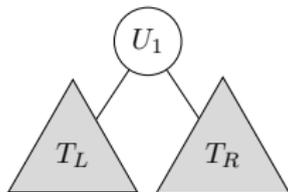


El rango de la raíz U_1 es uniforme entre 1 y n :

$$\Pr_n(|T_L| = k) = \frac{1}{n}, \forall k \in \{0, \dots, n-1\}.$$

Árboles Binarios de Búsqueda II

Dados U_1, \dots, U_n iid de $[0, 1]$, construimos un ABB de n nodos,



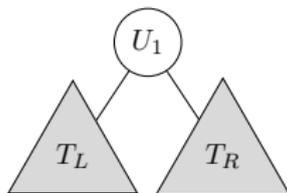
El rango de la raíz U_1 es uniforme entre 1 y n :

$$\Pr_n(|T_L| = k) = \frac{1}{n}, \forall k \in \{0, \dots, n-1\}.$$

\Rightarrow se puede usar esto para producir ABB aleatorios (distribución ABB).

Árboles Binarios de Búsqueda II

Dados U_1, \dots, U_n iid de $[0, 1]$, construimos un ABB de n nodos,



El rango de la raíz U_1 es uniforme entre 1 y n :

$$\Pr(|T_L| = k) = \frac{1}{n}, \forall k \in \{0, \dots, n-1\}.$$

\Rightarrow se puede usar esto para producir ABB aleatorios (distribución ABB).

Proposición

La distribución ABB de tamaño n no es uniforme.

Árboles Binarios de Búsqueda III

Sea $\omega(t) = \omega_\gamma(t)$ la cantidad de veces que un árbol “pattern” γ se encuentra en el árbol t , que pueden no ser disjuntas:

Proposición (Flajolet, Gourdon, Martínez '97)

Sea $p(t) = \Pr_n(t)$ donde $n = |t|$. La función generatriz bivariada

$$F(z, u) := \sum_t p(t) z^{|t|} u^{\omega(t)}.$$

satisface

$$\partial_z F(z, u) = F^2(z, u) + (u - 1)p(\gamma)|\gamma|z^{|\gamma|-1}.$$

Árboles Binarios de Búsqueda III

Sea $\omega(t) = \omega_\gamma(t)$ la cantidad de veces que un árbol “pattern” γ se encuentra en el árbol t , que pueden no ser disjuntas:

Proposición (Flajolet, Gourdon, Martínez '97)

Sea $p(t) = \text{Pr}_n(t)$ donde $n = |t|$. La función generatriz bivariada

$$F(z, u) := \sum_t p(t) z^{|t|} u^{\omega(t)}.$$

satisface

$$\partial_z F(z, u) = F^2(z, u) + (u - 1)p(\gamma)|\gamma|z^{|\gamma|-1}.$$

Sketch de prueba.

Probar que $F_n(u) = [z^n]F(z, u)$ satisface una recurrencia. □

Árboles Binarios de Búsqueda III

Sea $\omega(t) = \omega_\gamma(t)$ la cantidad de veces que un árbol “pattern” γ se encuentra en el árbol t , que pueden no ser disjuntas:

Proposición (Flajolet, Gourdon, Martínez '97)

Sea $p(t) = \Pr_n(t)$ donde $n = |t|$. La función generatriz bivariada

$$F(z, u) := \sum_t p(t) z^{|t|} u^{\omega(t)}.$$

satisface

$$\partial_z F(z, u) = F^2(z, u) + (u - 1)p(\gamma)|\gamma|z^{|\gamma|-1}.$$

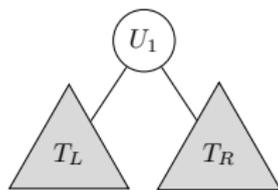
Sketch de prueba.

Probar que $F_n(u) = [z^n]F(z, u)$ satisface una recurrencia. □

Se puede probar que el pattern γ aparece en media $\sim \frac{2p(\gamma)}{(|\gamma|+1)(|\gamma|+2)}n$ veces.

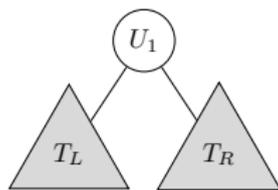
Quicksort IV

Recordamos: Dados U_1, \dots, U_n iid de $[0, 1]$, construimos un ABB de n nodos,



Quicksort IV

Recordamos: Dados U_1, \dots, U_n iid de $[0, 1]$, construimos un ABB de n nodos,



Quicksort corresponde a un ABB. Definimos $p_n(k) = \Pr_n(\text{costo} = k)$,

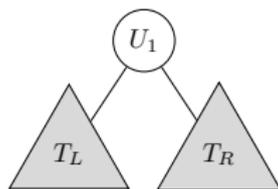
Proposición

La función generatriz bivariada $\sum p_n(k) z^n u^k$ satisface

$$F(z, u) = 1 + \int_0^z (F(zu, u))^2 dz.$$

Quicksort IV

Recordamos: Dados U_1, \dots, U_n iid de $[0, 1]$, construimos un ABB de n nodos,



Quicksort corresponde a un ABB. Definimos $p_n(k) = \Pr_n(\text{costo} = k)$,

Proposición

La función generatriz bivariada $\sum p_n(k) z^n u^k$ satisface

$$F(z, u) = 1 + \int_0^z (F(zu, u))^2 dz.$$

Ejercicio

Deducir de la fórmula de $F(z, u)$ los asintóticos de la media y segundo momento.

Para aprender más

 *Philippe Flajolet, Xavier Gourdon y Conrado Martínez*
Patterns in Random Binary Search Trees.

<https://www.cs.upc.edu/~conrado/research/papers/rsa-fgm97.pdf>

 *Philippe Flajolet y Robert Sedgewick*
Analytic Combinatorics.

<https://algo.inria.fr/flajolet/Publications/books.html>

 *Herbert S. Wilf*
Generatingfunctionology.

<https://www2.math.upenn.edu/~wilf/DownldGF.html>