

Direccionamiento Directo y Funciones de Enumeración

Año 2024

Introducción

Cuando vimos el tema *Distribución pseudoaleatoria de datos*, planteamos la necesidad de encontrar una función h , para que la técnica fuera aplicable. También vimos que, por lo general, esta función h no es inyectiva. Esto significa que, dada:

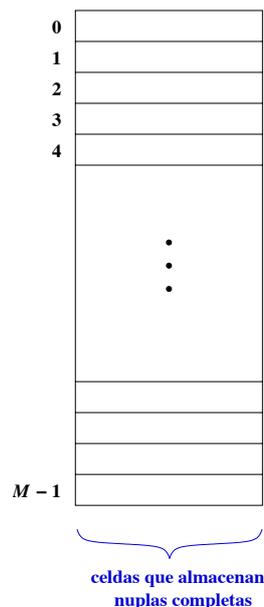
$$R \subseteq X \times Y$$

$$h : X \mapsto \mathbb{I}_{M-1} \cup \{0\}$$

existen al menos dos elementos distintos x_1 y x_2 en X tal que $h(x_1) = h(x_2)$. En este caso decíamos que x_1 y x_2 eran sinónimos bajo h . Por lo tanto, para aplicar la técnica era necesario contar no sólo con una función de pseudo-azar sino también con alguna técnica para manejar el rebalse que se produce por las colisiones. Así, la solución completa se planteaba como:

Pseudo-aleatorización + tratamiento de rebalse.

Por otra parte, supongamos por un momento que somos capaces de encontrar un función h inyectiva; es decir, una función tal que no existan dos elementos en X que sean sinónimos bajo $h : X \mapsto \mathbb{I}_{M-1} \cup \{0\}$. Entonces, si tal función existe, se puede simplificar la estructura de almacenamiento: *cada nupla (x, y) de la relación se almacena en la posición dada por $h(x)$* . Es decir, al no haber colisiones, se pueden reservar M celdas para almacenar $|X| = N$ nuplas y no existe la posibilidad de rebalse. En este caso a la función h se la denominaba *perfecta*. Cualquier operación sobre la estructura se podría realizar con costo uno; es decir, sólo basta con consultar una celda en las evocaciones y sólo modificar una celda en altas, bajas y modificaciones.



Sin embargo, si h además de ser inyectiva es sobreyectiva¹, $h : X \mapsto \mathbb{I}_{N-1} \cup \{0\}$, no sólo no hay posibilidad de rebalse sino que también bastaría con reservar N celdas para almacenar N nuplas. Aún

¹Si una función $f : A \mapsto B$ es total, inyectiva y sobreyectiva, entonces $|A| = |B|$.

Bajo estas condiciones y nombrando como DD al espacio reservado para la estructura, se puede notar que:

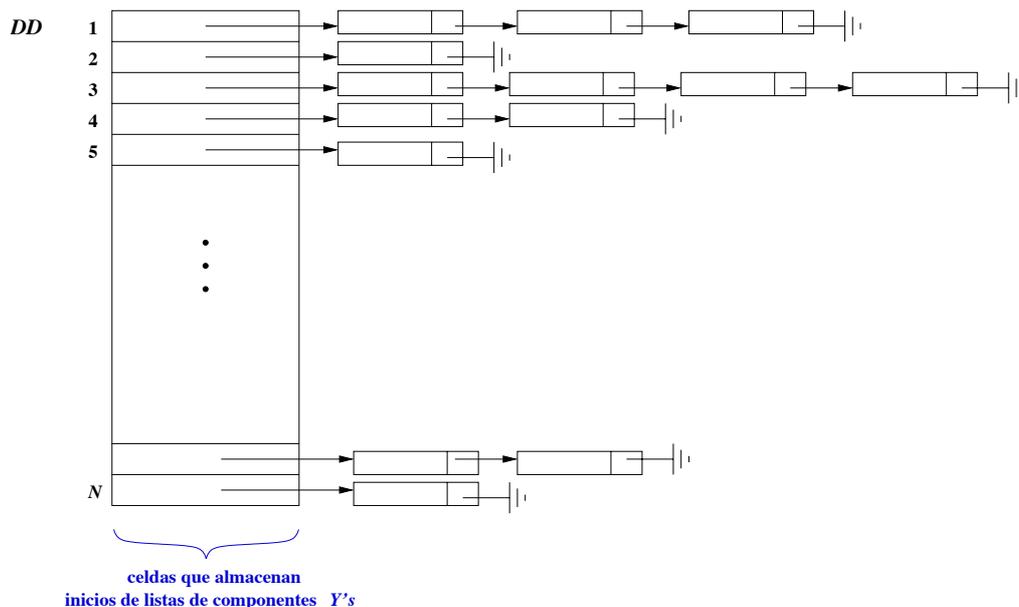
- No hay altas ni bajas.
- No es necesaria la rutina de Localización.
- Se necesitan $|X| = N$ celdas capaces de alojar los Y 's.
- La rutina de evocación asociativa no necesita el parámetro éxito y sería tan simple como:

$$\begin{aligned} & \text{Evocación Asociativa (in } x, \text{ out } y) \\ & y \leftarrow DD[e(x)] \end{aligned}$$

Sin embargo, es posible que no en toda situación se puedan cumplir completamente las condiciones previas planteadas. Afortunadamente, es posible admitir que no se cumplan algunas de las condiciones y aún así seguir teniendo algunas de las ventajas de esta estructura. Se analizan a continuación cómo se pueden relajar algunas de las condiciones y cómo afecta su no cumplimiento a la estructura.

Si X no determina funcionalmente a Y

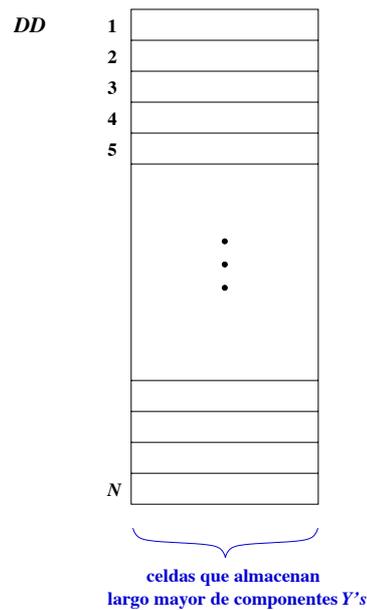
En este caso, es posible que existan nuplas de la forma (x, y) , (x, y') , (x, y'') , \dots ; es decir, nuplas con igual valor en la parte X . A su vez, no necesariamente todos los valores de X tendrán asociados el mismo número de Y 's distintos. Como se debe disponer de N celdas de tamaño fijo en DD , se puede considerar que la celda i mantiene el conjunto de todas las partes Y 's asociadas al x tal que $e(x) = i$. Una manera sencilla de lograrlo es almacenar en la posición $i = e(x)$ el inicio de una lista vinculada que mantenga los Y 's asociados a ese x .



Si las componentes Y 's no son codificables a largo contante

En este caso, es difícil decidir el tamaño de cada celda del almacenamiento contiguo de N celdas. Sin embargo, en algunas situaciones particulares se pueden considerar tamaños de celda adecuados:

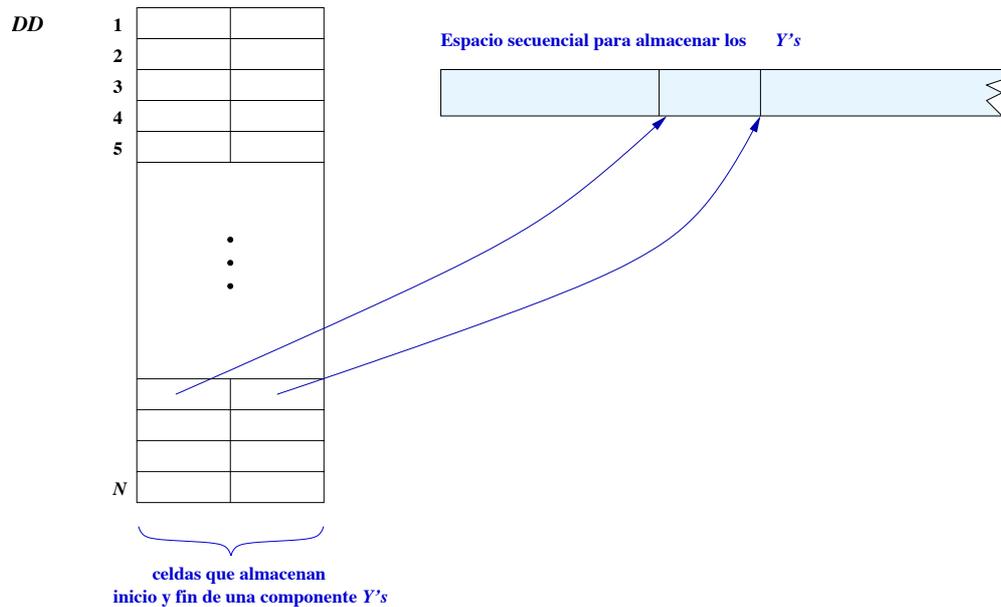
- Se conoce el largo máximo de los Y 's y que la varianza de los largos no es muy grande: así, basta con considerar que todas las celdas tengan como tamaño el largo máximo y no habrá desperdicio significativo de espacio.



- No se conoce el largo máximo de los Y 's o la varianza de los largos es demasiado grande: así, no es posible determinar un tamaño de celda fijo para el almacenamiento de los Y 's o si se fija, debido a que la varianza de los largos es demasiado grande, habrá mucho desperdicio de espacio.

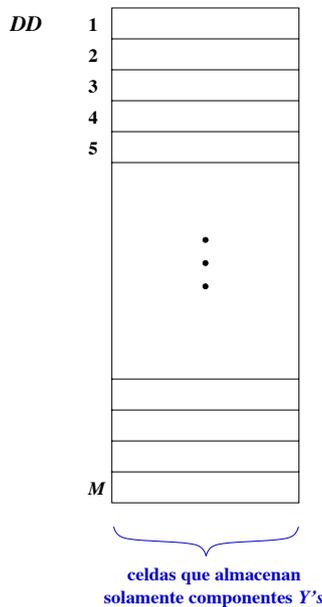
En este caso una posible solución sería contar con un espacio secuencial adicional capaz de alojar a todos los Y 's y que en DD se mantengan apuntadores al inicio y al fin del valor de Y correspondiente en el espacio secuencial adicional. Es decir, siendo la nupla (x, y) y $e(x) = i$, $DD[i]$ contendrá dónde se inicia y donde termina la representación de y en el espacio auxiliar. Además, si existe memorización previa y R es completa respecto de X y los Y 's se almacenan en el mismo orden de enumeración de X , se puede simplificar lo que se almacena en cada posición de DD , sólo guardando el inicio o el fin de cada componente Y ².

²Pensar cómo, sólo guardando el inicio de cada componente Y , se podría conocer dónde empieza y donde termina dicha componente Y en el almacenamiento secuencial adicional. Ídem para el caso en que sólo se guarde el fin de cada componente Y .



No existe una función de enumeración

Si no existe una función $e : X \rightarrow \mathbb{I}_{|X|}$, pero si existe una función de enumeración $e : X' \rightarrow \mathbb{I}_{|X'|}$ para $X' \supset X$, siendo $|X'| = M = \beta \cdot N = |X|$, donde $\beta > 1$.



En este caso, como $M > N$ algunas celdas no contendrán valores de Y válidos y por lo tanto podrán contener un valor inválido α , porque su valor de $x \in X$. Además, aún existiendo memorización previa y siendo R completa en X , se debe considerar que la rutina de evocación asociativa necesitará tener el parámetro *éxito* y así sería:

Evocación Asociativa (in x , out y , out éxito)

$\text{éxito} \leftarrow DD[e(x)] \neq \alpha$

If éxito Then

$y \leftarrow DD[e(x)]$

Ejemplo:

Sea $X = \{(d, m) / 1 \leq d \leq 31 \wedge 1 \leq m \leq 12 \wedge \text{"}d \text{ es un día válido en el mes } m\}$. Este conjunto no es fácilmente enumerable porque no es regular, porque por ejemplo: *no todos los meses tienen día 31 y febrero no tiene día 29, salvo que sea año bisiesto*. Es decir, hay meses que tienen 28 o 29 días, meses de 30 días y meses de 31 días.

Sin embargo, el conjunto $X' = \{(d, m) / 1 \leq d \leq 31 \wedge 1 \leq m \leq 12\} \supset X$, tal que $|X'| = 372 > |X| = 365$, es muy regular y por lo tanto, fácil de enumerar. Las siguientes posibilidades para la función $e : X' \rightarrow \mathbb{I}_M$ permiten enumerar a X' :

1. Por mes: $e(d, m) = 31(m - 1) + d$
2. Por día: $e(d, m) = 12(d - 1) + m$

En este caso, sólo 7 celdas no podrán contener un valor y válido, porque no corresponden a fechas de X . Así, el espacio necesario será de 372 celdas capaces de almacenar un valor de Y en lugar de las 365 celdas que se hubiesen necesitado si X hubiera sido enumerable.

No hay memorización previa

Si no hay memorización previa, es posible que se realice una evocación asociativa para un x que aún no se ha memorizado y por lo tanto se debe utilizar un valor inválido α que indique que aún no hay un valor y válido en la celda y por lo tanto la evocación asociativa también sería:

Evocación Asociativa (in x , out y , out éxito)

$\text{éxito} \leftarrow DD[e(x)] \neq \alpha$

If éxito Then

$y \leftarrow DD[e(x)]$

Además, deberían existir rutinas que permitan realizar *Altas, Bajas y Modificaciones*. Para todas ellas basta con acceder a la celda de $DD[e(x)]$ y verificar si es posible realizar la operación y , en caso de ser posible realizarla, modificar la celda como corresponda.

R no es completa con respecto a X

Si R no es completa respecto de X significa que $\Pi_X(R) \subset X$, lo cual implica que algunas celdas no contendrán valores de Y válidos porque su valor de X no aparece en ninguna nupla de R . Así, basta nuevamente con contar con un valor inválido α que indique que no hay una nupla con ese valor de X en R .

En este caso, también se admite contar con rutinas para *Altas, Bajas y Modificaciones*.

Tal como hemos visto, no siempre es posible obtener este tipo de funciones, porque de por sí no es sencillo encontrar una función que sea inyectiva. Para poder encontrar una función de enumeración es necesario que el conjunto X a enumerar tenga cierta regularidad.

Vamos a estudiar un método que nos permita encontrar una función de enumeración para aquellos casos en que tal función exista.

Definiciones

Informalmente, una función de enumeración e para un conjunto X , es una función que permite establecer un orden entre los elementos de dicho conjunto. Para ello, e asigna a cada elemento x de X un número i , con $1 \leq i \leq |X|$. Este número i es la posición que le corresponde a x de acuerdo al orden establecido por e .

Formalmente, una función de enumeración e para un conjunto X , con $|X| = n$, es una correspondencia biunívoca³ entre los elementos de X y los primeros n números naturales:

$$e : X \xrightarrow{\text{biunívoca}} \mathbb{I}_n$$

Esto implica que para cada $x \in X$ existe uno y sólo un $i \in \mathbb{I}_n$ que es imagen de x bajo e ; y para cada $i \in \mathbb{I}_n$ existe uno y sólo un $x \in X$ que es preimagen de i bajo e .

EJEMPLO 1

Sea $X = \{3, 4, 5, 6\}$, una posible función de enumeración es $e(x) = x - 2$.

En este caso $e(3) = 1$, $e(4) = 2$, $e(5) = 3$ y $e(6) = 4$, por lo que el orden de los elementos bajo e es: 3, 4, 5, 6.

Otra posible función de enumeración es $e(x) = 7 - x$. Esta función ordena los elementos de X en forma inversa a la anterior, dado que $e(6) = 1$, $e(5) = 2$, $e(4) = 3$ y $e(3) = 4$.

□

Si el conjunto es enumerable, generalmente existirá más de una función de enumeración, cada una estableciendo un orden, posiblemente distinto, sobre los elementos de X .

El conjunto dado en el ejemplo anterior es muy sencillo, y por lo tanto es muy simple obtener una función de enumeración para él. Veamos un ejemplo un poco más complejo.

EJEMPLO 2:

Sea $X = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\}$. Podríamos pensar que X son los subíndices de una matriz de 2×3 . El conjunto X puede también definirse por comprensión:

$$X = \{(r, s) / 1 \leq r \leq 2 \wedge 1 \leq s \leq 3\}$$

Para este conjunto una función de enumeración es $e(r, s) = 3r + s - 3$. Verifiquémoslo:

$$e : X \xrightarrow{\text{biunívoca}} \mathbb{I}_6$$

³Recordar que para ser una correspondencia biunívoca, debe ser función, total, inyectiva y suryectiva.

$$\begin{aligned} e(1, 1) &= 1 & e(2, 1) &= 4 \\ e(1, 2) &= 2 & e(2, 2) &= 5 \\ e(1, 3) &= 3 & e(2, 3) &= 6 \end{aligned}$$

□

En este ejemplo ya no es tan directo de entender cómo se obtiene la función de enumeración. Vamos a ver un método general que permite construir funciones de enumeración.

Construcción de funciones de enumeración

Supongamos que tenemos un conjunto X cuya estructura dificulta su enumeración. Lo que haremos será dividir X en partes más pequeñas $X^{(1)}, X^{(2)}, \dots, X^{(k)}$,⁴ que tengan una estructura más simple, de forma tal que su enumeración sea directa. La enumeración de X , se obtiene a partir de la enumeración de las partes $X^{(i)}$.

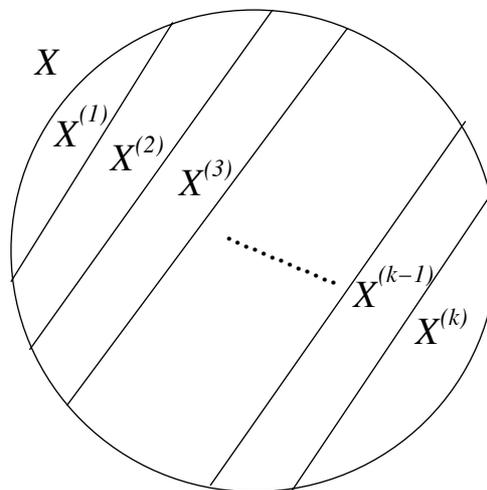
El primer punto que debemos resolver es cómo particionar X . Una forma de particionar es utilizando una función. Sea:

$$p : X \rightarrow \mathbb{I}_k$$

$$\text{entonces } X^{(i)} = \{x/x \in X \wedge p(x) = i\}$$

A la función p la llamaremos *función de partes*.

Gráficamente:



Notar que: lo que hemos hecho es particionar el conjunto X a través de una relación de equivalencia E inducida por p . Esta relación está definida de la siguiente manera:

$$xEy \Leftrightarrow p(x) = p(y)$$

Cada parte $X^{(i)}$ es una clase de equivalencia de E .

⁴ $X^{(1)}, X^{(2)}, \dots, X^{(k)}$ forman una partición del conjunto X .

EJEMPLO 3:

Para el ejemplo 2, una posible función de partes es $p(r, s) = r$. Esta función particiona X en:

$$X^{(1)} = \{(1, 1), (1, 2), (1, 3)\}$$

$$X^{(2)} = \{(2, 1), (2, 2), (2, 3)\}$$

Otra función de partes posible es $p(r, s) = s$. En este caso se particiona a X en tres subconjuntos:

$$X^{(1)} = \{(1, 1), (2, 1)\}$$

$$X^{(2)} = \{(1, 2), (2, 2)\}$$

$$X^{(3)} = \{(1, 3), (2, 3)\}$$

□

Hasta el momento hemos obtenido una partición de X en conjuntos cuya enumeración resulta directa.

Sea $e^{(i)}$ la función de enumeración para $X^{(i)}$. Entonces podemos enumerar u ordenar X de la siguiente forma: colocamos primero todos los elementos de $X^{(1)}$ ordenados de acuerdo a $e^{(1)}$, luego todos los elementos de $X^{(2)}$ ordenados de acuerdo a $e^{(2)}$, y así siguiendo hasta llegar a $X^{(k)}$. Gráficamente:

$$\underbrace{x_{11}, x_{12}, \dots, x_{1|X^{(1)}|}}_{X^{(1)} \text{ ordenado por } e^{(1)}}, \underbrace{x_{21}, x_{22}, \dots, x_{2|X^{(2)}|}}_{X^{(2)} \text{ ordenado por } e^{(2)}}, \dots, \underbrace{x_{k1}, x_{k2}, \dots, x_{k|X^{(k)}|}}_{X^{(k)} \text{ ordenado por } e^{(k)}}^5$$

Entonces, dado un elemento, para saber el número que le corresponde de acuerdo a este nuevo orden debemos hacer lo siguiente: *enumerarlo en su parte y sumarle la cardinalidad de los conjuntos que le han precedido*. Es decir, la función de enumeración para X se define de la siguiente manera:

$$e(x) = \sum_{i < p(x)} |X^{(i)}| + e^{(p(x))}(x)$$

Veamos ahora el desarrollo completo de esta técnica para algunos ejemplos.

EJEMPLO 4:

Sea $X = \{(r, s) / 1 \leq r \leq 2 \wedge 1 \leq s \leq 3\}$.

Primer paso: particionar X . Para ello vamos a usar como función de partes:

$$p : X \mapsto \mathbb{I}_2 : p(r, s) = r$$

Definimos formalmente la partición:

$$X^{(i)} = \{(r, s) / (r, s) \in X \wedge p(r, s) = i\}$$

$$\Rightarrow X^{(i)} = \{(r, s) / \underbrace{1 \leq r \leq 2 \wedge 1 \leq s \leq 3}_{(r,s) \in X} \wedge \underbrace{r = i}_{p(r,s)=r=i}\}$$

⁵ x_{ij} representa el elemento de la parte i al que le corresponde la posición j de acuerdo a $e^{(i)}$

Entonces podemos reemplazar cada aparición de r por i :

$$\Rightarrow X^{(i)} = \{(i, s) / 1 \leq s \leq 3\}, 1 \leq i \leq 2$$

Esto significa que, dentro de la parte $X^{(i)}$, la primer componente de todos los pares es la constante i . La variación de i que hemos puesto fuera de la definición de $X^{(i)}$, nos dice que la cantidad total de partes es 2.

Entonces, podemos eliminar por el momento la primer componente de cada par, dado que sabemos cuanto vale. Esto nos lleva a definir un conjunto que es isomorfo a $X^{(i)}$ y que tiene una trama más sencilla:

$$\Rightarrow X^{(i)'} = \{s / 1 \leq s \leq 3\}, 1 \leq i \leq 2$$

Este conjunto es enumerable por medio de la función identidad:

$$e^{(i)'}(s) = id(s) = s$$

También es directo el cálculo de su cardinalidad:

$$|X^{(i)'}| = 3$$

Esta función enumera $X^{(i)'}$, pero a nosotros nos interesa enumerar $X^{(i)}$. Si tenemos en cuenta cómo llegamos $X^{(i)'}$, la deducción de $e^{(i)}$ es directa: $e^{(i)}$ debe ser tal que deseche la primer componente:

$$e^{(i)}(r, s) = s$$

Además como $X^{(i)'}$ es isomorfo a $X^{(i)}$:

$$|X^{(i)'}| = |X^{(i)}| = 3$$

Gráficamente hemos seguido los siguiente pasos:

$$\begin{array}{c} X^{(i)} \leftrightarrow X^{(i)'} \\ \downarrow \\ e^{(i)} \leftrightarrow e^{(i)'} \end{array}$$

Ahora aplicamos la fórmula para enumerar X :

$$\begin{aligned} e(r, s) &= \sum_{i < p(r, s)} |X^{(i)}| + e^{(p(r, s))}(r, s) \\ &= \sum_{i=1}^{p(r, s)-1} |X^{(i)}| + e^{(r)}(r, s) \\ &= \sum_{i=1}^{r-1} 3 + s \\ &= (r-1) \cdot 3 + s \\ &\Rightarrow \boxed{e(r, s) = 3 \cdot r + s - 3} \end{aligned}$$

□

Analicemos ahora otro ejemplo un poco más complejo debido a que sus partes no tienen todas igual cardinalidad.

EJEMPLO 5:

Sea $X = \{(r, s) / 1 \leq r \leq s \leq N\}$, es decir X es el conjunto de subíndices de una matriz triangular superior.

Particionamos usando:

$$p : X \mapsto \mathbb{I}_N : p(r, s) = r$$

Luego definimos formalmente la partición y para llegar de (1) a (2) hacemos las sustituciones correctas en cada paso. En (2) llegamos a que todos los elementos de la parte $X^{(i)}$ tienen su primera componente constante igual a i y que la variación de i , que ahora va fuera, indica que existen N partes posibles.

Ahora, como ya hicimos en los ejemplos anteriores, pasamos a trabajar sobre un conjunto isomorfo a $X^{(i)}$ que es más simple en el cual eliminamos la primera componente del par por ya saber cuánto vale. Entonces, para enumerar dicho conjunto observamos que sus elementos van desde i hasta N , y nosotros necesitamos darles valores desde 1, por lo cual obtenemos una función de enumeración $e^{(i)'}$ enunciada en (3) que al valor i le hace corresponder el 1, al valor $i + 1$ le hace corresponder el 2 y así siguiendo. En este punto podemos ya deducir la cardinalidad del conjunto, la cual depende claramente de la parte que se esté considerando.

Como ya sabemos enumerar el conjunto $X^{(i)'}$ deducimos cómo enumerar a $X^{(i)}$, y a partir de $e^{(i)}$ podemos deducir también la cardinalidad de cada parte $X^{(i)}$ como se muestra en (4).

$$X^{(i)} = \{(r, s) / (r, s) \in X \wedge p(r, s) = i\} \quad (1)$$

$$= \{(r, s) / 1 \leq r \leq s \leq N \wedge r = i\}$$

$$= \{(i, s) / i \leq s \leq N\}, 1 \leq i \leq N \quad (2)$$

\Rightarrow

$$\begin{aligned} X^{(i)' } &= \{s / i \leq s \leq N\}, 1 \leq i \leq N \\ e^{(i)' } (s) &= s - i + 1 \\ |X^{(i)' }| &= N - i + 1 \end{aligned} \quad (3)$$

\Rightarrow

$$\begin{aligned} e^{(i)} (r, s) &= s - i + 1 \\ |X^{(i)}| &= N - i + 1 \end{aligned} \quad (4)$$



Aplicamos la fórmula para obtener la enumeración de X :

$$\begin{aligned}
 e(r, s) &= \sum_{i < p(r, s)} |X^{(i)}| + e^{(p(r, s))}(r, s) \\
 &= \sum_{i=1}^{p(r, s)-1} |X^{(i)}| + e^{(r)}(r, s) \\
 &= \sum_{i=1}^{r-1} (N - i + 1) + (s - r + 1) \\
 &= \sum_{i=1}^{r-1} (N + 1) - \sum_{i=1}^{r-1} i + (s - r + 1) \\
 &= (r - 1)(N + 1) - \frac{r(r - 1)}{2} + s - r + 1
 \end{aligned} \tag{5}$$

En (5) llegamos finalmente a la fórmula que nos permite enumerar a X , la cual también podríamos usar para calcular la cardinalidad de X (lo cual se deja como ejercicio).

□

EJEMPLO 6:

Sea $X = \{(d, m) / 1 \leq d \leq 31 \wedge 1 \leq m \leq 12 \wedge \text{"d es un día válido en el mes m"}\}$ el conjunto de fechas de un año y sea $X' = \{(d, m) / 1 \leq d \leq 31 \wedge 1 \leq m \leq 12\} \supset X$. Queda como ejercicio obtener, paso a paso, la función de enumeración $e : X' \rightarrow \mathbb{I}_M$, que se obtendría si se usa como función de partición cada una de las siguientes opciones:

1. Por mes: $p : X \rightarrow \mathbb{I}_M : p(d, m) = m$
2. Por día: $p : X \rightarrow \mathbb{I}_M : p(d, m) = d$

□

Conclusiones

Finalmente, podemos decir que en general para poder enumerar un conjunto se deben seguir siempre las mismas pautas:

1. Dividir el conjunto las veces que sean necesarias hasta que sus partes sean fácilmente enumerables.
2. Enumerar las partes resultantes y obtener la cardinalidad de cada parte.
3. Recorrer el camino inverso construyendo la o las funciones siempre de la misma manera:
 - Sumar la cardinalidad de todas las partes anteriores a aquélla en donde cae el elemento, y
 - Sumarle la enumeración en su parte.

Reconocimientos

El presente apunte se realizó tomando como base notas de las clases del **Ing. Hugo Ryckeboer** en la Universidad Nacional de San Luis.