Estructuras de Datos y Algoritmos II

Año 2025

Deducción de algunos esfuerzos para una Distribución pseudo-aleatoria de datos

Introducción

Vamos a desarrollar algunos de los esfuerzos para estructuras que utilizan una distribución pseudo-aleatoria de datos y mostraremos en detalle cada paso de la deducción.

Para cada uno de los esfuerzos planteados, primero haremos la deducción a posteriori, para luego usarlo en la deducción del esfuerzo a priori; ello porque es más sencillo realizar la deducción de los esfuerzos a priori, una vez que se conocen los esfuerzos a posteriori.

En todos los casos, por simplicidad, consideramos el caso de tener sólo una ranura por balde; o sea, $r = 1$ y consideramos estructuras sin bajas.

A estas estructuras en general tiene más sentido evaluarlas por Esperanza (esfuerzo medio) que por Máximo, ya que éste puede ser muy malo ($O(N)$, en el caso que todos los elementos caigan en una única lista), pero aquí analizaremos ambos.

Esfuerzos en Rebalse Separado

La función de costo utilizada es cantidad de celdas consultadas o baldes consultados.

a) A posteriori

Sean $n_1, n_2, \ldots, n_M$ los largos de la Lista 1, Lista 2 y Lista M respectivamente. En este caso lo único que se considera librado al azar es sobre qué elemento se realiza la consulta.

1. Esfuerzo máximo de búsqueda exitosa

El esfuerzo máximo de localización exitosa se obtendría al consultar el último elemento de la lista más larga; o sea:

$$\max(n_1, n_2, \ldots, n_M) \text{ baldes consultados}$$

2. Esfuerzo máximo de búsqueda no exitosa

El esfuerzo máximo de localización no exitosa se obtendría al consultar en la lista más larga. Como las listas son desordenadas para fracasar en ella se debe llegar hasta el final de la lista; entonces sería:

$$\max(n_1, n_2, \ldots, n_M) \text{ baldes consultados}$$

3. Esfuerzo medio de búsqueda no exitosa

Asumimos, como hipótesis, que existe igual probabilidad de caer en cualquiera de las listas al consultar por un elemento que no se encuentra en la estructura; o sea que trabajamos bajo hipótesis de isoprobabilidad de que la función h devuelva cualquiera de los M valores posibles.

La probabilidad de que el elemento corresponda a una determinada lista, si la función distribuye uniformemente, es $\frac{1}{M}$. Al buscar en cada lista por un elemento no presente, se llega hasta el final de la lista habiendo consultado todos los baldes. Por ejemplo, si es la Lista i entonces se consultan n_i baldes [1]. Por lo tanto, el esfuerzo medio de búsqueda no exitosa a posteriori vendría dado por la siguiente fórmula:

$$\sum_{i=1}^{M} \frac{1}{M} \cdot n_i = \frac{1}{M} \sum_{i=1}^{M} n_i = \frac{N}{M} = \rho$$

siendo ρ el factor de carga de dicha estructura.

En la primera igualdad sólo sacamos como factor común a $\frac{1}{M}$, porque no está afectado por la sumatoria. En la segunda igualdad vemos que sumar, sobre todas las listas, la cantidad de elementos presentes da la cantidad de nuplas instaladas, o sea N. En la tercera sólo reemplazamos $\frac{N}{M}$ por su definición.

Entonces, independientemente de los largos particulares de cada lista, el esfuerzo medio de búsqueda no exitosa sólo depende de la cantidad de listas y de la cantidad de nuplas a alojar.

4. Esfuerzo medio de búsqueda exitosa

Asumiendo que existe igual probabilidad de que cada una de las N nuplas sea consultada; es decir trabajando bajo isoprobabilidad de consulta, la probabilidad de consultar un elemento sería $\frac{1}{N}$.

El vector de costo que se plantearía en este caso sería:

$$\underbrace{(1, \ldots, n_1, 1, \ldots, n_2, \ldots, 1, \ldots, n_M)}_{N \text{ costos}}$$

Por lo tanto, la fórmula del esfuerzo podría plantearse como: $\sum_{i=1}^{N} \frac{1}{N} \cdot c_i$, la cual da la idea de un recorrido completo sobre el vector de costo.

Si analizamos una lista particular, por ejemplo la lista j, se podría plantear que:

$$\sum_{k=1}^{n_j} \frac{1}{N} \cdot k = \frac{1}{N} \frac{n_j(n_j+1)}{2} = \frac{n_j}{N} \frac{(n_j+1)}{2}$$

Si ahora consideramos el resultado de esta sumatoria, podemos observar que el primer término corresponde a la probabilidad de que una búsqueda exitosa caiga en la lista j, mientras que el segundo término correspondería al esfuerzo medio de evocación exitosa en una lista vinculada desordenada de n_j elementos.

Esta fórmula está agrupando aquellos productos $(costo \times probabilidad)$ que corresponden a la lista j, en este caso, en sumatorias más pequeñas y más sencillas de resolver.

Entonces, si ahora queremos rearmar la sumatoria más grande debemos volver a sumar pero ahora sobre todas las posibles listas, obteniendo así el esfuerzo medio de búsqueda exitosa:

[1] En caso de que el final de la lista se indique manteniendo una celda al final, con un valor discernible que lo indique, la cantidad de baldes consultados se aumentará en una unidad; o sea $n_i + 1$.

$$
\begin{aligned}
\sum_{j=1}^{M}\sum_{k=1}^{n_j}\frac{1}{N}\cdot k &= \sum_{j=1}^{M}\frac{1}{N}\,\frac{n_j(n_j+1)}{2} = \\
&= \frac{1}{2N}\sum_{j=1}^{M} n_j(n_j+1) = \frac{1}{2N}\sum_{j=1}^{M}({n_j}^2+n_j) = \\
&= \frac{1}{2N}\left(\sum_{j=1}^{M}{n_j}^2+\sum_{j=1}^{M}n_j\right) = \frac{\sum_{j=1}^{M}{n_j}^2}{2N}+\frac{N}{2N} = \\
&= \frac{\sum_{j=1}^{M}{n_j}^2}{2N}+\frac{1}{2} = \frac{1}{2N}\sum_{j=1}^{M}{n_j}^2+\frac{1}{2}
\end{aligned}
$$

En la primera igualdad hemos reemplazado la sumatoria más interna por su resultado. En la segunda igualdad hemos sacado factor común (dado que no varía dentro de la sumatoria). En la tercera distribuimos el producto que se realiza dentro de la sumatoria. Luego, en la cuarta igualdad, distribuimos la sumatoria respecto de la suma. Entonces, en la quinta igualdad, utilizamos que $\sum_{j=1}^{M} n_j = N$. Por último, en la sexta igualdad simplificamos los N's.

Mientras más parejas sean las longitudes de las listas menor será el esfuerzo medio, porque las longitudes están involucradas en una progresión cuadrática.

b) A priori

Para plantear los esfuerzos a priori debemos tener en cuenta todos las formas posibles en que se puede armar un rebalse separado con M listas y N nuplas, y consideramos la estructura sin bajas.

En el caso de los esfuerzos máximos se deberá pensar en el peor caso, sobre todos los posibles, y en caso de esfuerzos medios podría verse como una esperanza de los esfuerzos medios a posteriori para cada una de las estructuras posibles; o sea:

$$
\sum_{\substack{(n_1,n_2,\ldots n_M)\\ \text{tales que } \sum_{j=1}^{M} n_j = N}} \text{Prob}(n_1,n_2,\ldots,n_M)\cdot \text{Esf. Medio}(n_1,n_2,\ldots,n_M)
$$

1. Esfuerzo máximo de búsqueda exitosa

El esfuerzo máximo de consulta exitosa lo obtendría al consultar por el último elemento de la lista más larga posible; y la lista más larga posible, al tener N nuplas a alojar, ocurre si todas las nuplas caen en la misma lista. Entonces, dicho esfuerzo máximo será de N baldes consultados [2].

2. Esfuerzo máximo de búsqueda no exitosa

El esfuerzo máximo de consulta no exitosa lo obtendría al consultar por un elemento no presente, pero cuyo valor de función es el de la lista más larga; y la lista más larga posible, al tener N nuplas a alojar, ocurre si todas las nuplas caen en la misma lista. Entonces, dicho esfuerzo máximo será también de N baldes consultados.

[2] Dicho máximo ocurrirá con una probabilidad $\left(\frac{1}{M}\right)^N$, si la función h distribuye uniformemente.

3. Esfuerzo medio de búsqueda no exitosa

Para obtener esfuerzo medio a priori, debemos calcular la esperanza de los esfuerzos medios a posteriori. O sea, para cada distribución posible de los N elementos debemos obtener su esfuerzo medio a posteriori, el cual ya hemos calculado previamente.

La probabilidad de que ocurra una distribución particular $n_1, n_2, ..., n_M$ viene dada por una distribución multinomial y el esfuerzo medio de evocación no exitosa para esa distribución es ρ baldes consultados; por lo tanto tendríamos la siguiente fórmula:

$$\sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{j=1}^{M} n_j = N}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot \text{Esf. Medio}(n_1, n_2, \ldots, n_M) =$$

$$= \sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{j=1}^{M} n_j = N}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot \rho =$$

$$= \rho \cdot \sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{j=1}^{M} n_j = N}} \text{Prob}(n_1, n_2, \ldots, n_M) = \rho$$

En la primera igualdad sólo sustituimos el esfuerzo medio de búsqueda no exitosa (fracaso) por su valor, en la segunda igualdad sacamos factor común ρ porque su valor no varía en la sumatoria y en la siguiente igualdad aplicamos, algo ya sabido de probabilidades, que la suma de las probabilidades sobre todo el espacio de probabilidad es igual a 1.

Vemos que el resultado es nuevamente el factor de carga. De alguna manera esto lo habíamos adelantado cuando obtuvimos el esfuerzo medio de fracaso a posteriori, porque éste no dependía de la distribución de elementos que se tenía realmente, sino sólo de N y de M.

4. Esfuerzo medio de búsqueda exitosa

Para plantear dicho esfuerzo se pueden seguir, algebraicamente, dos caminos alternativos que nos permiten ambos deducirlo. El primero que veremos es el más simple, pero el planteo es indirecto relacionando esfuerzos medios de éxito con esfuerzos medios de fracaso, y el segundo es matemáticamente más complejo pero es el planteo directo.

1º. Para este primer planteo debemos recordar la relación existente entre esfuerzos medios de éxito y fracaso en aquellas estructuras que no reestructuran. Es decir, aquellas estructuras en las cuáles cuando se ubica un elemento éste nunca cambia de posición, como producto de las futuras altas. Este comportamiento está presente por ejemplo en listas desordenadas, árboles binarios ordenados y las distintas técnicas de manejo de rebalse en distribuciones pseudo-aleatorias de datos; y no lo está en listas ordenadas o árboles binarios balanceados, en las cuales para dar de alta se puede tener que reestructurar.

Cuando no hay reestructuración se puede ver que para encontrar un elemento se debe recorrer el mismo camino que se realizó cuando se lo dio de alta, y el costo del alta a su vez es aquél que realizó la búsqueda que fracasó más posiblemente una constante (cuya presencia depende de la estructura particular que se esté utilizando). Por lo tanto, el esfuerzo medio de alta [3] se puede ver como:

$$\text{Esf. de Alta}(N) = \begin{cases} \text{Esf. Medio de Fracaso}(N-1) \text{ o} \\ \\ \text{Esf. Medio de Fracaso}(N-1) + 1 \end{cases}$$

En el caso particular de lista vinculada desordenada, por simplicidad, suponemos que las altas se realizan al final [4], y que el fin de lista se indica teniendo un puntero al siguiente en nil en la última celda de la lista; la relación existente sería que: *el costo de buscar exitosamente un elemento es igual al costo de haber fracasado, antes de incorporarlo* [5], *más uno* (porque se detiene la búsqueda que fracasa en la celda que pasará a ser la anterior a aquella en la que se ubique el elemento). Entonces, si por ejemplo consideramos el i-ésimo elemento incorporado, tendríamos que el esfuerzo medio de buscar exitosamente el i-ésimo elemento a priori se relaciona con el esfuerzo medio de fracasar a priori en una estructura con $i-1$ elementos presentes.

Ahora podemos plantear formalmente esta relación entre esfuerzos medios:

$$\text{Esf. Medio de Éxito}(N) = \sum_{i=1}^{N} \frac{1}{N} (\text{Esf. medio de Alta}(i) + 1) = \frac{1}{N} \sum_{i=0}^{N-1} (\text{Esf. medio de Fracaso}(i) + 1)$$

Ya conocemos el esfuerzo medio de búsqueda que fracasa en un rebalse separado es ρ o $\dfrac{N}{M}$ y entonces reemplazando tendríamos:

$$\begin{aligned} \text{Esf. Medio de Éxito}(N) &= \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{i}{M} + 1 \right) = \frac{1}{NM} \left(NM + \sum_{i=0}^{N-1} i \right) = \\ &= \frac{1}{NM} \left(NM + \frac{N(N-1)}{2} \right) = 1 + \left(\frac{N-1}{2M} \right) = 1 + \frac{N}{2M} - \frac{1}{2M} = \\ &= \frac{\rho}{2} + 1 - \underbrace{\frac{1}{2M}}_{\text{término correctivo}} \end{aligned}$$

Esta deducción es a priori, porque en ningún momento decimos nada sobre una distribución particular de elementos; más aún, todo el razonamiento utilizado vale aunque los n_i obedezcan a reglas caprichosas.

[3] Se calcula el esfuerzo medio de inserción del elemento $N + 1$, teniendo ya almacenados N elementos.

[4] En caso de realizar las altas al inicio también es válido, aunque no es tan simple de observar.

[5] Entonces, estamos en una situación en que la estructura poseía un elemento menos (el que estamos por incorporar) respecto de la estructura luego que se hizo el alta.

2º. Para el segundo planteo del esfuerzo desarrollamos el esfuerzo medio a priori de acuerdo a su definición, o sea:

$$\sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot \text{Esf. Medio}(n_1, n_2, \ldots, n_M)$$

Para ello debemos recordar que la probabilidad de que ocurra una determinada secuencia de largos de listas $(n_1, n_2, \ldots, n_M)$, tales que $\sum_{i=1}^{M} n_i = N$, responde a una distribución multinomial:

$$\begin{aligned} \text{Prob}(n_1, n_2, \ldots, n_M) &= \begin{pmatrix} & N & & \\ & & & \\ n_1 & n_2 & \ldots & n_M \end{pmatrix} p_1^{n_1}\, p_2^{n_2} \ldots p_M^{n_M} = \\ &= \frac{N!}{n_1!\, n_2! \ldots n_M!}\, p_1^{n_1}\, p_2^{n_2} \ldots p_M^{n_M} \end{aligned}$$

Para el caso de $M = 2$ es equivalente a una distribución binomial:

$$\text{Prob}(n_1) = \frac{N!}{n_1!(N - n_1)!}\, p_1^{n_1}\, (1 - p_1)^{N - n_1}$$

Si queremos ver qué es lo que ocurre en un balde en particular:

o cae dentro del balde considerado o cae fuera

como vemos o el elemento cae en el balde considerado o cae fuera de él; en este último caso realmente no nos interesa en qué balde cayó. Por lo tanto, la distribución de probabilidad se transforma en *binomial*.

Partiendo desde la probabilidad multinomial tenemos que llegar a lo mismo:

$$\sum_{\substack{n_1 \text{ constante para} \\ \text{todas las combinaciones de} \\ (n_1, n_2, \ldots n_M)}} \begin{pmatrix} & N & & \\ & & & \\ n_1 & n_2 & \ldots & n_M \end{pmatrix} p_1^{n_1}\, p_2^{n_2} \ldots p_M^{n_M} =$$

$$= \frac{N!}{n_1!(N - n_1)!}\, p_1^{n_1}\, (1 - p_1)^{N - n_1} \sum_{\substack{\text{todas las formas} \\ \text{de repartir } (N - n_1) \\ \text{entre } n_2, \ldots n_M}} \frac{(N - n_1)!}{n_2! n_3! \ldots n_M!}\, p_2^{n_2} \ldots p_M^{n_M}$$

pero podemos ver que:

$$\sum_{\substack{\text{todas las formas}\\ \text{de repartir } (N-n_1)\\ \text{entre } n_2,\dots n_M}} \frac{(N-n_1)!}{n_2!\, n_3! \dots n_M!}\, p_2^{n_2} \dots p_M^{n_M} = (p_2 + \dots + p_M)^{N-n_1} = (1-p_1)^{N-n_1}$$

porque, aplicando sucesivamente que:

$$(a+b)^q = \sum_{r=0}^{q} \begin{pmatrix} q \\ r \end{pmatrix} a^r \cdot b^{q-r} \tag{1}$$

se obtiene la igualdad planteada anteriormente. Para ver cómo se obtendría, podemos mostrar uno de dichos pasos:

$$(p_2 + (p_3 + \dots + p_M))^{N-n_1} = \sum_{r=0}^{N-n_1} \binom{N-n_1}{r} p_2^r \cdot (p_3 + \dots + p_M)^{(N-n_1)-r}$$

Por lo tanto, reemplazando obtenemos que:

$$\begin{aligned} \sum_{\substack{n_1 \text{ constante para}\\ \text{todas las combinaciones de}\\ (n_1, n_2, \dots n_M)}} \begin{pmatrix} & N & & \\ n_1 & n_2 & \dots & n_M \end{pmatrix} p_1^{n_1}\, p_2^{n_2} \dots p_M^{n_M} &= \frac{N!}{n_1!(N-n_1)!}\, p_1^{n_1} (1-p_1)^{N-n_1} = \\ &= \begin{pmatrix} N \\ n_1 \end{pmatrix} p_1^{n_1} (1-p_1)^{N-n_1} \end{aligned}$$

Ahora podemos volver al planteo original para el esfuerzo medio de evocación exitosa a priori:

$$\sum_{\substack{(n_1, n_2, \dots n_M)\\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \dots, n_M) \cdot \text{Esf. medio}(n_1, n_2, \dots, n_M)$$

y como ya hemos calculado el esfuerzo medio de evocación a posteriori reemplazamos para obtener:

$$\sum_{\substack{(n_1, n_2, \dots n_M)\\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \dots, n_M) \cdot \left(\frac{1}{2N} \sum_{j=1}^{M} {n_j}^2 + \frac{1}{2} \right)$$

ahora, aplicando propiedad distributiva del producto respecto de la suma, veamos esta fórmula por términos:

1er término)

$$\sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot \frac{1}{2} = \frac{1}{2} \cdot \sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \ldots, n_M) = \frac{1}{2}$$

porque la suma de probabilidades sobre todo el espacio de probabilidades es 1.

2do término)

$$\frac{1}{2N} \sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot \sum_{j=1}^{M} {n_j}^2 = \frac{1}{2N} \sum_{j=1}^{M} \sum_{\substack{\text{todos los} \\ \text{casos posibles} \\ \text{con } n_j}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot {n_j}^2$$

La última igualdad representa a $M - 1$ sumas encadenadas, pero no es necesario escribirlas explícitamente. Estudiemos ahora el primer término de la sumatoria; es decir, cuando $j = 1$:

$$\sum_{\substack{\text{todos los} \\ \text{casos posibles} \\ \text{con } n_1}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot {n_1}^2 = \sum_{\substack{\text{todos los} \\ \text{casos posibles} \\ \text{con } n_1}} \binom{N}{n_1} p_1^{n_1} (1 - p_1)^{N - n_1} \, {n_1}^2$$

Ahora veamos como podemos resolver la sumatoria, viendo en general como se comporta para un $j = 1$:

$$\begin{aligned}
\sum_{\substack{\text{todos los} \\ \text{casos posibles} \\ \text{con } n_1}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot {n_1}^2 &= \sum_{\substack{\text{todos los} \\ \text{casos posibles} \\ \text{con } n_1}} \frac{N!}{n_1! \, n_2! \ldots n_M!} \, p_1^{n_1} \, p_2^{n_2} \ldots p_M^{n_M} \, {n_1}^2 = \\
&= \sum_{n_1=0}^{N} \sum_{\substack{(n_2, \ldots n_M) \\ \text{tales que} \\ \sum_{i=2}^{M} n_i = N - n_1}} \frac{N!}{n_1!(N - n_1)!} \, p_1^{n_1} \, {n_1}^2 \frac{(N - n_1)!}{n_2! \ldots n_M!} \, p_2^{n_2} \ldots p_M^{n_M} = \\
&= \sum_{n_1=0}^{N} \frac{N!}{n_1!(N - n_1)!} \, p_1^{n_1} \, {n_1}^2 \sum_{\substack{(n_2, \ldots n_M) \\ \text{tales que} \\ \sum_{i=2}^{M} n_i = N - n_1}} \frac{(N - n_1)!}{n_2! \ldots n_M!} \, p_2^{n_2} \ldots p_M^{n_M} =
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!}\, p_1^{n_1}\, {n_1}^2 (p_2 + \ldots + p_M)^{N-n_1} = \\
&= \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!}\, p_1^{n_1}\, {n_1}^2 (1-p_1)^{N-n_1} = \\
&= \sum_{n_1=1}^{N} \frac{N!}{n_1!(N-n_1)!}\, p_1^{n_1}\, {n_1}^2 (1-p_1)^{N-n_1} = \\
&= \sum_{n_1=1}^{N} \frac{N!}{n_1!(N-n_1)!}\, p_1^{n_1}\, n_1\, n_1 (1-p_1)^{N-n_1} = \\
&= p_1 \sum_{n_1=1}^{N} \frac{N!}{n_1!(N-n_1)!}\, p_1^{n_1-1}\, n_1\, n_1 (1-p_1)^{N-n_1} = \\
&= p_1 \sum_{n_1=1}^{N} \frac{N!}{n_1\,(n_1-1)!(N-n_1)!}\, p_1^{n_1-1}\, n_1\, n_1 (1-p_1)^{N-n_1} = \\
&= p_1 \sum_{n_1=1}^{N} \frac{N!}{(n_1-1)!(N-n_1)!}\, p_1^{n_1-1}\, n_1\, (1-p_1)^{N-n_1}
\end{aligned}
$$

Hacemos un cambio de variables, haciendo $n_1 - 1 = i$ (o lo que es lo mismo $n_1 = i + 1$), y por lo tanto, obtenemos:

$$
\begin{aligned}
p_1 \sum_{n_1=1}^{N} \frac{N!}{(n_1-1)!(N-n_1)!}\, p_1^{n_1-1}\, n_1\, (1-p_1)^{N-n_1} &= p_1 \sum_{i=0}^{N-1} \frac{N!}{i!\,(N-(i+1))!}\, p_1^{i}\, (i+1)\, (1-p_1)^{N-i-1} \\
&= p_1\, N \sum_{i=0}^{N-1} \frac{(N-1)!}{i!\,((N-1)-i)!}\, p_1^{i}\, (i+1)\, (1-p_1)^{(N-1)-i} = \\
&= p_1\, N \sum_{i=0}^{N-1} \binom{N-1}{i} p_1^{i}\, (1-p_1)^{(N-1)-i}\, (i+1)
\end{aligned}
$$

distribuyendo los productos respecto de $(i+1)$, obtenemos:

$$
\begin{aligned}
& p_1\, N \sum_{i=0}^{N-1} \binom{N-1}{i} p_1^{i}\, (1-p_1)^{(N-1)-i}\, (i+1) = \\
&= p_1\, N \left[\left(\sum_{i=0}^{N-1} \binom{N-1}{i} p_1^{i}\, (1-p_1)^{(N-1)-i}\, i \right) + \left(\sum_{i=0}^{N-1} \binom{N-1}{i} p_1^{i}\, (1-p_1)^{(N-1)-i} \right) \right] = \\
&= p_1\, N \left[\underbrace{\left(\sum_{i=0}^{N-1} \binom{N-1}{i} p_1^{i}\, (1-p_1)^{(N-1)-i}\, i \right)}_{\otimes} + 1 \right] =
\end{aligned}
$$

porque podemos ver que, aplicando **(??)**, obtenemos:

$$\sum_{i=0}^{N-1} \binom{N-1}{i} p_1^i \, (1-p_1)^{(N-1)-i} = (p_1 + (1-p_1))^{N-1} = 1^{N-1} = 1$$

además, podemos también observar que a la fórmula $\otimes$ le podemos realizar cambios del mismo tipo que aplicamos anteriormente, para llegar a:

$$\begin{aligned}
\sum_{i=0}^{N-1} \binom{N-1}{i} p_1^i \, (1-p_1)^{(N-1)-i} \, i &= \sum_{i=1}^{N-1} \frac{(N-1)!}{i!\,((N-1)-i)!} \, p_1^i \, (1-p_1)^{(N-1)-i} \, i = \\
&= p_1 \sum_{i=1}^{N-1} \frac{(N-1)(N-2)!}{i\,(i-1)!\,((N-1)-i)!} \, p_1^{i-1} \, (1-p_1)^{(N-1)-i} \, i = \\
&= p_1 \, (N-1) \sum_{i=1}^{N-1} \frac{(N-2)!}{(i-1)!\,((N-1)-i)!} \, p_1^{i-1} \, (1-p_1)^{(N-1)-i}
\end{aligned}$$

cambiando nuevamente de variable, sea ahora $n = i-1$ (o lo que es lo mismo $i = n+1$) y reemplazando obtenemos:

$$\begin{aligned}
& p_1 \, (N-1) \sum_{i=1}^{N-1} \frac{(N-2)!}{(i-1)!\,((N-1)-i)!} \, p_1^{i-1} \, (1-p_1)^{(N-1)-i} = \\
= \; & p_1 \, (N-1) \sum_{n=0}^{N-2} \frac{(N-2)!}{n!\,((N-2)-n)!} \, p_1^{n} \, (1-p_1)^{(N-2)-n} = \\
= \; & p_1 \, (N-1) \sum_{n=0}^{N-2} \binom{N-2}{n} p_1^{n} \, (1-p_1)^{(N-2)-n} = \\
= \; & p_1 \, (N-1) \, (p_1 + (1-p_1)^{N-2} = p_1 \, (N-1)
\end{aligned}$$

porque $(p_1 + (1-p_1)^{N-2} = 1$.

Pero, ahora vemos que para un j cualquiera hemos obtenido que:

$$\sum_{\substack{\text{todos los}\\ \text{casos posibles}\\ \text{con } n_j}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot {n_j}^2 = p_j \; N \, [p_j \; (N-1) + 1] = {p_j}^2 \; N \; (N-1) + p_j \; N$$

Por otra parte, ya habíamos establecido que la función h distribuía uniformemente y por lo tanto la probabilidad de caer en la *lista* 1 es la misma que la de caer en cualquiera de las listas; así tenemos que: $p_i = \dfrac{1}{M} \quad \forall i, 1 \leq i \leq M$.

Entonces, reemplazando tenemos:

$$\begin{aligned}
{p_j}^2 \; N \; (N-1) + p_j \; N &= \left(\frac{1}{M}\right)^2 N \; (N-1) + \frac{1}{M} \, N = \frac{1}{M^2} \, (N^2 - N) \; + \; \frac{N}{M} = \\
&= \frac{N^2}{M^2} \; - \; \frac{N}{M^2} \; + \; \frac{N}{M} = \rho^2 - \; \frac{\rho}{M} \; + \; \rho
\end{aligned}$$

Dado que el resultado obtenido para cada j no depende del j, todos los términos de la sumatoria serán iguales, entonces:

$$\sum_{j=1}^{M} \sum_{\substack{\text{todos los} \\ \text{casos posibles} \\ \text{con } n_j}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot {n_j}^2 = \sum_{j=1}^{M} \left(\rho^2 - \frac{\rho}{M} + \rho\right) = M \left(\rho^2 - \frac{\rho}{M} + \rho\right)$$

Así, reemplazando en la fórmula original obtenemos que:

$$\sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot \left(\frac{1}{2N} \sum_{j=1}^{M} {n_j}^2 + \frac{1}{2}\right) =$$

$$= \sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot \left(\frac{1}{2N} \sum_{j=1}^{M} {n_j}^2\right) + \sum_{\substack{(n_1, n_2, \ldots n_M) \\ \text{tales que } \sum_{i=1}^{M} n_i = N}} \text{Prob}(n_1, n_2, \ldots, n_M) \cdot \frac{1}{2} =$$

$$= \frac{1}{2N} M \left(\rho^2 - \frac{\rho}{M} + \rho\right) + \frac{1}{2} = \frac{1}{2N} M \rho^2 - \frac{1}{2N} M \frac{\rho}{M} + \frac{1}{2N} M \rho + \frac{1}{2}$$

$$= \frac{\rho}{2} - \frac{\rho}{2N} + \frac{1}{2} + \frac{1}{2} = \frac{\rho}{2} - \frac{N}{2MN} + 1 = \frac{\rho}{2} - \frac{1}{2M} + 1 = \frac{\rho}{2} + 1 - \underbrace{\frac{1}{2M}}_{\text{término correctivo}}$$

Con esto queda demostrado que con cualquiera de las dos formas, relacionando los esfuerzos medios de éxito y de fracaso o directamente deduciendo el esfuerzo medio de éxito, llegamos al mismo resultado.

Esfuerzos en Rebalse Abierto Realeatorizado Total

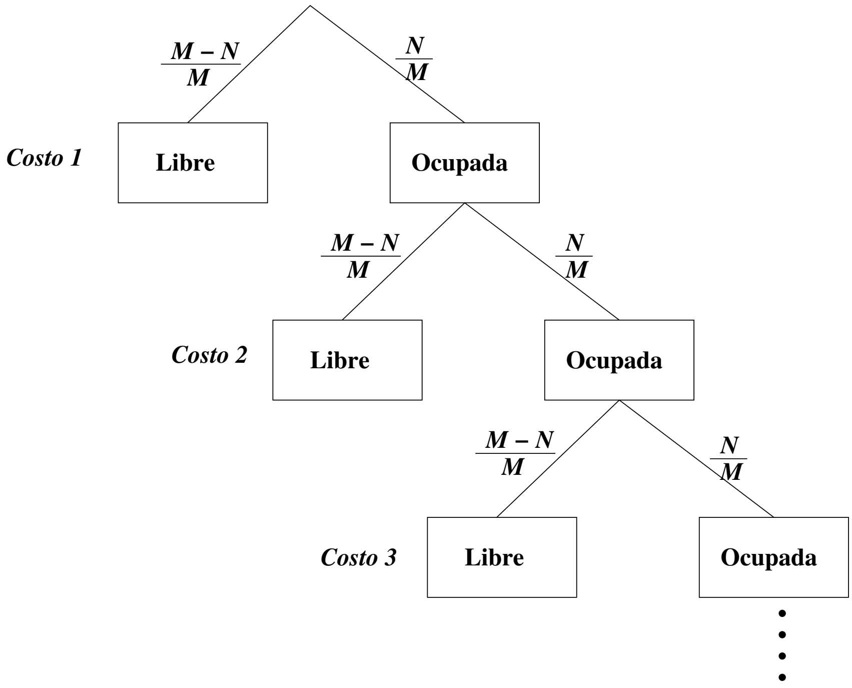
Vamos a deducir algunos de los esfuerzos medios a priori para dicho rebalse, suponiendo que la función h distribuye uniformemente y que en la estructura no se producen bajas.

Principalmente vamos a deducir el esfuerzo medio de evocación exitosa a priori; pero, para ello, haremos uso nuevamente de la relación existente entre esfuerzo medio de evocación exitosa y esfuerzo medio de alta (el que a su vez se relaciona con el esfuerzo de fracaso).

1. Esfuerzo medio de inserción a priori

Para la deducción de este esfuerzo debemos recordar la relación existente entre esfuerzos medios de inserción y fracaso en aquellas estructuras que no reestructuran. En esta estructura en particular se tiene que dar de alta un elemento es igual al costo de la búsqueda que fracasó cuando se lo incorporó (porque se detiene la búsqueda que fracasa en la celda virgen en donde se ubicará el elemento). Así, es indistinto hablar de esfuerzo medio de inserciń a priori o de esfuerzo medio de localización que fracasa a priori.

Si en la estructura hay N elementos presentes y estamos intentando insertar el elemento $N + 1$, podemos ver los casos que pueden ocurrir de la siguiente manera [6]:



el elemento se incorporará con *costo* 1 si la primera celda que visitó en la búsqueda era una celda virgen y, luego de tener N elementos presentes, esto puede ocurrir con probabilidad $\dfrac{M-N}{M}$ (porque hay sólo $M - N$ celdas vírgenes luego de haber ubicado N elementos en los M baldes disponibles). Puede ocurrir que en el primer intento se encuentre con una celda ocupada, con probabilidad $\dfrac{N}{M}$, y por lo tanto se deberá hacer un nuevo intento que tendrá una probabilidad $\dfrac{M-N}{M}$ de encontrar la siguiente celda virgen y con probabilidad $\dfrac{N}{M}$ la encontrará ocupada.

Si la función se comporta bien no habrá relación entre lo que ocurrió en un intento y lo que ocurrirá en el intento siguiente; por lo tanto, podemos considerarlos eventos independientes. Así, podemos plantear que el alta costará k si debió realizar $k - 1$ intentos, en que se encontró con celdas ocupadas, y en el intento k encontró una celda virgen. Dicho costo ocurrirá con probabilidad:

$$\left(\frac{N}{M}\right)^{k-1} \frac{M-N}{M}$$

Entonces, ahora podemos plantear el esfuerzo medio de inserción a priori que sería:

$$\sum_{i=1}^{\infty} i \left(\frac{N}{M}\right)^{i-1} \frac{M-N}{M} = \sum_{i=1}^{\infty} i\, \rho^{i-1}(1-\rho) = (1-\rho)\sum_{i=1}^{\infty} i\, \rho^{i-1}$$

Para resolver esta fórmula necesitamos analizar en detalle la sumatoria; pero como no conocemos

[6] Si se razonan los costos de localización que fracasa, ya se tienen N elementos presentes en la estructura y se está buscando uno que no está. La localización que fracasa siempre para en celdas vírgenes.

cuánto vale debemos tratar de resolverla. En realidad no sabemos cuánto vale $\sum_{i=1}^{\infty} i\,\rho^{i-1}$, pero sí sabemos cómo resolver una serie geométrica con razón $\rho < 1$; o sea, sabemos que $\sum_{i=0}^{\infty} \rho^i = \frac{1}{1-\rho}$.

Podemos así tratar de usar lo que sabemos para deducir el resultado de la sumatoria que nos interesa. Sea $S = \sum_{i=1}^{\infty} i\,\rho^{i-1}$, podemos ver que:

$$
\begin{array}{rll}
 & 1 + 2\rho + 3\rho^2 + 4\rho^3 + 5\rho^4 + \ldots\ldots\ldots & \big\} \longrightarrow \text{Suma de partida} \\
- & & \\
 & 1 + \ \rho + \ \rho^2 + \ \rho^3 + \ \rho^4 + \ldots\ldots\ldots & \big\} \longrightarrow \text{Serie geométrica de razón } \rho \\
\hline
 & \rho + 2\rho^2 + 3\rho^3 + 4\rho^4 + \ldots\ldots\ldots & \big\} \longrightarrow \text{Resultado de hacer la diferencia}
\end{array}
$$

pero:

$$
\rho + 2\rho^2 + 3\rho^3 + 4\rho^4 + \ldots = \rho\ \underbrace{(1 + 2\rho + 3\rho^2 + 4\rho^3 + \ldots)}_{\text{Suma de partida}}
$$

entonces podemos plantear la siguiente igualdad: $S - \sum_{i=0}^{\infty} \rho^i = \rho\, S$ y de aquí podemos obtener que:

$$
S = \rho\, S + \sum_{i=0}^{\infty} \rho^i = \rho\, S + \frac{1}{1-\rho}
$$

y agrupando las S obtenemos que:

$$
S - \rho\, S = (1-\rho)\, S = \frac{1}{1-\rho}
$$

y desde aquí podemos finalmente despejar a S:

$$
S = \sum_{i=1}^{\infty} i\,\rho^{i-1} = \frac{1}{(1-\rho)^2}
$$

Ahora, volviendo al desarrollo del esfuerzo medio de inserción a priori, tendríamos:

$$
(1-\rho) \sum_{i=1}^{\infty} i\,\rho^{i-1} = (1-\rho)\,\frac{1}{(1-\rho)^2} = \frac{1}{1-\rho}
$$

2. Esfuerzo medio de evocación exitosa a priori

Por ejemplo, si consideramos el i-ésimo elemento incorporado, podemos ver que el esfuerzo medio a priori de buscarlo exitosamente es el esfuerzo medio de insertarlo en una estructura con $i-1$ elementos presentes. Por lo tanto:

Elemento	*Factor de Carga*(ρ)	*Esfuerzo Medio de Inserción a priori*	*Esfuerzo Medio de Evocación a priori*
1	0	$\dfrac{1}{1-0}=1$	1
2	$\dfrac{1}{M}$	$\dfrac{1}{1-\dfrac{1}{M}}$	$\dfrac{1}{1-\dfrac{1}{M}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$\dfrac{N-1}{M}$	$\dfrac{1}{1-\dfrac{N-1}{M}}$	$\dfrac{1}{1-\dfrac{N-1}{M}}$

Ahora podemos plantear formalmente el esfuerzo medio de evocación exitosa, asumiendo isoprobabilidad de consulta:

$$\sum_{i=0}^{N-1}\frac{1}{N}\,\frac{1}{1-\dfrac{i}{M}}=\frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{1-\dfrac{i}{M}}=\frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{\dfrac{M-i}{M}}=\frac{1}{N}\sum_{i=0}^{N-1}\frac{M}{M-i}=\frac{M}{N}\sum_{i=0}^{N-1}\frac{1}{M-i}$$

Como tenemos que resolver ahora la sumatoria $\displaystyle\sum_{i=0}^{N-1}\frac{1}{M-i}$, la desarrollamos para ver cómo se comporta:

$$\sum_{i=0}^{N-1}\frac{1}{M-i}=\frac{1}{M}+\frac{1}{M-1}+\frac{1}{M-2}+\ \dots\ +\frac{1}{M-N+1}$$

nuevamente no es sencillo resolverla directamente; pero podemos ver que es una *serie armónica* recortada. Recordando que una *serie armónica* era: $H_k=\displaystyle\sum_{i=1}^{k}\frac{1}{i}$, y con H_k se denotaba el $k-$ésimo número armónico. Vemos entonces que si agregáramos los términos que forman el número armónico $H_{M-N}=\displaystyle\sum_{i=1}^{M-N}\frac{1}{i}=\frac{1}{M-N}+\frac{1}{M-N-1}+\ \dots\ +1$, completaríamos la sumatoria de manera tal de formar el número armónico H_M. Por lo tanto, podemos plantear que:

$$\sum_{i=0}^{N-1}\frac{1}{M-i}=H_M-H_{M-N}$$

Una serie armónica se aproxima por: $H_k \approx \ln k$, dado que realmente para H_k tenemos que: $H_k = \ln k + \gamma - O\left(\frac{1}{k}\right)$, siendo γ la constante de Euler ($\gamma = 0{,}5772156649\ldots$, $\gamma < 1$), y si k es suficientemente grande el término $O\left(\frac{1}{k}\right)$ será muy pequeño y se puede descartar. Así tenemos entonces que:

$$\sum_{i=0}^{N-1} \frac{1}{M-i} = H_M - H_{M-N} \approx \ln M - \ln(M-N) = \ln \frac{M}{M-N} = \ln\left(\frac{1}{1-\rho}\right)$$

Ahora volviendo al cálculo del esfuerzo medio de evocación exitosa tenemos:

$$\frac{M}{N}\sum_{i=0}^{N-1} \frac{1}{M-i} \approx \frac{1}{\rho}\ \ln\left(\frac{1}{1-\rho}\right) = \frac{1}{\rho}\ \left(-\ln\left(1-\rho\right)\right)$$

como $0 \leq \rho < 1$, $(1-\rho) < 1$, y por lo tanto $\ln(1-\rho) < 0$.

□

Reconocimientos

El presente apunte se realizó tomando como base notas de clases, de Estructuras de la Información y de Estructuras de Datos y Algoritmos, del Ing. Hugo Ryckeboer en la Universidad Nacional de San Luis.