

Los algoritmos genéticos, o lo que «no» es la evolución.

Mariano Vázquez Espí

Ondara (España), 2 de noviembre de 2002.

Me propongo acometer esta redacción bajo los dos principios fundamentales de la Biblioteca CF+S:[1]

1. No hemos venido aquí a sufrir.
2. La información está en la diferencia.

Y además intentaré ser breve y claro a la vez.[2]

Algoritmos genéticos simples.

Los *datos* de cualquier programa de ordenador pueden describirse como una *ristra de ceros o unos*[3]. Esa regla general puede aplicarse, en particular, a las soluciones para un problema concreto que pueda ser descrito algorítmicamente (¡no se asuste! ‘algorítmicamente’ significa por medio de un conjunto ordenado y bien especificado de operaciones aritméticas corrientes: sumas y multiplicaciones[4]). De hecho, un programa *que quiera ser útil* debe explorar distintas soluciones a un problema, ya sea secuencialmente, ya sea en paralelo. (El programa normalmente busca la ‘mejor’ solución.)

El denominado *código genético* (DNA) o *genoma* puede describirse con simpleza como una ristra de cuatro ‘sonidos’ distintos[5]. Se le puede mirar como un número en base cuatro (allí, donde una ristra de ceros y unos es un número en base dos[6]). Dada la cercanía analógica, era cuestión de tiempo que a alguien se le ocurriera un método elegante de generación de nuevas soluciones para un programa de ordenador: elíjanse dos soluciones previas, contémpense como una ristra de ‘genes’, y recombínense como si de DNA se tratara[7].

Cuadro 1

```
+-----+
| solución A: 00011001... |
| solución B: 10101000... |
| hija de AB: 00001000... |
| hija de BA: 10111001... |
+-----+
```

Lo único que hace falta para la recombinación es elegir un punto de corte (generalmente se escoge ‘al azar’):

Cuadro 2

solución A: 000 11001...
solución B: 101 01000...
hija de AB: 000 01000...
hija de BA: 101 11001...
^
corte

Si no lo ve claro no siga adelante, lea la siguiente nota al pie[8] y vuelva a empezar.

El método es extremadamente útil si se une a un método de selección que permita *diferenciar a* (o *elegir entre*) los miembros del conjunto de descendientes y progenitores. En problemas de optimación unidimensional, en los que a cada solución admisible es posible asignarle un *único* coste bien definido, las soluciones se pueden ordenar *totalmente* según su coste. Por ejemplo (con costes hipotéticos):

Cuadro 3

solución	coste	orden
A	5	3.
B	3	1.
AB	6	4.
BA	4	2.

Si el programa de ordenador busca la mejor solución (la de coste mínimo) puede seleccionar como progenitores a los ‘primeros de la fila’, B y BA en el ejemplo, y producir nueva descendencia (la asignación de costes sigue siendo hipotética):

Cuadro 4

solución	'DNA'	coste	orden
solución B	101010 00...	3	2.
hija de BA	101110 01...	4	3.
hija de BBA	101010 01...	2	1.
^			
corte			

Así pues, con estas simples reglas de *reproducción* y *selección* es posible construir un método muy general de aproximación a la solución óptima de un problema unidimensional. De hecho, un método que funciona y que se usa ya cotidianamente en ingeniería.

En la literatura, es corriente ver descrito el algoritmo anterior como un esquema de ‘selección natural’. Sin entrar en polémicas, resulta razonablemente preciso decir que el algoritmo emplea el *principio de selección del más apto*, siempre y cuando identifiquemos mejor adaptación con menor coste (o *del más fuerte*, si es que mayor fortaleza se puede identificar con menor coste, etc). Se trata de un *algoritmo genético* (GA en adelante).[9]

La relación entre coste y ‘código genético’ depende del problema en cuestión. Las reglas para pasar del ‘DNA’ al coste pueden ser complicadísimas pero, para lo que nos interesa aquí, basta con saber que, si el problema es descriptible algorítmicamente, a cada ‘DNA’ le corresponde un coste de forma unívoca.[10]

Tras el intenso uso que durante la última década se ha hecho de los GAs es posible extraer algunas conclusiones muy generales que, en principio, pueden aplicarse también a la selección del más apto (como concepto genérico).

1. No hay demostración disponible de que un GA conduzca a la mejor solución del problema. Las pruebas disponibles se limitan a asegurar que, cumplidas algunas condiciones técnicas ---no muy difíciles---, una población suficientemente grande y diversa mejora su aptitud media conforme las generaciones se suceden.[11]
2. La población tiende a ser, con el tiempo, uniformemente parecida, sin que, al final, se perciba diversidad. Si la población es óptima en sentido absoluto, la población puede ser *clónica* (formada por individuos idénticos).
3. Para alcanzar buenos resultados (soluciones muy cercanas a la óptima) no es necesario emplear mecanismos de mutación al azar, es decir, la alteración de aleatoria de un gen antes o después de la recombinación (en nuestro caso, elígase un *bit* al azar y si es ‘1’ cámbiase a ‘0’, o al revés). Basta con partir de una población suficientemente grande y diversa como para que por simple reproducción sexual la población evolucione hacia la región más prometedora del espacio de las soluciones. El denominado ‘operador mutación’ se emplea por conveniencia y con una frecuencia muy pequeña, pudiéndose considerar como un operador secundario[12].

En consecuencia, mi conjetura es que, en lo que se refiere al conjunto de fenómenos que usualmente se denotan como evolución biológica:

1. La selección del más apto no explica ni la aparición de nuevas especies, ni la diversidad en el interior de una misma especie, ni la diversidad entre ellas. Resulta muy razonable, incluso, dudar de que explique algo más que la variación de pequeños detalles morfológicos sobre una forma fundamental bien definida de antemano (como el color de las polillas británicas, o la forma y tamaño de los picos en las aves de las Galápagos).
2. La mutación al azar no parece que sea imprescindible en una explicación comprensiva de la evolución biológica: exista o no como hecho constatable, la importancia de su existencia es banal.

¿Es la vida un problema unidimensional?

Si lo fuera, la evolución, liderada por la selección del más apto, habría alcanzado (o alcanzará) el individuo ‘óptimo’ (¿de cada especie? ¿de cada tronco? ¿de cada raza?). Si *vivir fuera un problema intrínsecamente unidimensional*, bastará con mostrar un solo caso en que esta proposición no se cumpla, para mostrar así su falsedad general.

El caso está a la vista de cualquiera (y fue estudiado profusamente por Darwin): los animales domésticos. Se trata de seres vivos, se trata de especies (por ejemplo los perros), se trata de selección de los más aptos (bien que en un ambiente formado esencialmente por seres humanos, que son los agentes de la presión selectiva sobre las razas caninas, bien alimentadas y con suficiente territorio a su disposición), y sin embargo el resultado neto de esta evolución genética es una fascinante diversidad de razas que siguen siendo de la misma especie (son compatibles reproductivamente hablando).

Se puede opinar que el ejemplo es muy traído por los pelos, pero se trata de seres vivos que evolucionan delante de nuestras narices (hay razas caninas de apenas unas décadas de antigüedad). Y como cualquier explicación de la evolución con vocación generalista tiene que dar cuenta de todos los fenómenos englobados, este ejemplo no puede ser rechazado como una excepción.

Lo que pasa con los perros, pasa con los gatos, pero también con muchas otras especies domésticas y no domésticas, de suerte que, a poco que se medite sobre ello, parece que la conjetura razonable es que:

La vida, vivir, no es un problema unidimensional.

La alternativa, obvia, es mirar la vida, el vivir, desde una óptica multidimensional.

El problema de los problemas multidimensionales.

¿Qué es un problema multidimensional? Uno en que a cada solución se le asignan más de un ‘valor’ o ‘coste’: un jugador de baloncesto tiene que ser alto, pero también listo, ágil, fuerte, etc. Por eso, porque la elección de un jugador de baloncesto es un problema multidimensional, de vez en cuando surge un ‘base’ excepcionalmente bueno, aunque sea ‘bajito’.

¿Cómo sería un GA para problemas multidimensionales? Esencialmente igual, pero radicalmente diferente en algunos detalles. Nada cambia en las reglas de reproducción (con mutación o sin ella): se elige un punto de corte y se efectúa la recombinación. Pero ¿qué individuo es *mejor* que otro? ¿quién es el *más adaptado*? ¿cómo se *eligen* los progenitores? Aquí, como vamos a ver, se puede establecer, a lo más, un orden *parcial*, no *total*. Veámoslo con un ejemplo de problema bidimensional, en el que, a cada solución, se le asignan dos costes distintos (los valores siguen siendo hipotéticos).[13]

Cuadro 5

solución	coste 1	coste 2	orden
A	3	4	2
B	2	3	1
C	4	2	1
D	5	4	3
E	4	5	3
F	1	4	1
			clase

La regla de selección es:

Si una solución tiene costes menores que los de otra, y ninguno de sus costes es mayor que el de la otra, entonces y sólo entonces es mejor que la otra. O, lo que es lo mismo, la primera es mejor solución que la segunda. O, lo que es lo mismo, debe descartarse la segunda.

Con esa regla se ha ‘calculado’ la columna «orden» en el cuadro anterior. Para ello se han ido comparando las soluciones por parejas. En lo que sigue $x < y$ significará que x está mejor adaptado que y (tiene menores costes); y $x = y$ significará que x no está ni mejor ni peor adaptado que y (es decir, no se cumple ni que $x < y$ ni que $y < x$); $x = y$ puede leerse como ‘ x es equiparable con y ’ [14]. Al comparar cada una de las seis soluciones con las otras cinco se obtienen las siguientes relaciones:

Cuadro 6

Solución	Relaciones
A	B<A A=C A<D A<E F<A
B	B=C B<D B<E B=F
C	C<D C<E C=F
D	D=E F<D
E	F<E

Para establecer un orden sólo podemos operar con la regla de selección anterior, **descartar ‘el peor’, en apariencia sólo ligeramente distinta** a la ‘popular’: **elegir ‘lo mejor’**. La regla significa eliminar todas las soluciones que, en el cuadro anterior, aparecen a la **derecha** de la relación $<$ y quedarnos con el resto. Esta primera operación nos deja con las soluciones B, C y F que mantienen entre sí las relaciones $B=C$, $B=F$ y $C=F$. Relaciones que muestran que ninguna de las tres es mejor que las demás: todas son equiparables (técnicamente forman una clase de equivalencia); pero nótese que las tres son mejores que alguna de las soluciones descartadas (aunque, es notable, una de las tres es equiparable con una de las descartadas: $C=A$). En todo caso, concluimos que B, C y F son miembros de *primera clase*. [15]

Podemos repetir el proceso con las soluciones descartadas, las que no pertenecen a la primera clase, esto es con A, D y E.

Cuadro 7

solución	coste 1	coste 2
A	3	4
D	5	4
E	4	5

relaciones
A<D A<E
D=E

Soluciones descartadas: D y E. *Segunda clase*: A.

Usted puede repetir por tercera vez el proceso (es sencillo, puede hacerlo mentalmente), y concluir que D y E son miembros de *tercera clase*. Con esto termina el proceso de jerarquización de soluciones y, por tanto, el ‘cálculo’ de la columna *orden* (o *clase*) del cuadro 5.

Supongamos ahora que queremos sustituir nuestra actual generación por otra, formada con seis descendientes de progenitores ‘bien’ elegidos. Parece que sin duda estos últimos serán B, C y F. Tenemos suerte: en este caso podemos dejar que cada pareja posible (BC, CF y BF) tengan dos descendientes cada una y asunto concluido.

¿Cómo será la población resultante? Depende de la relación matemática entre cada uno de los dos costes y los ‘genes’ de cada individuo. En general, los descendientes pueden ser de cualquiera de las tres clases que, hasta ahora, hemos descubierto. De hecho, la nueva generación podría ser totalmente descartable frente a la primera clase de sus progenitores.

Todavía más, si dejáramos tener descendencia a progenitores de segunda y de tercera clase, nada hay que impida que alguno de los descendientes sea de la primera clase de la nueva generación. Todo depende, una vez más, de cómo sean las relaciones algorítmicas entre costes y genotipos en un caso concreto.[16]

Un truco «particular» para simular la evolución biológica

En general, existe un truco que permite asegurar que, casi siempre, la nueva generación sea equiparable a la anterior (aunque se requieren poblaciones más grandes que las del ejemplo). Sin entrar a describir detalles innecesarios, el truco consiste en emparejar a cada progenitor de primera clase con un individuo genéticamente próximo, lo que puede hacerse (aproximadamente) incluso sin conocer sus genes: basta elegir un individuo cuyos costes sean *cuasi*-iguales al del progenitor de primera clase: B con un *cuasi*B, C con un *cuasi*C y F con un *cuasi*F. Esto es tanto como decir que puede ser necesario llamar a miembros de segunda, tercera,... clase para emparentarse con los cabezas de serie de la primera clase (todo depende de si el cabeza de serie es el único de su ‘especie’ en la primera clase). Este es, en lo esencial, el truco de los criadores de perros y gatos ‘de raza’.

Hay *jerarquización* (pero incompleta), hay *descarte del peor* (salvo que el individuo sea el último que queda para formar pareja), pero *la selección del mejor adaptado* se ha difuminado hasta prácticamente desaparecer (sobre todo porque el concepto de ‘mejor’ en su versión absoluta ha desaparecido).

Si usted deja volar su imaginación no le costará admitir como conjeturas provisionales que, según se aplica este truco generación tras generación:

1. B, C y F pudieran ser la semilla de nuevas especies en la población, destinadas a incompatibilizarse reproductivamente, mediando tiempo suficiente.
2. La diversidad intraespecie disminuirá.
3. La diversidad interespecies se mantendrá.

Las dos últimas conjeturas cuadran razonablemente bien con muchos de los fenómenos observados actualmente en la población viva del planeta. Quedan por explicar, sin embargo, fenómenos tanto más cruciales para la vida como el de la génesis de organismos de ‘nivel superior’ (organismos pluricelulares a partir de células, por ejemplo).

Esta caricatura, también, deja sin describir aspectos más importantes que los aquí ilustrados: ¿qué es el ‘ambiente’ para un individuo concreto? ¿Y para una especie? ¿Puede cambiar una especie su ‘ambiente’? ¿Cómo afecta tal cambio a la multidimensionalidad de la vida (los costes) y, por tanto, a la división de la población en clases? Y, sobre todo, ¿cómo se elige pareja?[17]

Conclusión

El propósito final de esta redacción es concluir que **la evolución biológica, la que nos ha traído hasta aquí, no parece que pueda operar con la regla de la selección del mejor adaptado**, cualquiera que sea el nivel de organización de lo viviente que se ponga en observación. Que, además, la evolución por selección del mejor adaptado (en la simulación de un programa de ordenador) fuerza a una descripción unidimensional del fenómeno simulado y conduce a una población uniforme, casi con seguridad mejor adaptada que la población inicial, pero sin seguridad absoluta de que sea la mejor población entre las posibles (que en tales fenómenos estaría formada por copias de unos pocos clones diferentes).[18]

Apunte de algunas conclusiones laterales

1. El programa de ordenador[19] es quien conecta ‘genoma’ y coste, ‘genotipo’ y ‘fenotipo’[20]. En los GAs simples es una pieza constante. Es lo análogo al ribosoma celular[21]. Sin programa de ordenador (sin ribosoma) no hay fenotipo, no hay selección (de ninguna clase), casi me atrevería a decir que no hay vida. Si cambia el programa de ordenador, las mismas soluciones (las mismas ristras de ceros y unos en clave de datos, el mismo ‘genoma’) darán lugar a muy distintos fenotipos (si es que el nuevo programa funciona con ellos[22]).
 1. Llama la atención que los ribosomas sean razonablemente uniformes en todos los organismos vivos...
 2. Un óvulo fecundado: ¿de quién son los ribosomas que comienzan a leer el nuevo genoma? ¿No son de la madre? ¿Qué sería del DNA sin el ribosoma materno?
 3. Si usted tiene que elegir, qué considera más importante: ¿el programa de ordenador o los datos? ¿El ribosoma (proteínas y RNA) o el genoma (DNA)? ¿La persona que lee o el libro que lee?
2. Al principio he acudido a la muy pomposa expresión «ristra de ceros y unos», o lo que es lo mismo al concepto de *código genético*. Incluso lo he descrito en los primeros cuadros a fin de explicar detalles técnicos (corte, recombinación). Pero según el texto avanza, la necesidad de mantener el código genético a la vista ha ido disminuyendo hasta éste desaparecer[23]. De hecho, para la discusión del posible carácter multidimensional de la vida, el detalle del código genético se volvió completamente irrelevante: bastó con mostrar ejemplos descritos fenotípicamente (por sus costes). Esto es, desde luego, un simple hecho literario. Pero contrasta con los millones de dolares dedicados a describir cada nimio detalle del genoma humano, contrasta con las esperanzas creadas entre miles de personas enfermas, ilusionadas ahora con que, tras unos cuantos cambios en sus ‘ceros y unos’, van a verse libres de **todo mal**. Mientras tanto las preguntas fundamentales acerca de la *naturaleza y vida de la vida* siguen siendo apenas exploradas en los márgenes de la ciencia oficial.
3. Me encontré con los GAs redactando mi tesis doctoral sobre otro algoritmo de aproximación a problemas de optimación, el recocido simulado, en adelante SA (*simulated annealing*). Éste último es una analogía con la termodinámica de gases. En diseño de estructuras ambos algoritmos pueden considerarse equiparables: el GA se encamina más rápido hacia la región prometedora pero afina menos la solución óptima; el SA, por el contrario, es más parsimonioso, pero se acerca significativamente más a la solución óptima. El GA es una analogía con una cierta idea de la vida (el GA multidimensional parece una mejora sustancial de esa primera idea), mientras que el SA es una analogía con una cierta idea de la materia inanimada. Cuando mataba el tiempo con estas cosas, siempre encontré que el SA superaba, en cuanto a calidad de la solución encontrada, a un GA análogo.[24]

Lo que me interesa aquí subrayar, sin embargo, es que ambos algoritmos dibujan paradigmas evolutivos que pueden aplicarse a la solución de los *mismos* problemas. Esto me sugiere que podría existir una forma de ‘ver’ los ‘orígenes’ (el pasado, ese país extraño), en la que la evolución de la materia (desde las partículas más elementales hasta las moléculas grandes de la química inorgánica) *engarzara fácilmente* con la evolución biológica (desde las proteínas hasta la sociedad humana). Quizás se pueda construir, mediante el análisis de estas coincidencias y engarzes, un buen pedestal para una descripción *formalista* de la Naturaleza, más abarcadora y menos reduccionista que la del materialismo histórico tradicional.[25]

Referencias

Chomsky, Noam; & George A. Miller (1958) “**Finite State Languages**” (*Information and Control*, número 1, pp. 91-112)

David, Lawrence (ed.) (1987) **Genetic Algorithms and Simulated Annealing** (London: Pitman. Los Altos (CA): Morgan Kaufman)

Garey, Michael R.; and David S. Johnson (1979) **Computers and Intractability** (New York: Freeman. Se cita el ‘update’ de 1991 de la misma editorial.)

Goldberg, David E. (1989) **Genetic Algorithms in Search, Optimization & Machine Learning** (Reading (Mass.): Addison-Wesley)

Holland, John H. (1975) **Adaptation in Natural and Artificial Systems** (Cambridge (Mass): The MIT Press, 1992. [Edición por MIT Press de la edición original de *The University of Michigan*])

Holland, John H. (1992) “**Algoritmos genéticos**” (*Investigación y Ciencia*, número 192, pp. 38-45)

Kauffman, Stuart A. (1993) **The Origins of Order** (New York: Oxford University Press)

Ogayar, A.; Sánchez-Pérez, M. (1998) “**Prions: an evolutionary perspective**” (*Internatl. Microbiol.* 1:183-190)

Prusiner, S. B. (1995) “**El prión en la patología**” (*Investigación y Ciencia*, marzo 1995)

Vázquez Espí, Carlos et Mariano Vázquez Espí (1997) “**Discussion of ‘Sizing, Shape and Topology Design of Trusses Using Genetic Algorithm’ by S.D.Rajan**” (*J. of Structural Engineering*, vol. 123, pp. 375-376)

Vázquez Espí, Carlos et Mariano Vázquez Espí (1997a) “**Discussion of ‘Annealing Strategy for Optimal Structural Design’ by S.R.Tzan & C.P.Pantelides**” (*J. of Structural Engineering*, vol. 123, pp. 1277-1278)

Vázquez Espí, Mariano (1995) “**Un nuevo algoritmo para la optimación de estructuras: el recocido simulado**” (*Informes de la Construcción*, vol. 46, número 436, pp. 49-69)

Vázquez Espí, Mariano (1997) “**Discussion of ‘Fuzzy Controlled Genetic Algorithm Search for Shape Optimization’ by C.K.Soh & J.Yang**” (*J. of Computing in Civil Engineering*, v. 11, pp. 213-214)

Vázquez Espí, Mariano (1997a) “**Los límites de la técnica**” (*Ciudad y territorio*, vol XXIX, número 111, pp. 65-79 [Ahora también en *Boletín CF+S*, n.3, <http://habitat.aq.upm.es/boletin/n3/amvaz.html>])

Vázquez Espí, Mariano (1998) “**Discussion of ‘Genetic Algorithms-Based Methodologies for Design Optimization of Trusses’ by S.Rajeev & C.S.K.Krishnamoorthy**” (*J. Struct. Engineering*, v.124, pp. 979-980)

Vázquez Espí, Mariano (1998a) “**Valores, medidas y teoría de la decisión**” (*Archipiélago*, n.33, pp. 90-100)

Vázquez Espí, Mariano (1999) “**Ciudades sostenibles**” (*Boletín CF+S*, n.8, <http://habitat.aq.upm.es/boletin/n8/amvaz.html>)

Vázquez Espí, Mariano (2000) “**Cuantificación y toma de decisiones**” (en *Economía, ecología y sostenibilidad en la sociedad actual*, J.M.Naredo y F.Parra(eds). Madrid: Siglo XXI, pp. 175-192)

Fecha de referencia: 1-6-2003

1: Creo que es la primera vez que van a ser hechos públicos.

2: Para la brevedad, *no* seré parco en las notas al pie: ¡todo lo contrario! Pero espero que en una primera lectura pueda prescindirse totalmente de ellas (salvo de esta y la anterior). Mi consejo es que acuda sólo a estas notas al pie cuando necesite aclaraciones o mayores certidumbres sobre lo que se dice en el texto principal.

3: Normalmente resulta poco práctico hacer eso: los datos son ‘cosas’ como números de teléfono, nombres, direcciones: es decir ristras de caracteres corrientes. Pero a la hora de escribir un programa se puede *elegir* verlos así, o como números abstractos en formato hexadecimal, decimal, binario. Es en este último caso cuando se ven como ristras de ceros y unos; cualquier dato puede ‘mirarse’ así.

4: Claro que puede ser un conjunto tan numeroso que, en no pocas ocasiones, no puede llevarse a término si no es con la ayuda de una máquina de cálculo: ábacos, ordenadores analógicos o digitales, etc. En este sentido ‘programa de ordenador’ y ‘algoritmo’ pueden considerarse como sinónimos. Para la persona interesada la obra clave es [*Garey & Johnson, 1979*].

5: Los cuatro ‘sonidos’ son Adenina, Timina, Guanina y Citosina. Con ellos se pueden formar 64 ‘palabras’ distintas de tres sonidos cada una. Esas 64 palabras forman el ‘lenguaje genético’ que, conteniendo bastantes palabras sinónimas, ‘expresan’ 21 ‘ideas’ diferentes: las de 20 aminoácidos y la orden ‘STOP’ («parar la lectura»). A su vez, una ristra concreta de aminoácidos encadenados da lugar a una proteína. Cada proteína, única si se la piensa como una ristra ‘recta’, puede adoptar *cuasi*-innumerables formas según se pliegue o ‘curve’ en el espacio: cada forma concreta de una proteína

tiene propiedades concretamente diferentes.

Técnicamente se trata de un *lenguaje finito* según la clásica y útil definición de [Chomsky & Miller, 1958:91-94]. Aquí hay un detalle importante que no puedo dejar de subrayar. Con la definición anterior para el genoma (comúnmente admitida) se ‘maneja’ una fracción pequeña del total de la cadena de DNA. Para que las cosas cuadraran, al resto del DNA se le asignó primeramente un papel totalmente pasivo, de ‘relleno’. Sin embargo, puede que esa parte ‘muda’ del DNA forme parte de otros *lenguajes finitos* que todavía no hemos sido capaces de ‘escuchar’. En esa dirección apuntan algunas sugerencias de Máximo Sandín en su artículo en este mismo número.

6: Los números corrientes se escriben con diez cifras (del 0 al 9) y se dice que son de base 10. Un número escrito con sólo 5 cifras (A, B, C, D y E, por ejemplo ---como cifra puede servir cualquier cosa) está en base 5. Pero si el número representa un ‘número’ de cosas, el ‘número’ de cosas que un número representa es **independiente** de la base usada para escribir el número en cuestión. Sin embargo, esta ‘coincidencia’ es una particularidad del lenguaje aritmético.

7: Hasta donde sé, el primero en utilizar la regla fue John Holland, véase [Holland, 1975] y [Holland, 1992].

8: Es como cortar dos barajas de 40 cartas y mezclar la mitad de arriba de una de ellas con la mitad de abajo de la otra. Para que el ejemplo fuera perfecto sería necesario que los mazos en que quedan divididas ambas barajas fueran del mismo tamaño, de suerte que las dos nuevas barajas resultantes de la recombinación sigan teniendo 40 cartas. Es decir, ambas barajas debieran cortarse en el mismo punto, cosa que no puede hacerse sin contar las cartas.

9: Cosas tales como estructuras de edificación, motores de aviación, *chips*, trazados de redes (de agua, de telefonía, etc.), se diseñan (o se mejoran) con este método, véase [Holland, 1992]. En consecuencia, la *selección del más apto* (tal y como queda definida implícitamente por el método de los GAs) ha demostrado su ‘aptitud’ para resolver (aproximadamente) problemas de optimización unidimensional (o escalar). Éste es un **hecho** que difícilmente puede ser puesto en cuestión. Sin embargo, debe advertirse que no se trata del único método disponible ni, probablemente, del ‘mejor’ (aunque la elección de un método de resolución es ciertamente otro problema de *optimización*, bastante complejo por lo demás: se busca precisión, pero también velocidad, verificabilidad, etc.).

10: Por fijar ideas, piénsese en números en base dos de cuatro cifras (*bits*), tales como 0110 ó 0001. Una función de asignación de costes muy simple podría ser la suma de las cifras. Entonces, $\text{coste}(0110)=2$, $\text{coste}(0001)=1$. La solución de menor coste es, en este caso, obvia: $\text{coste}(0000)=0$. Un GA estándar obtiene esa solución en unas cuantas generaciones (en este caso, se trata de un algoritmo perfectamente superfluo: no lo necesitamos).

11: Sobre la fundamentación matemática de los algoritmos genéticos véase, por ejemplo, [Holland, 1975] o [Goldberg, 1989].

12: Precisamente Holland señala como una de las aportaciones de los GAs a la biología el haber suprimido el carácter *central* de las mutaciones, véase [Holland, 1975].

13: Volviendo al ejemplo de números binarios de cuatro cifras, podemos elegir como coste 1 la suma de las cifras pares: $\text{coste1}(0101)=2$, $\text{coste1}(0110)=1$, $\text{coste1}(0000)=0$. Se denomina ‘complemento’ al número binario que resulte de cambiar los ceros en unos y los unos en ceros. Por ejemplo, $\text{complemento}(0101)=1010$, $\text{complemento}(0110)=1001$, $\text{complemento}(0000)=1111$. Como coste 2 podemos elegir la suma de todas las cifras del complemento, de manera que $\text{coste2}(0101)=2$, $\text{coste2}(0110)=2$, $\text{coste2}(0000)=4$.

14: Volviendo al ejemplo de números binarios de cuatro cifras, la comparación entre, por ejemplo, 0101 y 0110 es como sigue:

número	coste 1	coste 2
0101	2	2
0110	1	2

luego $0110 < 0101$ puesto que $\text{coste1}(0110) < \text{coste1}(0101)$ mientras que $\text{coste2}(0110) = \text{coste2}(0101)$. Análogamente:

número	coste 1	coste 2
0101	2	2
0000	0	4

luego $0101 = 0000$, puesto que **no es cierto que** $0101 < 0000$ [ya que $\text{coste1}(0101) > \text{coste1}(0000)$] y **no es cierto que** $0000 < 0101$ [ya que $\text{coste2}(0000) > \text{coste2}(0101)$].

15: Volviendo al ejemplo de los números binarios de cuatro cifras, podemos especular sobre cuales números serán de *primera clase*. Las soluciones de menor coste 1 son de la forma $*0*0$, en donde $*$ significa que tanto da un 1 o un 0 en esa posición: no afecta al coste 1. Para todas las soluciones de esa forma, $\text{coste1}(*0*0)=0$, que es el menor valor posible para ese coste. La solución de menor coste 2 es, ciertamente, 1111, ya que $\text{coste2}(1111)=0$. Si comparamos $*0*0$ con 1111 tenemos que

número	coste1	coste2
$*0*0$	0	mayor o igual que 2
1111	2	0

en consecuencia $*0*0 = 1111$.

Puesto que cualquier número distinto de 1111 tiene un coste2 superior a 0, 1111 será equiparable a (no será ‘peor’ que) cualquier número cuyo coste1 sea 0 o 1. En consecuencia 1111 es de primera clase.

Para que un número de la forma $*0*0$ (sólo hay cuatro de esa forma) sea descartable tiene que existir un número ‘mejor’ con coste1 igual a cero y coste2 menor. Ese número ‘mejor’, de hecho, es también de la forma $*0*0$ (única forma que asegura que coste1 sea cero). Por ejemplo, 0000 es ‘peor’ que 1010, puesto que $\text{coste2}(0000)=4$ y $\text{coste2}(1010)=2$. Se ve así que 1010 es el ‘mejor’ del conjunto $*0*0$, y permite descartar a sus (otros tres) compañeros, todos ‘peores’ que él.

Puesto que cualquier número de forma distinta a *0*0 tiene un *coste1* mayor que 0, 1010 será equiparable a (no será ‘peor’ que) cualquier número cuyo *coste2* sea 0 ó 1. En consecuencia 1010 es de primera clase.

número	<i>coste1</i>	<i>coste2</i>
1111	2	0
1010	0	2

¿Existe algún otro miembro de primera clase? Para cualquier candidato hay que demostrar que **no** existe ningún número ‘mejor’ que él (lo que permitiría descartarle). Números prometedores serían aquellos con ambos costes iguales a 1. Para que el *coste1* sea 1 hay que asegurar un 0 en una posición par: formas *0** ó ***0. Para que el *coste2* sea 1 hay que asegurar tres 111 en cualquier posición. Cumplir ambas condiciones simultáneamente nos lleva a los números 1011 y 1110 equiparables tanto a 1111 como a 1010.

número	<i>coste1</i>	<i>coste2</i>
1111	2	0
1011	1	1
1110	1	1
1010	0	2

¿Habrán más miembros de primera clase? Repasemos siguiendo una senda creciente para *coste2*: con *coste2* igual a 0 sólo está 1111; con *coste2* igual a 1 tenemos los números con tres unos, pero cualquiera distinto a 1011 ó a 1110 tendría un *coste1* superior a 1 y, por tanto, sería descartable; con *coste2* igual a 2 están los números con sólo dos 1 en cualquier posición, pero cualquiera que no sea de la forma *0*0 (es decir distinto a 1010) sería descartable por tener un *coste1* superior a 0. **1111, 1011, 1110 y 1010 forman la primera clase.**

16: Volviendo al ejemplo de los números binarios de cuatro cifras, podemos ver cómo son algunos descendientes. Cada ‘embarazo’ puede especificarse dando los progenitores y el punto de corte. Por simplicidad podemos bautizar a los miembros de la primera clase con letras, y referirnos con ellas a ellos.

nombre	número	<i>coste1</i>	<i>coste2</i>
A	1111	2	0
B	1011	1	1
C	1110	1	1
D	1010	0	2

Ahora un ‘embarazo’ puede especificarse como (B,C,2), lo que significa:

$$10|11 + 11|10 = 10|10$$

Las posiciones de corte 0 ó 4 significan que no hay de hecho recombinación: el descendiente hereda todo el ‘genoma’ del primer o del segundo progenitor, respectivamente.

En este ejemplo, la primera clase es una población ‘genéticamente estable’: **cualquier posible ‘embarazo’ da como resultado un miembro de primera clase.** Podemos decir que la ‘evolución’ ha concluido, al menos mientras las funciones de coste (que de un modo implícito representan aquello que sea el ‘ambiente’ y sus ‘presiones’ selectivas) no cambien.

Hay una terna de detalles que merece la pena subrayar por su carácter pedagógico:

- Hay poca ‘diversidad’: cualquiera que sea el ‘embarazo’ siempre habrá un 1 en las posiciones impares (ningún miembro tiene un cero en esas posiciones). Si, debido a una ‘mutación’, un 1 en una posición impar se trastocara en 0, el nuevo descendiente no sería de primera clase pues, precisamente en este ejemplo, para ser de ‘primera clase’ hay que tener un 1 en esas posiciones. A poco que se reflexione, **este sencillo ejemplo debería ayudar a guardar la idea de que «las ‘mutaciones’ son imprescindibles» en el «baúl de los recuerdos».**
- La diversidad que hay (la posibilidad de heredar un 0 o un 1 en las posiciones pares) es ‘neutra’ desde el punto de vista de la ‘selección’: para ser de ‘primera clase’ en este ejemplo, cualquier valor vale en esas posiciones, su ‘mutación’ sería igualmente ‘neutra’.
- La población es de primera clase y es estable genéticamente **pero** ambas características **son independientes**. Imaginemos que esta población **no** fuera de primera clase: generación tras generación seguiría estancada en ‘su clase’ (cualquiera que sea). Este ejemplo muestra por qué no cabe esperar ninguna demostración futura acerca de la convergencia segura de un GA a la solución óptima de un problema: de hecho, el algoritmo puede estancarse. (La mutación ‘al azar’ de baja intensidad se utiliza como recurso técnico para que tales programas de ordenador no se estancuen con demasiada frecuencia. Pero este problema técnico puede resolverse también de otras muchas maneras: con mutaciones ‘deterministas’ que reintroducen alelos desaparecidos donde convenga; o mediante esquemas de selección que, menos ambiciosos, no eligen los progenitores sólo entre los mejores, dando por el contrario oportunidad de reproducirse a miembros de las clases inferiores, lo que mantiene mejor la diversidad de la población. En todo caso, la ‘mejora’ de la población se confía a los esquemas de selección y reproducción empleados en cada versión concreta: **el ‘sexo’ y el emparejamiento son lo importante**).

Para mejor explorar las posibles consecuencias de distintos ‘embarazos’, cambiemos las funciones de coste. Y para definir las implícitamente, fijemos de antemano los componentes de la ‘primera clase’ de las soluciones del problema (véase [Kauffman, 1992] para el uso y justificación de tales definiciones implícitas). De hecho omitiré el valor de los costes de cada miembro de la clase, pues ahora sólo me interesa observar si un ‘embarazo’ particular acaba con el parto de un miembro de primera clase. Sea la primera clase del hipotético problema como sigue:

```
Primera clase
nombre número
A      0110
B      1011
C      1110
D      1000
```

Veamos el resultado de algunos ‘embarazos’:

```
‘embarazo’ número observaciones
(A,D,1) 0000 descartable
(D,B,2) 1011 B, primera clase
(C,D,3) 1110 C, primera clase
(C,D,2) 1100 descartable
```

Una conclusión resulta evidente: progenitores de primera clase tendrán descendientes que no lo serán. ¿Y al revés? Considérese el siguiente ‘embarazo’ de progenitores que **no son** de primera clase:

(1111,0000,3) -> 1110

Como se ve, concluye con el parto de C, miembro de primera clase.

De todo lo anterior, no se deducen *ninguna* esperanza de que una adecuada elección de pareja pueda mejorar una población sujeta a presiones selectivas multidimensionales. Y puesto que la reproducción y la selección están en el núcleo de la explicación hoy ortodoxa de la evolución biológica, resulta enteramente razonable comenzar a dudar de ella.

17: Desde que me topé con el primer problema de optimación multidimensional de interés práctico (el cobro de mis honorarios como calculista de estructuras siendo, además, ecologista) he analizado en diversas ocasiones y con desigual profundidad algunas de estas cuestiones, véase [Vázquez, 1997a], [Vázquez, 1998a], [Vázquez, 1999] y [Vázquez, 2000].

18: Los clones han aparecido en el texto de forma tan espontánea como necesaria. Los jueguecitos que nos traemos con la clonación y con la ingeniería genética **son peligrosos**, tal y como se deduce del texto (a poco que se medite sobre ello), pero no es éste el momento adecuado para ahondar en ello...

19: Programa que, por otra parte, puede describirse a su vez como una ristra de ceros y unos, pero interpretada en clave diferente. Y que podría estar sujeto a evolución 'genética' a su vez. Quién sabe si la enorme porción de material genético 'neutro' (de función prácticamente desconocida para la explicación 'oficial') constituye un conjunto de 'programas' que controlan la expresión de proteínas de la porción de DNA dedicada a codificarlas.

20: El 'genotipo' son los ceros y unos del ejemplo de números binarios de cuatro cifras. El 'fenotipo' es aquello de lo que se observa que es distinto del 'genotipo': con unas 'gafas' de costes serían, precisamente, los valores de ambos costes. Pero unas 'gafas' de números decimales (habituales en programación), en vez de '0010' verían un '2'. En el caso de los seres vivos la conexión entre genotipo y fenotipo no es en absoluto tan directa, por el contrario en este caso ni siquiera se ha demostrado que a un genotipo concreto le corresponda la expresión de un único fenotipo. La conformación espacial de las proteínas da mucho margen para que actúen sobre el fenotipo (el ser vivo) 'fuerzas' distintas a las puramente 'genómicas': el escandaloso caso de las 'vacas locas' europeas apunta en esa dirección (cf. [Prusiner, 1995] y [Ogayar y Sánchez-Pérez, 1998]).

21: Organulo compuesto de proteínas y RNA, encargado de iniciar la 'lectura' del genoma, del DNA, sin cuyo concurso este último sería como un libro sin persona que lo lea.

22: En la programación corriente de ordenadores, conforme la complejidad de un programa aumenta, así lo hace la *ligazón* entre el programa y los 'datos' que es capaz de leer. Los programas muy simples son capaces de leer casi cualquier tipo de datos sin 'caerse' (el programa 'copiar un archivo', por ejemplo). Por el contrario los programas complejos bien concebidos, enfrentados a 'datos' arbitrarios suelen acabar en seguida con mensajes como «error de sintaxis». (Los programas mal concebidos o que funcionan en un entorno adverso, como *Windows The Worst*, enfrentados a 'datos' arbitrarios suelen acabar con un «El

programa X ha efectuado una operación no permitida. Windows va a terminar la sesión. Pulse Ok o Ver Detalles.» ---si se pulsa 'Ver Detalles' el mensaje siguiente suele ser algo como «El programa X ha efectuado una operación no permitida. Windows va a terminar la sesión. Los datos no archivados van a perderse. Pulse Ok.».)

23: Si bien los he mantenido en las notas al pie, por conveniencia a la hora de discutir detalles técnicos.

24: Sobre el recocido simulado véase [Vázquez, 1995]. Para la comparación genérica entre SA y GA puede consultarse [Davis, 1987], y para comparación en casos concretos en el contexto del diseño de estructuras véase [Vázquez et Vázquez, 1997], [Vázquez et Vázquez, 1997a], [Vázquez, 1997] y [Vázquez, 1998]. Un hecho significativo es que, en el algoritmo SA, la regla para elegir entre una nueva solución (obtenida al azar) y la solución anterior es como sigue: «elíjase **siempre** la nueva cuando sea mejor que la anterior, y si es peor elíjase la nueva, de todas formas, de vez en cuando». Se trata por tanto de un rescate ocasional de lo peor. El algoritmo 'escalada', que siempre elige la mejor entre la nueva y la anterior, es incapaz de encontrar la solución óptima (salvo en algunos problemas sencillos, técnicamente 'bien condicionados'). La novedad del SA (aparte de un montón de jerga técnica, tan difícil de entender como interesante) es la aplicación algorítmica de refranes bien conocidos, tales como «lo mejor es enemigo de lo bueno» o «vísteme despacio que tengo prisa».

25: Bajo el principio de que lo artificial es parte de lo natural, todos los fenómenos evolutivos de la cultura humana son, en última instancia, fenómenos enmarcados en la evolución biológica general. Suelo fijarme, por ello, en procesos evolutivos artificiales a la búsqueda de sugerencias para entender los procesos evolutivos no-artificiales (frecuentemente más misteriosos). Así, por ejemplo, estoy escribiendo esto con Word Perfect, un programa de 1990, cuya supervivencia no cabe explicarse por una 'mejor' adaptación, pues en principio habría que considerar el 'mejor' a MSWord (dada su popularidad) funcionando en el contexto (ambiente) de *Windows The Worst*. ¿Cómo es que Word Perfect sigue 'vivo' en mi viejo ordenador? La visión unidimensional de la evolución nada puede explicar. Ahora [un par de días después], las últimas correcciones están siendo hechas en Madrid, también con Word Perfect, pero en ordenadores más modernos, en un contexto *Linux* de última generación (pero capaz de albergar a criaturas primitivas). Puede que en unas pocas décadas *Windows The Worst* haya desaparecido (y con él, MSWord), mientras que Word Perfect (ésta misma versión de 1990 en la que escribo) siga habitando en quien quiera que suceda a *Linux*. Son estos fenómenos, estas «contingencias históricas» las que una 'teoría de la evolución' (o al menos una 'teoría de la historia') debiera intentar explicar...

Boletín CF+S > 21 -- El pasado es un país extraño > <http://habitat.aq.upm.es/boletin/n21/amvaz.html>