What is “epistasis”?

H. Van Hove   A. Verschoren

Abstract

In this note, we compare several descriptions of “epistasis”, which have been used in the literature and apply some techniques originating from linear algebra, to present a formal approach, which encompasses all of the previous ones.

0 Introduction

Although several variants of the genetic algorithm (GA) have been introduced in the past, basically “the” GA always acts on a set Ω of binary strings, of length l say, and aims to maximize a real valued fitness function $f : \Omega \to \mathbb{R}$, through the use of a suitable set of genetic operators, acting on successive populations sampled within Ω.

Most models use crossover and mutation as their fundamental operators. Crossover basically constructs new strings by combining pieces of given ones, whereas mutation randomly changes bits of a given string (with a very low probability). The GA starts from a randomly chosen initial population $P(0) \subseteq \Omega$ (a multiset, in general). It then picks couples of strings within $P(0)$, with a probability of being selected proportional to their fitness and applies crossover and mutation to this pair, thus producing a new population $P(1)$. This process is then iterated until some initially given criterion is satisfied.

The GA is a very robust and widely applicable algorithm, due to the fact that it does not directly act on the crude data of a problem, but on the encoding of these data into binary strings. As a side effect, it should be clear that the efficiency of the GA is highly dependent on the chosen encoding scheme and, in particular, on the interdependency of different bits in the strings resulting from it.

Received by the editors September 1996.
Communicated by A. Hoogewijs.
1991 *Mathematics Subject Classification* : 68Q27, 68T05,15A09.
Key words and phrases : epistasis, generalized inverse.

1 Epistasis

The (in)dependence of bits in a string or a population of strings, is what one usually refers to as *epistasis*. The term epistasis has been used in several publications, without giving a formal definition, however. As we just pointed out, it usually seems to mean that the presence of combinations of certain allele values in particular bit positions has a strong influence on the fitness of the string in which this combination occurs.

In [3] Rawlins describes two extreme cases of epistasis. The first is *maximum epistasis*, where no proper subset of bits is independent of the other bits. This means that for every bit, it is possible to find one or more bits whose value influences the contribution of this bit to the fitness of a string. The second is *zero epistasis* where all bits are independent.

2 The fitness function

Let us fix some notations that will be used throughout this text. We will count the bits of $s \in \Omega$ from right to left and from 0 to $l-1$, i.e., $s = s_0 \dots s_{l-1}$. The fitness function will be a real valued function $f : \Omega \to \mathbb{R}$. If $A = \{0, 1\}$ is the set of allele values and $I = \{0, \dots, l-1\}$ the index set of the bits, we can rewrite the condition for zero epistasis found in [3] as :

$$\exists g : I \times A \to \mathbb{R}, \forall s \in \Omega : f(s) = \sum_{i \in I} g(i, s_i). \tag{1}$$

So, every reasonable definition of *epistasis* as a real valued function $\varepsilon : \Omega \to \mathbb{R}$ will have to satisfy the condition $\varepsilon(s) = 0$ for all s in Ω, whenever such a function g exists.

On the other hand, in [2], Davidor proposes a definition for the *epistasis of a string* with respect to a population $P \subseteq \Omega$: if $|P| = N$ and if $s = s_0 s_1 \dots s_{l-1}$, then the epistasis $\varepsilon_{P,f}(s)$ of s with respect to the population P is given by

$$\varepsilon_{P,f}(s) = f(s) - \sum_{i=0}^{l-1} \frac{1}{N_i(s_i)} \sum_{t \in P_{i,s_i}} f(t) + \frac{l-1}{N} \sum_{t \in P} f(t).$$

Here P_{i,s_i} is the set of all strings $t \in P$ which have value s_i in position i and $N_i(s_i)$ is the cardinality of P_{i,s_i}. If we take $N = 2^l$ (i.e., $P = \Omega$), we obtain

$$\varepsilon(s) = \varepsilon_{\Omega,f}(s) = f(s) - \sum_{i=0}^{l-1} \frac{1}{2^{l-1}} \sum_{t \in \Omega_{i,s_i}} f(t) + \frac{l-1}{2^l} \sum_{t \in \Omega} f(t). \tag{2}$$

We will call this the *(global) epistasis* of s (with respect to the fitness function f).

3 A matrix representation

We can rewrite the definition of $\varepsilon(s)$ by using matrices. Denoting the binary representation of a positive integer $i \in \mathbb{N}$ by $b(i)$, we define $d(i,j)$ as the Hamming distance $H(b(i), b(j))$, for every pair of positive integers i, j, i.e., $d_{i,j}$ will be the number of bits in which $b(i)$ and $b(j)$ differ. If we put

$$\mathbf{e} = \begin{pmatrix} \varepsilon(00\ldots00) \\ \varepsilon(00\ldots01) \\ \vdots \\ \varepsilon(11\ldots11) \end{pmatrix}$$

and

$$\mathbf{f} = \begin{pmatrix} f(00\ldots00) \\ f(00\ldots01) \\ \vdots \\ f(11\ldots11) \end{pmatrix}$$

then clearly

$$\mathbf{e} = \mathbf{f} - \mathbf{E}\mathbf{f}, \tag{3}$$

where $\mathbf{E} = (e_{i,j}) \in \mathbb{R}^{2^l} \times \mathbb{R}^{2^l}$ with $e_{i,j} = 2^{-l}(l + 1 - 2d_{i,j})$.
If we define the (global) epistasis $\varepsilon(f)$ of f as $\sqrt{\sum_{s\in\Omega} \varepsilon(s)^2}$, then $\varepsilon(f) = ||\mathbf{e}||$.
Let us write $g : I \times A \to \mathbb{R}$ as a column vector

$$\mathbf{g} = \begin{pmatrix} g_{(l-1)1} \\ g_{(l-1)0} \\ \vdots \\ g_{11} \\ g_{10} \\ g_{01} \\ g_{00} \end{pmatrix}$$

where $g_{ia} = g(i,a)$, for every $(i,a) \in I \times A$.
Then we can rewrite (1) as:

$$\exists \mathbf{g} \in \mathbb{R}^{2l} : \mathbf{A}\,\mathbf{g} = \mathbf{f}, \tag{4}$$

where $\mathbf{A} = (a_{i,j}) \in \mathbb{R}^{2^l} \times \mathbb{R}^{2l}$ is defined as follows: if we encode a 0 as 01 and a 1 as 10, then the i-th row of $\mathbf{A}$ will be the encoded version of the binary number $i - 1$. For example, if $l = 3$, then

$$\mathbf{A} = \begin{pmatrix} 0\,1\,0\,1\,0\,1 \\ 0\,1\,0\,1\,1\,0 \\ 0\,1\,1\,0\,0\,1 \\ 0\,1\,1\,0\,1\,0 \\ 1\,0\,0\,1\,0\,1 \\ 1\,0\,0\,1\,1\,0 \\ 1\,0\,1\,0\,0\,1 \\ 1\,0\,1\,0\,1\,0 \end{pmatrix}$$

Alternatively, $\mathbf{A} = (a_{i,j})$ may be defined by putting for any $1 \le i \le 2^l, 1 \le j \le 2l$

$$a_{i,j} = \begin{cases} 1 - (((i-1)\mathrm{div}^{\lceil \frac{2l-j+1}{2} \rceil - 1} 2) \bmod 2) & \text{if j is even} \\ ((i-1)\mathrm{div}^{\lceil \frac{2l-j+1}{2} \rceil - 1} 2) \bmod 2 & \text{if j is odd} \end{cases}$$

Here, for any $x \in \mathbb{R}$, we let $\lceil x \rceil$ denote the smallest integer n with $n \ge x$ and div denotes integer division. Moreover div^k is inductively defined by:

$$\begin{cases} n\,\mathrm{div}^0 m = n \\ n\,\mathrm{div}^1 m = n\,\mathrm{div}\, m \\ n\,\mathrm{div}^k m = (n\,\mathrm{div}^{k-1} m)\mathrm{div}\, m \end{cases}$$

4 Generalized inverses

Recall (from [1], e.g.) that the *Moore-Penrose* (or *generalized*) *inverse* $\mathbf{A}^+$ of an arbitrary $p \times q$ matrix $\mathbf{A}$ is the (unique!) $q \times p$ matrix $\mathbf{X}$ with

$$\mathbf{AXA} = \mathbf{A} \tag{5}$$
$$\mathbf{XAX} = \mathbf{X} \tag{6}$$
$$(\mathbf{AX})^T = \mathbf{AX} \tag{7}$$
$$(\mathbf{XA})^T = \mathbf{XA} \tag{8}$$

It is easy to check that for any invertible matrix $\mathbf{A}$, we have $\mathbf{A}^+ = \mathbf{A}^{-1}$. Moreover, it is well known that a linear system $\mathbf{f} = \mathbf{Ag}$ has $\mathbf{A}^+\mathbf{f}$ as a solution, whenever solutions exist.
With this observation in mind, it then is clear that (4) is equivalent to

$$\mathbf{f} - \mathbf{AA}^+\mathbf{f} = \mathbf{0}. \tag{9}$$

Comparing (3) and (9), it thus seems natural to link $\mathbf{E}$ and $\mathbf{A}^+$. Actually we will see below that $\mathbf{E} = \mathbf{AA}^+$.
As a consequence, let us point out that this yields a straightforward proof of the main result of [4]:

6.Proposition The following assertions are equivalent:

(i) $\forall s \in \Omega : \varepsilon(s) = 0$;

(ii) $\exists g : I \times A \to \mathbb{R}, \forall s \in \Omega : f(s) = \sum_{i \in I} g(i, s_i)$.

Proof Since (i) is equivalent to $\mathbf{f} - \mathbf{Ef} = \mathbf{0}$, and (ii) is equivalent to $\mathbf{f} - \mathbf{AA}^+\mathbf{f} = \mathbf{0}$, this follows immediately from the equality $\mathbf{E} = \mathbf{AA}^+$. ■

In order to prove our main theorem, we will need the following results:

7. Lemma With notations as before, the matrix $\mathbf{E}$ is idempotent.

Proof Denote by $e^{(2)}_{i,j}$ the (i,j)-th element of $\mathbf{E}^2$, i.e.,

$$e^{(2)}_{i,j} = \sum_{k=0}^{2^l-1} \frac{l+1-2d_{i,k}}{2^l}\frac{l+1-2d_{k,j}}{2^l} \tag{10}$$

Expanding (10) and using

$$\sum_{i=0}^{2^l-1} d_{i,j} = \sum_{p=0}^{l} p\binom{p}{l} = l2^{l-1}$$

(which one easily verifies), we obtain:

$$e^{(2)}_{i,j} = \frac{1-l^2}{2^l} + 4\sum_{k=0}^{2^l-1} d_{i,k}d_{k,j}.$$

So, in order to prove our assertion, i.e., $e^{(2)}_{i,j}) = (e_{i,j}$, one thus easily reduces to verifying that

$$\sum_{k=0}^{2^l-1} d_{i,k}d_{k,j} = (l^2+l-2d_{i,j})2^{l-2}.$$

Let us introduce the following notations:

- $d(i,j|k)$ = the number of bits in i, which coincide with the corresponding ones in j, but not in k.
- $d(i|j,k)$ = the number of bits in j, which coincide with the corresponding ones in k, but not in i.

For example, if $i = 101011$, $j = 010111$ and $k = 100010$, then $d(i,j|k) = d(i|j,k) = 1$.
We then obtain

$$\sum_{k=0}^{2^l-1} d_{i,k}d_{k,j} = \sum_{k=0}^{2^l-1}(d(i,j|k)+d(i|j,k))(d(i,j|k)+d(j|i,k))$$

Now

$$\sum_{k=0}^{2^l-1} d(i,j|k)^2 = \sum_{m=0}^{l-d_{i,j}} m^2\binom{l-d_{i,j}}{m}2^{d_{i,j}} \tag{11}$$

$$= 2^{d_{i,j}}(l-d_{i,j})(l-d_{i,j}+1)2^{l-d_{i,j}-2} \tag{12}$$

$$\sum_{k=0}^{2^l-1} d(i,j|k)d(j|i,k) = \sum_{m=0}^{l-d_{i,j}}\sum_{n=0}^{d_{i,j}} m\binom{l-d_{ij}}{m}\binom{d_{i,j}}{n} \tag{13}$$

$$= ld_{i,j}2^{l-2} - d_{i,j}^2 2^{l-2} \tag{14}$$

and

$$\sum_{k=0}^{2^l-1} d(i|k,j)d(j|i,k) = \sum_{m=0}^{d_{i,j}}\binom{d_{i,j}}{m} m(d_{ij}-m)2^{l-d_{ij}} \tag{15}$$

$$= d_{i,j}^2 2^{l-1} - d_{i,j}^2 2^{l-2} - d_{i,j}2^{l-2} \tag{16}$$

So, it is clear that

$$\sum_{k=0}^{2^l-1} d_{i,k}d_{k,j} = (l^2 + l - 2d_{i,j})2^{l-2},$$

indeed, which finishes the proof. ∎

As a first consequence of the fact that $\mathbf{E}$ is idempotent (and symmetric), let us point out that

$$\varepsilon(f) = \sqrt{\mathbf{e}^T\mathbf{e}} = \sqrt{(\mathbf{f}-\mathbf{Ef})^T(\mathbf{f}-\mathbf{Ef})} = \sqrt{\mathbf{f}^T(\mathbf{I}-\mathbf{E})\mathbf{f}} = ||\mathbf{f}||_{\mathbf{I}-\mathbf{E}}$$

For any matrix $\mathbf{B}$, we denote by $\mathcal{R}(\mathbf{B})$ the range of $\mathbf{B}$, i.e., the vectorspace spanned by is columns.

8. Lemma With notations as before, $\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{E})$

Proof Let us first recall that, if we apply one of the following transformations to the columns of a matrix, the resulting matrix will have the same range as the original one:

- $(\mathbf{c}_1, \ldots, \mathbf{c}_n) \to (\mathbf{c}_{\sigma(1)}, \ldots, \mathbf{c}_{\sigma(n)})$ $(\sigma \in \mathcal{S}_n)$
- $(\mathbf{c}_1, \ldots, \mathbf{c}_n) \to (\alpha\mathbf{c}_1, \ldots, \alpha\mathbf{c}_n)$ $\alpha \in \mathbb{R}$
- $(\mathbf{c}_1, \ldots, \mathbf{c}_n) \to (\mathbf{c}_1 + \sum_{i=2}^n \alpha_i\mathbf{c}_i, \ldots, \mathbf{c}_n)$
- $(\mathbf{c}_1, \mathbf{c}_1, \ldots, \mathbf{c}_n) \to (\mathbf{c}_1, \ldots, \mathbf{c}_n)$

Let us now transform $\mathbf{E}$ using these rules.
We number the columns of $\mathbf{E}$ from left to right and from 0 to $2^l - 1$. To simplify calculations, we first multiply every column of $\mathbf{E}$ by 2^l.
(**i**) If for $0 \le i \le 2^{l-1} - 1$, we add to the i-th column $\mathbf{e}_i$ of $\mathbf{E}$ the column $\mathbf{e}_{\hat{i}}$ with $\hat{i} = (2^l - 1) - i$, the resulting column will only contain entries equal to $2(l+1) - 2l = 2$. Dividing these columns by 2 and removing all duplicates, except one, we thus obtain a new matrix $\mathbf{E}'$ of the form

$$\mathbf{E}' = (\mathbf{1}, (l+1)\mathbf{1} - 2\mathbf{d}_{2^{l-1}}, \ldots, (l+1)\mathbf{1} - 2\mathbf{d}_{2^l-1}),$$

where $\mathbf{1}$ is the column all of whose entries are equal to 1 and where $\mathbf{d}_i$ is the column

$$\mathbf{d}_i = \begin{pmatrix} d_{0,i} \\ \vdots \\ d_{2^l-1,i} \end{pmatrix}$$

Adding $-(l+1)\mathbf{1}$ to each of the columns (except the first one) and multiplying them by $-\frac{1}{2}$, we obtain a new matrix $\mathbf{E}_1$ with $\mathcal{R}(\mathbf{E}) = \mathcal{R}(\mathbf{E}_1)$ and given by

$$\mathbf{E}_1 = (\mathbf{1}, \mathbf{d}_{2^{l-1}}, \ldots, \mathbf{d}_{2^l-1})$$

(**ii**) If for $1 \le i \le 2^{l-2}$ we add $\mathbf{d}_{2^l-i}$ to $\mathbf{d}_{2^{l-1}+i-1}$, the resulting column will always be

$$\begin{pmatrix} l \\ \vdots \\ l \\ l-1 \\ \vdots \\ l-1 \end{pmatrix}$$

This is obvious, if we realize that except for the first bit, $2^l - 1$ and $2^{l-1} + i - 1$ are mutual bitwise complements for $1 \le i \le 2^{l-2}$. Adding $(l-1)\mathbf{1}$ to these columns and again removing identical ones, we obtain a new matrix $\mathbf{E}_2$, with $\mathcal{R}(\mathbf{E}_2) = \mathcal{R}(\mathbf{E})$ and given by

$$\mathbf{E} = \begin{pmatrix} 1\,1 & d_{0,2^{l-1}+2^{l-2}} & \dots & d_{0,2^l-1} \\ \vdots\,\vdots & \vdots & \ddots & \vdots \\ 1\,1 & d_{2^{l-1}-1,2^{l-1}+2^{l-2}} & \dots & d_{2^{l-1}-1,2^l-1} \\ 1\,0 & d_{2^{l-1},2^{l-1}+2^{l-2}} & \dots & d_{2^{l-1},2^l-1} \\ \vdots\,\vdots & \vdots & \ddots & \vdots \\ 1\,0 & d_{2^l-1,2^{l-1}+2^{l-2}} & \dots & d_{2^l-1,2^l-1} \end{pmatrix}$$

(**iii**) Repeating this procedure, it is now easy to see that we eventually obtain a matrix $\mathbf{M} = \mathbf{E}_l$, whose first column is $\mathbf{1}$ and whose other entries $m_{i,j}$ are just the number of zeros in the first j bits of i, and which still has the property that $\mathcal{R}(\mathbf{M}) = \mathcal{R}(\mathbf{E})$.
For example, if $l = 3$, then

$$\mathbf{M} = \begin{pmatrix} 1\,1\,2\,3 \\ 1\,1\,2\,2 \\ 1\,1\,1\,2 \\ 1\,1\,1\,1 \\ 1\,0\,1\,2 \\ 1\,0\,1\,1 \\ 1\,0\,0\,1 \\ 1\,0\,0\,0 \end{pmatrix}$$

To finish the proof, let us verify that $\mathcal{R}(\mathbf{M}) = \mathcal{R}(\mathbf{A})$, with $\mathbf{A} = (\mathbf{a}_0, \dots, \mathbf{a}_{2l-1})$ and $\mathbf{M} = (\mathbf{m}_0, \dots, \mathbf{m}_l)$. Recall from [4] that the set

$$\mathcal{B} = \{\mathbf{a}_{2i}; 0 \le i \le l-1\} \cup \{\mathbf{a}_1\}$$

is a basis for $\mathcal{R}(\mathbf{A})$.

For example, if $l = 3$, then $\mathcal{B}$ consists of the columns

$$\mathbf{a}_0 = \begin{pmatrix} 0\\0\\0\\0\\1\\1\\1\\1 \end{pmatrix} \mathbf{a}_1 = \begin{pmatrix} 1\\1\\1\\1\\0\\0\\0\\0 \end{pmatrix} \mathbf{a}_2 = \begin{pmatrix} 0\\0\\1\\1\\0\\0\\1\\1 \end{pmatrix} \mathbf{a}_4 = \begin{pmatrix} 0\\1\\0\\1\\0\\1\\0\\1 \end{pmatrix}$$

Let us create a equivalent basis $\mathcal{B}'$, by swapping the first two columns in $\mathcal{B}$ and adding the second column to the first one. The resulting basis now has $\mathbf{1}$ as a first column. The corresponding matrix will be denoted by $\mathbf{B} = (b_{i,j}) = (\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_l)$ and has the property that $\mathcal{R}(\mathbf{B}) = \mathcal{R}(\mathbf{A})$ For example, again with $l = 3$, we have

$$\mathbf{B} = \begin{pmatrix} 1\,0\,0\,0\\1\,0\,0\,1\\1\,0\,1\,0\\1\,0\,1\,1\\1\,1\,0\,0\\1\,1\,0\,1\\1\,1\,1\,0\\1\,1\,1\,1 \end{pmatrix}$$

It is clear that the rows of the submatrix $(\mathbf{b}_1, \ldots, \mathbf{b}_l)$ are just the binary encodings of the row number – this is essentially just the definition of $\mathbf{A}$.
It now follows that $\mathbf{m}_0 = \mathbf{b}_0$, $\mathbf{m}_1 = \mathbf{1} - \mathbf{b}_1$ and

$$\mathbf{m}_j = j\mathbf{1} - \sum_{k=0}^{j} \mathbf{b}_k$$

resp.,

$$\mathbf{b}_j = \mathbf{1} + \mathbf{m}_{j-1} - \mathbf{m}_j,$$

for $j \geq 2$ which shows that $\mathcal{R}(\mathbf{E}) = \mathcal{R}(\mathbf{M}) = \mathcal{R}(\mathbf{B}) = \mathcal{R}(\mathbf{A})$. ■

We may now finally prove:

9. Theorem With the above notations, we have: $\mathbf{E} = \mathbf{A}\mathbf{A}^+$.

Proof It is well known that $\mathbf{A}\mathbf{A}^+$ is the orthogonal projection on the range $\mathcal{R}(\mathbf{A}) \subseteq \mathbb{R}^{2l}$ of $\mathbf{A}$, and that a linear map is an orthogonal projection if and only if its corresponding matrix is idempotent and symmetric. Since the orthogonal projection on subspaces is unique, we thus have to prove that $\mathbf{E}^2 = \mathbf{E}$ and that $\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{E})$ – and this is just the content of the previous lemmas. ■

References

[1] A. Ben-Israel and T. Greville, Generalized Inverses: Theory and Applications, John Wiley & Sons,1974

[2] Y. Davidor, *Epistasis Variance: A viewpoint on GA-Hardness*, in: Foundations of Genetic Algorithms, ed. G.J.E. Rawlins, Morgan Kaufmann Publishers, 1991.

[3] G.J.E. Rawlins, Foundations of Genetic Algorithms, Morgan Kaufmann Publishers, 1991.

[4] H. Van Hove and A. Verschoren, *On Epistasis*, Computers and AI. (submitted).

University of Antwerp, RUCA
Department of Mathematics and Computer Science
Groenenborgerlaan 171
B-2020 Antwerp, Belgium
aver@ruca.ua.ac.be