Algebraic structures emerging from genetic inheritance

Quentin Ehret

Séminaire doctorants IRIMAS

January 16, 2023

IRIMAS UNIVERSITÉ HAUTE-ALSACE
Institut de Recherche en Informatique, Mathématiques, Automatique et Signal
Département de Mathématiques

Introduction

Main reference of this talk: Mary Lynn Reed, *Algebraic Structure of Genetic Inheritance*, Bull. Am. Math. Soc., 34 (2), 1997.

Contents

Genetic background

- **gene:** unit of hereditary information (ex: blood type gene).

Genetic background

- **gene:** unit of hereditary information (ex: blood type gene).
- **allele:** different forms of a gene (ex: blood type A, B, O).

Genetic background

- **gene:** unit of hereditary information (ex: blood type gene).
- **allele:** different forms of a gene (ex: blood type A, B, O).
- **chromosome:** DNA molecule with part (or all) of the genetic material of an organism.

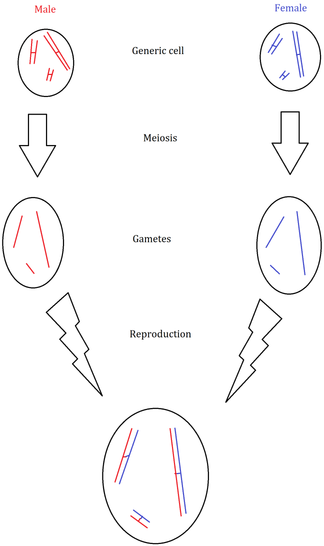
Genetic background

- **gene:** unit of hereditary information (ex: blood type gene).
- **allele:** different forms of a gene (ex: blood type A, B, O).
- **chromosome:** DNA molecule with part (or all) of the genetic material of an organism.
- humans are **diploid:** double set of chromosomes (one of each parent).

Genetic background

- **gene:** unit of hereditary information (ex: blood type gene).
- **allele:** different forms of a gene (ex: blood type A, B, O).
- **chromosome:** DNA molecule with part (or all) of the genetic material of an organism.
- humans are **diploid:** double set of chromosomes (one of each parent).
- **reproduction:**
  1. meiosis produces sex cells (gametes) carrying a single set of chromosomes;
  2. male and female gametes fuse $\rightsquigarrow$ produce new cells with double set of chromosomes.

Genetic background



Gametic algebras

- **Genotype:** alleles carried by chromosomes;
  Phenotype: alleles that express.

Gametic algebras

- **Genotype:** alleles carried by chromosomes;
  Phenotype: alleles that express.
- Here: single gene with two alleles A and B.

Gametic algebras

- **Genotype:** alleles carried by chromosomes;
  Phenotype: alleles that express.
- Here: single gene with two alleles A and B.
- Throught the process of reproduction, 3 possible genotypes: AA, BB (homozygotes) and AB (heterozygotes).

Gametic algebras

- **Genotype:** alleles carried by chromosomes;
  Phenotype: alleles that express.
- Here: single gene with two alleles A and B.
- Throught the process of reproduction, 3 possible genotypes: AA, BB (homozygotes) and AB (heterozygotes).
- What happens?

$\curvearrowright$	**A**	**B**
A	A	$\frac{1}{2}A + \frac{1}{2}B$
B	$\frac{1}{2}A + \frac{1}{2}B$	B

Gametic algebras

- **Genotype:** alleles carried by chromosomes;
  Phenotype: alleles that express.
- Here: single gene with two alleles A and B.
- Throught the process of reproduction, 3 possible genotypes: AA, BB (homozygotes) and AB (heterozygotes).
- What happens?

$\curvearrowright$	**A**	**B**
A	A	$\frac{1}{2}A + \frac{1}{2}B$
B	$\frac{1}{2}A + \frac{1}{2}B$	B

$\rightsquigarrow$ we have defined the **gametic algebra** on the basis $\{A, B\}$ with the above multiplication table.

Gametic algebras

- **Genotype:** alleles carried by chromosomes;
  Phenotype: alleles that express.
- Here: single gene with two alleles A and B.
- Throught the process of reproduction, 3 possible genotypes: AA, BB (homozygotes) and AB (heterozygotes).
- What happens?

$\curvearrowright$	**A**	**B**
A	A	$\frac{1}{2}A + \frac{1}{2}B$
B	$\frac{1}{2}A + \frac{1}{2}B$	B

$\rightsquigarrow$ we have defined the **gametic algebra** on the basis $\{A, B\}$ with the above multiplication table.

$\rightsquigarrow$ not associative: $A(AB) = \frac{3}{4}A + \frac{1}{4}B \neq (AA)B = \frac{1}{2}A + \frac{1}{2}B$.

Zygotic algebras

- For humans (for example), it is more complicated:

$$\text{cell with alleles AB} \xrightarrow{\textit{meiosis}} \begin{cases} \text{gamete carrying A with proba 0.5} \\ \text{gamete carrying B with proba 0.5.} \end{cases}$$

Zygotic algebras

- For humans (for example), it is more complicated:

$$\text{cell with alleles AB} \xrightarrow{\textit{meiosis}} \begin{cases} \text{gamete carrying A with proba 0.5} \\ \text{gamete carrying B with proba 0.5.} \end{cases}$$

- So in that case, AB shall be understood as $\frac{1}{2}A + \frac{1}{2}B$.

Zygotic algebras

- For humans (for example), it is more complicated:

$$\text{cell with alleles AB} \xrightarrow{meiosis} \begin{cases} \text{gamete carrying A with proba 0.5} \\ \text{gamete carrying B with proba 0.5.} \end{cases}$$

- So in that case, AB shall be understood as $\frac{1}{2}A + \frac{1}{2}B$.
- So $(AB)(AB) = \frac{1}{4}AA + \frac{1}{2}AB + \frac{1}{4}BB$.

Zygotic algebras

We therefore obtain an algebra on the basis $\{AA, AB, BB\}$ with multiplication given by

$\curvearrowright$	**AA**	**AB**	**BB**
AA	AA	$\frac{1}{2}(AA+AB)$	AB
AB	$\frac{1}{2}(AA+AB)$	$\frac{1}{4}AA+\frac{1}{2}AB+\frac{1}{4}BB$	$\frac{1}{2}(AB+BB)$
BB	AB	$\frac{1}{2}(AB+BB)$	BB

$\leadsto$ it is called the **Zygotic algebra**.

Non-associative algebras

Let $(V, +, \cdot)$ be a (finite dimensional) vector space over a field $\mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ for example).

Definition

Suppose that V is endowed with a bilinear map $* : V \times V \to V$, distributive with respect to $+$ (multiplication). Then $(V, +, \cdot, *)$ is called a **(non-associative) algebra***.*

Non-associative algebras

Let $(V,+,\cdot)$ be a (finite dimensional) vector space over a field $\mathbb{K}$ ($\mathbb{K}=\mathbb{R}$ for example).

Definition

Suppose that V is endowed with a bilinear map $*: V\times V\to V$, distributive with respect to $+$ (multiplication). Then $(V,+,\cdot,*)$ is called a **(non-associative) algebra**.

Suppose that $\{e_1,\cdots,e_n\}$ is a basis of V as $\mathbb{K}$-vector space:

$$\forall v\in V,\quad \exists\,(\lambda_i)_i\in\mathbb{K},\quad v=\sum_{i=1}^{n}\lambda_i e_i.$$

$\rightsquigarrow$ it is enough to define the multiplication of the basis of V.

Non-associative algebras

$$e_i * e_j = \sum_{k=1}^{n} C_{i,j}^{k} e_k \,, \quad C_{i,j}^{k} \in \mathbb{K} \,.$$

The multiplication is entirely determined by those n^3 structure constants.

Non-associative algebras

$$e_i * e_j = \sum_{k=1}^{n} C_{i,j}^k e_k , \quad C_{i,j}^k \in \mathbb{K} .$$

The multiplication is entirely determined by those n^3 structure constants.

Example: $V =< e_1, e_2 >$ with multiplication

$\curvearrowright$	$\mathbf{e_1}$	$\mathbf{e_2}$
$\mathbf{e_1}$	e_1	e_2
$\mathbf{e_2}$	e_2	e_2

$$C_{1,1}^1 = 1,\ C_{1,1}^2 = 0,\ C_{1,2}^1 = C_{2,1}^1 = 0,\ C_{1,2}^2 = C_{2,1}^2 = 1,\ C_{2,2}^1 = 0,\ C_{2,2}^2 = 1.$$

Associative algebras

Definition

Let $(V, *)$ *be a non-associative algebra. It is* **associative** *if* $*$ *satisfies*

$$a * (b * c) = (a * b) * c, \quad \forall a, b, c \in V.$$

Examples: $(\mathbb{R}, \times)$, $(M_n(\mathbb{R}), \text{matrix product})$, ...

Associative algebras

Definition

Let $(V,*)$ be a non-associative algebra. It is **associative** *if $*$ satisfies*

$$a*(b*c) = (a*b)*c, \quad \forall a,b,c \in V.$$

Examples: $(\mathbb{R},\times)$, $(M_n(\mathbb{R}), \text{matrix product}), \ldots$

Proposition

$(V,*)$ is associative if and only if its structure constants satisfy

$$\sum_{l=1}^{n} \left(C_{j,k}^{l} C_{i,l}^{p} - C_{i,j}^{l} C_{l,k}^{p} \right) = 0, \quad \forall\ 1 \leq i,j,k,p \leq n.$$

Commutative algebras

Definition

Let $(V, *)$ *be an algebra. It is* **commutative** *if* $*$ *satisfies*

$$a * b = b * a, \quad \forall a, b \in V.$$

Example: $(\mathbb{R}, \times)$; **Counterexample:** $(M_n(\mathbb{R}), \text{matrix product})$

Commutative algebras

Definition

Let $(V,*)$ *be an algebra. It is* **commutative** *if* $*$ *satisfies*

$$a*b=b*a,\quad \forall a,b\in V.$$

Example: $(\mathbb{R},\times)$; **Counterexample:** $(M_n(\mathbb{R}),\text{matrix product})$

Proposition

$(V,*)$ *is commutative if and only if its structure constants satisfy*

$$C_{i,j}^{k}-C_{j,i}^{k}=0,\quad \forall\ 1\le i,j,k\le n.$$

Lie algebras

Definition

Let $(V, *)$ *be an algebra. It is called* **Lie algebra** *if* $*$ *satisfies*

$$a * b = -b * a, \quad \forall a, b \in A. \tag{1}$$

$$0 = a * (b * c) + b * (c * a) + c * (a * b), \quad \forall a, b, c \in A. \tag{2}$$

Examples: $(V, * \equiv 0)$; $(M_n(\mathbb{R}), U * V = UV - VU)$

Jordan algebras

Definition

Let $(V, *)$ be an algebra. It is called **Jordan algebra** *it is commutative and if $*$ satisfies*

$$(a * b) * (a * a) = a * (b * (a * a)), \quad \forall a, b \in A.$$

Example: V associative $\Rightarrow (V, a * b = \frac{ab+ba}{2})$ is a Jordan algebra.

General gametic algebras

In many situations, the frequencies are not 0.5, but other recombination rules appear. We consider a population with n distincts alleles $(a_1, \cdots, a_n)$ of a given gene.

Definition

Take $\mathfrak{g} =< a_1, \cdots, a_n >$ *the (free) vector space on n generators.*

Consider the multiplication $a_i * a_j = \sum_{k=1}^{n} \gamma_{i,j}^{k} a_k$, *satisfying*

$$0 \leq \gamma_{i,j}^{k} \leq 1 \tag{3}$$

$$\sum_{k=1}^{n} \gamma_{i,j}^{k} = 1 \tag{4}$$

$$\gamma_{i,j}^{k} = \gamma_{j,i}^{k}. \tag{5}$$

Then $(\mathfrak{g}, *)$ *is called the* **general gametic algebra**.

General zygotic algebras

Denote $a_{ij} = a_i a_j$.

Definition

Take $\mathfrak{z} =< a_{ij} >_{i \leq j}$. *Consider the multiplication*

$a_{ij} * a_{pq} = \sum_{s=1}^{n} \sum_{k=1}^{s} \zeta_{(ij),(pq)}^{k,s} a_{ks}$, *satisfying*

$$0 \leq \zeta_{(ij),(pq)}^{k,s} \leq 1 \qquad (6)$$

$$\sum_{k,s=1}^{n} \zeta_{(ij),(pq)}^{k,s} = 1, \qquad i \leq j,\ p \leq q,\ k \leq s; \qquad (7)$$

$$\zeta_{(ij),(pq)}^{k,s} = \zeta_{(pq),(ij)}^{k,s}. \qquad (8)$$

Then $(\mathfrak{z}, *)$ *is called the* **general zygotic algebra**.

Links between the structures

Proposition

Consider the gametic algebra $\mathfrak{g}$ given by its structure constants $\gamma_{i,j}^{k}$. Define a new algebra $\mathfrak{z}$ with the following structure constants:

$$\zeta_{(ij),(pq)}^{k,s} = \begin{cases} \gamma_{i,j}^{k}\gamma_{p,q}^{s} + \gamma_{i,j}^{s}\gamma_{p,q}^{k}, & \text{if } k < s; \\ \gamma_{i,j}^{k}\gamma_{p,q}^{s} & \text{if } k = s. \end{cases} \quad (9)$$

Then, $\mathfrak{z}$ is a zygotic algebra.

Links between the structures

Proposition

Consider the gametic algebra $\mathfrak{g}$ given by its structure constants $\gamma_{i,j}^{k}$. Define a new algebra $\mathfrak{z}$ with the following structure constants:

$$\zeta_{(ij),(pq)}^{k,s} = \begin{cases} \gamma_{i,j}^{k}\gamma_{p,q}^{s} + \gamma_{i,j}^{s}\gamma_{p,q}^{k}, & \text{if } k < s; \\ \gamma_{i,j}^{k}\gamma_{p,q}^{s} & \text{if } k = s. \end{cases} \tag{9}$$

Then, $\mathfrak{z}$ is a zygotic algebra.

Those identities come from a construction called **commutative duplication:**

$$\mathfrak{z} = \frac{\mathfrak{g} \otimes \mathfrak{g}}{I}, \quad I = < x \otimes y - y \otimes x > .$$

It is a commutative algebra with multiplication

$$(a \otimes b) * (c \otimes d) = (ab \otimes cd).$$

Application: self-fertilization

- For a given population, we consider a gene having 2 alleles A, B and following the zygotic algebra rule of inheritance.

Application: self-fertilization

- For a given population, we consider a gene having 2 alleles A, B and following the zygotic algebra rule of inheritance.
- We have three possible genotypes: AA, AB, BB.

Application: self-fertilization

- For a given population, we consider a gene having 2 alleles A, B and following the zygotic algebra rule of inheritance.
- We have three possible genotypes: $AA,\ AB,\ BB$.
- Suppose that the first generation have a distribution

$$F_0 = \lambda AA + \mu AB + \epsilon BB, \quad \lambda, \mu, \epsilon \in \mathbb{R}\,.$$

$\rightsquigarrow$ what will be the state of the population after n steps of self-fertilization?

Application: self-fertilization

Let's compute the first step F_1.

$$F_1 = \lambda(AA * AA) + \mu(AB * AB) + \epsilon(BB * BB)$$

$$= \lambda AA + \mu\left(\frac{1}{4}AA + \frac{1}{2}AB + \frac{1}{4}BB\right) + \epsilon BB$$

$$= \left(\lambda + \frac{1}{4}\mu\right) AA + \frac{\mu}{2}AB + \left(\epsilon + \frac{1}{4}\mu\right) BB.$$

Application: self-fertilization

Let's introduce a sequence (u_n):

$$
\begin{aligned}
u_0 &= F_0 \\
u_1 &= F_1 - F_0 = \frac{1}{2}\mu\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right) \\
u_2 &= F_2 - F_1 = \frac{1}{4}\mu\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right) \\
&\vdots \\
u_n &= F_n - F_{n-1} = \frac{1}{2^n}\mu\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right).
\end{aligned}
$$

Application: self-fertilization

$$u_n = F_n - F_{n-1} = \frac{1}{2^n}\mu\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right).$$

Therefore we have

$$\sum_{i=1}^{n} u_i = \left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right)\left(\frac{1}{2} + \frac{1}{4} + \cdots \frac{1}{2^n}\right)\mu$$
$$= \mu\left(1 - \frac{1}{2^n}\right)\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right).$$

Application: self-fertilization

Then,

$$F_n = (F_n - F_{n-1}) + (F_{n-1} - F_{n-2}) + \cdots + (F_1 - F_0) + F_0.$$

Application: self-fertilization

Then,

$$F_n = (F_n - F_{n-1}) + (F_{n-1} - F_{n-2}) + \cdots + (F_1 - F_0) + F_0.$$
$$= u_n + u_{n-1} + \cdots + u_1 + F_0$$

Application: self-fertilization

Then,

$$
\begin{aligned}
F_n &= (F_n - F_{n-1}) + (F_{n-1} - F_{n-2}) + \cdots + (F_1 - F_0) + F_0. \\
&= u_n + u_{n-1} + \cdots + u_1 + F_0 \\
&= \frac{1}{2^n}\mu\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right) + \lambda AA + \mu AB + \epsilon BB
\end{aligned}
$$

Application: self-fertilization

Then,

$$\begin{aligned} F_n &= (F_n - F_{n-1}) + (F_{n-1} - F_{n-2}) + \cdots + (F_1 - F_0) + F_0. \\ &= u_n + u_{n-1} + \cdots + u_1 + F_0 \\ &= \frac{1}{2^n}\mu\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right) + \lambda AA + \mu AB + \epsilon BB \\ &= \left(\lambda + \frac{1}{2}\mu - \frac{\mu}{2^{n+1}}\right) AA + \frac{\mu}{2^n} AB + \left(\frac{1}{2}\mu + \epsilon - \frac{\mu}{2^{n+1}}\right) BB \end{aligned}$$

Application: self-fertilization

Then,

$$
\begin{aligned}
F_n &= (F_n - F_{n-1}) + (F_{n-1} - F_{n-2}) + \cdots + (F_1 - F_0) + F_0. \\
&= u_n + u_{n-1} + \cdots + u_1 + F_0 \\
&= \frac{1}{2^n}\mu\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right) + \lambda AA + \mu AB + \epsilon BB \\
&= \left(\lambda + \frac{1}{2}\mu - \frac{\mu}{2^{n+1}}\right) AA + \frac{\mu}{2^n} AB + \left(\frac{1}{2}\mu + \epsilon - \frac{\mu}{2^{n+1}}\right) BB \\
&\underset{n\to\infty}{\longrightarrow} \left(\lambda + \frac{\mu}{2}\right) AA + \left(\frac{\mu}{2} + \epsilon\right) BB.
\end{aligned}
$$

Application: self-fertilization

Then,

$$\begin{aligned} F_n &= (F_n - F_{n-1}) + (F_{n-1} - F_{n-2}) + \cdots + (F_1 - F_0) + F_0. \\ &= u_n + u_{n-1} + \cdots + u_1 + F_0 \\ &= \frac{1}{2^n}\mu\left(\frac{1}{2}AA - AB + \frac{1}{2}BB\right) + \lambda AA + \mu AB + \epsilon BB \\ &= \left(\lambda + \frac{1}{2}\mu - \frac{\mu}{2^{n+1}}\right) AA + \frac{\mu}{2^n}AB + \left(\frac{1}{2}\mu + \epsilon - \frac{\mu}{2^{n+1}}\right) BB \\ &\underset{n\to\infty}{\longrightarrow} \left(\lambda + \frac{\mu}{2}\right) AA + \left(\frac{\mu}{2} + \epsilon\right) BB. \end{aligned}$$

$\rightsquigarrow$ **self-fertilization kills heterozygotes!**

Last Slide of the Day

Thank you for your attention!