Stochastic matrices arising from genetic inheritance

Irene Paniello * , 1

Departamento de Estadística e Investigación Operativa, Universidad Pública de Navarra, 31006 Pamplona, Spain

A R T I C L E   I N F O

Article history:
Received 13 May 2010
Accepted 23 September 2010
Available online 23 October 2010
Submitted by R.A. Brualdi

AMS classification:
16W99
15B51
17D92

Keywords:
Genetic algebra
Coalgebra with genetic realization
Stochastic matrix

A B S T R A C T

We establish a connection between two different approaches that can be used to describe the same genetic situation. We focus on the notion of coalgebra with genetic realization introduced by Tian and Li and show how their structure constants can be used to characterize different stochastic processes related to the genetic inheritance.

© 2010 Elsevier Inc. All rights reserved.

0. Introduction

The study of population genetics have usually made use of a mathematical formulation to symbolically represent the process of genetic inheritance. Algebraic structures arising from this formalization were already studied by Etherington, Schafer and Holgate [6,14,7]. A nice survey of some algebraic methods used in genetics can be found in [13] where the author reviews various definitions and examples illustrating the use of algebras to describe genetic inheritance. There the reader can find a compilation of different algebraic structures which represent how the genetic information is transferred from one generation to the following ones.

The mathematical formulation we use here differs from that presented by most of the authors mentioned in [13]. The problem we are interested in also arises from population genetics but we are
mainly concerned to trace back the genetic information that a generation receives from their ancestors, that is starting from the genetic information about the present individuals of a population we are interested in tracing back the genetic information from the present generation to their ancestors.

A mathematical formalization of this situation can be given by using coalgebras with genetic realization. These objects were introduced in [17] in an attempt to provide an algebraic setting analogous to that of genetic algebras.

In this paper we will focus on the structure constants arising from the comultiplication rules of a coalgebra with genetic realization. Then attending to the underlying stochastic nature of the genetic processes, these structure constants are matricially represented in terms of square and cubic stochastic matrices. These matrices turn out to be helpful to describe not only the whole process of tracing back the genetic inheritance of a particular trait but also to provide models to describe the marginal inheritance.

The approach we follow in this paper is inspired by [17] and can be considered as an attempt to provide new tools to work with phylogenetic trees in determining the evolutionary behavior of genetic inheritance.

In a forthcoming paper we will deal with the connection between the decomposition of a coalgebra with genetic realization and the classification of the different states of the related stochastic processes.

To properly describe genetic inheritance it turns out to be fundamental to take into account the Mendelian or non-Mendelian nature of the process (see [16, 5.2.2]). An algebraic and stochastic characterization of a non-Mendelian genetic process can be found in [16]. This approach has recently been taken up again by a group of authors in [2,3,8].

Another different approach to genetic algebras can be found in [1] where the authors apply computational methods to study Bernstein algebras. We recall that Bernstein algebras are a class of nonassociative algebras which arise from the study of quadratic evolutionary operators and of Bernstein’s stationary principle [9,10].

The paper is organized as follows. We devote the first section to recall some basic algebraic notions which will be useful to describe coalgebras with genetic realization, paying special attention to their biological significance in a context of genetic inheritance.

In the second section we introduce the different types of square and cubic matrices which naturally appear related to the structure constants of genetic coalgebras. We also provide an stochastic interpretation of such matrices and consider the related dynamic processes.

Finally in the third section we give an example to illustrate the relation between the comultiplication constants of a coalgebra with genetic realization and the related square and cubic stochastic matrices. The chosen example is that of simple Mendelian inheritance for a single gene with two alleles A and a. We finish making a few comments on the connections between the coalgebraic and the stochastic approaches presented in the previous sections.

1. Noncoassociative coalgebras with genetic significance

We devote this section to recall some basic notions referred to coalgebras with the aim of applying them to a particular class of coalgebras having biological significance. Any basic notion on coalgebras can be found in [4,12,15]. Here we will primarily stress on the biological significance of these notions and therefore we only consider algebraic structures defined over the real numbers.

1.1. A coalgebra is a real vector space C with a linear map \( \Delta : C \to C \otimes C \) called comultiplication. A coalgebra C has a counit \( \epsilon \) if there exists a linear map \( \epsilon : C \to \mathbb{R} \) such that \( (\epsilon \otimes 1)\Delta = (1 \otimes \epsilon)\Delta \). A coalgebra C is coassociative if \((\Delta \otimes 1)\Delta = (1 \otimes \Delta)\Delta\) and it is said to be cocommutative if \( \tau \Delta = \Delta \)
where \( \tau : C \otimes C \to C \otimes C \) is the twist map given by \( \tau(a \otimes b) = b \otimes a \) for all \( a, b \in C \).

1.2. Let \((C, \Delta)\) be a coalgebra. A subspace \( D \) of \( C \) is a subcoalgebra if \( \Delta(D) \subseteq D \otimes D \). A subspace \( I \) of \( C \) is a left coideal (resp. a right coideal) if \( \Delta(I) \subseteq C \otimes I \) (resp. \( \Delta(I) \subseteq I \otimes C \)) and a coideal if \( \Delta(I) \subseteq C \otimes I + I \otimes C \) and \( \epsilon(I) = 0 \). For any coideal \( I \), the vector space \( C/I \) has a coalgebra structure with comultiplication induced from \( \Delta \).

The sum of a collection of subcoalgebras or left (right, two-sided) coideals is again a subcoalgebra or a left (right, two-sided) coideal. The intersection of subcoalgebras (left coideals, right coideals, two-sided coideals) is a subcoalgebra (not necessarily a left coideal, right coideal, two-sided coideal). If
$S$ is a subset of a coalgebra, the subcoalgebra generated by $S$ is the intersection of all subcoalgebras containing $S$.

A coalgebra $C$ is simple if it has no nonzero proper subcoalgebras and cosemisimple if it is a direct sum of simple subcoalgebras. The coradical $C_0$ of a coalgebra $C$ is the sum of all simple subcoalgebras of $C$. A coalgebra $C$ is pointed if every simple subcoalgebra is one-dimensional and connected if its coradical $C_0$ is one-dimensional. A coalgebra $C$ is irreducible if any two nonzero subcoalgebras have nonzero intersection. A subcoalgebra $D$ of a coalgebra $C$ is an irreducible component of $C$ if it is a maximal irreducible subcoalgebra.

1.3. Here we will be interested in a particular class of coalgebras which do not need to be neither coassociative or cocommutative or even have a nonzero counit. Following [17] we will consider a population space $\Omega$ which is a real vector space spanned by a finite set $\{e_1, \ldots, e_n\}$ where each generator represents a hereditary type of a particular trait whose inheritance we are considering and $\{e_i \otimes e_j\}_{i,j=1}^{n}$ is the set of all possible ordered pairs of progenitors.

1.4. A coalgebra with genetic realization is a real coalgebra $(C, \Delta)$ which admits a basis $\{e_1, \ldots, e_n\}$ such that the comultiplication constants $\beta_{ij}^k$ satisfy the following rules:

(i) $0 \leq \beta_{ij}^k \leq 1$ for all $i, j, k = 1, 2, \ldots, n$,
(ii) $\sum_{j=1}^{n} \beta_{ij}^k = 1$ for all $k = 1, 2, \ldots, n$,

where

$$\Delta(e_k) = \sum_{i,j=1}^{n} \beta_{ij}^k e_i \otimes e_j, \quad k = 1, 2, \ldots, n$$

(see [17, 3.4].)

1.5. We will say that the set $\{e_1, \ldots, e_n\}$ is a natural basis. This basis will usually consist of the hereditary types of a population with respect to a particular hereditary trait and therefore represents all the different genetic individuals in the population space. Thus, although when working with coalgebras it is broadly used Sweedler notation [4, p. 5], due to the significance of the basis elements we will restrict ourselves to work with natural basis.

1.6. For a better understanding of the sometimes necessary absence of coassociativity or cocommutativity in coalgebras with genetic realization we should take into account the biological consequences of satisfying these properties.

For a coalgebra with genetic realization coassociativity (i.e. satisfying the identity $(\Delta \otimes 1)\Delta = (1 \otimes \Delta)\Delta$) would mean the equivalence between two different ways of tracing back the genetic inheritance over the last two generations, from the present to the grandparental generation. However being noncoassociative (i.e. not necessarily coassociative) means that if we trace back the genetic information from a particular individual to its grandparental generation the result may differ depending on which choice we take when going backwards from the parental to the grandparental generation.

On the other hand cocommutativity is related to non-sex-linked inheritance. Those traits which are sex-linked will give rise to noncocommutative coalgebras while those non-sex-linked will have symmetric properties which will imply cocommutativity. Another important characteristic of coalgebras with genetic realization will be the absence of a nonzero counit. See [17] for further details on the biological significance of these properties.

1.7. Algebras arising in genetics have been studied by different authors and therefore several definitions of algebras with genetic significance can be found. We will say that a real algebra $A$ has genetic realization (equivalently that $A$ is an algebra with genetic realization) if $A$ has a basis $\{a_1, \ldots, a_n\}$ and a multiplication table

$$a_i a_j = \sum_{k=1}^{n} \gamma_{ijk} a_k$$

such that $0 \leq \gamma_{ijk} \leq 1$ for all $i, j, k$ and $\sum_{k=1}^{n} \gamma_{ijk} = 1$ for all $i, j = 1, \ldots, n$. 
Assuming that \(\{e_1, \ldots, e_n\}\) is a natural basis representing all the genetically distinct traits, the multiplication constant \(\gamma_{ijk}\) represents the probability that two individuals of types \(a_i\) and \(a_j\) produce an individual of type \(a_k\).

Algebras with genetic realization are nonassociative algebras which have been used in population genetics to formally describe the genetic processes occurring during sexual reproduction when the genetic information is passed on through generations. The algebra multiplication can here be understood as a representation of a forward dynamic phenomenon from parents to progeny.

1.8. It is a well-known result that the dual \(A^* = \text{Hom}_{\mathbb{R}}(A, \mathbb{R})\) of a finite dimensional real algebra \(A\) has a coideal structure with comultiplication

\[
\Delta : A^* \to (A \otimes A)^* \xrightarrow{\phi^{-1}} A^* \otimes A^*
\]

where \(\rho\) denotes the isomorphism \(A^* \otimes A^* \cong (A \otimes A)^*\) [4, 1.3.9]. Thus it is a very natural question whether coalgebras with genetic realization are just the duals of algebras with genetic realization. This question was already considered in [17] showing that coalgebras with genetic realization are not duals of genetic algebras and that they indeed represent a different genetic process which is that of a backward dynamic system from a present generation to the previous ones.

1.9. A baric coalgebra \((C, \Delta, \phi)\) is a coalgebra \((C, \Delta)\) with a nontrivial character \(\phi\), i.e. a nonzero linear map \(\phi : C \to \mathbb{R}\) such that \((\phi \otimes \phi)\Delta = \phi\).

A baric subcoalgebra \(D\) of a baric algebra \((C, \Delta, \phi)\) is a subcoalgebra \(D\) of \(C\) such that \(\phi_D\) is nonzero. A baric coideal is a coideal \(I\) of a baric coalgebra such that \(I\) is contained in \(\ker \phi\) and as a result the quotient coalgebra \(C/I\) is baric [17].

1.10. Finite dimensional coalgebras with genetic realization are baric coalgebras such that all the comultiplication constants are nonnegative [17, 4.3, 4.4]. Characters play an important role in the case of genetic algebras. Indeed the previously mentioned Bernstein algebras are a particular class of baric algebras where elements of baric weight 1 are precisely those elements reaching genetic equilibrium after one generation of random matings within the population. The number of necessary matings to reach the genetic equilibrium depends on the order of the Bernstein algebra [10].

2. Stochastic matrices

In this section we consider different types of stochastic matrices and study their connection with the genetic process under consideration. Since we will simultaneously work with square and cubic matrices, we will stress on the dimension of the matrices when necessary.

2.1. A (square) stochastic matrix is a nonnegative matrix \(P = (p_{ij})_{i,j=1}^n\) such that each row sum equals to one

\[
\sum_{j=1}^n p_{ij} = 1, \quad p_{ij} \geq 0
\]

These stochastic matrices are in correspondence to transition matrices of a particular class of Markov processes.

2.2. Let \(X = \{X_0, X_1, \ldots\}\) be a discrete time stochastic process where \(X_t\) is a discrete random variable defined on a finite or countably infinite state space which will be denoted by \(\{1, 2, \ldots\}\). A Markov stochastic process is a stochastic process such that its future behavior only depends on the present and not on its past history, that is, a stochastic process satisfying the Markov property

\[
P(X_n = i_n|X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = P(X_n = i_n|X_{n-1} = i_{n-1})
\]

where \(i_k \in \{1, 2, \ldots\}\) for \(k = 0, 1, \ldots, n\).

A Markov process \(X\) is called homogeneous if the transition probabilities \(P(X_{n+1} = i_{n+1}|X_n = i_n)\) do not depend on time \(n\), i.e.

\[
P(X_t = i_t|X_{t-1} = i_{t-1}) = P(X_{t+k} = i_{t+k}|X_{t+k-1} = i_{t+k-1})
\]

for all \(k\). We will denote by

\[
p_{ij} = P(X_{n+1} = j|X_n = i)
\]
the transition probabilities of an homogeneous Markov chain. The transition matrix of an homogeneous discrete time Markov chain \( X = \{X_0, X_1, \ldots \} \) with state space \( \{1, 2, \ldots \} \) and transition probabilities \( \{p_{ij}\}_{i,j=1}^{\infty} \) is

\[
P = \begin{pmatrix}
p_{11} & p_{12} & p_{13} & \cdots \\
p_{21} & p_{22} & p_{23} & \cdots \\
p_{31} & p_{32} & p_{33} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

If the set of states is finite \( \{1, 2, \ldots, N\} \) then \( P \) is a nonnegative \( N \times N \) matrix such that \( \sum_{j=1}^{N} p_{ij} = 1 \), i.e. \( P \) is a stochastic matrix. Conversely the \((i,j)\)-entry \( p_{ij} \) of any stochastic matrix \( P = (p_{ij})_{i,j=1}^{n} \) can be considered the transition probability from a state \( i \) to the state \( j \) for an homogeneous discrete time Markov process.

2.3. Let us denote by \((i,j,k)\) the cubic unit matrices, i.e. \((i,j,k)\) is a \( n \times n \times n \) cubic matrix whose \((i,j,k)\)th entry is equal to 1 and all other entries are equal to 0 [11]. A cubic matrix \( P = (p_{ijk})_{i,j,k=1}^{n} \) is an object with three indices \( i, j, k \) which can be uniquely written in the form

\[
P = (p_{ijk})_{i,j,k=1}^{n} = \sum_{i,j,k=1}^{n} p_{ijk}(i,j,k)
\]

2.4. Following [11] we will say that a cubic matrix \( P = (p_{ijk})_{i,j,k=1}^{n} \) is stochastic of type \((1,2)\) if \( p_{ijk} \geq 0 \) and \( \sum_{i,j,k=1}^{n} p_{ijk} = 1 \) for all \( k = 1, \ldots, n \).

Cubic stochastic matrices of types \((2,3)\) and \((1,3)\) can be defined analogously. However, though matrices of types \((1,2)\) and \((2,3)\) behave similarly, there exist some differences with respect to those of type \((1,3)\) (see [11] for more details).

2.5. Multidimensional matrices can be endowed with different nontrivial multiplications. In the case of cubic matrices some of these possible operations can be found in [11] together with the corresponding probabilistic interpretations.

For cubic stochastic matrices of type \((1,2)\) Maksimov introduces the following multiplication:

\[
(i,j,k) \cdot (m,n,r) = \begin{cases} (i,j,r), & \text{if } k = m; \\ 0, & \text{otherwise.} \end{cases}
\]

Given a finite physical system \( \{1, 2, \ldots, n\} \), a cubic stochastic matrix \( P = (p_{ijk})_{i,j,k=1}^{n} \) of type \((1,2)\) endowed with the above multiplication can be used to describe certain interactions involving three individuals. Indeed the \((i,j,k)\)-entry \( p_{ijk} \) can be seen as the probability for an individual of type \( k \) to have an ordered pair of parents of types \( i \) and \( j \) and the \((i,j,k)\)th entry of \( P^n \) can be seen as the probability of an individual of type \( k \) arising after \( n \) units of time (i.e. after \( n \) matings) from an ordered pair of ancestors of types \( i \) and \( j \).

2.6. Let \( P = (p_{ijk})_{i,j,k=1}^{n} \) be a cubic stochastic matrix of type \((1,2)\). Then we will say that the square matrix \( P_{(i)} = (p_{(ij)k})_{i,k=1}^{n} \), where \( p_{(ij)k} = \sum_{j=1}^{n} p_{ijk} \) is the \( i \)-accompanying matrix for \( P \). Similarly we will say that \( P_{(j)} = (p_{(i)jk})_{j,k=1}^{n} \), where \( p_{(i)jk} = \sum_{i=1}^{n} p_{ijk} \) is the \( j \)-accompanying matrix for \( P \). (Here the terminology is slightly different from that of [11] where only \( i \)-accompanying matrices are considered.)

Note that columns of accompanying matrices satisfy

\[
\sum_{i=1}^{n} p_{i(j)k} = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ijk} = 1,
\]

\[
\sum_{j=1}^{n} p_{(i)jk} = \sum_{j=1}^{n} \sum_{i=1}^{n} p_{ijk} = 1,
\]

and therefore, though accompanying matrices are not stochastic in the sense of 2.1, their transposes are indeed stochastic.
2.7. Let us next consider the $k$th layer of a cubic stochastic matrix $P = (p_{ijk})_{i,j,k=1}^n$ of type (1,2). Then we get a nonnegative $n \times n$ matrix

$$P_k = \begin{pmatrix} p_{11k} & p_{12k} & p_{13k} & \cdots & p_{1nk} \\ p_{21k} & p_{22k} & p_{23k} & \cdots & p_{2nk} \\ p_{31k} & p_{32k} & p_{33k} & \cdots & p_{3nk} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n1k} & p_{n2k} & p_{n3k} & \cdots & p_{nnk} \end{pmatrix}$$

such that

$$\sum_{i,j} p_{ijk} = 1.$$

2.8. Recall that given two categorical random variables $X$ and $Y$ taking values on finite sets $[r] = \{1, 2, \ldots, r\}$ and $[s] = \{1, 2, \ldots, s\}$ respectively, the cross-classification of $m$ independent and identically distributed realizations according to these two discrete criteria produces a random integer-valued array called a 2-way contingency table whose coordinate entry $n_{ij}$ is the number of times the label combination or cell $(i, j)$ is observed in the sample. The probability that a given cell appears in the sample is the joint probability

$$p_{ij} = P(X = i, Y = j), \quad i \in [r], \quad j \in [s].$$

Marginal probabilities are usually denoted by:

$$p_{i+} = \sum_{j=1}^s p_{ij} = P(X = i), \quad i \in [r],$$

$$p_{+j} = \sum_{i=1}^r p_{ij} = P(Y = j), \quad j \in [s].$$

2.9. **Proposition.** Let $P = (p_{ijk})_{i,j,k=1}^n$ be a cubic stochastic matrix of type (1,2). Then:

(i) The accompanying matrices $P_{(i)}$ and $P_{(j)}$ are the transposes of two square stochastic matrices.

(ii) For any $k \in \{1, 2, \ldots, n\}$, the square matrix $P_k = (p_{ijk})_{i,j=1}^n$ contains the joint probabilities of the cross-classification of a finite set of states $\{1, 2, \ldots, n\}$ according to two discrete criteria.

**Proof.** (i) follows from 2.6 and (ii) from 2.4 and 2.7. □

2.10. Let now $\Omega$ be a population space spanned by a finite set $\{e_1, \ldots, e_n\}$ where each $e_i, i = 1, \ldots, n$, represents a type of a particular hereditary trait and consider $(C, \Delta)$ a coalgebra over $\Omega$ with natural basis $\{e_1, \ldots, e_n\}$ and comultiplication given by

$$\Delta(e_k) = \sum_{i,j=1}^n \beta_{ij}^k e_i \otimes e_j, \quad i, j = 1, 2, \ldots, n,$$

where $\beta_{ij}^k$ is the probability of an individual of type $k$ to have an ordered pair of progenitors of types $i$ and $j$. Clearly $(C, \Delta)$ has a genetic realization.

Let $P = (p_{ijk})_{i,j,k=1}^n$ be the cubic matrix defined by the comultiplication constants, i.e. $p_{ijk} = \beta_{ij}^k$ for all $i, j, k = 1, \ldots, n$. It follows from 1.4 that $P$ is a cubic stochastic matrix of type (1,2). Moreover there is a correspondence between the biological significance of the $p_{ijk}$ and that mentioned in 2.5.

Consider now a fixed basis element $e_k$. Then we can write

$$\Delta(e_k) = \sum_{i=1}^n \left( \sum_{j=1}^n \beta_{ij}^k e_i \otimes e_j \right).$$
where $\sum_{j=1}^{n} \beta_{ij}^k$ is exactly the $(i, k)$-entry of the $i$-accompanying matrix $P(i)$ for $P$. Thus the $i$-accompanying matrix $P(i) = (p_{i(j)k})_{j,k=1}^{n}$ for $P$ is the transpose of a square stochastic matrix and each $p_{i(j)k}$ represents the probability of a individual of type $k$ to have a male progenitor of type $i$ independently of the female progenitor type. Similarly one has

$$\Delta(e_k) = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \beta_{ij}^k e_i \otimes e_j \right),$$

where $\sum_{i=1}^{n} \beta_{ij}^k$ is the $(j, k)$-entry of the $j$-accompanying matrix $P(j)$ for $P$ and therefore the entry $p_{i(j)k}$ of the $j$-accompanying matrix $P(j) = (p_{i(j)k})_{i,k=1}^{n}$ for $P$ represents the probability of a individual of type $k$ to have a female progenitor of type $j$ independently of the male progenitor type.

Finally it is straightforward to check that any $k$-layer of the cubic matrix $P$ gives the joint probability table of the cross-classification of the observed parents for those individuals of type $k$ in the filial generation. Moreover for a fixed generation type $k$, the marginal probabilities of the corresponding $k$-layer contingency table can be understood as the following conditional probabilities

$$p_{i+k} = \sum_{j=1}^{n} p_{i(j)k} = P(\text{father} = i | \text{filial type } k), \quad i = 1, \ldots, n,$$

$$p_{+j+k} = \sum_{i=1}^{n} p_{i(j)k} = P(\text{mother} = j | \text{filial type } k), \quad j = 1, \ldots, n,$$

where the “$+$” in the subscript represents a summation over the corresponding index values and $p_{i+k}$ (resp. $p_{+j+k}$) is the $(i, k)$-entry of the $i$-accompanying matrix for $P$ (resp. the $(j, k)$-entry of the $j$-accompanying matrix).

The above remarks allow us to settle the following result.

2.11. **Theorem.** Let $(C, \Delta)$ be a coalgebra with genetic realization and comultiplication given by

$$\Delta(e_k) = \sum_{i,j=1}^{n} \beta_{ij}^k e_i \otimes e_j, \quad k = 1, 2, \ldots, n.$$

Then the comultiplication constants define a cubic stochastic matrix $P = (p_{ijk})_{i,j,k=1}^{n}$ of type (1,2) such that $p_{ijk} = \beta_{ij}^k$, $i, j, k = 1, \ldots, n$. Moreover the accompanying matrices for $P$ describe the probability distributions relating the genetic type of the filial generation to that of the corresponding male or female progenitor.

**Proof.** It follows from 2.10. □

2.12. The study of the $P_k$ matrices of the cubic stochastic matrix $P$ of type (1,2) given by the comultiplication constants of a coalgebra with genetic realization allow us to known whether for a fixed individual of type $k$ there exists a relation between the genetic types of its progenitors. Indeed we just need to compute the rank of $P_k$ and apply the following proposition (see [5, 1.1.2]).

2.13. **Proposition** ([5, 1.1.2]). Let $X$ and $Y$ be two categorical random variables with values on $[r] = \{1, 2, \ldots, r\}$ and $[s] = \{1, 2, \ldots, s\}$ respectively. Then $X$ and $Y$ are independent if and only if the $r \times s$ matrix $(p_{ij})$ has rank 1, where $p_{ij} = P(X = i, Y = j)$ for $i \in [r]$ and $j \in [s]$.

3. **An example and further comments**

In this section we first compute the coalgebras and matrices discussed in the previous sections for a particular genetic situation and then we make some comments on the connections between the two presented approaches.
3.1. Example. Simple Mendelian inheritance for a single gene with two alleles

As a basic example we consider simple Mendelian inheritance for a single gene with two alleles \(A\) and \(a\) where zygotes have three possible genotypes \(AA, Aa\) and \(aa\) \([13, p. 107]\). Assuming simple Mendelian inheritance both alleles \(A\) and \(a\) are passed on with equal frequency.

Following \([17]\) let \(\mathcal{Z}\) be the three dimensional real vector space spanned by \(\{AA, Aa, aa\}\), i.e. the corresponding zygotic space. The zygotic algebra \((\mathcal{Z}, m)\) for simple Mendelian inheritance is the three dimensional real algebra with natural basis \(\{e_{11} = AA, e_{12} = Aa, e_{22} = aa\}\) and multiplication table (see \([13, p. 109]\)):

\[
\begin{array}{c|ccc}
  & AA & Aa & aa \\
\hline
AA & AA & \frac{1}{2}(AA + Aa) & Aa \\
Aa & \frac{1}{2}(AA + Aa) & \frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa & \frac{1}{2}(Aa + aa) \\
aa & Aa & \frac{1}{2}(Aa + aa) & aa \\
\end{array}
\]

Denoting the multiplication constants by

\[m(e_{ij} \otimes e_{kl}) = \sum_{r \leq s} \gamma_{ij,kl,rs} e_{rs}, \quad i \leq j, \; k \leq j, \; i, j, k, l = 1, 2,\]

it is straightforward to check that the following conditions are satisfied:

\[0 \leq \gamma_{ij,kl,rs} \leq 1, \quad \sum_{r,s=1, r \leq s}^2 \gamma_{ij,kl,rs} = 1, \quad \gamma_{ij,kl,rs} = \gamma_{kl,ij,rs},\]

for all \(i \leq j, \; k \leq j, \; i, j, k, l = 1, 2\), showing that this algebra has a genetic realization.

Consider now the dual coalgebra \((C, \Delta)\) of \((\mathcal{Z}, m)\). Now \(C\) is a three dimensional vector space with basis \(\{\eta_{11} = AA^*, \eta_{12} = Aa^*, \eta_{22} = aa^*\}\) and using 1.8 it is easily checked that the comultiplication \(\Delta\) is given by:

\[
\begin{align*}
\Delta(AA^*) &= AA^* \otimes AA^* + \frac{1}{2}(AA^* \otimes Aa^* + Aa^* \otimes AA^*) + \frac{1}{4}Aa^* \otimes Aa^*, \\
\Delta(Aa^*) &= \frac{1}{2}(AA^* \otimes Aa^* + Aa^* \otimes AA^*) + (AA^* \otimes aa^* + aa^* \otimes AA^*) \\
&\quad + \frac{1}{2}Aa^* \otimes Aa^* + \frac{1}{2}(Aa^* \otimes aa^* + aa^* \otimes Aa^*), \\
\Delta(aa^*) &= \frac{1}{4}Aa^* \otimes Aa^* + \frac{1}{2}(Aa^* \otimes aa^* + aa^* \otimes Aa^*) + aa^* \otimes aa^*.
\end{align*}
\]

Note that though all the comultiplication constants are nonnegative, the sum of those corresponding to the same basis element is not equal to one. However being this basis the dual of the algebra natural basis one expects it to be natural if such a basis exists (i.e. to retain the biological significance). Thus \((C, \Delta)\) is not a coalgebra with genetic realization.

Finally we consider the corresponding zygotic coalgebra \((\mathcal{Z}, \Delta)\). The population space is now the zygotic space spanned by \(\{e_1 = AA, e_2 = Aa, e_3 = aa\}\) and for each zygote the comultiplication constants are determined by the probability distribution of zygotes in the parental generation:
\[ \Delta(AA) = \frac{4}{9} AA \otimes AA + \frac{2}{9} (AA \otimes Aa + Aa \otimes AA) + \frac{1}{9} Aa \otimes Aa, \]
\[ \Delta(Aa) = \frac{1}{9} (AA \otimes Aa + Aa \otimes AA) + \frac{1}{9} Aa \otimes Aa \]
\[ + \frac{2}{9} (AA \otimes aa + aa \otimes AA) + \frac{1}{9} (Aa \otimes aa + aa \otimes Aa), \]
\[ \Delta(aa) = \frac{4}{9} aa \otimes aa + \frac{2}{9} (Aa \otimes aa + aa \otimes Aa) + \frac{1}{9} Aa \otimes Aa. \]

This coalgebra is cocommutative since we are considering non-sex-linked inheritance and it is also easy to see that coassociativity fails since, for example, \((\Delta \otimes 1)\Delta(AA) \neq (1 \otimes \Delta)\Delta(AA)\).

It is straightforward to check that this coalgebra has a genetic realization since the comultiplication constants are nonnegative and satisfy the conditions given in 1.4.

The comultiplication constants \(\beta_{i,j,k}^{2} \) determine a cubic \(3 \times 3 \times 3\) stochastic matrix whose \(k\)-layers are:

\[
\begin{pmatrix}
\frac{4}{9} & \frac{2}{9} & 0 \\
\frac{2}{9} & \frac{1}{9} & 0 \\
0 & 0 & 0
\end{pmatrix},
\begin{pmatrix}
0 & \frac{1}{9} & \frac{2}{9} \\
\frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\
\frac{2}{9} & \frac{1}{9} & 0
\end{pmatrix},
\begin{pmatrix}
0 & 0 & 0 \\
0 & \frac{1}{9} & \frac{2}{9} \\
\frac{2}{9} & \frac{4}{9} & 0
\end{pmatrix}
\]

On the other hand the \(i\) and \(j\)-accompanying matrices of the coalgebra \((Z, \Delta)\) are respectively:

\[
P_{(i)} = \begin{pmatrix}
\frac{6}{9} & \frac{3}{9} & 0 \\
\frac{3}{9} & \frac{3}{9} & \frac{3}{9} \\
0 & \frac{1}{9} & \frac{6}{9}
\end{pmatrix},
P_{(j)} = \begin{pmatrix}
\frac{6}{9} & \frac{3}{9} & 0 \\
\frac{3}{9} & \frac{3}{9} & \frac{3}{9} \\
0 & \frac{3}{9} & \frac{6}{9}
\end{pmatrix}.
\]

We should remark here that the coincidence and symmetry of the accompanying matrices is just due to the intrinsic nature of the example we are considering. The coincidence follows from the non-sex-linked inheritance and the symmetry from the fact that both alleles are passed on with equal probability.

The \((i, k)\)-entry of the \(i\)-accompanying matrix gives the probability for a individual of type \(k\) to have a male progenitor of type \(i\) without any restriction on the female progenitor type. Similarly the entries of the \(j\)-accompanying matrix give the probabilities relating the filial and the female progenitor types independently of the male progenitor type.

The transposes of both accompanying matrices can be considered as the transition matrices of two homogeneous discrete Markov processes. Thus both matrices can be used to trace back the genetic inheritance from the filial to the parental generations. Classifying the states of the underlying Markov chain we will obtain, for example, which genotypes can be reached from a fixed one or which ones are accessible or communicated. Each class of states will have a different biological significance.

We also note that in this example the accompanying matrices are indeed double stochastic matrices since not only each row but also the columns sum equal to one.

### 3.2. Further comments

There exists a well-known developed structure theory for coalgebras which allows us to decompose particular classes of finite dimensional coalgebras into simple subcoalgebras.
Considering subcoalgebras and coideals as the basic pieces resulting from coalgebra decomposition, the following notions make sense for a coalgebra \((C, \Delta)\) with genetic realization (see 1.2). Let \(\{e_1, \ldots, e_n\}\) be a natural basis of \(C\) where each generator \(e_i\) represents a hereditary type of a particular trait. We will say that a subcoalgebra \(D\) of \(C\) has genetic realization if \(D\) is spanned by a subset \(\{e_{i_1}, \ldots, e_{i_k}\}\) of the natural basis. Note that being \(D\) a subcoalgebra the equality \(\Delta(D) \subseteq D \otimes D\) holds and therefore the parental genotypes of any individual in \(D\) also yield in \(D\). Thus subcoalgebras spanned by elements of the natural basis represent genetically invariant subpopulations.

If instead of subcoalgebras we consider right coideals (i.e. subspaces \(D\) of \(C\) such that \(\Delta(D) \subseteq D \otimes C\)) then it follows from
\[
\Delta(D) \subseteq D \otimes C, \\
(\Delta \otimes 1)\Delta(D) \subseteq D \otimes C \otimes C, \\
(\Delta \otimes 1 \otimes 1)(\Delta \otimes 1)\Delta(D) \subseteq D \otimes C \otimes C \otimes C,
\]
that right coideals represent subpopulations which are genetically invariant with respect to the male progenitor. Similarly left coideals will be in correspondence with subpopulations which are genetically invariant with respect to the female progenitor. Moreover direct sums and intersections of subcoalgebras or coideals will allow us to handle a finite number of genetic traits simultaneously.

On the other hand splitting the population space into genetically invariant subpopulations (with respect to a particular trait or a finite number of traits) will be equivalent to algebraically decompose the corresponding coalgebra with genetic realization.

Thus we expect that using algebraic tools similar to those use in coalgebra decomposition we should be able to genetically classify the individuals of any population into genetically distinct subpopulations and then the notions introduced in 1.2 will represent different types of subpopulations. Moreover such classification should allow us to establish a correspondence between the coalgebraic biological classes and different stochastic states, proving tools to work with multidigit matrices.

References