

Populationsgenetik

**Eine Zusammenfassung von Bernhard Kabelka
zur Vorlesung von Prof. Länger im WS 2003/04**

Version 1.0, 5. Juni 2005

Es sei ausdrücklich betont, dass

- (1) dieses Essay ohne das Wissen und die Mitarbeit von Prof. Länger entstanden ist,
- (2) trotz großer Anstrengungen seitens des Autors, eine möglichst fehlerfreie und vollständige Zusammenfassung zu liefern, sich Fehler eingeschlichen haben könnten (sollte jemand einen Fehler entdecken, so bittet der Autor um Benachrichtigung, vorzugsweise per eMail an bernhard@kabelka.net),
- (3) die Lektüre dieser Zusammenfassung keinesfalls den persönlichen Besuch der Vorlesung bzw. das Studium des Skriptums ersetzen, sondern bestenfalls ergänzen kann.

Die aktuelle Version dieser Datei ist erhältlich unter:

- <http://fsmat.at/~bkabelka/math/nawi/download/popgen.pdf> (Online-Version mit Hyperlinks)
- <http://fsmat.at/~bkabelka/math/nawi/download/popgen.ps.gz> (Druck-Version ohne Hyperlinks)

Inhaltsverzeichnis

Einleitung	1
1 Mendels Kreuzungsversuche	1
2 Hardy-Weinberg-Gleichgewicht	1
2.1 Zwei-Allele-Modell	1
2.2 Züchtung reinerbiger Individuen	2
2.3 n -Allele-Modell	3
3 Geschlechtsgebundene Vererbung	3
4 Phänotypsysteme	3
5 Markov-Ketten	4
6 Mutation	4
7 Wright-Modell	5
8 Genetische Algebren	5
9 Fisher-Wright-Selektionsmodell	6
10 Genotypen verwandter Individuen	6
Index	7

Einleitung

Eine **Population** ist eine Zusammenfassung von Individuen derselben Art in einem gemeinsamen Lebensraum. Die Populationsgenetik befasst sich in erster Linie mit Aussagen über Gen- und Genotyphäufigkeiten (und auch deren zeitlichem Verhalten).

1 Mendels Kreuzungsversuche

Gregor Johann **Mendel** (1822 – 1884) unternahm einige Versuche in Bezug auf Vererbung und stellte so drei (Grund-)Regeln der Vererbungslehre auf:

Kreuzt man reinerbig violettblühende und weißblühende Erbsenrassen miteinander, so besteht die erste Tochtergeneration aus lauter violettblühenden Pflanzen (**Uniformitätsregel**).

Kreuzt man nun diese wieder miteinander, so ergeben sich in der zweiten Tochtergeneration sowohl violett- als auch weißblühende Pflanzen, und zwar im Verhältnis 3 : 1 (**Regel vom charakteristischen Aufspaltungsverhältnis**).

Betrachtet man nun zwei reinerbige Erbsenrassen – die eine mit gelben, runden Samen, die andere mit grünen, kantigen Samen – so ergeben sich in der ersten Tochtergeneration wieder lauter Pflanzen mit gelben, runden Samen. In der zweiten Generation entstehen Pflanzen mit gelben, runden, bzw. gelben, kantigen, bzw. grünen, runden, bzw. grünen, kantigen Samen im Verhältnis 9 : 3 : 3 : 1. Die Vererbung dieser beiden Merkmale ist also voneinander unabhängig (**Unabhängigkeitsregel**).

2 Hardy-Weinberg-Gleichgewicht

Im Folgenden wollen wir immer – sofern nicht anders vermerkt – eine sogenannte **ideale Population** betrachten, d. h. eine Population für die gilt:

- „sehr große“ Individuenzahl (damit die relative Häufigkeit von Genen, etc. ungefähr gleich der Wahrscheinlichkeit ist)
- keine Selektion
- keine Mutation
- keine Migration
- Zufallspaarung
- getrennte Generationen (d. h. Elterngeneration ausgestorben, wenn Tochtergeneration im fortpflanzungsfähigen Alter)

2.1 Zwei-Allele-Modell

Wir betrachten ein festes Merkmal, das durch ein bestimmtes Gen fixiert wird. Die beiden möglichen Gene seien A und a .

Tritt das Gen A mit der relativen Häufigkeit p , das Gen a mit der relativen Häufigkeit $q = 1-p$ auf, sowie die drei möglichen Genotypen AA , Aa bzw. aa mit den relativen Häufigkeiten D , H , bzw. R ($D + H + R = 1$), so ergibt sich durch einfach Rechnung:

$$p = D + \frac{H}{2} \quad \text{bzw.} \quad q = R + \frac{H}{2}$$

d. h. aus den Genotyphäufigkeiten folgen die Genhäufigkeiten. Die Umkehrung gilt jedoch nicht.

Der **Satz von Hardy-Weinberg** sagt aus, dass die Genhäufigkeiten ab der ersten, die Genotyphäufigkeiten ab der zweiten Generation konstant bleiben. Das sich einstellende Gleichgewicht nennt man **Hardy-Weinberg-Gleichgewicht**, für das gilt:

$$D = p^2, \quad H = 2pq, \quad R = q^2$$

Ein Tripel (D, H, R) mit $D, H, R > 0$ und $D + H + R = 1$ ist genau dann ein Hardy-Weinberg-Gleichgewicht, wenn gilt:

$$\begin{aligned} (D, H, R) &= (p^2, 2pq, q^2) \\ \iff \exists a, b \geq 0 : a + b = 1 : (D, H, R) &= (a^2, 2ab, b^2) \\ \iff H^2 &= 4DR \end{aligned}$$

Geometrisch lassen sich die Hardy-Weinberg-Gleichgewichte mit dem sogenannte **de-Finetti-Diagramm** beschreiben, indem man ausnützt, dass die Summe der Normalabstände eines beliebigen inneren Punktes von den Seiten eines gleichseitigen Dreiecks gleich der Höhe des Dreiecks ist. Man kann so das Paar (D, H, R) wie in [Abbildung 1\(a\)](#) gezeigt beschreiben.

Benützt man obige Bedingung $H^2 = 4DR$, so erhält man als mögliche Orte für die Hardy-Weinberg-Gleichgewichte die **Hardy-Weinberg-Parabel** (siehe [Abbildung 1\(b\)](#)).

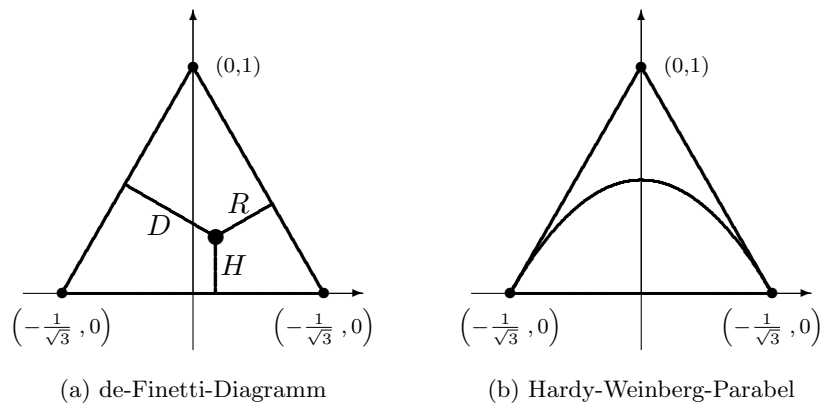


Abbildung 1: Das Hardy-Weinberg-Gleichgewicht

2.2 Züchtung reinerbiger Individuen

Die Rechnungen, die für obige Resultate angestellt wurden, können benützt werden, um ein Verfahren zur Züchtung reinerbiger Individuen zu entwickeln:

Möchte man beispielsweise – in Fortsetzung des Beispiels aus [Abschnitt 1](#) – reinerbige, violettblühende Erbsen züchten, so führt man nur eine Zufallspaarung zwischen violettblühenden Pflanzen durch, und entfernt die weißblühenden. So erhält man in der k -ten Generation homozygote Individuen (also jene mit Genotyp Aa) mit der relativen Häufigkeit von

$$H_k = \frac{2H_0}{kH_0 + 2} \searrow 0 \quad \text{für } k \rightarrow \infty$$

2.3 n -Allele-Modell

Im Allgemeinen sollte man sich natürlich nicht nur auf zwei Gene beschränken, sondern auch den Fall mehrerer möglicher Gene in Betracht ziehen: Die möglichen Gene seien daher A_1, A_2, \dots, A_n , die mit der jeweiligen relativen Häufigkeit p_1, p_2, \dots, p_n auftreten. Der Genotyp $A_i A_j$ habe die relative Häufigkeit p_{ij} .

Durch ähnliche Vorgangsweise wie im Fall der zwei Allele (siehe [Abschnitt 2.1](#)) erhält man auch hier den Satz von Hardy-Weinberg. Diesmal befindet sich die Population im Hardy-Weinberg-Gleichgewicht genau dann, wenn gilt:

$$\begin{aligned} p_{ij} &= p_i \cdot p_j && \forall i, j \\ \iff \exists a_1, a_2, \dots, a_n \geq 0 : \sum_{k=1}^n a_k &= 1 : && p_{ij} = a_i \cdot a_j \\ \iff p_{ij}^2 &= p_{ii} \cdot p_{jj} && \forall i, j \end{aligned}$$

3 Geschlechtsgebundene Vererbung

Gewisse Gene sind mit den Chromosomen gekoppelt. Es gibt daher (je nach Geschlecht) unterschiedliche Genotypen (zuerst für den Fall zweier Allele):

- Männchen: Genotyp A bzw. a mit den relativen Häufigkeiten P bzw. Q
- Weibchen: Genotyp AA bzw. Aa bzw. aa mit den relativen Häufigkeiten D bzw. H bzw. R und den Genhäufigkeiten p bzw. q

Auch hier kann man ermitteln, wie sich die Gen- bzw. Genotyphäufigkeiten beim Übergang von einer Generation zur nächsten ändern:

$$P' = p, \quad Q' = q, \quad D' = Pp, \quad H' = Pq + Qp, \quad R' = Qq$$

Davon ausgehend kann man Differenzgleichungen für die Gen- bzw. Genotyphäufigkeiten in der k -ten Generation bestimmen, und erhält schließlich Lösungen, die (für $k \rightarrow \infty$) mit abnehmenden Oszillationen gegen ein Hardy-Weinberg-Gleichgewicht streben.

Diese Aussagen gelten sinngemäß auch für den Fall mit n Allelen, wo sich schlussendlich ergibt:

$$P_i^{(\infty)} = \frac{1}{3} \cdot P_i^{(0)} + \frac{2}{3} \cdot p_i^{(0)}$$

4 Phänotypsysteme

Unter einem **Phänotypsystem** versteht man eine Äquivalenzrelation auf den Genotypen, wobei zwei Genotypen genau dann äquivalent sind, wenn sie denselben Phänotyp, d. h. dasselbe äußere Erscheinungsbild hervorrufen.

Nun stellt sich die Frage, ob man jedem Gen einen Faktor zuordnen kann, sodass zwei Genotypen genau dann zum selben Phänotyp gehören, wenn sie dieselben Faktoren besitzen. Ist also α ein Phänotypsystem, so ist eine **Faktorvereinigungsdarstellung** (kurz: **FU-Darstellung**) eine Abbildung $f : G \rightarrow s^F$ (wobei F die Menge der Faktoren sei), sodass gilt

$$\alpha = \left\{ (x, y) \in G_2^2 \mid \bigcup_{z \in x} f(z) = \bigcup_{z \in y} f(z) \right\}$$

wobei G_2 die Menge aller ein- und zwei-elementigen Teilmengen von G bezeichne.

Besitzt ein Phänotypsystem eine solche Faktorvereinigungsdarstellung, so sagt man, es sei ein **Faktorvereinigungssystem** (oder kurz **FU-System**).

Man kann zeigen, dass ein Phänotypsystem α genau dann ein FU-System ist, wenn α die Einschränkung der kleinsten Kongruenzrelation auf $(2^G \setminus \{\emptyset\}, \cup)$ ist, die α umfasst. Diese kleinste Kongruenzrelation θ , die α umfasst, ist gegeben durch die transitive Hülle von

$$\{(x \cup z, y \cup z) \mid (x, y) \in \alpha, z \in G\}$$

Ein Phänotypsystem α ist also genau dann ein FU-System, wenn gilt: $\alpha = \theta \cap G_2^2$

Ist α ein FU-System, so erhält man eine FU-Darstellung von α vermöge der Abbildung $f : G \rightarrow 2^F$ mit $F := (2^G \setminus \{\emptyset\}) / \theta$ und

$$f(x) := \{y \in F \mid y \not\preceq [\{x\}]_\theta\}$$

Es genügt auch, in der Definition von $f(x)$ anstelle von $y \in F$ nur jene y zu betrachten, die \wedge -irreduzibel sind, wobei ein Element a einer halbgeordneten Menge (H, \leq) genau dann \wedge -irreduzibel ist, wenn gilt:

$$\nexists b, c \in H \setminus \{a\} : a = b \wedge c$$

5 Markov-Ketten

Gegeben sei ein System S , das sich zu jedem der Zeitpunkte $0, 1, 2, \dots$ in genau einem der Zustände $1, 2, \dots, n$ befinden kann, wobei der Zustand zum Zeitpunkt $t+1$ nur vom Zustand zum Zeitpunkt t abhängt. Sei a_{ij} die Übergangswahrscheinlichkeit vom Zustand j in den Zustand i und x_{ij} die Wahrscheinlichkeit, dass sich S zum Zeitpunkt j im Zustand i befindet. Dann ist $A := (a_{ij})$ eine sogenannte **stochastische Matrix** (d. h. die Spaltensummen sind jeweils gleich 1), und es gilt:

$$\mathbf{x}_k = A \cdot \mathbf{x}_{k-1} = A^k \cdot \mathbf{x}_0$$

Existiert sogar ein $m \geq 1$, sodass $A^m > 0$ (d. h. jede Komponente der Matrix A^m ist größer als 0), dann konvergiert \mathbf{x}_k für $k \rightarrow \infty$ gegen den auf Komponentensumme 1 normierten, eindeutig bestimmten, Eigenvektor von A zum Eigenwert 1.

Die Existenz einer positiven ganzen Zahl n mit $A^n > 0$ lässt sich für nicht-negatives A (d. h. jede Komponente der Matrix A ist nicht-negativ) wie folgt überprüfen: Man ordnet der Matrix A einen (gerichteten) Graphen mit den Knoten $1, 2, \dots, n$ zu, bei dem genau dann eine gerichtete Kante von i nach j führt, wenn $a_{ij} > 0$ ist. Existiert dann in dem so erhaltenen Graphen einerseits zu jedem Paar (i, j) von Knoten eine gerichtete Kantenfolge von i nach j und andererseits mindestens zwei gerichtete Kantenfolgen mit zueinander teilerfremden Längen, so existiert ein $n \in \mathbb{N}^+$ mit $A^n > 0$.

6 Mutation

Hier kommen Markov-Ketten (siehe [Abschnitt 5](#)) zum Einsatz: Sei a_{ij} die Wahrscheinlichkeit einer Mutation des Gens A_j zum Gen A_i . Dann ist die Genverteilung in der k -ten Generation gegeben durch

$$\mathbf{b}_k = A^k \cdot \mathbf{b}_0$$

wobei \mathbf{b}_0 der Vektor der Genhäufigkeiten zu Beginn ist.

Im Zwei-Allele-Fall ergibt sich mit positiven Mutationsraten

$$\begin{aligned} \alpha &= \mathbb{P}[\text{Mutation } A \rightarrow a] \\ \beta &= \mathbb{P}[\text{Mutation } a \rightarrow A] \end{aligned}$$

im Grenzwert Genhäufigkeiten von

$$\mathbb{P}[A] = \frac{\beta}{\alpha + \beta} \quad \text{bzw.} \quad \mathbb{P}[a] = \frac{\alpha}{\alpha + \beta}$$

7 Wright-Modell

In diesem Abschnitt betrachten wir eine Population, die ideal ist bis auf die Populationsgröße (d. h. die Anzahl der Individuen ist eher klein, nämlich N). Man nimmt an, dass die Zufallspaarung wie eine Ziehung mit Zurücklegen aus dem sogenannten **Genpool** der Eltern-generation erfolgt.

Bezeichnet X_k die Zufallsvariable der absoluten Häufigkeit des Gens A in der k -ten Generation, so ergibt sich

$$\mathbb{E}[X_{k+1}] = \mathbb{E}[X_k]$$

sofern man beachtet, dass unter der Voraussetzung $X_k = j$ die Zufallsvariable X_{k+1} binomialverteilt mit den Parametern $\left(2N, \frac{j}{2N}\right)$ ist.

Man kann nun zeigen, dass mit Wahrscheinlichkeit 1 eines der beiden Gene ausstirbt. Für den Grenzwert $X_\infty := \lim_{k \rightarrow \infty} X_k$ gilt nämlich

$$\mathbb{P}[X_\infty = j] = \begin{cases} 1 - \frac{\mathbb{E}[X_0]}{2N} & \text{falls } j = 0 \\ 0 & \text{sonst} \\ \frac{\mathbb{E}[X_0]}{2N} & \text{falls } j = 2N \end{cases}$$

8 Genetische Algebren

Unter einer **Algebra** versteht man einen Vektorraum $(A, +, \cdot)$ über \mathbb{R} , der zusätzlich noch mit einer Ringmultiplikation $*$ versehen ist, sodass $(A, +, *)$ (mit derselben Addition wie vorhin) ein Ring ist, und außerdem die zwei Multiplikationen im folgenden Sinn verträglich sind:

$$\alpha \cdot (x * y) = (\alpha \cdot x) * y = x * (\alpha \cdot y) \quad \forall \alpha \in \mathbb{R}, \quad \forall x, y \in A$$

Eine **genetische Algebra** (auch **Algebra mit genetischer Realisierung** genannt) ist dann eine Algebra

$$A = \left\{ \sum_{j=1}^n \alpha_j \cdot a_j \mid \alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R} \right\}$$

mit den Basisvektoren a_1, a_2, \dots, a_n , für die gilt

$$a_i * a_j = \sum_{k=1}^n \alpha_{ijk} \cdot a_k \quad \forall i, j \in \{1, 2, \dots, n\}$$

wobei

$$\begin{aligned} \alpha_{ijk} &\geq 0 & \forall i, j, k \in \{1, 2, \dots, n\} \\ \sum_{k=1}^n \alpha_{ijk} &= 1 & \forall i, j \in \{1, 2, \dots, n\} \end{aligned}$$

Eine Algebra A heißt **barisch**, wenn es einen nicht-trivialen Homomorphismus von A auf \mathbb{R} gibt.

Eine Algebra A heißt **dibaris**, wenn es einen Homomorphismus von A auf die sogenannte **Geschlechtsdifferenzierungsalgebra** S gibt, wobei

$$S = \{ \alpha \cdot w + \beta \cdot m \mid \alpha, \beta \in \mathbb{R} \}$$

mit

$$m * m = w * w = 0 \quad \text{und} \quad m * w = w * m = \frac{1}{2} \cdot m + \frac{1}{2} \cdot w$$

Ist die Algebra A dibaris, so ist A^2 barisch.

9 Fisher-Wright-Selektionsmodell

In diesem Abschnitt betrachten wir eine Population, die ideal ist bis auf Selektionseinflüsse. Zwischen dem Neugeborenen- und dem Erwachsenenstadium findet also eine Selektion statt, sodass ein Individuum mit Genotyp AA , Aa bzw. aa nur mit Wahrscheinlichkeit f , g bzw. h (die sogenannten **Fitnessparameter**) überlebt. Daraus ergibt sich eine **mittlere Fitness** $\Phi(p)$ der Population von $\Phi(p) = fp^2 + 2gpq + hq^2$.

Der **Fundamentalsatz der natürlichen Selektion** sagt nun aus, dass beim Übergang von einer Generation auf die nächste die mittlere Fitness auf jeden Fall zunimmt, sofern sich die Genhäufigkeiten p und q ändern. Es lässt sich außerdem zeigen, dass die Genhäufigkeit von A konvergiert, und zwar (je nach Wert von f , g bzw. h) gegen 0 (null), 1 (eins) oder $\bar{p} = \frac{h-g}{f-2g+h}$.

Analoge Aussagen gelten wieder im Fall von n Allelen.

10 Genotypen verwandter Individuen

In diesem Abschnitt sollen nur jene Individuen, die direkt auf den Abstammungslinien liegen, als verwandt betrachtet werden, nicht jedoch „angeheiratete“ Personen.

Definiert man dann für jedes $k, l \geq 0$ eine Matrix B_{kl} als Matrix der Wahrscheinlichkeit, dass ein Individuum den Genotyp $A_i A_j$ besitzt, wenn ein Nachkomme l -ten Grades der Eltern k -ten Grades den Genotyp $A_s A_t$ besitzt, so erhält man:

$$\begin{aligned} B_{k0} &= B_{0k} = B^k & \forall k \in \mathbb{N} \\ B_{k,l} &= B^{k+l-1} & \forall (k, l) \in \mathbb{N}^+ \times \mathbb{N}^+ \setminus \{(1, 1)\} \\ B_{1,1} &= G \end{aligned}$$

mit

$$B = \begin{pmatrix} p & \frac{p}{2} & 0 \\ q & \frac{1}{2} & p \\ 0 & \frac{q}{2} & q \end{pmatrix} \quad \text{bzw.} \quad G = \begin{pmatrix} \frac{(p+1)^2}{4} & \frac{p \cdot (p+1)}{4} & \frac{p^2}{4} \\ \frac{(p+1) \cdot q}{2} & \frac{p \cdot q}{2} & \frac{(q+1) \cdot p}{2} \\ \frac{q^2}{4} & \frac{q \cdot (q+1)}{4} & \frac{(q+1)^2}{4} \end{pmatrix}$$

Man kann so beispielsweise die Wahrscheinlichkeit des Auftretens von Erbkrankheiten berechnen. Erbliche Defekte treten nämlich meistens im Zusammenhang mit rezessiven Genen, d. h. beim Genotyp aa auf. Mit Hilfe obiger Formel kann man dann ausrechnen, wie hoch die Wahrscheinlichkeit ist, dass ein Individuum an der Erbkrankheit leidet, wenn z. B. ein Onkel daran erkrankt ist. Es zeigt sich, dass die Wahrscheinlichkeit, den Genotyp aa zu haben, am höchsten ist, wenn dieser Genotyp auch bei einem Bruder bzw. einer Schwester auftritt.

Index

\wedge -irreduzibel, 4

Algebra, 5

barische, 5

dibarische, 6

genetische, 5

mit genetischer Realisierung, 5

Aufspaltungsverhältnis, charakteristisches,
1

de-Finetti-Diagramm, 2

Faktorvereinigungsdarstellung, 3

Faktorvereinigungssystem, 4

Fitness, mittlere, 6

Fitnessparameter, 6

FU-Darstellung, 3

FU-System, 4

Fundamentalsatz der natürlichen Selekti-
on, 6

Genpool, 5

Geschlechtsdifferenzierungsalgebra, 6

Hardy-Weinberg-Gleichgewicht, 2

Hardy-Weinberg-Parabel, 2

Markov-Kette, 4

Mendel, 1

Mutation, 4

Phänotypsystem, 3

Population, 1

ideale, 1

Satz von Hardy-Weinberg, 2

stochastische Matrix, 4

Unabhängigkeitsregel, 1

Uniformitätsregel, 1

Vererbung, geschlechtsgebundene, 3

Wright-Modell, 5

Züchtung reinerbiger Individuen, 2