

Selektion

Vorbemerkungen zu deterministischer Dynamik

M. Birkner, Stochastic population models

JGU Mainz, WS 2025/26

Inhalt

Hardy-Weinberg-Gleichgewicht

Allelhäufigkeiten unter Selektion

- Ein deterministisches Modell für Selektion

- Verschiedene Szenarien von (diploider) Selektion

- Gerichtete Selektion

- Balancierende Selektion

Hardy-Weinberg-Gleichgewicht (1908)



Godfrey Harold Hardy
(1877–1947)



Wilhelm Weinberg
(1862–1937)

Idealisierte Population („Mathematisches Modell“): sehr groß, diploid, hermaphroditisch. An einem Genort gebe es zwei verschiedene Allele, A und a .

Annahme: „Neutralität“, d.h. Reproduktionserfolg unabhängig vom Genotyp.

Genotypenhäufigkeiten heute:

Genotyp	AA	Aa	aa	$(x_{AA} + x_{Aa} + x_{aa} = 1)$
Anteil	x_{AA}	x_{Aa}	x_{aa}	

Das bedeutet für die Allelhäufigkeiten

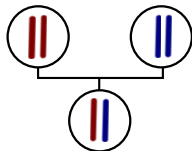
Allel	A	a
Anteil	$p_A = x_{AA} + \frac{1}{2}x_{Aa}$	$p_a = \frac{1}{2}x_{Aa} + x_{aa}$

(offenbar auch $p_A + p_a = 1$)

Wir nehmen weiter an:

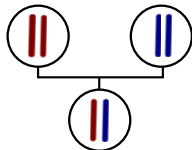
„rein zufällige Paarungen“,

Mendelsche Segregation



Genotyphäufigkeiten in der nächsten Generation?

Heute:	Genotyp	AA	Aa	aa
	Anteil	x_{AA}	x_{Aa}	x_{aa}
	Allel	A		a
	Anteil	$p_A = x_{AA} + \frac{1}{2}x_{Aa}$		$p_a = \frac{1}{2}x_{Aa} + x_{aa}$



Nächste Generation:

Genotyp	AA	Aa	aa	Allel	A	a
Anteil	x'_{AA}	x'_{Aa}	x'_{aa}	Anteil	p'_A	p'_a

$$x'_{AA} = x_{AA}^2 + 2x_{AA}x_{Aa} \cdot \frac{1}{2} + x_{Aa}^2 \cdot \frac{1}{2} \cdot \frac{1}{2} = (x_{AA} + \frac{1}{2}x_{Aa})^2 = p_A^2$$

Analog:

$$x'_{Aa} = 2(x_{AA} + \frac{1}{2}x_{Aa})(x_{aa} + \frac{1}{2}x_{Aa}) = 2p_A p_a$$

$$x'_{aa} = x_{aa}^2 + 2x_{Aa}x_{aa} \cdot \frac{1}{2} + x_{Aa}^2 \cdot \frac{1}{2} \cdot \frac{1}{2} = (x_{aa} + \frac{1}{2}x_{Aa})^2 = p_a^2$$

Wir sehen:

Anteile heute: x_{AA}, x_{Aa}, x_{aa} (Genotypen),

$$p_A = x_{AA} + \frac{1}{2}x_{Aa}, p_a = \frac{1}{2}x_{Aa} + x_{aa} \text{ (Allele)}$$

Anteile nächste Generation:

$$x'_{AA} = p_A^2, x'_{Aa} = 2p_A p_a, x'_{aa} = p_a^2 \text{ (Genotypen),}$$

$$p'_A = x'_{AA} + \frac{1}{2}x'_{Aa} = p_A^2 + p_A p_a = p_A(p_A + p_a) = p_A,$$

$$p'_a = \frac{1}{2}x'_{Aa} + x'_{aa} = p_A p_a + p_a^2 = p_a(p_A + p_a) = p_a \text{ (Allele)}$$

Also:

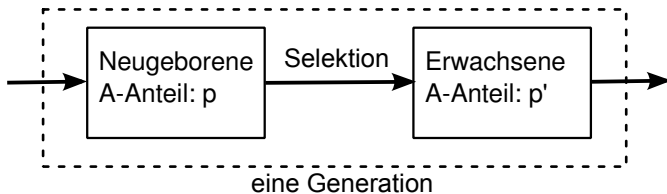
- ▶ Allelhäufigkeiten sind konstant über die Generationen.
- ▶ Unabhängig von den ursprünglichen Genotyphäufigkeiten stellt sich für die Genotypen AA, Aa, aa nach einer Generation das Verhältnis

$$p^2 : 2p(1 - p) : (1 - p)^2$$

ein und ändert sich dann nicht mehr:

Hardy-Weinberg-Gleichgewicht

Ein Modell für Selektion



Genotyp	AA	Aa	aa
(rel.) Fitness	w_{AA}	w_{Aa}	w_{aa}

Interpretation: Die Chance eines Nachkommen vom Typ AA , bis zum Reproduktionsalter zu überleben, ist proportional zu w_{AA} , etc.

Beispiel 1:

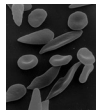
	AA	Aa	aa
	1	0,95	0,9



(nachteilige
Flügel­färbung bei
Callimorpha dominula)

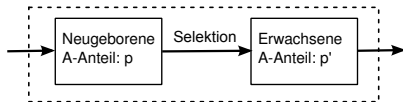
Beispiel 2:

	AA	Aa	aa
	1	1,17	0



(Sichelzellanämie und
endemische Malaria)

Ein Modell für Selektion: die nächste Generation



Genotyp	<i>AA</i>	<i>Aa</i>	<i>aa</i>
(rel.) Fitness	w_{AA}	w_{Aa}	w_{aa}
Anteil heute	p^2	$2pq$	q^2
rel. Anteil nächste Gen.	$p^2 w_{AA}$	$2pq w_{Aa}$	$q^2 w_{aa}$
Anteil nächste Gen.	$\frac{p^2 w_{AA}}{w_{\text{ges}}}$	$\frac{2pq w_{Aa}}{w_{\text{ges}}}$	$\frac{q^2 w_{aa}}{w_{\text{ges}}}$

wobei

p ($= p_A$) Anteil des *A*-Allels,

$q = 1 - p$ ($= p_a$) Anteil des *a*-Allels,

$$w_{\text{ges}} = p^2 w_{AA} + 2pq w_{Aa} + q^2 w_{aa}$$

die „Gesamtfitness“ der (heutigen) Population

Umparametrisierung

Wir schreiben statt w_{AA} , w_{Aa} , w_{aa} lieber

Genotyp	AA	Aa	aa
(rel.) Fitness	1	$1 - hs$	$1 - s$

wobei

s : Selektionskoeffizient, $s \leq 1$,

h : (Koeffizient des) „Heterozygoteneffekt(s)“.

Beispiel 1:

	AA	Aa	aa
	1	0,95	0,9

$s = 0,1$, $h = 0,5$

Beispiel 2:

	AA	Aa	aa
	1	1,17	0

$s = 1$, $h = -0,17$

Verschiedene Szenarien von (diploider) Selektion

Genotyp	AA	Aa	aa
(rel.) Fitness	1	$1 - hs$	$1 - s$

$s = 0$: neutraler Fall

$s \neq 0$ (wir betrachten den Fall $0 < s \leq 1$, sonst vertausche die Rollen von A und a)

$h = 1$: A rezessiv

$h = 0$: A dominant

$0 < h < 1$: unvollständige Dominanz

$h < 0$: Überdominanz (Aa „am fittesten“)

$h > 1$: Unterdominanz

(Bem.: $h = 1/2$: „additiver Fitnessseffekt“, ist mathematisch besonders angenehm und häufig für $s \approx 0$ gerechtfertigt)

Fragen

Verändern sich die Allelhäufigkeiten im Laufe der Zeit?

Wenn ja, wie?

Wie sieht es nach sehr langer Zeit aus?

Wird sich das A -Allel durchsetzen?

Mathematisch formuliert:

Sei p_n der Anteil A -Allele in der Population in Generation n ,
 $n = 0, 1, 2, \dots$

Wie verhält sich p_n , wenn n nach ∞ strebt?

Die Antwort hängt (hauptsächlich) von h ab ...

Änderung des A -Anteils als Funktion des aktuellen Anteils

Anteil heute: p relative Fitness: $\frac{AA}{1} \mid \frac{Aa}{1 - hs} \mid \frac{aa}{1 - s}$

Anteil in der nächsten Generation:

$$N(p) := \frac{p^2 \cdot 1 + \frac{1}{2} \cdot 2pq \cdot (1 - hs)}{p^2 + 2pq(1 - hs) + q^2(1 - s)}$$

(mit $q = 1 - p$)

Änderung über eine Generation

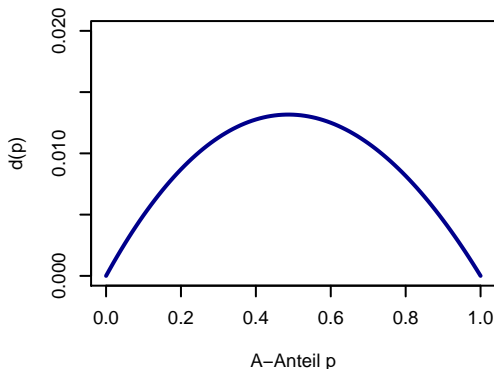
$$\begin{aligned} d = N(p) - p &= \frac{p^2 \cdot 1 + \frac{1}{2} \cdot 2pq \cdot (1 - hs)}{p^2 + 2pq(1 - hs) + q^2(1 - s)} - p \\ &= \frac{spq(ph + q(1 - h))}{p^2 + 2pq(1 - hs) + q^2(1 - s)} \end{aligned}$$

[Details der Rechnung ggf. an der Tafel]

$0 \leq h \leq 1$: gerichtete Selektion (Diagramm für $s=0,1$, $h=0,5$)

Änderung des A-Anteils als Funktion von p :

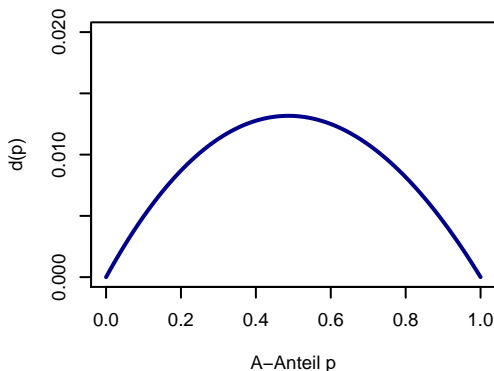
$$d(p) = \frac{sp(1-p)(ph + (1-p)(1-h))}{p^2 + 2p(1-p)(1-hs) + (1-p)^2(1-s)}$$



Wir sehen: $d(p) > 0$ außer für $p \in \{0, 1\}$, d.h. der A-Anteil nimmt stets zu.

$0 \leq h \leq 1$: gerichtete Selektion (Diagramm für $s=0,1$, $h=0,5$)

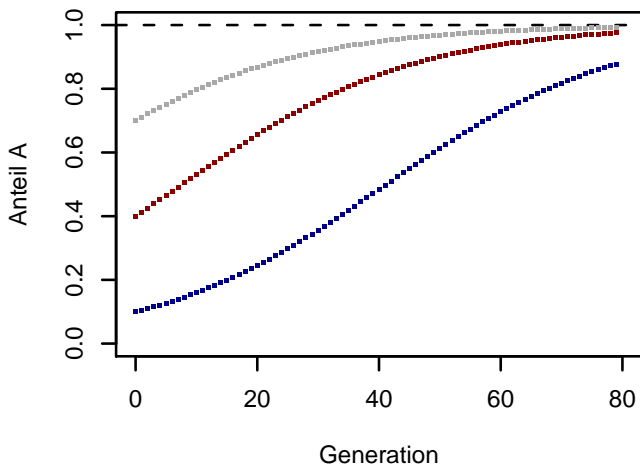
Änderung des A-Anteils $d(p)$ als Funktion von p :



$d(p) > 0$, d.h. der A-Anteil nimmt stets zu (außer für $p \in \{0, 1\}$) :

$$p_n = N(p_{n-1}) = p_{n-1} + d(p_{n-1}) > p_{n-1}$$

$0 \leq h \leq 1$: gerichtete Selektion, Konvergenz des A-Anteils



Anteil A in Abhängigkeit der Generation für verschiedene Startwerte ($s=0,1$, $h=0,5$)

$0 \leq h \leq 1$: gerichtete Selektion, Konvergenz des A-Anteils

$p_n = p_{n-1} + d(p_{n-1})$, also für $0 < p_1 < 1$:

$p_1 < p_2 < p_3 < \dots \leq 1$, andererseits kann p_n für $\varepsilon > 0$ nicht für alle n unterhalb $1 - \varepsilon$ liegen (denn die Funktion $d(\cdot)$ ist strikt positiv im Intervall $[p_1, 1 - \varepsilon]$), somit

Satz

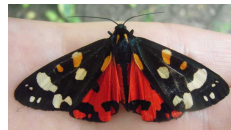
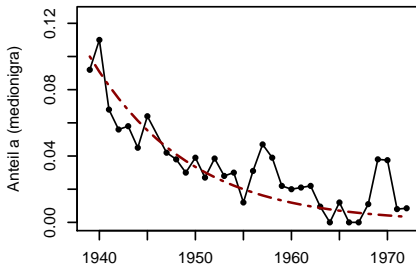
Im Fall $0 \leq h \leq 1$ gilt für jedes $p_1 \in (0, 1]$: $p_n \rightarrow p^* = 1$ für $n \rightarrow \infty$

Biologische Interpretation

Im Fall $0 \leq h \leq 1$ setzt sich stets Allel A durch (sofern es zu Beginn vorhanden ist). Die Population wird auf lange Sicht nur aus AA-Individuen bestehen.

Beispiel: Das *medionigra*-Allel (*a*) in einer Population von *Callimorpha dominula* (dt. Schönbär) in Oxford

Beobachteter Anteil *a* 1939–1972 und Modellvorhersage



Nach Kap. 3 in John H. Gillespie, *Population genetics : a concise guide*, Johns Hopkins Univ. Press, 1998, siehe auch Ford, E.B. and P.M. Sheppard, The *medionigra* polymorphism of *Panaxia dominula*. *Heredity* 24:112–134, 1969.

Bemerkung. Die Modellkurve passt recht gut, das beweist allerdings nicht, dass tatsächlich gerichtete Selektion für die beobachteten Änderungen des *a*-Anteils verantwortlich ist — es könnte andere Effekte geben, die z.T. kontrovers in der Literatur diskutiert werden.

$h < 0$: balancierende Selektion

Erinnerung: Rel. Fitness $1 : 1 - hs : 1 - s$, also bedeutet $h < 0$, dass Heterozygote der „fitteste“ Typ sind.

(Erneut:) Änderung als Funktion von p :

$$d(p) = \frac{sp(1-p)(ph + (1-p)(1-h))}{p^2 + 2p(1-p)(1-hs) + (1-p)^2(1-s)}$$

Für welche p ist $d(p) = 0$?

$p = 0$, $p = 1$, oder p Lösung von

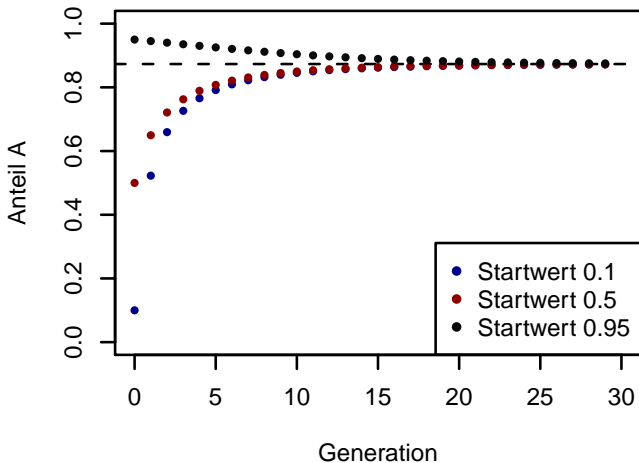
$$ph + (1-p)(1-h) = 0,$$

$$\text{d.h. } p = \hat{p} = \frac{1-h}{1-2h}.$$

Im Beispiel: $h=-0,17$, $\hat{p}=0,87$ (also $1-\hat{p}=0,13$)

$h < 0$: balanzierende Selektion

Zeitliche Entwicklung des A-Anteils bei verschiedenen Startwerten ($s=1$, $h=-0,17$)



$h < 0$: balanzierende Selektion, Konvergenz

$$p_n = p_{n-1} + d(p_{n-1}), \quad \hat{p} = \frac{1-h}{1-2h}$$

Man kann (leicht) zeigen, dass $p_1 < p_2 < p_3 < \dots \leq \hat{p}$ wenn

$0 < p_1 < \hat{p}$ und

$p_1 > p_2 > p_3 > \dots \geq \hat{p}$ wenn $\hat{p} < p_1 < 1$,

somit

Satz

Im Fall $h < 0$ gilt für jedes $p_1 \in (0, 1)$:

$$\lim_{n \rightarrow \infty} p_n = \hat{p} = \frac{1-h}{1-2h}.$$

(Die Folge p_n konvergiert gegen \hat{p} .)

Biologische Interpretation

Im Fall $h < 0$ bleiben beide Allele in der Population erhalten, das genaue Verhältnis hängt von h ab (das hier die Stärke der Überdominanz misst).



Übrigens: Der Schönbär war Schmetterling des Jahres 2010

www.bund-nrw-naturschutzstiftung.de/schmetterling2010.htm