

theoretische Biologie (SS 2017)

Inhaltsverzeichnis

1	Vorlesung 06.04.2017	1
1.1	Begriffe und Konzepte	1
2	Vorlesung 13.04.2017	1
2.1	Begriffe und Konzepte	1
3	Vorlesung 20.04.2017	1
4	Vorlesung 27.04.2017	2
4.1	Teil 1: Dynamische Systeme	2
4.2	Qualitative Analyse von DS	3
4.3	Teil 2: Genkonzept	4
5	Vorlesung 04.05.2017	4
6	Vorlesung 11.05.2017	5
6.1	Teil 1: Populationsdynamik	5
6.2	Teil 2: Diskussion zu den Vorträgen beim mitteldeutschen Bioinformatik-Meeting 2017	5
7	Vorlesung 18.05.2017	6
7.1	Musterbildung	6
8	Vorlesung 01.06.2017	7
8.1	Teil 1: Musterbildung	7
8.2	Teil 2: Cat Coat Colors	7
8.3	Vergleich Übungsaufgaben	7
9	Vorlesung 08.06.2017	8
9.1	Teil 1: Fitnesslandschaften	8
9.1.1	Genetische Algorithmen	8
9.1.2	3D-Strukturen	9
9.1.3	Optimierung auf Landschaften	9
9.1.4	Autokorrelationsfunktionen	9
9.2	Übung farbliche Ausprägung Katzenfell und beteiligte Gene . . .	10

1 Vorlesung 06.04.2017

1.1 Begriffe und Konzepte

- Begriffsbildung am Beispiel Information (Was ist Information? [Prüfungsrelevant!])
- Vorlesungsunterlagen siehe ¹
- Begriffsbildung am Beispiel Gen [Prüfungsrelevant!]
 - Welche Überschneidungen, welche Differenzen?
 - Welche Genkonzepte gibt es? (zu lesen: siehe ² und ³)

2 Vorlesung 13.04.2017

2.1 Begriffe und Konzepte

- GWAS (Prof. Markus Scholz)
- Diskussion zum Begriff Struktur

3 Vorlesung 20.04.2017

- Gendefinition im Kontext der Messtechnik⁴
- random mating, rezessive und dominante Epistasie ???

¹<http://www.bioinf.uni-leipzig.de/Leere/SS17/TBio/concepts.pdf>

²http://www.bioinf.uni-leipzig.de/Leere/SS17/TBio/Gerstein07_gene_definition.pdf

³http://www.bioinf.uni-leipzig.de/Leere/SS17/TBio/Stadler09_gene_definition.pdf

⁴http://www.bioinf.uni-leipzig.de/Leere/SS17/TBio/gene_definition.pdf

4 Vorlesung 27.04.2017

4.1 Teil 1: Dynamische Systeme

- **diskrete Zeit: "Generationen"**

X_1, X_2, \dots

$$X_n = F(X_{n-1}) =: X_{n-1} + \overbrace{f(X_{n-1})}^{\text{Änderung des Zustandes}}$$
$$X_{n+1} = F(X_n) = F(F(X_{n-1}))$$

Beispiel:

$$X_n = (1 + \underbrace{a}_{\text{effektive Vermehrungsrate}}) \cdot x_{n-1} = \text{Geburtenrate} - \text{Sterberate}$$

Anfangsbedingung: $X_{t_0} = X_0$

Bedingung: effektive Vermehrungsrate a verändert sich nicht

Lösung: $X_n = (1 + a)^n \cdot x_0$

im allgemeinen mit zeitlich variablen Vermehrungsraten: $X_n = \prod_{i=0}^{n-1} (1+a_i) \cdot x_0$

3 verschiedene Resultate:

- $+\infty$ für $a > 0$
- x_0 für $a = 0$
- 0 für $a < 0$

$$X_n = X_{n-1} + a \cdot \underbrace{X_{n-1}}_{f(X_{n-1})}$$

$$f(X_{n-1}) = X_{n-1} \cdot r(X_{n-1})$$

mit $r(0)=\text{const.}$ entspricht autonomer Wachstumsrate $[\lim(x \rightarrow 0)r(x) \in R_0^+]$

- **kontinuierliche Zeit**

$$x(t + \Delta t) = x(t) + f(x(t)) \cdot \Delta t$$

$$\frac{x(t + \delta t) - x(t)}{\delta t} = f(x(t))$$

$$\lim(\delta t \rightarrow 0) \frac{x(t + \delta t) - x(t)}{\delta t} = \frac{\delta x}{\delta t} \hat{=} \text{zeitlicher Ableitung von } x$$
$$= \dot{x} = f(x)$$

Beispiel:

$$\dot{x} = a \cdot x, x(0) = x_0$$

$$\frac{dx}{dt} = a \cdot x$$

$$\frac{dx}{a \cdot x} = dt$$

$$\int_{x_0}^{x(t)} \frac{1}{a \cdot x} \cdot dx = \int_0^t 1 \cdot dt = t$$

$$\dot{x} = f(x) \Rightarrow \int_{x_0}^{x(t)} \frac{1}{f(x)} = \int_0^1 dt = t$$

$$\frac{1}{a} \int \frac{1}{x} dx = \frac{1}{a} \cdot \ln(x)$$

$$\frac{1}{a} \ln(x(t)) - \frac{1}{a} \ln(x_0) = a \cdot t$$

$$\ln(x(t)) = at + \ln(x_0)$$

$$x(t) = e^{at} \cdot x_0$$

Wie machen wir das Model realistischer?

$f(x)$ und $r(x)$ muss für sehr große x dann ≤ 0 werden.

$$\dot{x} = f(x) = x \cdot (a - bx)$$

Übungsaufgabe:

1. Löse $\dot{x} = x(a - bx)$ mit $x(0) = x_0$
2. Löse $x' = x + x(a - bx)$ mit $x(0) = x_0$

4.2 Qualitative Analyse von DS

1. Fixpunkte: keine zeitliche Veränderung ($x' = x, \dot{x} = 0$)
d.h. diskret und kontinuierlich, $f(x)=0$
Welche Fixpunkte gibt es? im Beispiel $x(a-bx)=0$

(a) $x=0 \rightarrow$ Population ausgestorben

(b) $a-bx=0 \rightarrow x = \frac{a}{b}$

Störung: $x(0) = \underbrace{\hat{x}}_{\text{Fixpunkt}} + \epsilon$ mit sehr kleinem ϵ

$$\dot{x} = f(x) = f(\hat{x} + \epsilon) = \dot{\epsilon}$$

mit $x = \hat{x} + \epsilon$

$$\dot{x} = \frac{\delta \hat{x}}{\delta t} + \dot{\epsilon}$$

$$\dot{\epsilon} = f(\hat{x} + \epsilon)$$

mit Taylorreihenentwicklung: $0 = f(\hat{x}) + \epsilon \frac{\delta f}{\delta x}(\hat{x}) + O(\epsilon^2)$

Für sehr kleine Störungen:

$$\dot{\epsilon} = \frac{\delta f}{\delta x}(\hat{x}) \cdot \epsilon + \cancel{O(\epsilon^2)}$$

Linearisierung der Differentialgleichung $x = f(x)$ in der Nähe eines Fixpunktes

$$\hat{x}: \epsilon(x) = e^{[\frac{\delta f}{\delta x}(\hat{x})] \cdot t}$$

$$\epsilon_0 = x_0 - \hat{x}$$

$\epsilon_0 \leftarrow$ initiale Störung

- Störung wird gedämpft wenn $\frac{\delta f}{\delta x}(\hat{x}) < 0 = \text{STABIL}$

- Störung eskaliert wenn $\frac{\delta f}{\delta x}(\hat{x}) > 0 = \text{INSTABIL}$

im Diskreten Fall?

$$x' = x + f(x) \text{ mit } x = \hat{x} + \epsilon$$

$$\hat{x}' + \epsilon' = \hat{x} + \epsilon + f(\hat{x} + \epsilon) = f(\hat{x}) + \epsilon \cdot \frac{\delta f}{\delta x}(\hat{x}) + \text{Rest}(\epsilon)$$

$$\epsilon' = \epsilon(1 + \frac{\delta f}{\delta x}(\hat{x})) \text{ mit:}$$

- $\epsilon \rightarrow 0$ wenn $|1 + \frac{\delta f}{\delta x}(\hat{x})| < 1$
- $\epsilon \rightarrow \infty$ wenn $|1 + \frac{\delta f}{\delta x}(\hat{x})| > 1$

jetzt Mehrdimensional:

- Räuber x: $f_x(x, y) = x(-a + by - cx)$
- Beute y: $f_y(x, y) = y(+d - ex - gy)$

Fixpunkte:

- $f_x(x, y) = 0$
- $f_y(x, y) = 0$

Stabilität:

gegeben durch

- $\frac{\delta f_x}{\delta x}(\hat{x}, \hat{y}) \quad \frac{\delta f_x}{\delta y}(\hat{x}, \hat{y})$
- $\frac{\delta f_y}{\delta x}(\hat{x}, \hat{y}) \quad \frac{\delta f_y}{\delta y}(\hat{x}, \hat{y})$

Übungsaufgabe 2:

Bestimme die Fixpunkte von Räuber-Beute-Modell für $a, b, c, d, e, g > 0$

Welche Fixpunkte gibt es immer? Wieviele sind das?

4.3 Teil 2: Genkonzept

- Unterschiede und Überschneidungen zwischen den beiden in den Papern vorgestellten Genkonzepten (siehe Vorlesung 13.04.2017) **[Prüfungsrelevant]**

5 Vorlesung 04.05.2017

- Vorlesung entfallen wegen: Mitteldeutschen Bioinformatik-Meeting 2017⁵

⁵<http://me17.bioinf.uni-leipzig.de/>

6 Vorlesung 11.05.2017

6.1 Teil 1: Populationsdynamik

6.2 Teil 2: Diskussion zu den Vorträgen beim mitteldeutschen Bioinformatik-Meeting 2017

7 Vorlesung 18.05.2017

7.1 Musterbildung

8 Vorlesung 01.06.2017

8.1 Teil 1: Musterbildung

8.2 Teil 2: Cat Coat Colors

8.3 Vergleich Übungsaufgaben

9 Vorlesung 08.06.2017

9.1 Teil 1: Fitnesslandschaften

evol. Theorie: Wachstumsrate in Räuber-Beute-Modellen + Reproduktion mit Variation + "Survival of the fittests" → Wachstumsrate einer Population in einer gege. Umgebung

Günther Wagner: Messtheorie von Fitness (measurement theorie)

X... Suchraum (genotyp—phänotyp), allgemein irgendeine Repräsentation der betrachteten Taxa

Ähnlichkeitsstruktur σ

Fitnessfunktion: $f : x \rightarrow R$

mit R =totale geordnete Menge ($f_1, f_2 \in R : f_1 < f_2, f_1 > f_2, f_1 = f_2$)

Begründer: Sewall Wright (~ 1930)

siehe zurückliegendes Bild: Individuum hat höhere Wahrscheinlichkeit Erbgut in nächste Generation zu übertragen ("Verbesserung")

9.1.1 Genetische Algorithmen

Idee: benutze künstliche Evolution um Optimierungsprobleme zu lösen

1. Population $A \subseteq X$
2. Nachfolgerpopulation von Kandidaten $C(A)$
3. Selektiere die Besten bezüglich Fitnessfunktion $x \in C(A)$
4. zurück zu 1.

genetische Algorithmen, evolutionäre Programmierung (Rechenberg, Schwefel $\sim 1960/70$)

geg.: RNA oder Proteinsequenz α ges.: Alle möglichen Strukturen x , die α einnehmen kann → Menge x von Konfigurationen

Energiefunktion $f : X \rightarrow R$

z.B. Loop basiertes Energiemodell für RNA Sekundärstrukturen

Lenskis E.Coli Zucht⁶

X... Menge von Gen oder Genomsequenzen

σ ... Mutationen (hauptsächlich Substitution, Insertion, Deletion)

(X, σ) ... Suchraum \Leftrightarrow Graphen über $\{A, G, T, C\}^n$

mit n =Sequenzlänge

mit Kanten=Hammingdistanz 1 (| Levensteindistanz 1)

9.1.2 3D-Strukturen

Proteinstruktur = $(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)$ mit \vec{x}_1 =3D-Koordinaten für Atom 1

Constraint: Bindungswinkel, Bindungslängen

X... alle möglichen 3D-Einbettungen des Proteins

Nachbar: $\|\vec{x} - \vec{x}'\| = \sum_i |\vec{x}_i - \vec{x}'_i| < \epsilon$ für gegebenes $\epsilon > 0$

Wenn RNA Sekundärstrukturen?

X... Menge aller erlaubten Strukturen $[(,),.]$

$x \sim y$ wenn x und y sich durch ein Basenpaar unterscheiden \Rightarrow Graph

Beispiel:

9.1.3 Optimierung auf Landschaften

\rightarrow max, min finden

Wie misst man Rauheit?

Minimum: $\hat{x} \in X$ sodass $\forall y$ Nachbar von $\hat{x} : f(\hat{x}) \leq f(y)$

für metrischen (kontinuierlichen) Raum: $\forall y : |\hat{x} - y| < \epsilon$

Maximum: $f(\hat{x}) \geq f(y)$

Was möchte man messen?

- # lokale Minima, nur gut bei kleinen Instanzen, daher sampeln! (zufällige x wählen und bestimmen ab Minimum)
- mittlere Länge von adaptiven walks
 $x_0, x_1, x_2, \dots, x_l$ sodass x_i Nachbar von x_{i-1} ist und $\underbrace{f(x_i) > f(x_{i-1}))}_{*}$ $i=1 \dots l$
* für Fitness (für Energie $<$)
- Alternative: gradient walks (Weg des stetigen Anstiegs), Distanz zum "nächstgelegenen lokalen Min/Max

9.1.4 Autokorrelationsfunktionen

$x_0, x_1, x_2, \dots, x_l$ sodass x_i Nachbar von x_{i-1}

betrachte Folge der Funktionswerte $f(x_0), f(x_1), \dots$ betrachte das als Signal (Zeitreihe)

$$\varrho(\tau) = \frac{\langle f(x_t) \cdot f(x_{t+\tau}) \rangle_t - \langle f \rangle_t^2}{\langle f^2 \rangle_t - \langle f \rangle_t^2}$$

$$\begin{aligned} &< f_t > \dots \text{Mittelwert über die } f(x) \\ &< f_t > := \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} f(x_t) \\ &< f(x_t) \cdot f(x_{t+\tau}) > := \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} f(x_t) f(x_{t+\tau}) \end{aligned}$$

X... Menge von Gen oder Genomsequenzen

σ ... Graph, regulär (jedes x hat gleich viele Nachbarn D)

A... Adjazenz von (X, σ)

$$\varrho(\tau) = \frac{(f(\frac{1}{D} \cdot A)^\tau \cdot f) - (f)^2}{(f^2) - (f)^2}$$

\Rightarrow leichter so auf Graphen als direkt auf Fitness (Funktion)

Korrelationslänge:

Funktion in Abhängigkeit der Verschiebung von τ

$$L_c = \sum_{\tau=0}^{\infty} \varrho(\tau)$$

$$(f) = \frac{1}{|X|} \sum_{x \in X} f(x)$$

$$(f^2) = \sum_{x \in X} f(x)^2$$

$$(f, \frac{1}{D} \cdot A \cdot f) := \sum_{x \in X} \sum_{y \in X} \frac{1}{D} f(y) \cdot A_{yx} \cdot f(x)$$

wenn $(f) = 0$ folgt vereinfachte Gleichung

$$(f^2) = 1 \rightarrow \varrho(\tau) = \underbrace{\langle f(x_t), f(x_{t+\tau}) \rangle}_{\text{Graphstruktur}} = (f, \frac{1}{D} \cdot A \cdot f)$$

Beispiel:

Kostenfunktion in Fall 1 ändert sich stärker als in Falls 2

$$\Rightarrow L_R \simeq 2L_T$$

?(Länge Korrelationslängen \rightarrow Lange Wege zum nächsten Minimum \rightarrow gut!)

9.2 Übung farbliche Ausprägung Katzenfell und beteiligte Gene