A. Wakolbinger, Stochastische Modelle der Populationsgenetik, SoSe 2016

Ergänzungsblatt zur Vorlesung 3:

Zur Dynamik austauschbarer Koaleszenten in diskreter Zeit

Im Folgenden seien $n \le N \in \mathbb{N}$ fest, und
$\mathcal{N} = (\mathcal{N}_1, \dots, \mathcal{N}_N)$ eine austauschbare $S_{N,N}$-wertige ZV (der *Kinderzahlvektor*).
A sei eine zufällige Abbildung von $[N]$ nach $[N]$, für die f.s. gegeben $\mathcal{N}$ gilt:

$$\# A^{-1}(i) = \mathcal{N}_i,\ i \in [N].$$

$(\mathcal{N}^{g-1}, A^g)$, $g \in \mathbb{Z}$, seien unabhängige Kopien von $(\mathcal{N}, A)$.

Unabhängig davon sei $J = (J(h))_{h\in[N]}$ eine rein zufällige Permutation von $[N]$.
Für $k = 1, 2, \dots$ setzen wir $A^0_{-k} := A^{-k+1} \circ \cdots \circ A^0$. Wir denken bei $A^0_{-k}(i)$ an den Ahnen in Generation $-k$ von Individuum i aus der Generation 0 (und ergänzen die Definition durch $A^0_0 := \mathrm{id}\,|_{[N]}$).
Für $k = 0, 1, \dots$ und $n \le N$ setzen wir

$$\mathscr{C}_k := \{(A^0_{-k} \circ J)^{-1}(i) : i \in [N]\}, \qquad \mathscr{C}^n_k := \mathscr{C}_k \cap [n].$$

Wir bemerken: $\mathscr{C}_k$ wird erzeugt von der Relation $h_1 \overset{k}{\sim} h_2 \longleftrightarrow A^0_{-k}(J(h_1)) = A^0_{-k}(J(h_2)))$, $h_1, h_2 \in [N]$, und $\mathscr{C}^n_k$ wird erzeugt von derselben Relation auf $[n]$.

Lemma 1. Für jedes $k \ge 0$ gilt
(i) $\mathscr{C}^n_k$ ist eine austauschbare Partition von $[n]$.
(ii) Gegeben $\mathscr{C}^n_k = \xi$ (mit $\#\xi =: b$) ist $\{(A^0_{-k} \circ J)(h) : h \in [n]\}$ eine rein zufällige b-elementige Teilmenge von $[N]$.

Lemma 2. Sei $b \ge a \in \mathbb{N}$ und $\mathcal{P} := B_1, \dots, B_a$ eine Partition von $[b]$ mit $b_\ell := \#B_\ell$, $\ell = 1, \dots, a$. Sei $\nu \in S_{N,N}$, und sei $\mathcal{F} = (F_1, \dots, F_N)$ eine Partition von $[N]$ mit $\#F_i = \nu_i$, $i \in [N]$. Es sei $\mathcal{A} := (\mathcal{A}_1, \dots, \mathcal{A}_b)$ eine rein zufällige Stichprobe (ohne Zurücklegen) des Umfangs b aus $[N]$. Wir setzen $\ell \sim \ell'$, wenn $\mathcal{A}_\ell$ und $\mathcal{A}_{\ell'}$ in dasselbe Partitionselement von $\mathcal{F}$ fallen. Dann ist die Wahrscheinlichkeit, dass die zufällige Relation $\sim$ die Partition $\mathcal{P}$ induziert, gleich

$$\sum_{i_1,\dots,i_a \in [N]\,\text{paarw.versch}} \frac{1}{(N)_{b\downarrow}} \prod_{\ell=1}^{a} (\nu_{i_\ell})_{b_\ell\downarrow}$$

Dabei ist für $x \in \mathbb{R}$ und $a \in \mathbb{N}$ $\qquad (x)_{a\downarrow} := x(x-1)\cdots(x-a+1)$
die *a-te fallende Faktorielle von x*.

Korollar zu Lemma 1 und Lemma 2. $(\mathscr{C}^n_k)_{k\in\mathbb{N}_0}$ ist eine Markovkette mit Übergangswahrscheinlichkeiten

$$\mathbf{P}(\mathscr{C}^n_{k+1} = \eta | \mathscr{C}^n_k = \xi) = \frac{(N)_{a\downarrow}}{(N)_{b\downarrow}} \mathbf{E}\left[\prod_{\ell=1}^{a} (\mathcal{N}_\ell)_{b_\ell\downarrow}\right]$$

wobei $a := \#\eta \le \#\xi =: b$, und η so aus ξ hervorgeht, dass (für jedes $\ell = 1, \dots, a$) b_ℓ Partitionselemente von ξ zu einem Partitionselement von η vereinigt werden (mit $b_1 + \cdots + b_a = b$).

Wir nennen $(\mathscr{C}_k)_{k\in\mathbb{N}_0}$ den *austauschbare Koaleszenten (exchangeable coalescent)* zur Kinderzahlverteilung $\mathscr{L}(\mathcal{N})$ bei Populationsgröße N, und $(\mathscr{C}^n_k)_{k\in\mathbb{N}_0}$ den entsprechenden *austauschbaren n-Koaleszenten.*