

Betrachte einen Λ - n -Koaleszenten, längs dessen Kanten (bis zum MRCA) sich mit Rate $\theta/2$ Mutationen (gemäß dem infinite sites-Modell) ereignen

S_n = Gesamtanzahl Mutationen

Frequenzspektrum:

$\xi_i^{(n)}$ = # Mutationen, die in i der n Blätter sichtbar sind
($i = 1, \dots, n - 1$)

Für $\Lambda = \text{Beta}(2 - \alpha, \alpha)$ mit $\alpha \in (1, 2)$ gilt

$$\frac{S_n}{n^{2-\alpha}} \longrightarrow \frac{\theta}{2} \frac{\alpha(\alpha - 1)\Gamma(\alpha)}{2 - \alpha}$$
$$\frac{\xi_i^{(n)}}{n^{2-\alpha}} \longrightarrow \frac{\theta}{2} \alpha(\alpha - 1)^2 \frac{\Gamma(i + \alpha - 2)}{i!}$$

für $n \rightarrow \infty$

(J. Berestycki, N. Berestycki, J. Schweinsberg, Beta-coalescents and continuous stable random trees, Ann. Probab. 35, 1835–1887 (2007); dasselbe Resultat mit geänderten Konstanten, wenn nur $\Lambda(dx) = f(x)dx$ und $f(x) \sim cx^{1-\alpha}$ für $x \downarrow 0$ vorausgesetzt wird)

$$\frac{\xi_i^{(n)}}{S_n} = \frac{\xi_i^{(n)}/n^{2-\alpha}}{S_n/n^{2-\alpha}} \rightarrow \frac{\theta \alpha(\alpha-1)^2 \frac{\Gamma(i+\alpha-2)}{i!}}{2} / \frac{\theta \alpha(\alpha-1)\Gamma(\alpha)}{2-\alpha}$$

für $n \gg 1$ also typischerweise

$$\frac{\xi_i^{(n)}}{S_n} \approx \frac{(2-\alpha)\Gamma(i+\alpha-2)}{i!\Gamma(\alpha-1)} =: a_i$$

Verteilung und Momente von S_n , $\xi_i^{(n)}$ sind nicht explizit bekannt, aber man kann mittels einer Rekursionsformel $\mathbb{E}[S_n]$, $\mathbb{E}[\xi_i^{(n)}]$ numerisch bestimmen und somit auch

$$\frac{\mathbb{E}[\xi_i^{(n)}]}{\mathbb{E}[S_n]} =: b_i$$

(M.B., J. Blath, B. Eldon, Statistical properties of the site-frequency spectrum associated with Lambda-coalescents, arXiv:1305.6043, 2013)

(Erinnerung: Für den Kingman-Koaleszenten gilt

$$\mathbb{E}[\xi_i^{(n)}] = \theta/i, \mathbb{E}[S_n] = \theta \sum_{i=1}^{n-1} 1/i \sim \theta \log n)$$

Einar Árnason, Mitochondrial Cytochrome *b* DNA Variation in the High-Fecundity Atlantic Cod: Trans-Atlantic Clines and Shallow Gene Genealogy, *Genetics* 166, 1871–1885 (2004)

berichtet genetische Variabilität an einem 250 Bp langen Stück des mitochondrialen Genoms in einer Stichprobe von $n = 1278$ atlantischen Kabeljaus:

$S = 39$ segregierende Positionen

i	1	2	3	4	5	6	7	8	9	10	11+
ξ_i	17	4	4	2	4	1	0	1	0	0	6
Km	5.0	2.5	1.7	1.3	1.0	0.8	0.7	0.6	0.6	0.5	24.2
Sb_i	18.5	4.6	2.3	1.5	1.0	0.8	0.6	0.5	0.4	0.4	8.4
Sa_i	20.3	4.9	2.4	1.5	1.0	0.8	0.6	0.5	0.4	0.3	5.8

Km.: Erwartung unter Kingman-Koaleszent,

Sb_i : Erwartung unter Beta($2 - \alpha, \alpha$)-Koaleszent ($\alpha = 1.48$),

Sa_i : Asymptotische Formel für Beta($2 - \alpha, \alpha$)-Koaleszent ($\alpha = 1.48$),

(Es gibt allerdings in den Daten Widersprüche zum IMS-Modell.)