Übungen zu „Stochastische Modelle der Populationsgenetik“

8. Sei (X_i) ein zeitlich inhomogener Münzwurf, mit $p_i := \mathbf{P}(X_i = 1) := \frac{1}{i}$ für $i = 1, 2, \ldots$. Zeigen Sie: Für die Anzahl K von Erfolgen bei den ersten n Würfen gilt

$$\mathbf{P}(K = k) = s_n^{(k)}/n!, \qquad k = 1, \ldots, n.$$

Dabei ist $s_n^{(k)}$ die Anzahl der n-Permutationen mit k Zyklen. (Tipp: Stellen Sie eine Beziehung zur Hoppe-Urne her.)

9. Es seien $Z_1, \ldots, Z_K$ die Größen der absteigend nach ihrem Alter geordneten Allele in einer Stichprobe vom Umfang n im Infinite-Alleles-Modell. Der Einfachheit halber betrachten wir den Fall $\theta = 1$. Begründen Sie: Z_1 ist uniform verteilt auf $\{1, \ldots, n\}$; gegeben $Z_1 = z$ mit $z < n$ ist Z_2 uniform verteilt auf $\{1, \ldots, n - z\}$; u. s. w.

10. Rekapitulieren Sie, warum ein Standard-Yule Prozess, der zur Zeit 0 mit einem Individuum startet, zur Zeit t eine Geom(e^{-t})-verteilte Größe hat.

11. $T_1 < T_2 < \cdots$ seien die Punkte eines homogenen Poissonprozesses auf $\mathbb{R}_+$ mit Intensität $\theta > 0$. Zu den Zeiten T_i starten unabhängige Standard-Yule Prozesse. Für $t > 0$ sei $K_t := \max\{i : T_i < t\}$ und (für $1 \le i \le K_t$) sei Z_t^i die Größe des i-ten Yule-Prozesses zur Zeit t. Begründen Sie, dass für jedes $t > 0$ und $n \in \mathbb{N}$ gilt:

a) Gegeben $\sum_{i:T_i<t} Z_t^i = n$ ist $(Z_t^1, \ldots, Z_t^{K_t})$ so verteilt wie die nach ihrem Alter geordneten Allelgrößen in einer n-Stichprobe im Infinite Alleles Modell.

b) Für $j = 1, 2, \ldots$ sei $B_j(t)$ die Anzahl der i aus $\{1, \ldots, K_t\}$ mit $Z_t^i = j$. Zeigen Sie: Die $B_j(t)$ sind unabhängige Poissonverteilte Zufallsvariable mit Parameter $\frac{\theta}{j}(1 - e^{-t})^j$. (Hinweis: Denken Sie sich die Punkte T_i “markiert” mit Z_t^i und verwenden Sie einen bekannten Satz über das “unabhängige Markieren” von Poissonpunkten.)

Folgern Sie aus a) und b) (mit Grenzübergang $t \to \infty$): Sind $B_1, \ldots, B_n$ unabhängig mit Pois$(\frac{\theta}{j})$-verteiltem B_j, dann ist die unter $\{\sum_{j=1}^n jB_j = n\}$ bedingte Verteilung von $(B_1, \ldots, B_n)$ gleich der Ewens(n, θ)-Verteilung.