

Aufgabe 11.1 (Wieviele verschiedene Farben sieht man nach n -maligem Herausgreifen aus der Paintbox? (Vgl. Aufg. 9.2)) Seien $p_1 \geq p_2 \geq \dots > 0$ Zahlen mit $\sum_{i=1}^{\infty} p_i = 1$, weiter seien $X_1, X_2, \dots$ u.i.v. $\mathbb{N}$ -wertige Zufallsvariablen mit $\mathbb{P}(X_1 = i) = p_i$, $i \in \mathbb{N}$. Für $n \in \mathbb{N}$ sei

$$K(n) := \#\{X_1, X_2, \dots, X_n\} = \sum_{i=1}^{\infty} G_i(n) \quad \text{mit } G_i(n) := \mathbb{1}(X_k = i \text{ für ein } k \leq n)$$

Weiter sei für $r \in \mathbb{N}$

$$K_r(n) := \#\{i \in \mathbb{N} : \#\{1 \leq k \leq n : X_k = i\} = r\} = \sum_{i=1}^{\infty} \mathbb{1}(F_i(n) = r)$$

mit $F_i(n) := \#\{1 \leq k \leq n : X_k = i\}$.

a) Sei $(R_t)_{t \geq 0}$ ein Poissonprozess mit Rate 1, der unabhängig von den X_i ist. Dann sind die Zufallsvariablen $G_1(R_t), G_2(R_t), \dots$ unabhängig und es gilt $\mathbb{P}(G_i(R_t) = 0) = \exp(-tp_i)$, $i \in \mathbb{N}$. Auch die Zufallsvariablen $F_1(R_t), F_2(R_t), \dots$ sind unabhängig und $F_i(R_t) \sim \text{Pois}(tp_i)$.

b) Zeigen Sie für $t \geq 0$ ist

$$\begin{aligned} \mathbb{E}[K(R_t)] &= \sum_{i=1}^{\infty} (1 - e^{-tp_i}) & \text{Var}[K(R_t)] &= \sum_{i=1}^{\infty} e^{-tp_i} (1 - e^{-tp_i}) \\ \mathbb{E}[K_r(R_t)] &= \frac{t^r}{r!} \sum_{i=1}^{\infty} p_i^r e^{-tp_i} & \text{Var}[K_r(R_t)] &= \frac{t^r}{r!} \sum_{i=1}^{\infty} p_i^r e^{-tp_i} \left(1 - \frac{t^r p_i^r}{r!} e^{-tp_i}\right) \end{aligned}$$

c) Sei $\Phi(t) := \sum_{i=1}^{\infty} (1 - e^{-tp_i})$, $t \geq 0$ und es gebe $0 < \gamma < 1, c \in (0, \infty)$ mit

$$\lim_{t \rightarrow \infty} \frac{\Phi(t)}{ct^\gamma} = 1 \tag{1}$$

Dann gilt

$$\frac{K(R_t)}{t^\gamma} \xrightarrow[t \rightarrow \infty]{\mathbb{P}} c \quad \text{und} \quad \frac{K(n)}{n^\gamma} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} c$$

(Tatsächlich gelten diese Konvergenzen auch fast sicher, falls Sie Lust haben, zeigen Sie auch dies.)

d) Es gelte $p_i \sim c_1 i^{-\beta}$ für $i \rightarrow \infty$ mit einem $\beta > 1$ und $c_1 \in (0, \infty)$. Dann gilt Bedingung (1) aus Teil c) mit $\gamma = 1/\beta$ und $c = c_1^{1/\beta} \Gamma(1 - 1/\beta)$.

e) Unter den Voraussetzungen von Teil d) gilt auch für jedes $r \in \mathbb{N}$

$$\frac{K_r(R_t)}{t^{1/\beta}} \xrightarrow[t \rightarrow \infty]{\mathbb{P}} \frac{c_1^{1/\beta}}{\beta} \frac{\Gamma(r - 1/\beta)}{r!} \quad \text{und} \quad \frac{K_r(n)}{n^{1/\beta}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{c_1^{1/\beta}}{\beta} \frac{\Gamma(r - 1/\beta)}{r!}$$

(Auch diese Konvergenzen gelten fast sicher, falls Sie Lust haben, zeigen Sie auch dies.)

Hinweise. Ein entscheidender Gedanke für a) und b) kann sein, einen Poissonprozess Ξ auf $\mathbb{R} \times \mathbb{N}$ mit Intensitätsmaß Lebesgue-Maß $\otimes \sum_{i \geq 1} p_i \delta_i$ zu betrachten, und damit $F_i(R_t)$ darzustellen als $\Xi([0, t] \times \{i\})$. Für (dann nützliche) Eigenschaften allgemeiner Poissonprozesse siehe z.B. Aufgabe 5.3.

c) Verwenden Sie b). Um von $K(R_t)$ auf $K(n)$ zu schließen ist sozusagen noch eine „De-Poissonisierung“ notwendig. Die entscheidende Beobachtung ist, dass $R_t/t \rightarrow 1$ für $t \rightarrow \infty$ (f.s.) gilt.

d) Hier kann man entweder allgemeine Theorie verwenden (das technische Stichwort lautet „Taubersatz“, siehe z.B. Kapitel XIII.5 in W. Feller, *An introduction to probability theory and its applications*, Bd. 2, Wiley, 1971) oder „von Hand“ folgendermaßen vorgehen: Überlegen Sie, dass für $t \rightarrow \infty$ gilt

$$\begin{aligned} \sum_{i=1}^{\infty} (1 - e^{-tp_i}) &\sim \sum_{i=1}^{\infty} (1 - e^{-c_1 t i^{-\beta}}) = (c_1 t)^{1/\beta} \sum_{i=1}^{\infty} \frac{1}{(c_1 t)^{1/\beta}} \left(1 - \exp\left(-\left(i/(c_1 t)^{1/\beta}\right)^{-\beta}\right)\right) \\ &\sim (c_1 t)^{1/\beta} \int_0^{\infty} (1 - \exp(-x^{-\beta})) dx \end{aligned}$$

sowie

$$\int_0^{\infty} (1 - \exp(-x^{-\beta})) dx = \int_0^{\infty} (1 - \exp(-y)) \beta^{-1} y^{-1-1/\beta} dy = \int_0^{\infty} e^{-y} y^{-1/\beta} dy = \Gamma(1 - 1/\beta)$$

(Substitution $y = x^{-\beta}$, also $\frac{dy}{dx} = -\beta x^{-\beta-1} = -\beta y^{1+1/\beta}$, dann partielle Integration).

e) Hier kann man wieder entweder auf die allgemeine Theorie zurückgreifen (siehe etwa Theorem 2 in Kapitel XIII.5 des genannten Buchs von W. Feller) oder ähnlich wie in d) von Hand abschätzen. Es ist

$$\begin{aligned} \frac{t^r}{r!} \sum_{i=1}^{\infty} (c_1 i^{-\beta})^r e^{-c_1 t i^{-\beta}} &= \frac{(c_1 t)^{1/\beta}}{r!} \sum_{i=1}^{\infty} \frac{1}{(c_1 t)^{1/\beta}} \left(\frac{i}{(c_1 t)^{1/\beta}}\right)^{-\beta r} \exp\left(-\left(i/(c_1 t)^{1/\beta}\right)^{-\beta}\right) \\ &\sim \frac{(c_1 t)^{1/\beta}}{r!} \int_0^{\infty} x^{-\beta r} \exp(-x^{-\beta}) dx = \frac{(c_1 t)^{1/\beta}}{r!} \int_0^{\infty} \frac{y^{r-1-1/\beta}}{\beta} e^{-y} dy \\ &= \frac{c_1^{1/\beta}}{\beta} \frac{\Gamma(r - 1/\beta)}{r!} t^{1/\beta} \end{aligned}$$

Falls Sie „stecken bleiben“ sollten: Dies alles (und noch viel mehr) findet sich in dem Artikel von Alexander Gnedin, Ben Hansen und Jim Pitman, Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws, *Probab. Surv.* 4 (2007), 146–171. (Für den direkten Vergleich mit Proposition 17 dort: In der Formel für $\vec{\nu}(x)$ auf S. 153 lese man $\vec{\nu}(x) = \#\{i \in \mathbb{N} : p_i \geq x\}$, unter der Annahme aus Teil d) also $\vec{\nu}(x) \sim c_1^{1/\beta} x^{1-\beta}$ für $x \downarrow 0$.)

Eine Anwendung dieser Resultate auf die an den Blättern eines Λ - n -Koaleszenten beobachtete genetische Variabilität findet sich in dem Artikel von Julien Berestycki, Nathanaël Berestycki und Vlada Limic, Asymptotic sampling formulae for Λ -coalescents, *Ann. Inst. H. Poincaré Probab. Statist.* 50(3), 715–731, (2014), siehe speziell Theorem 8 dort.

Aufgabe 11.2 (Überlebenswahrscheinlichkeit eines kritischen Galton-Watson-Prozesses) Wir betrachten einen zeitdiskreten Galton-Watson-Prozess wie in Aufgabe 10.2 im Spezialfall einer $\text{Geom}_{1/2}$ -verteilten Nachkommenzahl, d.h. wir setzen $Z_0 = 1$ und $Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i}$ für $n \in \mathbb{N}$, wobei die $X_{n,i}$ u.i.v. sind mit $\mathbb{P}(X_{1,1} = k) = 2^{-(k+1)}$, $k \in \mathbb{N}_0$. Sei $\varphi_{Z_n}(s) := \mathbb{E}[s^{Z_n}]$.

Durch explizite Rechnung mit der Erzeugendenfunktion lassen sich in diesem Fall weitergehende präzise Ergebnisse über die Überlebenswahrscheinlichkeit und die Nachkommenverteilung in Generation n bedingt auf $\{Z_n \neq 0\}$ elementar beweisen.

a) Zeigen Sie, dass dann für $s \in [0, 1]$ und $n \in \mathbb{N}$

$$\varphi_{Z_n}(s) = \frac{n - (n-1)s}{n+1 - ns}$$

sowie für die k -te Ableitung, $k \in \mathbb{N}$,

$$\varphi_{Z_n}^{(k)}(s) = \frac{k!n^{k-1}}{(n+1 - ns)^{k+1}}$$

gilt.

b) Berechnen Sie mit Hilfe von a) die Überlebenswahrscheinlichkeit nach n Generationen und zeigen Sie, dass

$$\lim_{n \rightarrow \infty} n\mathbb{P}(Z_n > 0) = 1$$

c) Folgern Sie aus a) und b) mindestens eine der folgenden Aussagen:

$$\lim_{n \rightarrow \infty} \mathbb{E}[\exp(-\lambda Z_n/n) \mid Z_n > 0] = \frac{1}{1 + \lambda} \quad \text{für } \lambda \geq 0$$

$$\lim_{n \rightarrow \infty} \mathbb{E}[(Z_n/n)^k \mid Z_n > 0] = k! \quad \text{für } k \in \mathbb{N}_0$$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{Z_n}{n} \geq t \mid Z_n > 0\right) = e^{-t} \quad \text{für } t \geq 0$$