Prof. Dr. Matthias Birkner | JGU Mainz

Stochastische Modelle der Populationsbiologie: Übungen und Ergänzungen | **WS 2022/23**

Blatt 5

Aufgabe 5.1 R.S. Singh, R.C. Lewontin und A.A. Felton berichteten in ihrem Artikel Genetic heterogeneity within electrophoretic "allele" of xanthine dehydrogenase in *drosophila pseudoobscura*, *Genetics* 84:609–629, (1976) die Ergebnisse einer Untersuchung der genetischen Variabilität innerhalb einer Stichprobe (einer gewissen Taufliegen-Art) mittels Gel-Elektrophorese.

In einer Stichprobe der Größe $n = 146$ fanden sie $k = 37$ verschiedene Typen, das beobachtete Typenhäufigkeitsspektrum war $b_1 = 20, b_2 = 3, b_3 = 7, b_5 = 2, b_6 = 2, b_8 = 1, b_{11} = 1, b_{68} = 1$ (d.h. 20 Typen kamen nur einmal vor, 3 Typen kamen zweimal vor, etc.; hier nicht aufgeführte Einträge sind $= 0$).

(a) Nehmen wir an, das infinitely-many-alleles-Modell träfe zu. Welchen Wert des Mutationsparameters θ würden Sie anhand der Beobachtungen schätzen (z.B. mittels des Maximum-Likelihood-Schätzers)?

(b)* Wie „gut" ist der Ihr bzw. der ML-Schätzer hier? Schätzen Sie beispielsweise mittels einer Simulationsstudie seine Streuung und seinen Mittelwert ein. Ist er unverzerrt?

Aufgabe 5.2 (Eine Einladung zur Martingaltheorie) Falls Sie sie nicht schon kennen, machen Sie sich mit den Anfangsgründen der Theorie der Martingale (in diskreter Zeit) vertraut, beispielsweise anhand des Steilkurses über Martingale in diskreter Zeit auf der Webseite der Vorlesung (`https://www.staff.uni-mainz.de/birkner/SMPB2223/Steilkurs_Martingale.pdf`), der dort genannten Quellen oder anderer guter Bücher über Stochastik.

Damit ausgestattet betrachten Sie die sogenannte Pólya-Urne: Eine Urne enthalte zu Beginn $R_0 = r \in \mathbb{N}$ rote und $G_0 = g \in \mathbb{N}$ grüne Kugeln. In jedem Zug wird eine der aktuell in der Urne befindlichen Kugeln rein zufällig gezogen und dann zusammen mit einer weiteren Kugel derselben Farbe zurückgelegt. Es bezeichne R_n die Anzahl roter und G_n die Anzahl grüner Kugeln in der Urne nach n Zügen sowie $A_n = R_n/(R_n + G_n) = R_n/(r + g + n)$ den Anteil der roten Kugeln nach n Zügen. Weshalb konvergiert die zufällige Folge $(A_n)_{n=0,1,2,...}$?

Aufgabe 5.3 (Eine Einladung zu allgemeinen Poissonprozessen) Sei E ein metrischer Raum (z.B. $E = \mathbb{R}^d$ mit Euklidischem Abstand), μ ein σ-endliches Maß auf E (d.h. es gibt eine aufsteigende Folge $E_1 \subset E_2 \subset \cdots \subset E$ messbarer Teilmengen mit $\mu(E_n) < \infty$ für $n \in \mathbb{N}$ und $\cup_{n=1}^{\infty} E_n = E$).

Ein zufälliges Maß Ξ auf E heißt ein Poissonprozess mit Intensität(smaß) μ, falls gilt

(i) Für messbares $B \subset E$ ist $\Xi(B)$ Poisson($\mu(B)$)-verteilt.

(ii) Sind $B_1, \dots, B_n \subset E$ paarweise disjunkte (messbare) Teilmengen von E, $n \in \mathbb{N}$, so sind $\Xi(B_1), \dots, \Xi(B_n)$ unabhängig.

Ein solches Ξ besitzt dann eine Darstellung $\Xi = \sum_i \delta_{X_i}$ als Summe seiner Atome.

(a) Es gelte $\mu(E) < \infty$, sei $N \sim$ Poisson($\mu(E)$), unabhängig davon seien $X_1, X_2, \dots$ u.i.v. mit Verteilung $\mu(\cdot)/\mu(E)$. Zeigen Sie: $\Xi = \sum_{i=1}^{N} \delta_{X_i}$ ist in diesem Fall ein Poissonprozess mit Intensität μ.

(b) Wie kann man diese Konstruktion auf den Fall übertragen, dass μ σ-endlich ist mit unendlicher Gesamtmasse $\mu(E) = \infty$?

Sei im Folgenden $\Xi = \sum_i \delta_{X_i}$ ein Poissonprozess auf E mit Intensität μ. Zeigen Sie:

(c) Für $f \in L^1(\mu)$ gilt

$$\mathbb{E}\left[\int_E f(x)\,\Xi(dx)\right] = \int_E f(x)\,\mu(dx)$$

(Betrachten Sie beispielsweise den Fall von Indikatorfunktionen.)

(d) Für messbare $f : E \to [0, \infty)$ gilt

$$\mathbb{E}\left[\exp\Big(-\int_E f(x)\,\Xi(dx)\Big)\right] = \exp\Big(\int_E (1 - e^{-f(x)})\,\mu(dx)\Big)$$

(e) Sei $f : E \to E'$ messbar, $\Xi' := \Xi \circ f^{-1} = \sum_i \delta_{f(X_i)}$. Dann ist Ξ' ein Poissonprozess auf E' mit Intensität $\mu' = \mu \circ f$.

(f) Sei W ein (weiterer) metrischer Raum, $Y_1, Y_2, \ldots$ u.i.v. W-wertige Zufallsvariablen unabhängig von Ξ mit Verteilung ν, $\tilde{\Xi} := \sum_i \delta_{(X_i, Y_i)}$. Dann ist $\tilde{\Xi}$ ein Poissonprozess auf $E \times W$ mit Intensität $\mu \otimes \nu$.

(g) Sei $E = (0, \infty)$, $\mu(dx) = (\theta/x)e^{-x}dx$ mit einem $\theta > 0$, $\Xi = \sum_i \delta_{X_i}$ der zugehörige Poissonprozess. Dann besitzt $Z := \sum_i X_i$ die Gamma(θ)-Verteilung.

Hinweis. Die Lösungen (und noch viel mehr) finden Sie natürlich in vielen guten Büchern über Stochastik, z.B. in Chapter 15 in O. Kallenberg, *Foundations of modern probability*, 2nd Ed., Springer, 2021 oder J.F.C. Kingman, *Poisson processes*, Oxford University Press, 1993. Probieren Sie es erst einmal selbst, und schlagen Sie nach, falls Sie „stecken bleiben".

Aufgabe 5.4 (Erwartete Fixationszeit im (zeitdiskreten) Moran-Modell mit zwei Typen) Wir betrachten das Moran-Modell aus Aufgabe 4.2. a) mit Populationsgröße N, d.h. ein Cannings-Modell, in dem die Verteilung der Nachkommensvektoren die einer rein zufällig gewählten Permutation von

$$\big(2, 0, \underbrace{1, 1, \ldots, 1}_{(N-2)\text{-mal}}\big)$$

ist. Wir nehmen weiter an, dass es zwei genetische Typen 0 und 1 gibt, die jeweils vom Elter zum Kind vererbt werden. In Generation 0 seien $X_0^{(N)} = x_0^{(N)} \in \{0, 1, \ldots, N\}$ viele Individuen vom Typ 1, $X_g^{(N)}$ bezeichne die Anzahl Individuen vom Typ 1 in Generation $g \in \mathbb{N}_0$.

a) Argumentieren Sie, dass $(X_g^{(N)})_{g \in \mathbb{N}_0}$ eine Markov-Kette ist und stellen Sie die Übergangsmatrix auf.

b) Sei $T_{\text{fix}}^{(N)} := \inf\{g \in \mathbb{N}_0 : X_g^{(N)} = 0 \text{ oder } X_g^{(N)} = N\}$ die Zeit (in Generationen), bis einer der beiden Typen aus der Population verschwunden ist und

$$f_N(x) := \mathbb{E}_x[T_{\text{fix}}^{(N)}], \quad x \in \{0, 1, \ldots, N\}$$

Stellen Sie ein Gleichungssystem für $f_N(\cdot)$ auf, z.B. mit Hilfe von Aufgabe 3.1.

c) Zeigen Sie: Falls $x_0^{(N)}/N \to p \in (0, 1)$ für $N \to \infty$, so gilt

$$\lim_{N \to \infty} \frac{1}{N^2} \mathbb{E}_{x_0^{(N)}}[T_{\text{fix}}^{(N)}] = -p \log(p) - (1 - p) \log(1 - p)$$

Hinweis: Betrachten Sie $g_N(x) := f_N(x + 1) - f_N(x)$, stellen Sie anhand von b) ein Gleichungssystem für $g_N(\cdot)$ auf und verwenden Sie, dass $H_k := \sum_{j=1}^k 1/j = \log(j) + \gamma + O(1/k)$ (mit $\gamma \doteq 0{,}57721566...$ der Euler–Mascheroni-Konstante).