Stochastik-Praktikum **WS 2023-2024**

Prof. Dr. Matthias Birkner

Jan Lukas Igelbrink

Blatt 12

Aufgabe 1 (2 + 2 + 2 + 3 = 9 Punkte)

Wir betrachten eine Irrfahrt auf $S = \{1, \ldots, n\}$ mit reflektierendem Rand, konkret eine Markovkette $(X_k)_{k=0,1,2,\ldots}$, deren Übergangswahrscheinlichkeiten $p_{i,j} = P(X_{k+1} = j \mid X_k = i)$ gegeben sind durch

$$p_{i,i+1} = p_{i,i-1} = p_{i,i} \text{ für } i \in \{2, \ldots, n-1\} \text{ und } p_{n,n-1} = p_{1,2} = 1 \text{ (alle anderen } p_{i,j} \text{ sind 0)}$$

a) Simulieren Sie die Anzahl an Schritten, die die Irrfahrt braucht, bis sie erstmalig n oder 1 bei Start in $\frac{n}{2}$ erreicht. Stellen Sie Ihre Resultate folgendermaßen dar: Für $n = 10, 20, 30, \ldots, 100$ machen Sie jeweils 1.000 Simulationen und tragen die beobachtete mittlere Anzahl an benötigten Schritten als Funktion von n in einem Graphen auf.

b) Gehen Sie wie in Teil **a)** vor, aber starten sie die Irrfahrt in einem uniform aus S gewählten Punkt.

c) Schätzen Sie die Wahrscheinlichkeit $p(x)$, dass ein solcher Irrfahrer bei Start in $x = 1, \frac{n}{10}, \frac{2n}{10}, \ldots, n$ die 1 vor n trifft. Simulieren Sie dafür für $n = 10, 50, 100$ jeweils 1.000 Mal und tragen Sie für jedes n die geschätzte Wahrscheinlichkeit $p(x)$ als Funktion von x in einem Graphen auf.

d) Schätzen Sie die Wahrscheinlichkeit $q(n)$, dass ein Irrfahrer bei Start in einem uniform aus S gewählten Punkt die 1 vor n trifft. Simulieren Sie dafür für $n = 10, 50, 100$ jeweils 1.000 Mal und tragen Sie die geschätzte Wahrscheinlichkeit $q(n)$ als Funktion von n in einem Graphen auf.

Aufgabe 2 (2 + 2 + 2 + 3 = 9 Punkte) Das Wright-Fisher-Modell ist ein einfaches mathematisches Populationsmodell aus der Evolutionsbiologie. Jede Generation besteht aus N Individuen mit Genotypen aus $\{A, a\}$. Die Generation G_n bildet sich aus der Generation G_{n-1}, indem jedes Individuum aus G_n unabhängig den Genotyp eines der Individuen aus der Vorgängergeneration G_{n-1} erbt, wobei der Vorfahre basierend auf einer Gleichverteilung gewählt wird. Wenn X_{n-1} die Zahl der Individuen mit Genotyp A in G_{n-1} ist, so gilt dann

$$X_n \sim \mathrm{Bin}\left(N, \frac{X_{n-1}}{N}\right). \tag{1}$$

a) Simulieren Sie 8 mal 20 Generationen des Wright-Fisher Modells mit $N = 20$ und $X_0 = 10$. Plotten Sie die Verläufe gemeinsam in einen Plot.

b) Die Zustände 0 und N sind absorbierend. Sei

$$\tau := \inf\{n : X_n \in \{0, N\}\} \text{ und } f_k = P(X_\tau = N \mid X_0 = k) \text{ für } k \in \{0, 1, \ldots, N\}$$

Schätzen Sie f_k für $N = 20$ und $k = 1, 2, ..., 19$ mithilfe von 100 Simulationen, simulieren Sie jeweils bis zur Absorption. Plotten Sie Ihre Schätzungen für f_k in Abhängigkeit von k.

Beim Wright-Fisher-Modell mit Selektion vererbt jedes Individuum i aus G_{n-1} seinen Genotyp mit Wahrscheinlichkeit $w_i/(w_1 + ... + w_N)$, wobei $w_i = 1 + s$ für Typ A, $s > 0$, und $w_i = 1$ für Typ a. Somit

$$X_n \sim \mathrm{Bin}\left(N, (1+s)\frac{X_{n-1}}{N + sX_{n-1}}\right). \tag{2}$$

c) Bestimmen Sie wie in **b)** die Wahrscheinlichkeiten f_k für $N = 20$, $k = 1, 2, ..., 19$ und $s = 0.05, 0.1, 0.2, 0.4$. Erstellen Sie für jeden Wert von s einen separaten Plot wie in **b)** und zeichnen Sie alle vier Plots in ein Bild.

Beim Wright-Fisher-Modell mit Mutation können sich die Genotypen nach der Vererbung durch Mutation verändern. Hierbei findet die Mutation $a \to A$ mit Wahrscheinlichkeit γ_a und die Mutation $A \to a$ mit Wahrscheinlichkeit γ_A statt. Im Fall ohne Selektion gilt dann

$$X_n \sim \mathrm{Bin}\left(N, (1-\gamma_A)\frac{X_{n-1}}{N} + \gamma_a \frac{N - X_{n-1}}{N}\right). \tag{3}$$

d) Wegen der Mutation findet keine Absorption mehr statt. X_n besitzt nun eine eindeutige, nicht-triviale Gleichgewichtsverteilung. Schätzen Sie diese für $N = 20$, $(\gamma_A, \gamma_a) = (0.05, 0.03)$, $s = 0$ und $X_1 = 10$, indem Sie das Wright-Fisher-Modell 1000 mal 100 Generationen lang simulieren und ein Histogramm von X_{100} erstellen.

Bonus für Interessierte (nicht bepunktet): Bestimmen Sie die invariante Verteilung numerisch mithilfe von Matrixinversion. Ergänzen Sie das Histogramm von **d)** um Ihre numerische Lösung.

Bonusaufgabe (2+2 = 4 Punkte)
In dieser Aufgabe wollen wir uns mit dem berühmten Algorithmus von Propp und Wilson zur Simulation der Gleichgewichtsverteilung einer Markovkette beschäftigen. Eine leicht verdauliche Ausarbeitung der Ideen rund um diesen Algorithmus finden Sie unter `https://www.stat.berkeley.edu/~aldous/206-RWG/RWGpapers/propp_wilson.pdf` (siehe Seite 5). Denken wir an eine Markovkette $(X_t)_{t\in\mathbb{N}}$ mit diskretem endlichen Zustandsraum S und Übergangswahrscheinlichkeiten $p_{a,b} := \mathbf{P}(X_{t+1=a} | X_t = b)$. Ohne Beschränkung der Allgemeinheit nehmen wir $S = \{1, \ldots, n\}, n \in \mathbb{N}$ an.
Die Übergangsdynamik lässt sich auch folgendermaßen verstehen: Für $a \in S$ setzen wir

$$f^a : (0,1) \to S, \quad u \mapsto \min\left\{x \in S : \sum_{y \le x} p_{a,y} > u\right\}. \tag{4}$$

Es seien die $(U_t)_{t\in\mathbb{N}}$ uniform auf $(0,1)$ und unabhängig. Der Übergang von Schritt t nach $t+1$ ist dann gegeben durch

$$X_{t+1} = f^{X_t}(U_{t+1}). \tag{5}$$

Wir wollen uns jetzt mit dem einfachen Beispiel von $n = 10$, $p_{i,i+1} = p_{i,i-1} = p_{i,i}$ für $i \in \{2, \ldots, 9\}$ und $p_{10,9} = p_{1,2} = 1$ beschäftigen.

a) Simulieren Sie die Gleichgewichtsverteilung, indem Sie 1.000 uniform aus S gewählte Startpunkte X_0 wählen und die Markovkette einmal 2 und einmal 100 Übergangsschritte machen lassen. Veranschaulichen Sie die (empirische) Verteilung von X_2 respektive X_{1000} dann mittels zweier Histogramme.

b) Verwenden Sie den Algorithmus von Propp und Wilson zur Simulation der Gleichgewichtsverteilung. Folgen Sie dafür dem Algorithmus auf Seite fünf oben für alle möglichen Startzustände $1, \ldots, 10$. Erstellen Sie ein Histogramm, nach wie vielen Schritten der Algorithmus terminiert.