Übungsblatt 9 – Coupling, Balls into Bins and Poissonisation

Randomisierte Algorithmik

Aufgabe 1 – Coupling einer Irrfahrt

Seien $X_1, X_2, \ldots \sim \mathcal{U}(\{-1, 1\})$ unabhängige Zufallsvariablen. Für $n \in \mathbb{N}_0$ sei $W_n := \sum_{i=1}^{n} X_i$. Man nennt $(W_n)_{n\in\mathbb{N}_0}$ eine Irrfahrt (engl. *random walk*).

Wir können auch eine verschobene Irrfahrt $(V_n)_{n\in\mathbb{N}_0}$ betrachten mit $V_n := W_n + 42$, die also Startpunkt $V_0 = 42$ statt $W_0 = 0$ hat. Wir wollen zeigen, dass die Wahl des Startpunktes typischerweise langfristig keine Rolle spielt.

Wir wollen dabei ohne Beweis verwenden, dass die Irrfahrt mit Wahrscheinlichkeit 1 jede ganze Zahl mindestens einmal besucht. Insbesondere gilt $\lim_{n\to\infty} \Pr[\max\{W_1, \ldots, W_n\} < c] = 0$ für alle $c \in \mathbb{N}$.

(i) Seien $S_1, S_2, \ldots \subseteq \mathbb{Z}$ beliebige Mengen. Zeige $\lim_{n\to\infty} |\Pr[W_n \in S_n] - \Pr[V_n \in S_n]| = 0$.
Hinweis: Konstruiere ein Coupling $(W'_n, V'_n)_{n\in\mathbb{N}_0}$ von $(W_n)_{n\in\mathbb{N}_0}$ und $(V_n)_{n\in\mathbb{N}_0}$ für das $\lim_{n\to\infty} \Pr[W'_n = V_{n'}] = 1$ gilt.

(ii) Zeige, dass das Ergebnis von Aufgabenteil (i) für eine Verschiebung von 43 statt 42 nicht in dieser Form gilt.

Aufgabe 2 – Coupling und Total Variation Distance

Seien X und Y zwei Zufallsvariablen mit Werten in $\mathbb{N}$. Der Totalvariationsabstand (engl. *total variation distance*) von X und Y (bzw. deren Verteilungen) ist definiert als[1]

$$d(X, Y) = \frac{1}{2} \sum_{i\in\mathbb{N}} |\Pr[X = i] - \Pr[Y = i]|.$$

(i) Zeige: Es gibt ein Coupling (X', Y') von X und Y sodass $\Pr[X' \neq Y'] = d(X, Y)$.

(ii) Zeige: Kein Coupling (X', Y') von X und Y erfüllt $\Pr[X' \neq Y'] < d(X, Y)$.

[1] Eine allgemeine Definition, die auch für kontinuierliche Wahrscheinlichkeitsräume funktioniert, findet man auf Wikipedia.

Aufgabe 3 – Eigenschaften der Poissonverteilung

Sei $X \sim \mathrm{Pois}(\lambda)$. Zeige:

(i) $\mathbb{E}[X] = \lambda$.

(ii) $\mathsf{Var}(X) = \lambda$.

(iii) Für $Y \sim \mathrm{Pois}(\rho)$ unabhängig von X gilt $X + Y \sim \mathrm{Pois}(\lambda + \rho)$.

(iv) Für $X' \sim \mathrm{Bin}(X, p)$ gilt $X' \sim \mathrm{Pois}(\lambda p)$.
Beachte: Hier wird also ein zweistufiges Zufallsexperiment durchgeführt. Das Ergebnis X des ersten ist ein Parameter des zweiten.

Aufgabe 4 – Poissonisierte Bloom-Filter

Wir betrachten ein Poisson-Modell von Bloom-Filtern, gehen also davon aus, dass jede Position im Array unabhängig von andere Positionen $\mathrm{Pois}(\alpha k)$-verteilt oft als Hashwert vorkommt.

(i) Wir wählen wieder $\alpha k = \ln 2$. Wie lässt sich zeigen, dass der Anteil $\frac{Z}{m}$ der Nuller mit hoher Wahrscheinlichkeit nahe an $\frac{1}{2}$ liegt?

(ii) Wie ließe sich das Ergebnis in ein nicht-Poissonisiertes Modell übertragen?