Karlsruher Institut für Technologie

Stefan Walzer
ITI Prof. Sanders
Wintersemester 2023/2024

Übungsblatt 3 – Important Random Variables and How to Sample Them

Randomisierte Algorithmik

Aufgabe 1 – Ber(1/3) aus Ber(1/2)

Entwerfe einen Algorithmus der, gegeben eine Folge $B_1, B_2, \ldots \sim \mathrm{Ber}(1/2)$ von Zufallsbits in erwarteter Zeit $\mathcal{O}(1)$ ein Sample $B \sim \mathrm{Ber}(1/3)$ berechnet.

Aufgabe 2 – Ber(p) und $\mathcal{U}(\{1, \ldots, n\})$ aus $\mathcal{U}([0, 1])$

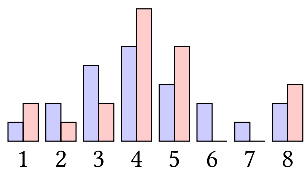
Wir nehmen nun ein Maschinenmodell an in dem reelle Zahlen verarbeitet werden können und indem wir $U \sim \mathcal{U}([0, 1])$ sampeln können. Zeige, dass wir auch $B \sim \mathrm{Ber}(p)$ für $p \in [0, 1]$ und $X \sim \mathcal{U}(\{1, \ldots, n\})$ für $n \in \mathbb{N}$ sampeln können.

Hinweis: Für den Rest des Blattes und des Semesters nehmen wir Voraussetzung und Ergebnis dieser Aufgabe als gegeben an.

Aufgabe 3 – Rejection Sampling Allgemein

Seien $\mathcal{D}_1$ und $\mathcal{D}_2$ Verteilungen auf einer endlichen Menge D. Wir nehmen an:

1. Wir können in Zeit $\mathcal{O}(1)$ ein Sample $X \sim \mathcal{D}_1$ erzeugen.
2. Wir können wir ein gegebenes $x \in D$ in Zeit $\mathcal{O}(1)$ die Wahrscheinlichkeiten $p_1(x) := \Pr_{X\sim\mathcal{D}_2}[X = x]$ und $p_2(x) := \Pr_{X\sim\mathcal{D}_1}[X = x]$ berechnen.
3. Es existiert $C > 0$ sodass für alle $x \in D$ gilt:
$$p_2(x) \leq C \cdot p_1(x).$$



Mögliches Histogramm für $\mathcal{D}_1$ (blau, links) und $\mathcal{D}_2$ (rot, rechts). Es gilt stets “rot $\leq 2 \cdot$ blau”, also ist Bedingung (3) mit $C = 2$ erfüllt.

Entwerfe einen Algorithmus, der in erwarteter Zeit $\mathcal{O}(C)$ ein Sample $Y \sim \mathcal{D}_2$ erzeugt.

Aufgabe 4 – $G \sim \mathrm{Geom}_1(p)$ mit Inverse Transform Sampling

Entwerfe einen Algorithmus der für gegebenes $p \in (0, 1]$ eine Zufallsvariable $G \sim \mathrm{Geom}_1(p)$ in Zeit $\mathcal{O}(1)$ sampelt.

Aufgabe 5 – Ziehen ohne Zurücklegen

Wir betrachten Algorithmen, die für $k, n \in \mathbb{N}$ mit $0 \leq k \leq n/2$ eine Menge $S \subseteq [n]$ der Größe k berechnen, die aus allen Teilmengen von $[n]$ der Größe k uniform zufällig gewählt ist.

(a) Warum dürfen wir $k \leq n/2$ ohne Beschränkung der Allgemeinheit annehmen?

(b) Beschreibe einen Algorithmus, der eine erwartete Laufzeit von $O(k \log k)$ hat.
Hinweis: Rejection Sampling und Suchbaum.

(c) **Bonus:** Entwerfe einen Algorithmus, der eine Worst-Case Laufzeit von $O(k \log k)$ hat.

(d) **Bonus:** Recherchiere, wie man eine Worst-Case Laufzeit von $O(k)$ erreichen kann:

`https://stackoverflow.com/a/67850443`