**Institut für angewandte Mathematik**
**Sommersemester 2013**
Andreas Eberle, Matthias Erbar / Jan Maas

# 7. Übungsblatt „Algorithmische Mathematik II"

**Abgabe bis Dienstag 13 Uhr in den Postfächern gegenüber der Bibliothek**

---

**1. (Zeitumkehr einer Markovkette)**

Sei $X_0, \ldots, X_n$ eine Markovkette mit Übergangswahrscheinlichkeiten $p(x, y)$, die in einer Gleichgewichtsverteilung $\mu$ gestartet wird. Weiter sei $Y_0 := X_n, \ldots, Y_n := X_0$.
Zeigen Sie, dass $Y_0, \ldots, Y_n$ eine Markovkette mit Übergangswahrscheinlichkeiten

$$\hat{p}(x, y) \;=\; \frac{\mu(y)\, p(y, x)}{\mu(x)}$$

und invarianter Verteilung $\mu$ ist. Wann gilt $\hat{p}(x, y) = p(x, y)$?

**2. (Coupon Sammler)**

Aus einer Urne mit $n$ Kugeln werden mit Zurücklegen Kugeln gezogen, und zwar so lange, bis jede Kugel einmal gegriffen wurde. Sei $T$ die Anzahl der nötigen Züge. Zeigen Sie:

$$E[T] \;=\; n\left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1}\right) \;=\; O(n \log n).$$

*Hinweis: Betrachten Sie dazu die Zufallsvariablen* $1 = T_1 < T_2 < \cdots < T_n = T$, *die „Erfolgsmomente", zu denen eine vorher noch nicht gegriffene, neue Kugel gezogen wird. Was ist die Verteilung und der Erwartungswert von* $T_{i+1} - T_i$*? Wie bestimmt sich folglich der Erwartungswert von* $T$*?*

**P. (Simulation von Wahrscheinlichkeitsverteilungen; Aufgabe zählt doppelt)**

a) **Direktes Verfahren:**

Definieren Sie eine Mathematica-Funktion `sampleBin[n,p]`, die eine Stichprobe von der Binomialverteilung mit Parametern $n$ und $p$ ausgehend von einer Zufallszahl aus $(0, 1)$ mithilfe des direkten Verfahrens simuliert.

Ermitteln Sie mithilfe der Mathematica-Funktion `Timing[ ]` den Zeitaufwand für die Simulation von 1000 Stichproben mithilfe Ihres Verfahrens für $p = 1/2$ und verschiedene Werte von $n$ (z.B. $n = 10, 100, 1000, \ldots$). Überlegen Sie sich, wie Sie das direkte Verfahren geschickt anwenden können, um auch für große $n$ bessere Ergebnisse zu erzielen.

b) Seien $a, b \in \mathbb{Z}$, $a < b$, und $p \in (0,1)$. Überlegen Sie sich: Ist $U$ eine auf $(0,1)$ gleichverteilte Zufallsvariable, dann ist

$$X := a + \lfloor (b-a+1) \cdot U \rfloor$$

auf $\{a, a+1, \ldots, b\}$ gleichverteilt und

$$X := \lceil \log_{1-p} U \rceil = \lceil -\log_{1/(1-p)} U \rceil$$

geometrisch verteilt mit Parameter $p$. Hierbei bezeichnet $\lceil x \rceil$ den oberen ganzzahligen Anteil von $x$. Definieren Sie eine Mathematica-Funktion `sampleq[a,b,λ]`, die auf effiziente Weise eine Stichprobe von der Wahrscheinlichkeitsverteilung $Q_{a,b,\lambda}$ auf $\mathbb{Z}$ mit Gewichten

$$q(k) = \begin{cases} 1/Z, & \text{für } a \le k \le b \\ \mathrm{e}^{-\lambda(k-b)}/Z, & \text{für } k > b \\ \mathrm{e}^{-\lambda(a-k)}/Z, & \text{für } k < a. \end{cases}$$

erzeugt. Hierbei ist $\lambda > 0$ und

$$Z := b - a + 1 + \frac{2}{\mathrm{e}^{\lambda} - 1}$$

die Normierungskonstante, die $q$ zur Massenfunktion einer Wahrscheinlichkeitsverteilung normiert.

c) **Acceptance–Rejection–Verfahren:**

Die Hauptschwierigkeit bei der Konstruktion eines effizienten AR-Verfahrens zur Simulation der Binomialverteilung ist das Auffinden einer leicht zu simulierenden Referenzverteilung, durch die sich die Binomialverteilung ohne zu große Verluste abschätzen läßt. Dazu verwenden wir die oberen Schranken

$$p_{n,p}(k) \le \frac{2}{\sqrt{2\pi np(1-p)}} \qquad \text{und} \qquad p_{n,p}(k) \le \mathrm{e}^{-2(k-np)^2/n}, \qquad k = 0, \ldots, n,$$

für die Gewichte der Binomialverteilung. Die erste Abschätzung folgt aus der Stirlingschen Formel. Überlegen Sie sich, daß die zweite Abschätzung aus der Bernstein-Ungleichung folgt.

Definieren Sie eine Mathematica-Funktion `arsampleBin[n,p]`, die eine Stichprobe von der Binomialverteilung mit Parametern $n$ und $p$ mithilfe eines Acceptance-Rejection-Verfahrens mit Referenzverteilung $Q_{a,b,\lambda}$ erzeugt.

Wählen Sie hierbei

$$a = \left\lfloor np - \frac{1}{2}\sqrt{n \log n} \right\rfloor \qquad \text{und} \qquad b = \left\lceil np + \frac{1}{2}\sqrt{n \log n} \right\rceil.$$

Bestimmen Sie einen geeigneten Wert von $\lambda$, sodass sich eine Abschätzung der Massenfunktionen mit einer Konstanten ergibt, die nur langsam in $n$ wächst.

Untersuchen Sie den Zeitaufwand in Abhängigkeit von $n$ experimentell und theoretisch, und vergleichen Sie mit *Aufgabenteil a)*.