Institut für angewandte Mathematik
Sommersemester 2009
Andreas Eberle, Martin Slowik

5. Übungsblatt „Algorithmische Mathematik II"

Abgabe bis Dienstag, 14 Uhr, in der Mathematikbibliothek (LWK)

---

1. (Tom Bayes in Bandrika)
Tom Bayes befindet sich auf seiner Bandrika-Reise genau in der Situation, wie Sie in *Aufgabe 1 vom 4. Übungsblatt*. Im Gegensatz zu Ihnen geht er jedoch davon aus, dass die Antwort Osten mit Wahrscheinlichkeit ε korrekt ist. Zeigen Sie:

a) Egal welche Antwort Tom auf seine erste Frage bekommt, glaubt er weiterhin, dass die Antwort Osten mit Wahrscheinlichkeit ε richtig ist.

b) Sind die ersten beiden Antworten identisch (OO oder WW), so glaubt Tom, dass die Antwort Osten mit Wahrscheinlichkeit ε richtig ist.

c) Nach drei gleichen Antworten beurteilt Tom die Situation folgendermaßen:

$$P\big[\,\text{Osten ist korrekt}\,|\,\text{OOO}\,\big] \;=\; \frac{9\varepsilon}{11-2\varepsilon}, \quad P\big[\,\text{Osten ist korrekt}\,|\,\text{WWW}\,\big] \;=\; \frac{11\varepsilon}{9+2\varepsilon}$$

Welche Werte ergeben sich für $\varepsilon = \frac{9}{20}$?

2. (Geburtenverteilung)
Angenommen, die Anzahl der Geburten an einem Tag in einem Krankenhaus ist Poissonverteilt mit Parameter λ. Weiterhin sei jede Geburt, unabhängig von anderen Geburten und von der Gesamtzahl der Geburten, mit Wahrscheinlichkeit p ein Junge und mit Wahrscheinlichkeit $q = 1 - p$ ein Mädchen. Es wird mit J die Anzahl der Jungen und mit M die Anzahl der Mädchen bezeichnet.

a) Zeigen Sie:

$$P[J = j, M = m] \;=\; \frac{(\lambda p)^j \mathrm{e}^{\lambda p}}{j!} \cdot \frac{(\lambda q)^m \mathrm{e}^{\lambda q}}{m!}.$$

b) Folgern Sie, dass J und M unabhängig und Poissonverteilt mit Parameter λp bzw. λq sind.

3. (Unabhängigkeit und Zahlentheorie)
Sei $s > 1$. Die *Riemannsche Zeta–Funktion* ist definiert durch

$$\zeta(s) \;:=\; \sum_{n=1}^{\infty} n^{-s}.$$

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{N}$ eine Zufallsvariable mit Werten in $\mathbb{N}$ und Verteilung

$$P[X = n] \;=\; \frac{n^{-s}}{\zeta(s)}.$$

Sei E_m das Ereignis "X ist teilbar durch m". Zeigen Sie:

a) Für alle $m \in \mathbb{N}$ ist $P[E_m] = m^{-s}$.

b) Die Ereignisse E_p, wobei p eine Primzahl ist, sind unabhängig.

c) Berechnen Sie $P\big[\bigcap_{p \text{ Primzahl}} E_p^c\big]$, und folgern Sie die *Eulersche Formel*

$$\frac{1}{\zeta(s)} \;=\; \prod_{p \text{ Primzahl}} \Big(1 - \frac{1}{p^s}\Big).$$

d) Die Wahrscheinlichkeit, dass X durch keine Quadratzahl außer 1 teilbar ist, beträgt $1/\zeta(2s)$.

*e) Sei Y unabhängig von X mit derselben Verteilung, und sei H der größte gemeinsame Teiler von X und Y. Sei B_p das Ereignis, dass X und Y beide durch p teilbar sind. Was hat das Ereignis $\bigcap B_p^c$ mit H zu tun? Zeigen Sie:

$$P[H = n] \;=\; \frac{n^{-2s}}{\zeta(2s)}.$$

4. (Ballot Theorem)

a) Betrachte einen random walk, der in $a = 0$ startet. Für $\lambda \in \mathbb{Z}$ sei

$$T_\lambda(\omega) \;=\; \min\{n \in \mathbb{N} \mid S_n(\omega) = \lambda\}.$$

Insbesondere ist T_0 die *erste Rückkehrzeit zum Startpunkt*. Zeigen Sie für $\lambda > 0$:

$$P[T_0 > n \text{ und } S_n = \lambda] \;=\; P[T_\lambda = n] \;=\; \frac{\lambda}{n}\, P[S_n = \lambda].$$

(Die Aussagen von Aufgabe 4 vom 4. Übungsblatt dürfen vorausgesetzt werden.)

b) Bei einer Wahl erhält Kandidat A α Stimmen und Kandidat B β Stimmen, $\beta < \alpha$. Angenommen, die Stimmen werden in "völlig zufälliger " Reihenfolge ausgezählt. Zeigen Sie: Die Wahrscheinlichkeit, dass A während der Stimmenauszählung stets in Führung liegt, ist $\frac{\alpha-\beta}{\alpha+\beta}$.

P. (Simulation von Binomialverteilungen, Abgabe bis 26.5.)

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[λ,a,b]**, die auf effiziente Weise eine Stichprobe von der Wahrscheinlichkeitsverteilung $Q_{\lambda,a,b}$ auf $\mathbb{Z}$ mit Gewichten

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

erzeugt. Hierbei ist $\lambda > 0$, und

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

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

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

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 $Q_{\lambda,a,b}$ erzeugt. Wählen Sie hierbei

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

Bestimmen Sie einen geeigneten Wert von λ, 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 a).