Übungen Algorithmen der Bioinformatik II
Wintersemester 2004/05

Stefan Posch, Birgit Möller, Denis Williams

Institut für Informatik
Universität Halle

# Blatt 6

**Aufgabe 6.1** In einer klinischen Untersuchung werden verschiedene Patienten und und Patientinnen hinsichtlich des Auftretens von zwei Symptomen A und B untersucht. Wir erhalten bei Untersuchung von $N$ Fällen folgende relative Häufigkeiten:

| A | B | Häufigkeiten |
|---|---|---|
| ja | ja | 0.457 |
| ja | nein | 0.133 |
| nein | ja | 0.193 |
| nein | nein | 0.217 |

(a) Wir nehmen für einen Moment diese relativen Häufigkeiten als Verbundwahrscheinlichkeiten für die zwei Symptome an, die wir jeweils als Zufallsvariable modellieren.
Zeigen Sie, daß diese zwei Zufallsvariablen nicht statistisch unabhängig sind.

(b) Skizzieren Sie das prinzipielle Vorgehen, um die statistische Unabhängigkeit zu beweisen oder zu widerlegen, wenn wir nicht mehr die relativen Häufigkeiten mit den Verbundwahrscheinlichkeiten gleichsetzen. (Sondern die zugehörigen absoluten Häufigkeiten als $N$ unabhängige Beobachtungen betrachten.)

(c) In der Zwischenzeit wurde eine neue diagnostische Methode entwickelt, die die Untersuchung auf ein drittes Symptom S ausdehnt. Wir erhalten nun folgende Tabelle von relativen Häufigkeiten aus dem Labor:

| S | A | B | Häufigkeiten |
|---|---|---|---|
| ja | ja | ja | 0.448 |
| ja | ja | nein | 0.112 |
| ja | nein | ja | 0.112 |
| ja | nein | nein | 0.028 |
| nein | ja | ja | 0.009 |
| nein | ja | nein | 0.021 |
| nein | nein | ja | 0.081 |
| nein | nein | nein | 0.189 |

Wir betrachten wiederum diese relativen Häufigkeiten (illegalerweie!) als Verbundwahrscheinlichkeiten. Zeigen Sie, daß A und B bedingt statistisch unabhängig sind, wenn S gegeben ist.
Interpretieren Sie diese Unabhängigkeit durch eine Hypothese über kausale Zusammenhänge zwischen den drei untersuchten Symptomen.

Abgabe: 25.11.04

**Aufgabe 6.2** Betrachten Sie wie in Aufgabe 4.3 den N-fachen Wurf einer Münze mit der Wahrscheinlichkeit $p$ für das Ereignis "Kopf". Als a priori Verteilung für $p$ wählen Sie wieder ein Beta-Verteilung mit den Hyperparametern $\alpha$ und $\beta$.

Sie haben unter obigen Annahmen nach der Beobachtungen von $k$ Ereignissen Kopf bei $N$ Experimenten die a posteriori Verteilung als $Be(p|\tilde{\alpha}, \tilde{\beta})$ bestimmt.

Ein Freund oder eine Freundin von Ihnen hat mit derselben Münze ebenfalls $N$ Würfe durchgeführt. Nach der Beobachtung von $l \neq k$ Kopf-Ereignissen kommt er oder sie auf **diesselbe** a posteriori Verteilung $Be(p|\tilde{\alpha}, \tilde{\beta})$.

Kann dies mit rechten Dingen zugehen? Wie und unter welchen Bedingnungen kann das passieren?

**Aufgabe 6.3** Wir betrachten die absolute Häufigkeit $N_b$ der vier Basen in einer DNA-Sequenz der Länge $N$ und modellieren sie als eine vektorwertige Zufallsvariable $\vec{X} = (N_A, N_C, N_G, N_T)$. Weiterhin werden alle Positionen der DNA-Sequenz als statistisch unabhängig angenommen, wobei jede Position mit demselben DNA-Würfel mit Parameter $\vec{\theta}$ ausgewürfelt wird. Damit ist also $\vec{X}$ multinomialverteilt. Als a prior Verteilung für $\vec{\theta}$ nehmen wir eine Dirichlet-Verteilung mit Parameter $\vec{\alpha}$ an.

Geben Sie die prädiktive Verteilung $P(\vec{Y}|\vec{X})$ an. $\vec{Y}$ ist eine neue Beobachtung der absoluten Häufigkeiten der vier Basen in einer (neuen) DNA-Sequenz der Länge $M$.

*Hinweis: Nuzten Sie denselben "Trick" wie in der Vorlesung bei der Betrachtung der prädiktiven Verteilung einer Normalverteilung. Betrachten Sie zudem die Konstanten beim Beweis zur a posterori Verteilung genauer.*

---

Abgabe: 25.11.04