Albert-Ludwigs-Universität Freiburg Sommersemester 2009

Statistische Methoden der Datenanalyse

Markus Schumacher

Übung I

Matthew Beckingham und Markus Warsinsky

29.4.2009

Anwesenheitsaufgaben

Aufgabe 1 *Betrügt Sie Ihr alter Freund/Ihre alte Freundin?*

Sie treffen einen alten Freund in einem Freiburger Biergarten. Ihr Freund schlägt vor, dass derjenige die Bierchen bezahlt, der aus einem Kartenstapel die geringere Karte zieht. Die genauen Regeln hierfür sind unwichtig. Sie machen das Spielchen mit, ihr Freund gewinnt und Sie müssen zahlen. Auf den darauffolgenden Tagen setzten Sie das Spielchen fort und jedesmal müssen Sie die Zeche zahlen. Langsam fangen sie an argwöhnisch zu werden und überlegen, ob alles mit rechten Dingen zu geht oder Ihr alter Freund sie reinlegt.

(i) Wie entwickelt sich die Wahrscheinlichkeit dafür, dass ihr Freund ein kleiner Betrüger ist mit der Anzahl der von Ihnen bezahlten Rechnungen (1,2,3,4,5,10,15,20) ? Nehmen Sie dafür an, daß die Wahrscheinlichkeit für die Hypothese, daß Ihr Freund Sie übers Ohr haut, nur 5 % beträgt, da es ja ein alter Freund von Ihnen ist (aber man weiß ja nie). Falls er Sie betrügt, so gehen wir davon aus, dass er es an jedem Abend macht. Falls das Spielchen lediglich vom Zufall abhängt, nehmen wir natürlicherweise an, daß die Wahrscheinlichkeit, dass Sie die allabendliche Rechnung bezahlen müssen 50% ist.

(ii) Stellen Sie die gleichen Überlegungen für den Fall an, dass Sie ihm mit 1% bzw. 50% zutrauen, Sie zu hintergehen! Was beobachten Sie?

Aufgabe 2 *Wahrscheinlichkeitsdichtefunktion der Kumulativfunktion*

Gegeben sei eine Zufallsvariable x mit Wahrscheinlichkeitsdichtefunktion $f(x)$. Zeigen Sie, dass die Wahrscheinlichkeitsdichtefunktion der Kumulativfunktion $F(x) = \int_{-\infty}^{x} f(x')dx'$ gegeben ist durch $g(F(x)) = 1$.

Aufgabe 3 *Geometrische Verteilung*

Nehmen Sie ein Zufallsexperiment mit nur zwei möglichen Ergebnissen an, z.B. beim Würfeln eine sechs zu würfeln, oder eben nicht. Die Erfolgswahrscheinlichkeit sei p. Betrachtet werden soll nun das Ereignis, dass im r-ten Versuch, also z.B. beim r-ten mal Würfeln erstmals ein Erfolg eintritt.

(i) Zeigen Sie, dass die Wahrscheinlichkeit für dieses Ereignis gegeben ist durch

$$P(r;p) = p \cdot (1-p)^{r-1}. \tag{1}$$

(ii) Zeigen Sie, dass sich charakteristische Funktion zu dieser Zufallsverteilung ergibt zu

$$\phi(t) = \frac{p \cdot e^{it}}{1-(1-p) \cdot e^{it}}. \tag{2}$$

(iii) Zeigen mittels der charakteristischen Funktion, dass Mittelwert und Varianz dieser sogenannten geometrischen Verteilung gegeben sind durch

$$\mu = \frac{1}{p} \qquad V = \frac{1-p}{p^2}. \tag{3}$$

Aufgabe 4 *Orthogonaltransformation von Zufallsvariablen*

Gegeben seien n Zufallsvariablen $x_1, \ldots, x_n$ und die zugehörige Kovarianzmatrix $V_{ij} = \mathrm{cov}\,[x_i, x_j]$, die nicht notwendigerweise diagonal sein muss. Oft ist es von Vorteil, eine Variablentransformation in n neue Variablen $y_1, \ldots, y_n$ vorzunehmen, deren Kovarianzmatrix $U_{ij} = \mathrm{cov}\,[y_i, y_j]$ diagonal ist.

Im folgenden soll gezeigt werden, dass dies immer mit einer linearen Transformation

$$y_i = \sum_{j=1}^{n} A_{ij} x_j \tag{4}$$

erreicht werden kann.

(i) Berechnen Sie die Kovarianzmatrix der transformierten Variablen U_{ij} aus V_{ij} und den A_{ij}.

(ii) Was muss für die Matrix A gelten, damit U diagonal ist?

(iii) Was folgt aus der Symmetrie der Kovarianzmatrix für die Eigenvektoren der Matrix?

(iv) Konstruieren Sie die Matrix A und überprüfen Sie, dass damit U Diagonalgestalt hat.

(v) Wodurch sind die Varianzen der y_i gegeben?

(vi) Was gilt für AA^T? Wie nennt man eine solche Matrix?

(vii) Für den Fall $n = 2$ kann die Kovarianzmatrix der Zufallsvariablen x_1, x_2 ausgedrückt werden als

$$V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \tag{5}$$

Bestimmen Sie die Transformationsmatrix A.
<u>Tipp:</u> Parametrisieren Sie die Eigenvektoren von V durch einen Winkel θ als

$$\vec{r}_+ = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \qquad \vec{r}_- = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}, \tag{6}$$

und bestimmen Sie θ, indem sie zuerst $\tan 2\theta$ ausrechnen und möglichst weit vereinfachen.

Hausaufgaben

Aufgabe 5 *Wechseln oder nicht?* **2 Punkte**

In einer Spielshow gibt es drei Türen. Hinter einer Tür wartet ein Gewinn, hinter den beiden anderen Türen befinden sich Nieten. Nachdem der Kandidat eine Tür ausgewählt hat, öffnet der Moderator eine Tür mit einer Niete (aber nicht die Tür, die der Kandidat gewählt hat). Der Kandidat darf erneut eine der beiden übrigen Türen wählen.

Sollte er bei der ursprünglichen Tür bleiben oder besser die andere Tür wählen, um seine Gewinnchance zu erhöhen? Benutzen Sie explizit das Bayes-Theorem, um Ihre Antwort zu begründen.

Aufgabe 6 *Teilchenidentifikation* **3 Punkte**

Betrachten Sie zwei verschiedene Teilchenstrahlen. Strahl 1 hat einen Anteil von 10^{-4} von Elektronen, der Rest besteht aus Photonen. Strahl 2 weisst ein Verhältnis von Elektronen zu Photonen von 10^{-4} auf und die restliche Strahlzusammensetzung ist unbekannt.

Die Teilchen werden in einer zweilagigen Spurkammer nachgewiesen. Jede der beiden Lagen kann beim Durchgang eines Teilchens ansprechen, entsprechend kann es 0, 1 oder 2 Treffer geben. Die Wahrscheinlichkeiten für 0, 1 oder 2 Treffer betragen:

$$\begin{array}{ll} P(0|e) = 0.001 & P(0|\gamma) = 0.99899 \\ P(1|e) = 0.010 & P(1|\gamma) = 0.001 \\ P(2|e) = 0.989 & P(2|\gamma) = 0.00001 \end{array}$$

Beim Durchgang eines Teilchens wird nun ein Treffer beobachtet.

(i) Wie groß ist die Wahrscheinlichkeit im Fall von Strahl 1, dass es sich um ein Photon gehandelt hat, und wie groß ist das Verhältnis zwischen Signal und Untergrund?

(ii) Wie ändern sich diese Werte für Strahl 2? Was für Probleme treten hier auf?

Aufgabe 7 *Quotient zweier Gaussverteilter Variablen* **4 Punkte**

Die Zufallsvariablen x und y seien unabhängig und identisch normalverteilt mit der folgenden Wahrscheinlichkeitsdichtefunktion:

$$f(x,y) = f(x)f(y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right). \tag{7}$$

Bestimmen Sie die Wahrscheinlichkeitsdichtefunktion $g(u)$ des Quotienten $u = y/x$.
Tipp: Führen Sie eine geeignete Hilfsfunktion, $v(x,y) = x$ ein, transformieren $f(x,y)$ in die kombinierte Wahrscheinlichkeitsdichtefunktion $h(u,v)$ und marginalisieren Sie über v.

Aufgabe 8 *Kombination zweier Gleichverteilungen* **4 Punkte**

Gegeben seien zwei unabhängige Zufallsvariablen x und y, die beide gleichverteilt sind zwischen 0 und 1, d.h. ihre Wahrscheinlichkeitsdichtefunktion ist gegeben durch

$$g(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{sonst} \end{cases} \tag{8}$$

und analog durch $h(y)$.

(i) Benutzen Sie die Mellin-Faltung von g und h, um zu zeigen, dass die Wahrscheinlichkeitsdichtefunktion $f(z)$ von $z = xy$ gegeben ist durch:

$$f(z) = \begin{cases} -\ln z & 0 < x < 1 \\ 0 & \text{sonst.} \end{cases} \tag{9}$$

(ii) Zeigen Sie dasselbe, indem Sie eine zweite Funktion $u = x$ einführen und die gemeinsame Wahrscheinlichkeitsdichtefunktion von z und u bilden, und schliesslich über u integrieren.
Tipp: Jacobi-Determinante!

(iii) Zeigen Sie, dass die Kumulativverteilung gegeben ist durch

$$F(z) = z(1 - \ln z). \tag{10}$$

Aufgabe 9 *Exponentialverteilung* **3 Punkte**

Gegeben sei eine Zufallsvariable x mit der Wahrscheinlichkeitsdichtefuntion

$$f(x) = \begin{cases} C \cdot e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

(i) Bestimmen Sie die Normierung C.

(ii) Bestimmen Sie die Kumulativverteilung $F(x) = \int_{-\infty}^{x} f(x')dx'$.

(iii) Zeigen Sie durch explizites Nachrechnen, dass die Wahrscheinlichkeitsdichtefunktion für $F(x)$ flach ist.

Aufgabe 10 *Negative Binomialverteilung* **4 Punkte**

Betrachtet werden soll wieder ein Zufallsexperiment mit nur zwei möglichen Ergebnissen und Erfolgswahrscheinlichkeit p. Man führt es nun so lange durch, bis man insgesamt r Erfolge erzielt hat.

(i) Zeigen Sie, dass die Wahrscheinlichkeit, dabei insgesamt k Miserfolge zu haben gegeben ist durch

$$P(k; r,p) = \begin{pmatrix} k+r-1 \\ r-1 \end{pmatrix} p^r (1-p)^k. \tag{11}$$

(ii) Zeigen Sie, dass sich die zugehörige charakteristische Funktion zu

$$\phi(t) = \left(\frac{p}{1-(1-p)e^{it}} \right)^r \tag{12}$$

ergibt.

Tipp: Die folgende Reihenentwicklung könnte sich als nützlich erweisen:

$$(1-x)^{-m} = 1 + mx + \frac{m(m+1)}{2!}x^2 + \frac{m(m+1)(m+2)}{3!}x^3 + \dots$$

(iii) Zeigen mittels der charakteristischen Funktion, dass Mittelwert und Varianz dieser sogenannten negativen Binomialverteilung gegeben sind durch

$$\mu = r\frac{1-p}{p} \qquad\qquad V = r\frac{1-p}{p^2}. \tag{13}$$