5. Übungsblatt zur Vorlesung
Statistische Methoden der Datenanalyse
Abgabe: 24.11.2014, 18:00 Uhr

WS 2014/2015
Prof. W. Rhode

?[subtitle=Zweidimensionale Gaußverteilung]5 Eine zweidimensionale Gaußverteilung sei durch folgende Parameter gekennzeichnet:

$$\mu_x = 4, \quad \mu_y = 2, \quad \sigma_x = 3,5, \quad \sigma_y = 1,5 \quad \text{und} \quad Cov_{xy} = 4,2 \tag{1}$$

a) Wie groß ist der Korrelationskoeffizient?

b) Zeigen Sie, dass die Kurven konstanter Wahrscheinlichkeitsdichte Ellipsen sind. Zeichnen Sie die Ellipse, bei der $f(x, y)$ auf das $1/\sqrt{e}$-fache des Maximums abgefallen ist.

c) Zeichnen Sie die Werte μ_x, μ_y, $\mu_x \pm \sigma_x$ und $\mu_y \pm \sigma_y$ in Ihrer Zeichnung ein.

d) Wie lang sind die Hauptachsen der Ellipse und welchen Winkel bilden sie mit den Koordinatenachsen? Zeichnen Sie diese Werte in die Zeichnung ein.

e) Geben Sie eine Rotationsmatrix M an, so dass die Variablen $(x', y')^T = M(x, y)^T$ unkorreliert sind. Wie groß sind $\sigma_{x'}$ und $\sigma_{y'}$? Zeichnen Sie diese Werte in die Zeichnung ein.

f) Wie lauten die bedingten Wahrscheinlichkeitsdichten $f(x|y)$ und $f(y|x)$? Zeichnen Sie diese Werte in die Zeichnung ein.

g) Wo liegen die bedingten Mittelwerte $E[x|y]$ und $E[y|x]$? Zeichnen Sie diese Werte in die Zeichnung ein.

?[subtitle=Zwei Populationen] Gegeben seien zwei Populationen von jeweils 10000 Punkten in einer Ebene. Die Population P_0 sei eine zweidimensionale unkorrelierte Gaußverteilung mit:

$$\mu_x = 0, \quad \mu_y = 4, \quad \sigma_x = 1.5 \quad \text{und} \quad \sigma_y = 1.5,$$

während P_1 gegeben ist durch eine Gaußverteilung in x mit

$$\mu_x = 4 \quad \text{und} \quad \sigma_x = 3.5,$$

und eine Gaußverteilung in y, deren Mittelwert von x abhängt:

$$\mu_{y|x} = 2 + 0.35(x - 4) \quad \text{und} \quad \sigma_{y|x} = 0.9$$

a) Stellen Sie die beiden Populationen zusammen in einem zweidimensionalen Histogramm dar.

b) Berechnen Sie die Stichproben-Mittelwerte und -Varianzen von x und y sowie die Kovarianz und den Korrelationskoeffizienten für die Einzelpopulationen und die Gesamtheit.

c) Erzeugen Sie eine weitere Population mit den Eigenschaften der Population P_0 mit jedoch nur 1000 Punkten. Erstellen sie anschließend ein ROOT-File mit drei TTrees und füllen Sie die drei Populationen in diese Trees.

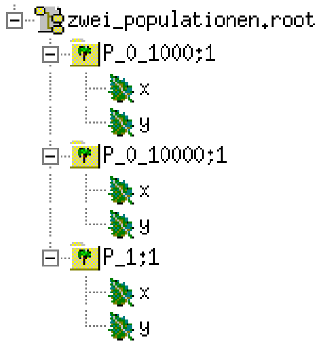

Abbildung 1: Gewünschte Struktur

?[subtitle=Mehrere Messungen]5

a) Die Lichtgeschwindigkeit ist in drei Experimenten unabhängig voneinander mit den folgenden Ergebnissen gemessen worden:

$$299798000 \pm 5000 \text{ m/s}$$
$$299789000 \pm 4000 \text{ m/s}$$
$$299797000 \pm 8000 \text{ m/s}$$

Welches ist das beste kombinierte Ergebnis und dessen Fehler?

b) Bei einer sehr langlebigen radioaktiven Quelle werden in einer Minute 389 Zerfälle und in der nächsten 423 gemessen. Welches ist das beste kombinierte Ergebnis und dessen Fehler? (*Hinweis:* es ist nicht 405,3. Warum nicht?)

?[subtitle=Teilchenspuren]5 In einem Teilchenphysikexperiment stehen 2 Ebenen von Driftkammern senkrecht zur z-Achse an den Positionen z_1 und z_2. Die Kammer an der Position z_1 misst die Position x_1 und die Kammer an der Position z_2 die Position x_2 eines hindurchfliegenden geladenen Teilchens. Diese Messungen sind unabhängig voneinander und weisen die Fehler σ_{x_1} und σ_{x_2} auf.

a) Berechnen Sie die Geradengleichung

$$x = a_1 \cdot z + a_2,$$

die die Bewegung des Teilchens in der x–z-Ebene beschreibt, sowie die Fehler, die Kovarianzmatrix und den Korrelationskoeffizienten von a_1 und a_2.

b) Die Messungen in den beiden Driftkammerebenen bei z_1 und z_2 sollen nun verwendet werden, um die Position des Teilchens im nächsten Detektorelement vorherzusagen. Dies sei eine weitere Driftkammerebene parallel zu den ersten beiden bei $z = z_3$. Berechnen Sie also mit Hilfe der in (a) bestimmten Geradengleichung die Position x_3 und den Fehler σ_{x_3}.

c) Wie ändert sich der Fehler von x_3, wenn Sie fälschlich die Korrelation zwischen a_1 und a_2 nicht berücksichtigen?