Übungen zur Vorlesung „Statistik“

Abgabetermin: Montag, 12.November 01, in der Vorlesung

Aufgabe 9 : **a)** Wie wahrscheinlich ist es, dass eine mit Kopf beginnende faire Münzwurfserie der Länge n auf Adler endet, und dass noch dazu die Summe der Anzahlen der Köpfe am Anfang und der Adler am Ende ≥ 7 ist?

b) Es geht um die Hypothese, dass zwei reelle Stichproben aus ein-und derselben Verteilung stammen. John Tukey hat dazu 1959 folgenden Test vorgeschlagen: “Falls in der einen Stichprobe k_1 Werte kleiner sind als der kleinste Wert in der anderen Stichprobe, und in der anderen Stichprobe k_2 Werte größer sind als der größte Wert in der einen Stichprobe, und falls $k_1 + k_2 \geq 7$, dann verwerfe man die Hypothese auf dem 5% - Niveau. Falls $k_1 + k_2 \geq 11$, dann verwerfe man die Hypothese auf dem 1%-Niveau. (Dabei ist angenommen, dass die beiden Stichprobenumfänge annähernd gleich sind.)” Was hat die in a) formulierte Aufgabe mit Tukeys Test zu tun?

Aufgabe 10 : S. Olsen, J. Simpson und L. Eden (A Bayesian analysis of a multiplicative treatment effect in weather modification, Technometrics 17 (1975), 161-166) diskutieren die Analyse der Daten aus einem Experiment über die künstliche Erzeugung von Regen durch „cloud seeding“. Eine Wolke wurde als „behandlungswürdig“ eingestuft, wenn sie gewissen Kriterien genügte; für jede behandlungswürdige Wolke wurde eine Zufallsentscheidung getroffen, ob man sie nun tatsächlich behandeln sollte. Die nicht behandelten Wolken dienten als Kontrollgruppe. Die folgende Tabelle enthält den Regenfall aus 26 behandelten und 26 Kontroll-Wolken. Fertigen Sie Q-Q-Plots für Regenfall gegen Regenfall und log Regenfall gegen log Regenfall an. Welche Hinweise auf mögliche Effekte des „cloud seeding“ kann man daraus ablesen?

Behandelte Wolken:

129,6	31,4	2745,6	489,1	430,0	302,8	119,0	4,1
92,4	17,5	200,7	274,7	274,7	7,7	1656,0	978,0
198,6	703,4	1697,8	334,1	118,3	255,0	115,3	242,5
32,7	40,6						

Kontroll-Wolken:

26,1	26,3	87,0	95,5	327,4	0,01	17,3	24,4
11,5	321,2	68,5	81,2	47,3	28,6	830,1	345,5
1202,6	36,6	4,9	4,9	41,1	29,0	163,0	244,3
147,8	21,7						

Aufgabe 11 : Es seien α, β mit $0 < \alpha < 1 < \beta$ fest, X sei ein Schätzer für einen reellen Parameter ϑ mit der Eigenschaft, dass $[\alpha X, \beta X]$ ein Konfidenzintervall für ϑ zum Niveau $\frac{2}{3}$ ist. Es seien $X_1, X_2, \ldots$ unabhängig und wie X verteilt; M_n bezeichne den Stichprobenmedian von $X_1, \ldots, X_n$.
Zeigen Sie: $J_n := [\alpha M_n, \beta M_n]$ ist eine Folge von Konfidenzintervallen für ϑ, deren Niveau exponentiell schnell gegen 1 konvergiert.
(**Hinweis**: Man darf sich beim Beweis auf folgendes - schon auf Jakob Bernoulli zurückgehendes - Resultat über die großen Abweichungen beim Münzwurf stützen: Ist S_n die Anzahl der Erfolge beim n-maligen Münzwurf und ist $p_0 < \frac{1}{2}$, dann gibt es ein $c = c_{p_0} < 1$ so, dass für alle $p \leq p_0$ gilt:

$$\mathrm{Ws}_p(S_n \geq n/2) \leq c^n.$$

Salopp ausgedrückt: Die Wahrscheinlichkeit, mit einer „schlechten“ Münze mehr als 50% Erfolge zu haben, nimmt exponentiell ab mit der Anzahl der Versuche.)

Die folgende Aufgabe führt - auf den Spuren der Vorlesung - in die Welt der Bootstrap-Schätzung ein. Wenn Sie mit Bootstrap-Statistiken arbeiten wollen, können Sie entweder dazu eigene Routinen entwerfen oder auf vorgefertigte Bibliotheken zurückgreifen. Für R gibt es eine Zusatzbibliothek namens boot, die Sie sich beispielsweise von CRAN herunterladen können (http://cran.r-project.org/src/contrib/boot_1.2-5.tar.gz). Wenn Sie R im Mathecluster benutzen, laden Sie sie mittels > library(boot, lib.loc = “ ~ birkner/lib/R ”)

Aufgabe 12 : In der R-library “MASS” findet sich der Datensatz “galaxies”; es handelt sich dabei um die gemessenen Geschwindigkeiten in km/sec von 82 Galaxien aus der Region Corona Borealis. (Siehe dazu auch: Roeder, K. (1990) Density estimation with confidence sets exemplified by superclusters and voids in galaxies, J. Am. Statist. Assoc. 85, 617-624.)

9172	9350	9483	9558	9775	10227	10406	16084	16170	18419
18552	18600	18927	19052	19070	19330	19343	19349	19440	19473
19529	19541	19547	19663	19846	19856	19863	19914	19918	19973
19989	20166	20175	20179	20196	20215	20221	20415	20629	20795
20821	20846	20875	20986	21137	21492	21701	21814	21921	21960
22185	22209	22242	22249	22314	22374	22495	22746	22747	22888
22914	23206	23241	23263	23484	23538	23542	23666	23706	23711
24129	24285	24289	24366	24717	24990	25633	26690	26995	32065
32789	34279								

a)

Nehmen wir an, es handelt sich bei galax $\leftarrow$ galaxies/10000 um eine Stichprobe aus einer Verteilung ν. Finden Sie i) ein exaktes - ii) ein “basic bootstrap” - iii) ein mit der Stichprobenstreuung studentisiertes bootstrap - 95 % Konfidenzintervall für den Median von ν.
b) Schätzt man die Dichte mitttels plot(density(galax)), dann könnte man auf die Idee kommen, ν sei gleich $0.12 \cdot \mathfrak{N}(1, 0.15) + 0.47 \cdot \mathfrak{N}(2, 0.15) + 0.35 \cdot \mathfrak{N}(2.34, 0.15) + 0.06 \cdot \mathfrak{N}(3.3, 0.22)$. Gesetzt, es wäre so. Ziehen Sie 1000 mal je 82 Werte aus dieser Verteilung, und notieren Sie, wie oft die drei Konfidenzintervalle den “wahren Median” überdecken.