R. Höpfner: Schätzer und Tests, Sommersemester 2012

Übungsblatt 7

Abgabe per mail an `hoepfner@uni-mainz.de` bis **FR 06.07.12**

Besprechung voraussichtlich DI 10.07.12

Aufgabe 7.1 (Tests im nichtparametrischen Zweistichproben-Lokationsmodell I):

Im nichtparametrischen Zweistichproben-Lokationsmodell $(Z*)$

$$(Z*) \qquad \left(I\!R^N, \mathcal{B}(I\!R^N), \mathcal{P} := \left\{ \bigotimes_{j=1}^{m} F(\cdot - \Delta) \otimes \bigotimes_{j=m+1}^{N} F(\cdot) \;:\; \Delta \in I\!R\,,\, F \in \mathcal{F}_C \right\} \right)$$

mit $m = N - m = 10$ betrachte man zum Niveau $\alpha = 0.05$ den van-der-Waerden-Test, den Mediantest und den Wilcoxontest als Tests für

$$\mathbf{H} : \Delta \leq 0 \quad \text{gegen} \quad \mathbf{K} : \Delta > 0 \ .$$

a) Auf jedem der durch

$$F \in \left\{ \mathcal{N}(0,1)\,,\, R(0,1)\,, \text{Doppelexponential-}\,, \text{Cauchy-}\,, \text{symmetrisch stabile V. mit Index } \tfrac{1}{3} \right\}$$

festgelegten eindimensionalen Submodellen von $\mathcal{P}$ erstelle man einen vergleichenden Plot der durch Simulation approximativ berechneten Gütefunktionen der drei Tests. Was fällt auf ?

b) Zusätzlich beziehe man im Fall

$$F = \mathcal{N}(0,1)$$

den Zweistichproben-t-Test und seine Gütefunktion in den Vergleich mit ein.

(Abgabe: Graphiken, Programm)

Hinweis: Die betrachteten Rangtests sind mit einer nichtfallenden Funktion $a(\cdot) = a_N(\cdot) : \{1, \ldots, N\} \to I\!R$ jeweils von Gestalt

$$\phi \;=\; \left(1_{(c_\alpha, \infty)} + \gamma_\alpha\, 1_{\{c_\alpha\}} \right) \circ S \quad , \quad S = \sum_{n=1}^{m} a_N(R_j) \ .$$

R_j ist der Rang der Beobachtung X_j in der Gesamtstichprobe $(X_1, \ldots, X_N)$. Unter $\Delta = 0$ ist das Tupel $(R_1, \ldots, R_N)$ gleichverteilt auf der Gruppe aller Permutationen von $(1, \ldots, N)$, und man findet die Verteilung von S unter $\Delta = 0$, die kritischen Werte c_α sowie die Konstanten γ_α teils vertafelt (z.B. Hajek, Nonparametric Statistics, 1969, S. 159–167), teils im Computer (z.B. `help(pwilcox)` unter R); die Verteilung von S unter $\Delta = 0$ im Fall des Mediantests (hier notwendig $m = N - m$) ist hypergeometrisch $\mathcal{H}(N, m, m)$.

Aufgabe 7.2 (Tests im nichtparametrischen Zweistichproben-Lokationsmodell II):

Im nichtparametrischen Zweistichproben-Lokationsmodell $(Z*)$ mit beliebigem m und N berechne man die kritischen Werte c_α und γ_α für einseitige Tests der Gestalt

$$\varphi \;:=\; \left(1_{(c_\alpha,\infty)} + \gamma_\alpha 1_{\{c_\alpha\}}\right) \circ S \;\;,\;\; S \;:=\; \sum_{i=1}^{m} a_N(R_i) \quad \text{für } a_N : \{1, \ldots, N\} \to \mathbb{R} \text{ nichtfallend}$$

zum Niveau α per Simulation.

a) Man schreibe eine Funktion in den Argumenten α, m, N, L, die $c_\alpha \in \mathbb{R}$ und $\gamma_\alpha \in (0,1)$ mittels einer grossen Zahl L von Simulationsläufen approximativ berechnet:
In jedem Simulationslauf ziehe man eine zufällige Permutation der Zahlen $\{1, \ldots, N\}$ und berechne die Ränge der Beobachtungen $X_1, \ldots, X_m$ sowie den Wert der Teststatistik S (Befehle `sample` und `rank`). Daraus gewinnt man die empirische Verteilung von S; aus dieser bestimmt man empirische Quantile.
Für kleines α werden die empirischen Quantile nur sehr langsam konvergieren (für $\alpha = 0.05$ sollte man an $L = 10^6$ denken ...). Daher prüfe man durch Variation von L, inwieweit die errechneten Werte bereits 'stabil' sind.
b) Für $\alpha \in \{0.25, 0.1, 0.05, 0.025\}$ und $m = N - m = 10$ berechne man nach a) kritische Werte fuer den Wilcoxon-Test und vergleiche das Ergebnis mit den in Hajek (1969) tabellierten Werten. Man vergleiche auch mit den in `R` bereitgestelten Werten (siehe `dwilcox` etc.).
c) Für $\alpha \in \{0.25, 0.1, 0.05\}$ und $m = N - m = 17$ berechne man nach a) kritische Werte fuer den Fisher-Yates Test.
Abgabe: Programme, empirische Verteilung / Quantile zum Fisher-Yates Test als Graphik.

Aufgabe 7.3 (Tests im nichtparametrischen Zweistichproben-Lokationsmodell III):

Wir wollen annehmen, dass die Körperlängen X bzw. Y neugeborener Jungen bzw. Mädchen bis auf einen Verschiebungsparameter Δ derselben Verteilung $F \in \mathcal{F}_C$ folgen. Eine Untersuchung von $n = 17$ Zwillingspärchen ergab die folgenden Werte:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Y_i	49	53	51	47	54	51	47	45	51	49	51	50	53	51	48	46	54
X_i	50	55	51	49	54	52	47	47	50	51	52	49	55	52	49	50	54

In welchem der folgenden Modelle ist die Aussage 'Jungen sind bei ihrer Geburt in Mittel schwerer als Mädchen' statistisch signifikant zu einem der folgenden Niveaus $\alpha = 0.25$, $\alpha = 0.1$ oder $\alpha = 0.05$:

a) Zweistichproben-Normalverteilungsmodell (ZN'), unter der Voraussetzung gleicher Varianz in beiden Teilstichproben ?

b) Zweistichproben-Lokationsmodell $(Z*)$? Mit welchen Tests ?

(Abgabe: kurze Text/Formelzeilen als pdf)