TU DORTMUND Sommersemester 2019

Fakultät Statistik 09.07.2019

Prof. Dr. Christine Müller
M.Sc. Dennis Malcherczyk

Robuste Statistik

Abschlussprojekt

Sie sollen beim Abschlussprojekt eigenständig eine Simulationsstudie zur Güteeigenschaft von statistischen Testverfahren basierend auf Vorzeichen-Tiefen durchführen. Dazu wird Ihnen ein individuelles Modell zugeteilt (siehe Tabelle unten für die Zuordnung), aus welchem sowohl unter H_0 als auch unter H_1 Daten simuliert werden sollen und die Ablehnungsrate verschiedener Testfahren ermittelt werden soll.

Aufgabenstellung:
Simulieren Sie jeweils für Stichprobenumfänge $N \in \{15, 20, 50\}$ Realisationen Ihres jeweiligen Modells sowohl unter der Nullhypothese

$$H_0 : \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

als auch unter verschiedenen Alternativen. Unter der Alternativen sollten z.B. die folgenden Bereiche A, B, C von Parametern untersucht werden:

$$\begin{aligned}
a)\, \boldsymbol{\beta} \in A &:= [-1, 3] \times [-1, 3] \times \{1\} \setminus \{(1\ 1\ 1)^\top\}; \\
b)\, \boldsymbol{\beta} \in B &:= [-1, 3] \times \{1\} \times [-1, 3] \setminus \{(1\ 1\ 1)^\top\}; \\
c)\, \boldsymbol{\beta} \in C &:= \{1\} \times [-1, 3] \times [-1, 3] \setminus \{(1\ 1\ 1)^\top\}.
\end{aligned}$$

Berechnen Sie jeweils für verschiedene Testverfahren die Testentscheidung unter den verschiedenen Parametern $\boldsymbol{\beta}$ (zum Niveau $\alpha = 0.05$). Wiederholen Sie eine Simulation mindestens $M = 1000$ mal und berechnen Sie die jeweilige relative Anzahl von Ablehnungen der Nullhypothese zur Berechnung des Fehlers erster bzw. zweiter Art. Interpretieren Sie Ihre Ergebnisse und stellen Sie diese geeignet dar.

Zum Testen soll die Vorzeichen Tiefe für $K \in \{3, 4, 5\}$ verwendet werden. Für Berechnungen der Tiefen in linearer Zeit und für die Quantile der asymptotischen Verteilung von der reskalierten Tiefe unter H_0

$$N \left(d_K(E_1, \ldots, E_N) - \left(\frac{1}{2}\right)^{K-1} \right)$$

für unabhängig, identisch verteilte Fehler $E_1, \ldots, E_N$ mit $P(E_n > 0) = P(E_n < 0) = \frac{1}{2}$ werden Ihnen `R`-Dateien zur Verfügung gestellt.

Ziel der Simulationsstudie ist es, in Abhängigkeit des Stichprobenumfangs N festzustellen, welches K für Ihr gegebenes Modell empfohlen werden sollte oder ob überhaupt eine globale Empfehlung geliefert werden kann. Dabei kann es interessant sein, weitere Stichprobenumfänge als die oben gegebenen zu betrachten oder die Güteeigenschaften der Testverfahren auf weiteren Mengen zu untersuchen.

Zuteilung zu den Modellen:

Folgende Modelle sollen jeweils betrachtet werden (die Zuordnung erfolgte zufällig):

$Y_n = \beta_0 - \beta_1 t_n^2 + \beta_2 t_n^4 + e_n$	Felix Fesca
$Y_n = \beta_0 + \beta_1 t_n^2 + \beta_2 t_n^3 + e_n$	Lars Schroeder
$Y_n = \beta_0 + \beta_1 t_n + \beta_2 t_n^3 + e_n$	Jonas Heiner
$Y_n = \beta_0 + \beta_1 (t_n - \beta_2 + 1) \exp(-(t_n - \beta_2 + 1)^2) + e_n$	Hendrik Dohme

Tabelle 1: Modelle mit Regressoren im Intervall $[-2, 2]$

$Y_n = \beta_0 + \beta_1 t_n + \beta_2 \frac{1}{t_n} + e_n$	Lilia Michailov
$Y_n = \beta_0 + \beta_1 t_n + \beta_2 \sqrt{t_n} + e_n$	Robin Brauckmann
$Y_n = \beta_0 + \beta_1 t_n + \beta_2 \sin(\pi t_n) + e_n$	Sujin Park
$Y_n = \beta_0 + \beta_1 (t_n - \beta_2) \exp(-t_n) + e_n$	Markus Jansen

Tabelle 2: Modelle mit Regressoren im Intervall $[0, 4]$

Dabei sollen die e_n und die t_n zufällig generiert werden. Die Fehler sollen Realisationen von unabhängigen $\mathcal{N}(0, \sigma^2)$-verteilten Zufallsvariablen. Ist Ihr Modell in Tabelle 1 gelistet, sollen die Regressoren Realisationen von unabhängigen $U(-2, 2)$-verteilten Zufallsvariablen sein. Falls Ihr Modell in Tabelle 2 steht, sollen die Regressoren Realisationen von unabhängigen $U(0, 4)$-verteilten Zufallsvariablen sein. Dabei beschreibt $U(a, b)$ die stetige Gleichverteilung auf dem Intervall $[a, b]$. Beachten Sie, dass für die Berechnung der Vorzeichen-Tiefen die Regressoren t_n, $n = 1, \ldots, N$ eine vorgegebene Ordnungsstruktur erfüllen sollen (z.B. die kanonische Ordnungsrelation). Die Varianz σ^2 der Normalverteilung sollen Sie selbst festlegen. Es kann interessant sein, verschiedene Varianzen in den Simulationen zu betrachten.

Formalia:

- Abgabefrist: 16.09.2019
- Zur Abgabe gehören:
  - Ein ausreichend dokumentierter und lauffähiger `R`-Code, mit dem die Ergebnisse reproduziert werden können, d.h. setzen Sie Seeds für den Zufallszahlengenerator.
  - Ein kurzer schriftlicher Bericht, welcher mindestens folgende Aspekte abdeckt:
    - Erläuterung der Aufgabenstellung und Ziele des Projekts
    - *Kurze* Beschreibung der verwendeten Testverfahren (Definition der Vorzeichen-Tiefe, reskalierte Form, vorliegendes Regressionsmodell, Voraussetzungen an die Fehler und Regressoren, Entscheidungsregel)
    - Überblick über den Ablauf der Simulationsstudie
    - Darstellung der Ergebnisse (falls möglich sollten grafische Darstellungen verwendet werden)
    - Interpretation der Ergebnisse (zum Beispiel: Einhaltung des Signifikanzniveaus für verschiedene Stichprobenumfänge, Optimalität der Güte unter den Verfahren, Einfluss des Stichprobenumfangs N, Erklärungen der Resultate durch das vorliegende Modell, indem Güteunterschiede durch Vorzeichenstrukturen visualisiert werden (analog zum Aufgabenblatt 14))
- Die Abgabe des Berichts und des `R`-Codes erfolgen per E-Mail an dennis.malcherczyk@tu-dortmund.de elektronisch.
- Der Bericht kann auch auf Englisch verfasst werden