Übungen zur Vorlesung *Schätzer und Tests* **SoSe 2011**

Blatt 7

Prof. Dr. Reinhard Höpfner Dipl.-Math. Michael Diether

Aufgabe 1

Nach Definition heißt eine Familie von i.i.d. Zufallsvariablen $X_1, X_2, \ldots$ *symmetrisch stabil mit Index* α, falls gilt

$$\text{für alle } n \geqslant 1 \text{ gilt:} \qquad \mathcal{L}\left(n^{-\frac{1}{\alpha}}(X_1 + \ldots + X_n)\right) = \mathcal{L}(X_1).$$

Symmetrisch stabile Verteilungen existieren zu jedem Stabilitätsindex $\alpha \in (0, 2]$ und besitzen charakteristische Funktionen der Form

$$Ee^{itX} = \int e^{itx} F(\mathrm{d}x) = \exp(-c|t|^\alpha), \qquad t \in \mathbb{R} \tag{🌍}$$

mit Skalierungskonstanten $c > 0$ (insbesondere ist die Cauchyverteilung symmetrisch stabil zum Index $\alpha = 1$ und die zentrierte Normalverteilung symmetrisch stabil zum Index $\alpha = 2$). Nach Chambers, Mallows and Stuck (JASA 1976) simuliert man eine nach (🌍) mit $c = 1$ verteilte Zufallsvariable X folgendermaßen:

1. Es seien Ψ standard-exponentialverteilt und Φ gleichverteilt auf $(-\frac{\pi}{2}, \frac{\pi}{2})$ unabhängig.
2. Für $\alpha \neq 1$ leistet
$$\frac{\sin(\alpha\Phi)}{(\cos\Phi)^{\frac{1}{\alpha}}} \left(\frac{\cos\big((1-\alpha)\Phi\big)}{\Psi}\right)^{\frac{1-\alpha}{\alpha}} \tag{☯}$$
das Gewünschte.
3. Für $\alpha = 1$ ist eine Cauchyverteilung mit Skalierungsparameter 1 zu simulieren.

Überzeugen Sie sich für die Fälle $\alpha = \frac{1}{4}$ und $\alpha = \frac{3}{4}$ von dieser Methode auf folgende Weise:

a) Simulieren Sie $N = 10000$ i.i.d. Zufallszahlen nach (☯). Plotten Sie die empirische charakteristische Funktion zu diesen Beobachtungen, und vergleichen Sie sie mit (🌍) für $c = 1$.

b) Ziehen Sie in $N = 10000$ Simulationsläufen je $n = 1000$ i.i.d. Zufallszahlen X nach (🌍). Berechnen Sie $Y := n^{-\frac{1}{\alpha}} \sum_{j=1}^n X_j$ und betrachten Sie ein Histogramm aus N dieser Y-Werte. Vergleichen Sie dies mit einem Histogramm aus N zusätzlich gezogenen X-Werten.

Hinweis: Bei der Berechnung der empirischen charakteristischen Funktionen können Sie getrennt mit Real- und Imaginärteil arbeiten. Zeichnen Sie insbesondere jeweils einen Plot für Real- und Imaginärteil.

Aufgabe 2

Seien $m, N \in \mathbb{N}$ und sei $a : \{1, \ldots, N\} \to \mathbb{R}$ eine monoton wachsende Abbildung. Wir wollen in dieser Aufgabe im Zweistichproben-Lokationsmodell (Umfang der 1. Stichprobe m, Umfang der 2. Stichprobe $N - m$) die kritischen Werte zu den Rangsummentests der Gestalt

$$S := \sum_{i=1}^m a(R_i), \qquad \text{mit} \quad R_j := \sum_{k=1}^N \mathbb{1}_{\{X_k \leqslant X_j\}}$$

empirisch berechnen; die Nullhypothese lautet also „$\mathbf{H}: \; X_1, \ldots, X_N \sim F \in \mathcal{F}_c$ i.i.d.“

Bitte wenden!

Gehen Sie hierzu folgendermaßen vor:

1. Ziehen Sie eine zufällige Permutation der Menge $\{1, \ldots, N\}$ und berechnen Sie die zugehörigen Ränge, sowie die Teststatistik. Hierbei können die `R`-Befehle `sample` und `rank` nützlich sein.

2. Wiederholen Sie den ersten Schritt genügend oft und berechnen Sie die kritischen Werte c_α und γ_α zum Signifikanzniveau α. Die auf diese Weise ermittelten empirischen kritischen Werte scheinen nur recht langsam zu konvergieren; eine Anzahl von 1.000.000 Wiederholungen liefert jedoch (für unsere Zwecke) brauchbare Approximationen.

Berechnen Sie die kritischen Werte zum Signifikanzniveau $\alpha = 0{,}05$ für

a) den Wilcoxontest, $a(x) = x$;

b) den van der Waerden-Test, $a(x) = \Phi^{-1}\big(\frac{x}{N+1}\big)$;

c) den Mediantest, $a(x) = \mathbb{1}_{(\frac{1}{2},1]}\big(\frac{x}{N+1}\big)$.

Aufgabe 3

Im nichtparametrischen Zweistichproben-Lokationsmodell

$$\mathcal{P} = \left(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mathcal{P} := \left\{ \bigotimes_{j=1}^{m} F(\cdot - \Delta) \otimes \bigotimes_{j=m+1}^{N} F(\cdot) \colon \Delta \in \mathbb{R},\ F \in \mathcal{F}_C \right\}\right)$$

mit $m = N - m = 10$ betrachten wir zum Niveau $\alpha = 0{,}05$ den van-der-Waerden-Test, den Mediantest und den Wilcoxontest als Tests für

$\mathbf{H} : \Delta \leqslant 0$ gegen $\mathbf{K} : \Delta > 0.$

Erstellen Sie auf jedem der durch

- $F = \mathcal{N}(0, 1)$
- $F = U(0, 1)$
- $F =$ Cauchyverteilung
- $F =$ symmetrisch stabile Verteilung mit Index $\frac{1}{3}$
- $F =$ Doppelexponentialverteilung mit Lageparameter 0 und Rate 1 (vgl. Blatt 3, Aufg. 1)
- $F =$ Student'sche t-Verteilung mit 2 bzw. mit 3 Freiheitsgraden

festgelegten eindimensionalen Submodelle von $\mathcal{P}$ einen vergleichenden Plot der durch Simulation approximativ berechneten Gütefunktionen der drei Tests. Hierbei können Sie die empirisch ermittelten kritischen Werte aus Aufgabe 2 verwenden. Bei der Simulation der stabilen Verteilung folgen Sie dem Vorgehen aus Aufgabe 1.

Was fällt auf?

Abgabe: Per E-Mail an `dietherm@uni-mainz.de` bis Montag, 18. 7. 2011, 23.59 Uhr