8. Übung am 16. Mai 2022

UV Angewandte Statistik (405.170)

Übungsaufgabe 43 (Fortsetzung Aufgabe 42)**.** Angenommen, $F, F_1, F_2, \dots$ sind absolut stetige Verteilungsfunktionen mit Wahrscheinlichkeitsdichten $f, f_1, f_2, \dots$. Weiters bezeichne $\mu, \mu_1, \mu_2, \dots$ die entsprechenden Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$.
Zeigen Sie: Aus $\lim_{n\to\infty} TV(\mu_n, \mu) = 0$ folgt $\lim_{n\to\infty} \|f_n - f\|_1 = 0$.

Übungsaufgabe 44. Gehen Sie den R-Code confidence.band_dn.R durch und überlegen Sie sich, was hier gemacht wird. Erweitern Sie den Code so, dass die Variable 'ok' gleich 0 ist und genau dann auf 1 gesetzt wird, wenn $F \in C_n$ gilt.
Modifizieren Sie weiters den Code so, dass die Überdeckungswahrscheinlichkeit des Konfidenzbands für die Exponentialverteilung mit Parameter $\theta = 0.5$ simuliert/überprüft, und dass zusätzlich die Überdeckungswahrscheinlichchkeit des Intervalls $[L_n(1), U_n(1)]$ für den Parameter $\theta = F(1)$ (approximativ) berechnet wird.

Übungsaufgabe 45. Zeigen Sie mittels Simulationen in R (konkrete Wahl einer stetigen cdf F), dass für kleine Sample Size n das asymptotische Konfidenzband gemäß Satz 4.19 Überdeckungswahrscheinlichkeit weit kleiner als $1-\alpha$ haben kann (i.e., für kleine Stichroben sollte dieses Band nicht verwendet werden). Simulieren Sie für Ihr Beispiel zusätzlich die Überdeckungswahrscheinlichkeit des exakten Bands.

Übungsaufgabe 46. Verwenden Sie den Satz von Massart, um ein asymptotisches Konfidenzband für beliebige (nicht notwendigerweise stetige) Verteilungsfunktionen herzuleiten. Überprüfen Sie mittels Simulationen die Überdeckungswahrscheinlichkeit des erhaltenen Konfidenzbands für mindestens zwei verschiedene unstetige Verteilungen†.

Übungsaufgabe 47 (Mischverteilung)**.** Sei $X_1 \sim \mathcal{N}(\mu_1, 1), X_2 \sim \mathcal{N}(\mu_2, 1)$, und $Z \sim A(p)$ mit $p \in (0, 1)$. Die drei Zufallsvariable X_1, X_2, Z seien unabhängig. Wir betrachten die Zufallsvariable

$$Y = ZX_1 + (1-Z)X_2,$$

$Y_1, \dots, Y_n$ sei eine Zufallsstichprobe von Y. Überlegen Sie Sich, wie sich die Parameter μ_1, μ_2, p ausgehend von $Y_1, \dots, Y_n$ schätzen lassen, und überprüfen Sie die Güte der Schätzer mit Hilfe von Simulationen in R.

Übungsaufgabe 48 (Schadenssummen)**.** Der Datensatz Insurance.RData enthält für 100.000 Sturmschäden einer Versicherung die ausbezahlte Summe, wobei selbige maximal 5.000 Euro (Deckungssumme im Vertrag) beträgt obwohl die echte Schadenshöhe höher sein kann†. Angenommen, wir wissen, dass die echte Schadenshöhe S logarithmisch normalverteilt mit Parameter μ, σ^2 ist. Schätzen für diesen Fall μ und σ^2 und damit $F_S(10000)$ und $F_S(100000)$, wobei F_S wie gewohnt die Verteilungsfunktion von S bezeichnet.

---

†i.e. für Zufallsvariable deren Verteilungsfunktion mindestens eine Unstetigkeitsstelle hat

†diese Fälle sind im Datensatz mit 'censored=1' markiert, d.h. in diesen Fällen ist nur bekannt, dass die echte Schadenshöhe mehr als 5.000 Euro betrug