technische universität
dortmund

LS 4

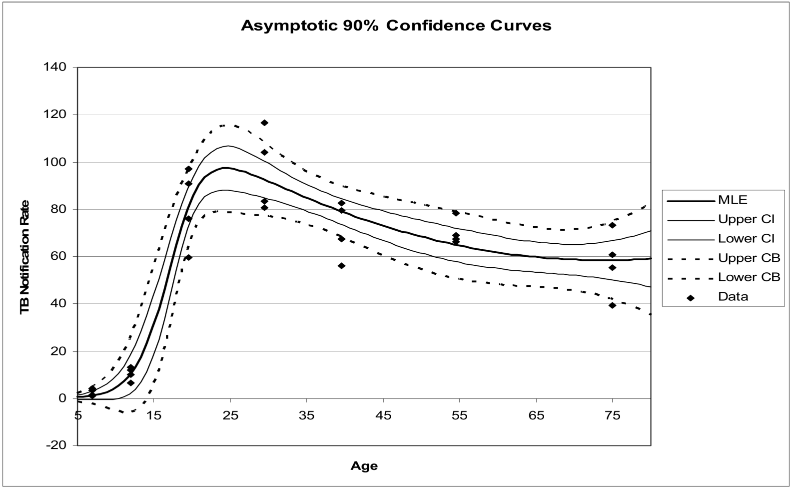
Bootstrapping Ansätze zur Bestimmung von Konfidenzbändern für Verteilungsfunktionen

Dennis Richter

3. Januar 2021

Lehrstuhl IV
Informatik

Motivation und Problemstellung



Motivation und Problemstellung

- Konfidenzintervalle sind visuelle Hilfestellung zum interpretieren der Daten
- die Intervalle gelten jeweils immer nur für einen Messwert
- manchmal möchte man solche Schätzungen für die ganze Regressionsfunktion haben, dann benötigt man Konfidenzbänder
- Es gibt viele Verschiedene Methoden zu Bestimmung von Konfidenzbändern
- neben analytischen Verfahren, kann auch Bootstrap eingestetzt werden -> einfacher, aber ähnlich gut
- zwar theoretische ansätze in der Literatur aber wenig tatsächliche Implementierungen und damit Emipirischer Vergleich zu herkömlichen Methoden
- (Aufgabe ist also vorstellung der Methoden, umsetztung in Omnet und empirischer Vergleich zu anderen CBs)

PIBA - vier Stichpunkte (anstatt Gliederungsfolie)

- Was ist das Problem? -> Implementierung von Bootstrap-Ansätzen zur Bestimmung von Konfidenzbändern
- Idee? -> Implementierung einiger Ansätze im Kontext von OMNeT++, Integration in die IDE nach Möglichkeit und Bewertung anhand von Simulationen
- Vorteil? -> Boostrap ist ein simples und allgemeines Verfahren und einsetzbar in ungewissen situationen -> die arbeit gibt einen erfahrungswert
- Aktion? -> zuerst die genannten und weitere Ansätze genauer recherchieren und vorstellen, dann in C++ im Kontext von OMNeT++ implementieren und nach Möglichkeit in die IDE integrieren, dann testen der Methoden anhand von einfachen Beispielen in Form einer OMNeT++ Simulation (z.B. ...)

Grundlagen

- Konfidenzintervalle

  $P\left(\theta_L \leq \theta_0 \leq \theta_U\right) = 1 - \alpha$

  $\theta_L, \theta_U = \hat{\theta} \mp z_{\alpha/2}\sqrt{\mathbf{V}(\hat{\theta})}$

  $\forall x \,:\, P\left(y_L(x) \leq \eta(x, \theta_0) \leq y_U(x)\right) \geq 1 - \alpha$

  $y_L(x), y_U(x) = \eta(x, \hat{\theta}) \mp z_{\alpha/2}\sqrt{\left(\frac{\partial\eta(x,\theta)}{\partial\theta}\right)_{\hat{\theta}}^T \mathbf{V}(\hat{\theta}) \left(\frac{\partial\eta(x,\theta)}{\partial\theta}\right)_{\hat{\theta}}}$

- Konfidenzbänder

  $P\left(\forall x \,:\, y_L(x) \leq \eta(x, \theta_0) \leq y_U(x)\right) \geq 1 - \alpha$

  $y_L(x), y_U(x) = \eta(x, \hat{\theta}) \mp \sqrt{\chi_p^2(a)\left(\frac{\partial\eta(x,\theta)}{\partial\theta}\right)_{\hat{\theta}}^T \mathbf{V}(\hat{\theta}) \left(\frac{\partial\eta(x,\theta)}{\partial\theta}\right)_{\hat{\theta}}}$

Grundlagen

- analytische Verfahren zur Bestimmung von Konfidenzbändern

Grundlagen

- Basic-Sampling-Methode

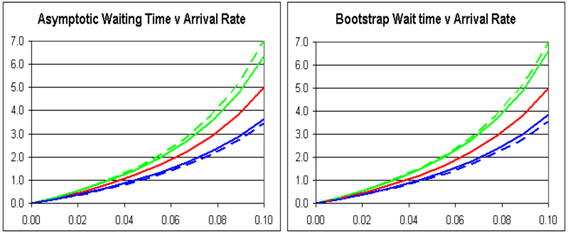
  for $j = 0$ to B **do**
    for $i = 0$ to n **do**
      Draw sample y_{ij} from $F(.)$
    end for
    Calculate statistic $s_j = s(y_j)$
  end for
  Form the EDF $G_n(.|s)$

- Bootstrap-Methode

  Require: Random sample $y = (y_1, y_2, ...y_n)$ from $F(.)$
  Form the EDF $F_n(.|y)$
  for $j = 0$ to B **do**
    for $i = 0$ to n **do**
      Draw sample y^*_{ij} from $F_n(.|y)$
    end for
    Calculate statistic $s^*_j = s(y^*_j)$
  end for
  Form the EDF $G_n(.|s*)$

Verwandte Arbeiten

- Cheng, Russell. (2005). Bootstrapping simultaneous confidence bands. 8 pp.-. 10.1109/WSC.2005.1574257.
- Cheng, Russell. (2015). Bootstrap confidence bands and goodness-of-fit tests in simulation input/output modelling. 16-30. 10.1109/WSC.2015.7408150.



Weitere:

- Govind, Nirmal & Roeder, Theresa. (2006). Estimating Expected Completion Times with Probabilistic Job Routing. 1804-1810. 10.1109/WSC.2006.322958.
- Wang, Xing & Wang, Xin & Sun, Zhaonan. (2009). Comparison on Confidence Bands of Decision Boundary between SVM and Logistic Regression. 272-277. 10.1109/NCM.2009.281.

2 Lösungsansatze

- Parametric Bootstrap
- Non-Parametric Bootstrap

Anwendnungs Beispiel

Av. Age	Year	1980	1986	1993	2000
2	0-4	1.26	2.78	0.63	0.34
7	5-9	3.53	4.10	1.31	0.91
12	10-14	11.98	13.14	9.86	6.53
19.5	15-24	90.82	97.12	75.85	59.46
29.5	25-34	83.45	116.62	104.00	80.85
39.5	35-44	55.98	67.28	79.33	82.66
54.5	45-64	66.32	78.53	69.10	67.27
75	65+	39.42	55.35	60.76	73.20

Abbildung: Morocco Pulmonary TB notifications per 100,000

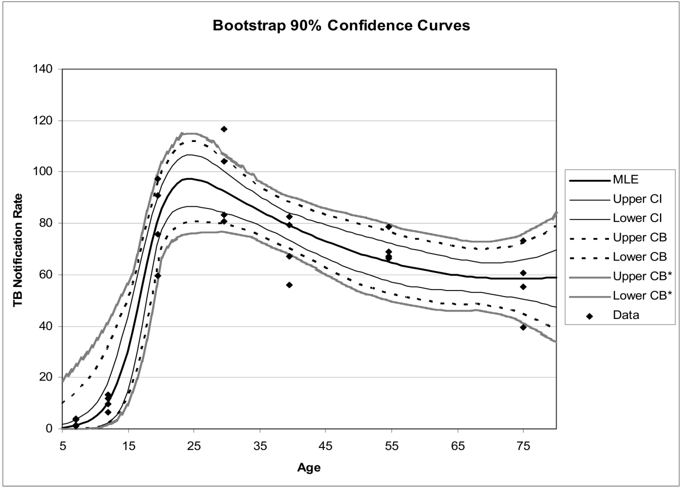
Parameter	MLE
θ_1	10.17
θ_2	153.79
θ_3	17.26
θ_4	-2.58
θ_5	0.455
θ_6	0.01746

Abbildung: MLE's for the Morocco TB Model

Als Modell wurde gewählt:

$$y_j = (\theta_2 + \theta_4 x_j + \theta_6 x_j^2)\frac{exp(\theta_5(x_j - \theta_3))}{1+exp(\theta_5(x_j - \theta_3))} + \epsilon_j$$

Anwendnungs Beispiel



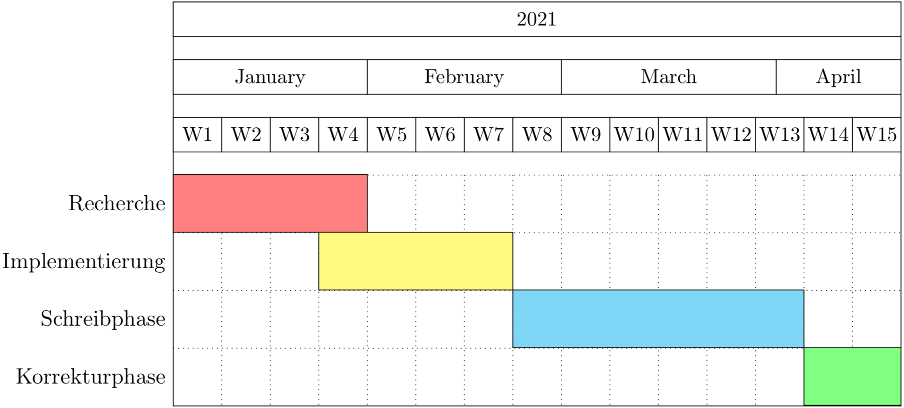
Umsetzung

- genauere Recherche zu Konfidenzbändern und Bootstrapping-Ansätze zur Bestimmung von Konfidenzbändern (mehr dimensionale Parameter)
- Recherche zu Parameterstudien, Auswertung und Darstellung in Kontext von OMNeT++
- Erstellung einfacher Beispiele als Grundlage (z.B. für die Evaluation)
- Implementierung der Verfahren in C++ und Integration in die OMNeT++ IDE
- Anwendung der Verfahre anhand der erstellten Beispiele

Geplante Evaluation

-

Zeitplan


2021
January
February
March
April
W1
W2
W3
W4
W5
W6
W7
W8
W9
W10
W11
W12
W13
W14
W15
Recherche
Implementierung
Schreibphase
Korrekturphase

Zusammenfassung, dazu einfach PIBA-Folie (Folie 2) nochmals auflegen