

Eine gewöhnliche Maximum-Likelihood Schätzung ist stets parametrisch und basiert auf einer Verteilungsannahme, die die Wahrscheinlichkeit des Auftretens der beobachteten Daten widerspiegeln soll. Schätzwerte für die Parameter des postulierten Modells werden ermittelt, indem das Produkt dieser Wahrscheinlichkeiten (Likelihood) maximiert wird, was die Daten bestmöglich erklären soll. Ist das zugrundeliegende Modell jedoch so kompliziert, dass es von einer der bekannten Verteilungen nicht repräsentiert werden kann, so ist die parametrische Schätzung stets eine Approximation. Ein simulationsbasiertes Modell - das die relative Häufigkeit des Auftretens einer Beobachtung nicht-parametrisch zu generieren vermag - bietet hier die Möglichkeit einer von Verteilungsannahmen unabhängigen Schätzung (1). Diese Freiheit kann jedoch nur durch den Preis der den Zufallsprozessen einer Simulation inhärenten Ungenauigkeit, und somit durch eine nicht-exakte Likelihood, erkauft werden. Da eine solche "simulierte Likelihood" fehlerbehaftet ist, bedarf es methodischer Anstrengungen und zusätzlicher Fehlerabschätzungen, um eine verlässliche Schätzung gewährleisten zu können.

Anhand eines Beispiels aus der Infektionsepidemiologie sollen die Möglichkeiten und Grenzen dieser simulationsbasierten Schätzung aufgezeigt werden. Datengrundlage liefern dabei altersabhängige Infektionsintensitäten unter ca. 900 Bewohnern in sieben afrikanischen Dörfern, die an der dort endemischen Parasitenerkrankung Onchocercose (Flussblindheit) erkrankt sind. Die Problemstellung bestand in der Schätzung epidemiologisch wichtiger Parameter aus solchen Querschnittsdaten, deren große Varianz eine nicht-parametrische Modellierung erforderlich machte. Die zwölf Parameter erlauben, den Infektionsdruck pro Dorf, die Heterogenität in der Bevölkerung und die Immunitätslage im Menschen zu erklären. Die Bestimmung von Konfidenzintervallen erweist sich als schwierig, wird jedoch anhand der auf einem Likelihood-Ratio-Test beruhenden Profile-Likelihood [2] vorgeschlagen. Die in diesem Beitrag gezeigte Methodik wurden bislang kaum eingesetzt und könnte sich als Prototyp für ein neues Verfahren etablieren. Aus theoretischer Sicht gibt es für eine simulationsbasierte Schätzung einige Angriffspunkte; auch die Vor- und Nachteile gegenüber MCMC sollten erörtert werden.

Literatur

- 1) Diggle PJ & Gratton RJ (1984): Monte Carlo Methods of Inference for Implicit Statistical Models. J. R. Statist. Soc., B: 46 (2), 193-227.
- 2) Azzalini A: Statistical Inference Based on the likelihood. First edition, Chapter 4.3. Monographs on Statistics and Applied Probability 68. Chapman & Hall, London, 1996

Anschrift

Institut für Medizinische Biometrie, Universität Tübingen
Westbahnhofstr. 55, 72070 Tübingen
<http://www.uni-tuebingen.de/biometry/>
hans-peter.duerr@uni-tuebingen.de