

Simulationen (in R)

Propädeutikum im Sommersemester 2016

Paul Fink

Institut für Statistik, LMU

08.04.2016

Statistische Analyseverfahren beruhen in der Regel auf einer Annahme darüber, wie die zu analysierenden Daten entstanden sind.

Beispiele: $z_i \stackrel{iid}{\sim} \text{Po}(\lambda)$ oder $y_i \sim \text{N}(\mathbf{x}_i' \boldsymbol{\beta}, \sigma_\epsilon^2)$

Ziel: Etwas über die durch das Modell beschriebene Verteilung zu lernen:

- Verteilungsparameter
- Strukturparameter

Frage-/Zielstellungen:

- Unter welchen Datensituationen funktioniert ein Modell gut?
- Bei einer gegebenen Datensituation, welches ist das *optimale* Modell?
- Wie gravierend sind Verletzungen von Annahmen?

Verhalten von Modellen oft analytisch nicht berechenbar

⇒ Überprüfung der Fragestellung anhand von Simulationen

- ① Datengenerierender Prozess: (Verteilungs-)Modell zur Datenerzeugung
- ② Statistische(s) Verfahren zur Untersuchung
- ③ Performance-Kriterium/Kriterien

Variation von Elementen aus 1. und 2. ergibt m Szenarien

Pro Szenario:

Erzeugung von k Datensätzen und Evaluierung der/des statistischen Verfahren darauf

Konkrete Wahl von k und m situationsabhängig

- Was ist als fest anzunehmen (Verfahren/Parameter/Daten)?
- Was soll/muss flexibel sein?
- Welche Strategie verwendet man für die flexiblen Parameter?

- Nicht zu viele Parameter!
Exponentielle Explosion der Komplexität
- Fairer Vergleich von mehreren Verfahren auf gleicher Datenbasis
- Umsetzung in Software: Code in Funktionen statt viel Copy-Paste;
bei Fehler(n) muss man nur an einer Stelle etwas ausbessern
- Beim Ziehen von Zufallszahlen vorher Seed setzen

Kleiber, C. und Zeileis, A. (2013). Reproducible econometric simulations, *Journal of Econometric Methods* 2(1): 89–99, [[DOI](#)]

Beispielhafte Simulation in R mit Hervorhebung der folgenden Aspekte:

- Modularisierung
- Nachvollziehbarkeit
- Wiederverwendbarkeit von Code

- Nur Code verwenden, der ohne Interaktion funktioniert!
- Vorher den Code unbedingt testen!! (`source("Dateiname")`)
- Grundlegende Anleitung im [Cip-Wiki](#)
- Zugriff am Institutsserver nur auf eigenes Laufwerk

Nützliche Kommandozeilen Befehle

Synopsis	Erklärung	Beispiel
<code>cd <i>Ordner</i></code>	In Unterordner <i>Ordner</i> wechseln	<code>cd dummy, cd ..</code>
<code>mkdir <i>Name</i></code>	Order <i>Name</i> im aktuellen Verzeichnis erstellen	<code>mkdir dummy</code>
<code>ls -l</code>	Zeigt alle Ordner und Dateien in Liste an	<code>ls -l</code>
<code>mv <i>alt neu</i></code>	Verschiebt eine Datei pfadweise von <i>alt</i> nach <i>neu</i>	<code>mv old.R dummy/new.R</code>
<code>cp <i>alt neu</i></code>	Kopiert eine Datei pfadweise von <i>alt</i> nach <i>neu</i>	<code>cp old.txt new.txt</code>
<code>rm <i>Name</i></code>	Löscht Datei <i>Name</i> unwideruflich	<code>rm new.txt</code>
<code>rm -r <i>Name</i></code>	Löscht Ordner <i>Name</i> und alle seine Inhalte unwiderruflich	<code>rm -r dummy</code>
<code>cat <i>Name</i></code>	Gibt Inhalt der (Text-)Datei <i>Name</i> auf Konsole aus ohne Datei zu blockieren	<code>cat old.txt</code>