

Bericht zum Stand des Promotionsverfahrens

Sehr geehrte Damen und Herren des Fakultätsrates,

im Folgenden berichte ich gemäß § 5 Abs. 5 der Promotionsordnung über den Stand meines Promotionsverfahrens.

Thema der Arbeit

vorläufiger Titel der Dissertation:

State Space Models for Epidemiological Surveillance of COVID-19: Computational Methods and Applications to German data

Seit Beginn der COVID-19-Pandemie beschäftige ich mich intensiv mit den Herausforderungen bei der Analyse und Interpretation epidemiologischer Daten. In meiner Arbeit untersuche ich die Anforderungen, die solche Daten an statistische Modelle stellen, und entwickle Zustandsraummodelle (State Space Models, SSMs), um die komplexen Dynamiken und Unsicherheiten während einer Epidemie besser abzubilden und verständlich zu machen.

Ein Schwerpunkt meiner Forschung liegt auf der effizienten Durchführung von Inferenz und Vorhersagen mit SSMs. Dazu werden in der Literatur Simulationstechniken wie die Cross-Entropy-Methode (CE-Methode) und Efficient Importance Sampling (EIS) verwendet, wobei unklar ist, unter welchen Umständen welche Methode zu bevorzugen ist. Ich analysiere die theoretischen Grundlagen dieser Methoden, insbesondere durch den Beweis von neuen zentralen Grenzwertsätzen, und liefere Hinweise, warum EIS in der Praxis oft zu besseren Ergebnissen führt als die CE-Methode – vor allem wegen der geringeren asymptotischen Varianz, aber nicht wegen der optimalen Vorschlagsverteilung. Durch eigene Simulationsstudien und algorithmische Weiterentwicklungen vertiefe ich diese Erkenntnisse.

Um die Praxistauglichkeit von SSMs zu demonstrieren, wende ich sie in drei verschiedenen Fallstudien auf deutschen COVID-19-Daten an. Ich modelliere den komplexen Meldeprozess von Infektionsfällen und berücksichtige typische Artefakte wie Feiertageffekte und Meldungsverzögerungen. Außerdem entwickle ich ein Modell, das den Austausch von Fällen zwischen Regionen explizit abbildet und für Vorhersagen genutzt wird, die ich mit Echtzeitdaten des ECDC ForecastHub validiere. Schließlich konstruiere ich ein Modell für die verzögerte Meldung von Hospitalisierungen, das ich zur Echtzeit-Schätzung der Hospitalisierungsinzidenz einsetze und mit den Vorhersagen des deutschen NowcastHub vergleiche.

Die Arbeit leistet damit sowohl einen theoretischen als auch einen praktischen Beitrag zur statistischen Modellierung von Infektionskrankheiten. Die Arbeit ist öffentlich auf GitHub einsehbar: <https://github.com/stefanheyder/dissertation>.

Bislang erzielte Ergebnisse

Die Ergebnisse unterteilen sich in theoretische und angewandte Ergebnisse.

Ein zentraler Bestandteil der Analyse von SSMs ist Importance Sampling, eine Monte-Carlo Simulationstechnik. Ich vergleiche zwei Methoden für Importance Sampling: die Cross-Entropy-Methode (CE-Methode) und Efficient Importance Sampling (EIS). Beide Ansätze dienen dazu, optimale Vorschlagsverteilungen für die Simulation zu bestimmen, wurden bisher jedoch in unterschiedlichen wissenschaftlichen Disziplinen verwendet. Ich liefere erstmals eine theoretisch fundierte Gegenüberstellung dieser Methoden, beweise zentrale Grenzwertsätze und analysiere die Konsistenz sowie die asymptotische Varianz der Verfahren. Besonders liefere ich erste Hinweise darüber, warum EIS in vielen praktischen Anwendungen eine geringere Varianz aufweist und damit effizientere Schätzungen ermöglicht, als die CE-Methode. Ergänzend dazu entwickle ich neue Algorithmen um die CE-Methode effizient für SSMs anzuwenden.

Anschließend demonstriere ich die praktische Relevanz von SSMs in drei Fallstudien zu deutschen COVID-19 Daten.

Reporting Delays und Wochentageffekte Ich entwickle ein Zustandsraummodell, das die Verzerrungen durch Meldeverzögerungen und Wochentageffekte in den COVID-19-Falldaten in Deutschland explizit modelliert. Das Modell erlaubt es, die Entwicklung der Epidemie auf Tagesebene zu rekonstruieren und liefert bereinigte Wachstumsfaktoren, die für die Bewertung von Maßnahmen und die Interpretation des Infektionsgeschehens entscheidend sind. Durch die explizite Modellierung von Verzögerungen und Artefakten können auch fehlerhafte oder fehlende Datenperioden, wie z.B. die Weihnachtszeit, behandelt werden.

Regionaler Wachstumsfaktor Für die Analyse auf regionaler Ebene entwickle ich ein Modell, das die wöchentliche Fallzahl in allen Landkreisen Deutschlands betrachtet und den Austausch von Fällen zwischen Regionen berücksichtigt. Die regionalen Wachstumsfaktoren werden abhängig modelliert, wobei Pendlerdaten zur Schätzung der Abhängigkeiten herangezogen werden. Das Modell ermöglicht es, lokale Ausbrüche zu identifizieren, regionale Unterschiede zu analysieren und Vorhersagen auf Landkreisebene zu erstellen, die über einfache Basismodelle hinausgehen.

Nowcasting von Hospitalisierungen Ich erweitere ein bestehendes Modell zur Echtzeit-Schätzung der Hospitalisierungsinzidenz ("ILM-prop") um ein flexibles SSM, welches die Verzögerungen im Meldeprozess und die altersabhängige Hospitalisierungswahrscheinlichkeit explizit abbildet. Das Modell liefert für verschiedene Altersgruppen und Zeitverzögerungen Vorhersagen und Unsicherheitsintervalle. Die Ergebnisse der Vorhersage werden anhand des deutschen COVID-19 NowcastHubs evaluiert.

Begründung der Verzögerung bei der Bearbeitung

Seit Oktober 2024 bin ich in Vollzeit bei der singularIT GmbH in Leipzig angestellt. Neben der Beschäftigung in Vollzeit war es nicht leicht Zeit für die Fertigstellung der Dissertation freizumachen.

Weitere Schritte und Zeitplan

Die Dissertation ist größtenteils fertiggestellt - inhaltlich fehlen dort nur noch ein Unterkapitel (4.2 zum Regionalen Wachstumsfaktor), sowie eine finale Einordnung der Ergebnisse. Gerne können Sie sich vom aktuellen Fortschritt einen eigenen Eindruck verschaffen, die aktuelle Version ist unter <https://github.com/stefanheyder/dissertation/blob/main/thesis.pdf> verfügbar. Nach Fertigstellung dieser letzten Textpassagen erwarte ich nur noch Feedback von meinem Betreuer, Prof. Thomas Hotz, bevor ich die Arbeit einreiche.

Für Anfang September habe ich zwei Wochen (bereits genehmigten) Urlaub zur Fertigstellung der letzten Schritte eingeplant. Dann fehlt nur noch das Feedback meines Doktorvaters. Ich rechne also damit, im Oktober, spätestens im November dieses Jahres meine Arbeit einzureichen.

Vielen Dank für Ihr Verständnis.

Mit freundlichen Grüßen

Stefan Heyder