



Athenata

Effektive Nutzung der Vorteile Bayesianischer Statistik

Athenata



Warum Bayesianische Statistik?

- Der Einsatz herkömmlicher Statistik in der Life Science Branche trägt wesentlich zu nicht-reproduzierbaren Resultaten und zu Verlusten von vielen Milliarden EUR jährlich bei, vgl. z.B.
 - DOI: [10.1371/journal.pbio.1002165](https://doi.org/10.1371/journal.pbio.1002165)
 - DOI: [10.1038/d41586-019-00857-9](https://doi.org/10.1038/d41586-019-00857-9)
 - DOI: [10.1038/533452a](https://doi.org/10.1038/533452a)



Warum Bayesianische Statistik?

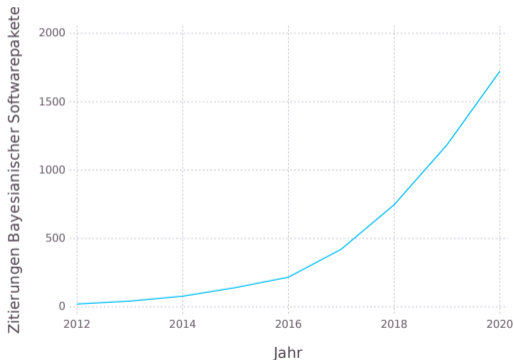
- Durch Bayesianische Methoden werden neue Lösungen möglich
 - Auswertungen bei vielen Prädiktoren und wenig Daten
 - Sog. hierarchische Modelle ermöglichen einzelne Betrachtung verschiedener Gruppen, ohne eine komplette Unabhängigkeit zwischen diesen anzunehmen
 - Guter Umgang mit Multikollinearität



Die Anzahl jährlicher Zitierungen Bayesianischer Softwarepakete nimmt stark zu und sind zum Beispiel in den Bereichen Immunologie und Mikrobiologie besonders relevant (<https://doi.org/10.1002/ail2.62>).

Athenata

Verbreitung





Verbreitung

Bayesianische Methoden verbreiten sich erst jetzt immer mehr aufgrund von:

- hoher benötigter Rechenleistung
- verbesserten Algorithmen
- Trägheit der Lehre



Verbreitung - Zulassungsbehörden

FDA: „The Bayesian approach is also frequently useful in the absence of prior information. First, the approach can accommodate adaptive trials (e.g., interim analyses, change to sample size, or change to randomization scheme) and even some unplanned, but necessary trial modifications. Second, the Bayesian approach can be useful for analysis of a complex model when a frequentist analysis is difficult to implement or does not exist. Other potential uses include adjustment for missing data, sensitivity analysis, multiple comparisons, and optimal decision making (Bayesian decision theory).“

<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials>



Verbreitung - Zulassungsbehörden

EMA: „When more empirical approaches are used, appropriate statistical methods can be applied for comparison and for quantification of uncertainty (precision of estimated effects) between groups (e.g. a Bayesian framework or model-based meta-analysis).“

https://www.ema.europa.eu/en/documents/scientific-guideline/adopted-reflection-paper-use-extrapolation-development-medicines-paediatrics-revision-1_en.pdf



Beispiele

- **Neue Zusammenhänge erkennen:** Mit Bayes-Methoden konnten quantitative Unterschiede in der Nutzung verschiedener Antikörper-Gene zwischen Gesunden und HCV-Infizierten gezeigt werden, die mit herkömmlichen Null-Hypothesen-Signifikanz-Tests nicht sichtbar sind (<https://doi.org/10.1093/bioinformatics/btaa174>).
- **Vorhandene Methoden prüfen:** Durch eine Analyse die auf Bayesianischer Statistik beruht, konnte gezeigt werden, dass ein klassisches Modell des Selbstabbaus von Trypsin auf Oberflächen nicht haltbar ist (<https://doi.org/10.1021/acsomega.0c01109>).



Vorteile

- **Datensparsamkeit:** Gerade, wenn der Datenerhebungsprozess (kosten-) aufwändig ist, ist eine optimale Nutzung der verfügbaren Datenmenge besonders wichtig. Bei unseren Methoden müssen z. B. unvollständige Datenpunkte nicht weggeschmissen werden.
- **Expertenwissen:** Die Aufnahmen von Expertenwissen, welches nicht in den Daten enthalten ist, kann die Analysequalität entscheidend verbessern.



Vorteile

- **Unsicherheiten:** Bayesianische Methoden quantifizieren Unsicherheiten, wodurch ein falsches Gefühl von Sicherheit vermieden wird. Nur mit einem präzisen Bild der Unsicherheit sollten datenanalytische Auswertungen bei der Entscheidungsfindung berücksichtigt werden.
- **Lösbarkeit neuer Probleme:** Wie auf Folie 3 beschrieben, können Fragestellungen beantwortet werden, zu denen es sonst keine passenden Ansätze gibt.



Vorteile

- **Anpassungsgüte:** Genauere und zuverlässigere Ergebnisse durch die Möglichkeit, iterativ ein auf die Fragestellung zugeschnittenes Modell zu entwickeln.
- **Besseres Experiment Design:** Durch Bayesianisches Experiment Design können Dauer und Kosten von Experimenten optimiert werden. Die Erfolgswahrscheinlichkeit kann auch schon während des Experiments abgeschätzt werden, sodass es ggf. vorzeitig angepasst werden kann.



Vorteile

- **Robustheit:** Ein häufiges Problem in der klassischen Statistik ist die Durchführung von mehreren Hypothesentests auf einem Datensatz (multiple comparisons), weshalb p-Werte angepasst werden müssen. Bei Bayesianischen Modellen können alle relevanten Vergleiche in einem Modell betrachtet werden und das Problem der „multiple comparisons“ entfällt (DOI: 10.1080/19345747.2011.618213) .

Bayesianische Modelle sind generativ, das heißt, dass aus den inferierten Parameterverteilungen simulierte Daten generiert werden können. Durch einen Vergleich dieser simulierten Daten mit den tatsächlich beobachteten Daten können Modellannahmen effektiv getestet werden („posterior predictive checks“, DOI: 10.1111/rssa.12378). Eine solche Überprüfung ist bei klassischen Methoden nicht möglich.



Über uns - Gründung

Die Idee zur Gründung von Athenata stammt aus dem **Lehrstuhl für Bioinformatics and Computational Biophysics** der Universität Duisburg-Essen, der sich auf Datenanalyse mittels Bayesianischer Statistik spezialisiert hat. Unser Ziel ist es, **Unternehmen die Vorteile Bayesianischer Statistik zugänglich zu machen**. Dabei werden wir im Rahmen des EXIST-Programms durch das Bundesministerium für Wirtschaft und Energie und den Europäischen Sozialfonds gefördert.

Gefördert durch:



Bundesministerium
für Wirtschaft
und Klimaschutz



Zusammen.
Zukunft.
Gestalten.

aufgrund eines Beschlusses
des Deutschen Bundestages

Athenata



Über uns - Forschungsarbeiten

Predominance of Central Memory T Cells with High T-Cell Receptor Repertoire Diversity is Associated with Response to PD-1/PD-L1 Inhibition in Merkel Cell Carcinoma (DOI: 10.1158/1078-0432.CCR-19-2244)

Wichtige Schlussfolgerungen dieser Arbeit sind, dass:

1. es möglich ist, Korrelationen auch mit wenigen Daten zu identifizieren.
2. man sich bei fehlenden Daten auf geeignete Methoden verlassen sollte, anstatt Daten zu entfernen.



Über uns - Forschungsarbeiten

HAMdetector: A Bayesian regression model that integrates information to detect HLA-associated mutations (DOI: [10.1093/bioinformatics/btac134](https://doi.org/10.1093/bioinformatics/btac134))

Die Erkenntnisse aus dieser Arbeit sind:

1. Bayesianische Methoden sind besonders geeignet, um Informationen aus unterschiedlichen (unsicheren) Quellen zu kombinieren.
2. Modelle, die zusätzliche Daten berücksichtigen, funktionieren deutlich besser als Modelle, die diese Informationen ignorieren.
3. Modelle, die Annahmen treffen, die nicht gut zu den Daten passen, liefern oft unzuverlässige Ergebnisse.



Zusammenarbeit

- Unterstützung von Statistikabteilungen
 - Gemeinsam können wir evaluieren, ob die Vorteile Bayesianischer Statistik in den konkreten Anwendungsfällen genutzt werden können sowie Vorschläge zur Umsetzung liefern.
 - Wir entwickeln Modelle und Lösungen, die bei wiederkehrenden Fragestellungen eigenständig eingesetzt werden können.
 - Bereits mit konventioneller Statistik ausgewertete Experimente können erneut betrachtet werden, um ggf. neue Erkenntnisse zu gewinnen.
 - Bei vorhandenen Zweifeln können wir die Robustheit der Analyse prüfen und sicherstellen.
- Bei Bedarf können wir auch komplette Analysen übernehmen



Austausch

In einem kurzen Gespräch könnten wir genauer auf die einzelnen Punkte eingehen, weitere Beispiele besprechen und Fragen dazu beantworten.

Über ein Kennenlernen und einen Austausch über Ihre Ideen zur Bayesianischen Statistik sowie die für Sie relevanten Punkte der Datenauswertung würden wir uns freuen.

Melden Sie sich bei Interesse gerne bei uns, damit wir einen Termin vereinbaren können:

Mail: jan.schmitz@athenata.com Telefon: +49 1573 4858376