

Geodatenanalyse I: Monte Carlo Methoden

Kathrin Menberg



Stundenplan



Vorläufiger Stundenplan		
Datum	Thema	Dozent
20.10.2021	Einführung in die Programmierung mit Python	Gabriel Rau
25.10.2021	Univariate Statistik und statistisches Testen	Kathrin Menberg
01.11.2021	Feiertag	
08.11.2021	Umgang und Berechnung von Datensätzen	Gabriel Rau
15.11.2021	Bivariate und schließende Statistik	Kathrin Menberg
22.11.2021	Datenvisualisierung mit <i>matplotlib</i>	Gabriel Rau
29.11.2021	Multivariate Statistik	Kathrin Menberg
06.12.2021	Einführung in den Umgang mit Datensätzen	Gabriel Rau
13.12.2021	ausgefallen	Kathrin Menberg
20.12.2021	Weiterführender Umgang mit Datensätzen	Gabriel Rau
27.12.2021	Weihnachtsferien	
03.01.2022	Weihnachtsferien	
10.01.2022	Monte-Carlo Methoden	Kathrin Menberg
17.01.2022	Analyse und Visualisierung von Geodaten	Gabriel Rau
24.01.2022	Räumliche Interpolation	Kathrin Menberg
31.01.2022	Datenethik, Lizensierung und Entwicklungstools	Gabriel Rau
07.02.2022	Regressionsanalyse	Kathrin Menberg

Vorlesungsplan



Uhrzeit	Inhalt
10:00 – 10:30	Monte Carlo Methoden
10:30 – 11:15	Übung
11:15 – 11:30	<u>Pause</u>
11:30 – 12:15	Fortsetzung Übung
12:15 – 12:30	Diskussion und Reflexion

Lernziele



Am Ende der Stunde werden die Teilnehmer:

- mit dem Prinzip von Unsicherheiten und statistischen Zufallsexperimenten vertraut sein.
- ... einen Überblick über Methoden für (Pseudo-)
 Zufallsexperimente und Markov-Chain Monte Carlo Methoden haben.
- ... einfache Monte Carlo Simulationen in Python durchführen und auswerten können.

Unsicherheiten



"Nothing is certain, but our ignorance." (unbekannt)

Welche "Unsicherheiten" fallen Euch in Bezug auf Geodatenanalyse ein?

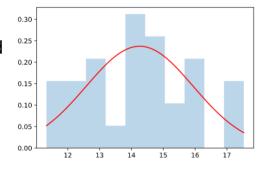
Arten von Unsicherheit



- Aleatorisch (aleatoric)
 - Statistische Unsicherheit
 - Unbestimmbare Größen, die in wiederholten Experimenten variieren
 - Komplexe Prozesse, die nicht erfassbar sind
 - USW.



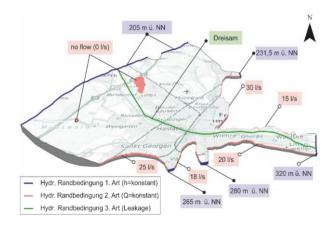
- Systematische Unsicherheit
- Größen, die in Experimenten nicht gemessen wurden.
- Vereinfachungen in numerischen Modellen
- Usw.



Quellen von Unsicherheiten



- Daten, Parameter
 - Messfehler, Messgenauigkeit
 - Szenarien
- Modelle
 - Annahmen (Randbedingungen usw.)
 - Vernachlässigte Prozesse
 - Auflösung, bzw. Diskretisierung
 - Numerische Fehler, Approximation
- Modellierer
 - Bedienungsfehler, Eingabefehler
- und viele mehr...



Ohmer et al. (2021)

Visualisierung von Unsicherheiten

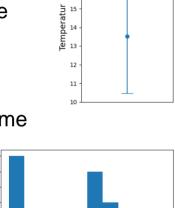


- Fehlerbalken
- Balkendiagramme
- Histogramme
- Boxplots
- Violinen Diagramme

Häufigkeit ™ ħ ਯ

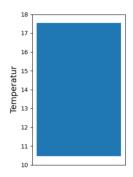
12

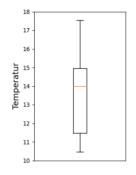
- Schwarmplot
- u.v.m.

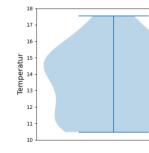


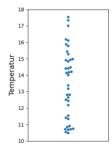
ıs 14 15 Temperatur

17







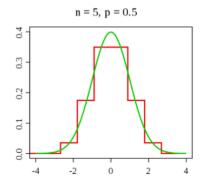


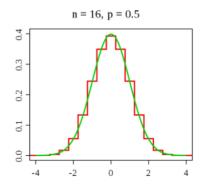
Zentraler Grenzwertsatz

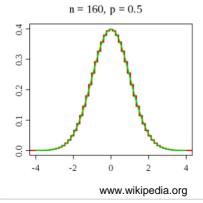


- engl. central limit theorem
- Für $n \to \infty$ nähert sich die empirische Verteilungsfunktion Z_n der Standardnormalverteilung $\Phi(z)$
 - ▶ Konvergenz gegen $\Phi(z)$

$$\lim_{n \to \infty} P(Z_n \le z) = \Phi(z)$$







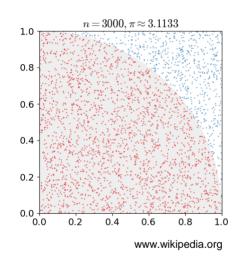
Monte Carlo Simulation



- ▶ Ziele (Beispiele)
 - Approximation von irrationalen Zahlen
 - Verteilungseigenschaften von Zufallsvariablen
 - Nachbildung komplexer Prozesse (z.B. Wetter- und Klimaphänomene)



- mögliche Inputs definieren
- 2. zufällige Inputs anhand einer Wahrscheinlichkeitsfunktion generieren
- 3. <u>deterministische</u> Berechnung mit den Inputs durchführen
- 4. Ergebnisse zusammenfassen

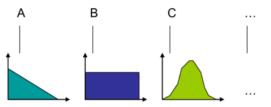


Monte Carlo Simulation



Input data & parameters

Assumed distributions

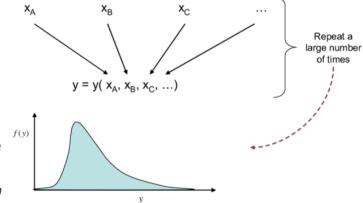




2. Take random sample

Obtain model output

Repeat process a large number of times to generate output distribution

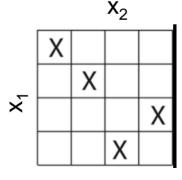


Henderson & Bui 2005

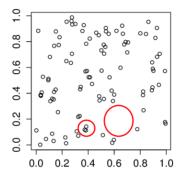


- einfache Zufallswerte (random values)
 - z.B. in Python über numpy.random

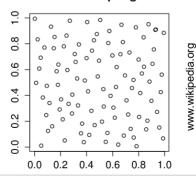
- "near-random" Zufallswerte
 - Latin Hypercube sampling
 - Orthogonal sampling
 - Sobol Sequenzen
 - **...**



Random Uniform



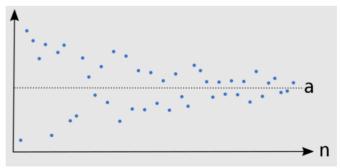
LH Sampling



Wie groß muss n sein?



- ▶ Konvergenz gegen unbekanntes $\Phi(z)$
- ▶ 1. Option: n sehr hoch ansetzen (z.B. 10⁷)
 - lange Rechenlaufzeiten!
- 2. Option: n schrittweise erh\u00f6hen und Ergebnisse beobachten



Übung 6: Monte Carlo Methoden



- Bestimmung der Unsicherheit von biologischen Abbauraten in einem kontaminierten Aquifer
 - Monte Carlo Simulation für das analytische Modell in Eq. (6):

$$\lambda = \frac{\Delta \delta^{13} C \cdot k_f \cdot i}{\varepsilon \cdot s \cdot n_e}$$

- Angaben zu Parametern in Tab. 1
- ► Brunnen G10m G30u
- Aufgaben in Jupyter Notebook: 06_Monte_Carlo_uebung



natural attenuation (MNA) at such field sites.

. Introduction

The contentionism of the robustness due to leaking point surveine from inductivity joint in a vollespone flower, and often require extensive form inductivity joint in a vollespone flower, and often require extensive long term monitoring and remailisation Gibnource-should read, 2010. As some continuous control of the co

During the hart decades stable compound-specific issues analysis (Cidi) has been estimational set the mine method for amenting the excursions of bindergardation in contaminated appellers (a.g. Simulation Schollers of the Contaminated appellers (a.g. Simulation Schollers of the Contaminated Appellers of th

Modia biologicalisis rate contains for explore ranging between 0.08 and 0.22 ⁻¹ were entirated. by unite treers based groundword until gain there two wells, byfoultic conditatives could be its orientate, which are in a similar range as is values derived from sieve analysis, a pumping test and a calibrated groundword from modist. These results forcely demonstrate the populoshilty of trace-based groundword entire for the elemination of its sits byformial conductivities in aquifore without pumping communicated groundworse. Finally, ascentifying pulsar just private using a Manter Code installation. These results induced by terministies of the entirely majority in general using a Manter Code installation. These results induced by terministies of the

the main limitations of the novel combined isotone approach for a successful implementation of monitore

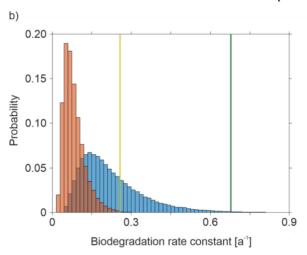
Impe/, vol. 01/2/10.1016/s) 300m/s, 2020.10075/
Received 5 June 2020; Received in revised form 26 October 2020; Accepted 12 December 2020 Available online 5 January 2021
1069-7722/0 2020 Bisevier B.V. All rights reserved.

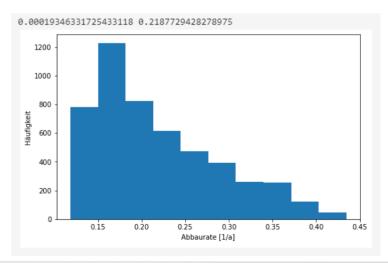
Corresponding author.
 E-mail address: blum@kit.edu (P. Blum).



Aufgabenbesprechung

► Monte Carlo Simulation (Würth et al. 2021, Abb. 3)





Literatur



- Bättig (2017): Angewandte Datenanalyse, 2. Aufl., Springer
 Spektrum
- Gelman et al. (2014): Bayesian Data Analysis, 2nd Ed., CRC
 Press
- Würth et al. (2021): Quantifying biodegradation rate constants of o-xylene by combining compound-specific isotope analysis and groundwater dating. Journal of Contaminant Hydrology, 238, 103757



