

# Geodatenanalyse I: Monte Carlo Methoden

Kathrin Menberg



# Stundenplan

Vorläufiger Stundenplan		
Datum	Thema	Dozent
20.10.2021	Einführung in die Programmierung mit <i>Python</i>	Gabriel Rau
25.10.2021	Univariate Statistik und statistisches Testen	Kathrin Menberg
01.11.2021	<i>Feiertag</i>	
08.11.2021	Umgang und Berechnung von Datensätzen	Gabriel Rau
15.11.2021	Bivariate und schließende Statistik	Kathrin Menberg
22.11.2021	Datenvisualisierung mit <u><i>matplotlib</i></u>	Gabriel Rau
29.11.2021	Multivariate Statistik	Kathrin Menberg
06.12.2021	Einführung in den Umgang mit Datensätzen	Gabriel Rau
13.12.2021	<i>ausgefallen</i>	Kathrin Menberg
20.12.2021	Weiterführender Umgang mit Datensätzen	Gabriel Rau
27.12.2021	<i>Weihnachtsferien</i>	
03.01.2022	<i>Weihnachtsferien</i>	
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, Lizenzierung und Entwicklungstools	Gabriel Rau
07.02.2022	Regressionsanalyse	Kathrin Menberg

# Vorlesungsplan

Uhrzeit	Inhalt
10:00 – 10:30	<b>Monte Carlo Methoden</b>
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.

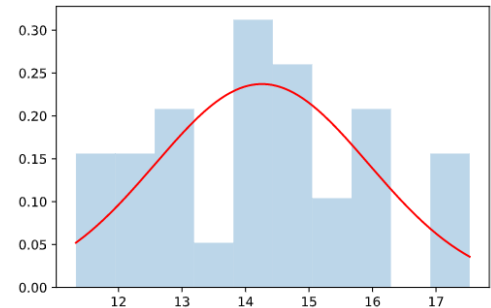
# Unsicherheiten

*„Nothing is certain, but our ignorance.“* (unbekannt)

- ▶ Welche „Unsicherheiten“ fallen Euch in Bezug auf Geodatenanalyse ein?

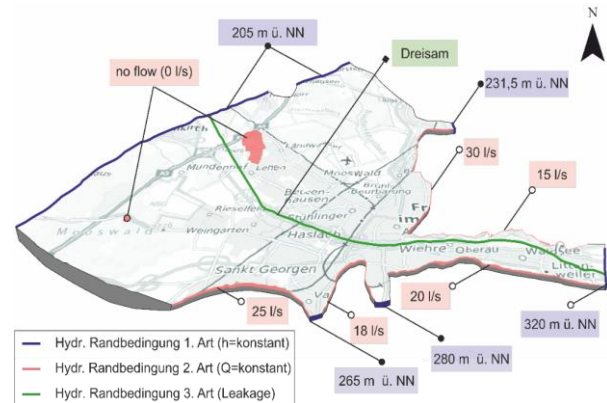
# Arten von Unsicherheit

- ▶ **Aleatorisch** (aleatoric)
  - ▶ Statistische Unsicherheit
  - ▶ Unbestimmbare Größen, die in wiederholten Experimenten variieren
  - ▶ Komplexe Prozesse, die nicht erfassbar sind
  - ▶ usw.
  
- ▶ **Epistemisch** (epistemic)
  - ▶ Systematische Unsicherheit
  - ▶ Größen, die in Experimenten nicht gemessen wurden.
  - ▶ Vereinfachungen in numerischen Modellen
  - ▶ Usw.



# Quellen von Unsicherheiten

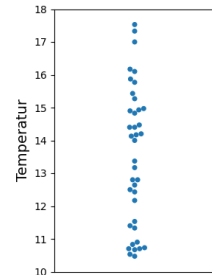
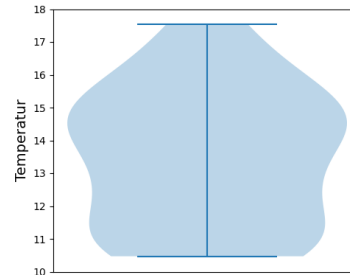
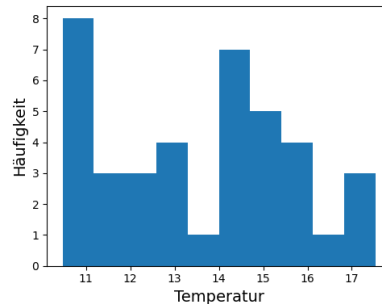
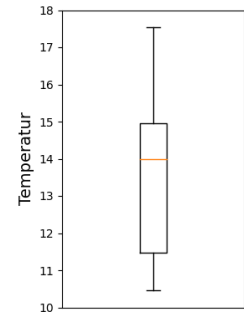
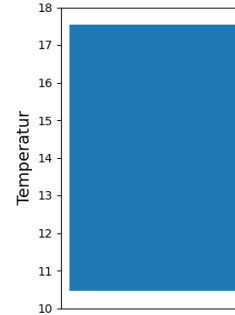
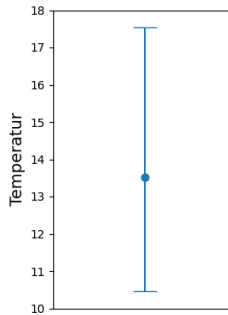
- ▶ Daten, Parameter
  - ▶ Messfehler, Messgenauigkeit
  - ▶ Szenarien
- ▶ Modelle
  - ▶ Annahmen (Randbedingungen usw.)
  - ▶ Vernachlässigte Prozesse
  - ▶ Auflösung, bzw. Diskretisierung
  - ▶ Numerische Fehler, Approximation
- ▶ Modellierer
  - ▶ Bedienungsfehler, Eingabefehler
- ▶ und viele mehr...



Ohmer et al. (2021)

# Visualisierung von Unsicherheiten

- ▶ Fehlerbalken
- ▶ Balkendiagramme
- ▶ Histogramme
- ▶ Boxplots
- ▶ Violinen Diagramme
- ▶ Schwarmplot
- ▶ u.v.m.

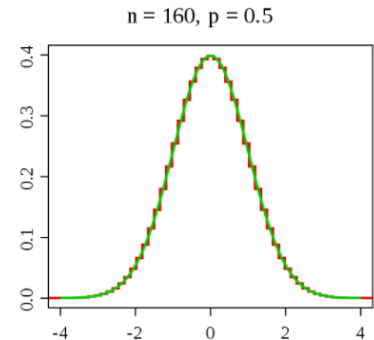
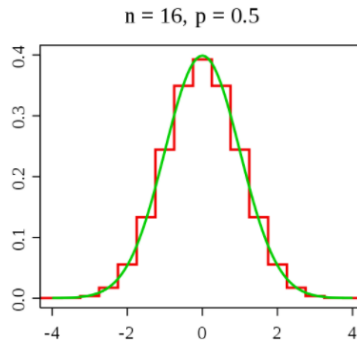
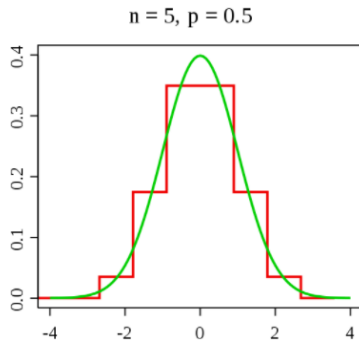




# Zentraler Grenzwertsatz

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

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z)$$



[www.wikipedia.org](http://www.wikipedia.org)

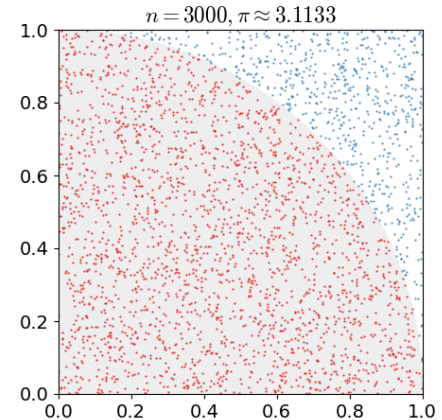
# Monte Carlo Simulation

## ► Ziele (Beispiele)

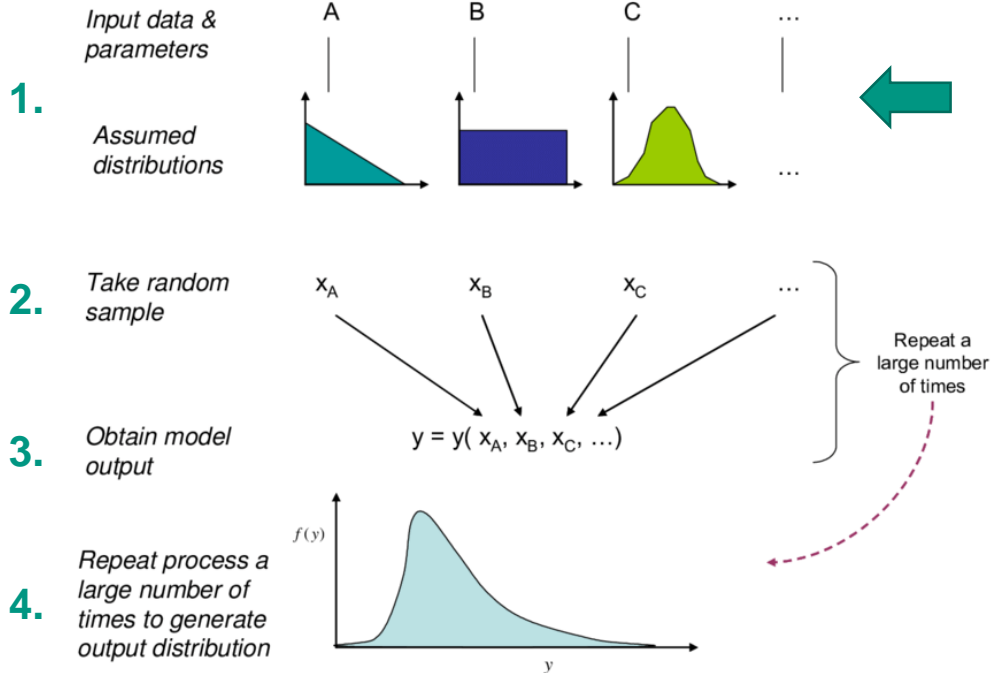
- Approximation von irrationalen Zahlen
- Verteilungseigenschaften von Zufallsvariablen
- Nachbildung komplexer Prozesse (z.B. Wetter- und Klimaphänomene)

## ► Prinzip:

1. mögliche Inputs definieren
2. zufällige Inputs anhand einer Wahrscheinlichkeitsfunktion generieren
3. deterministische Berechnung mit den Inputs durchführen
4. Ergebnisse zusammenfassen



# Monte Carlo Simulation

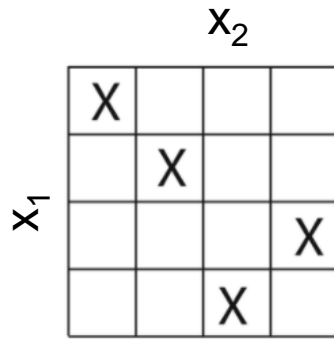


# Generieren von Zufallswerten

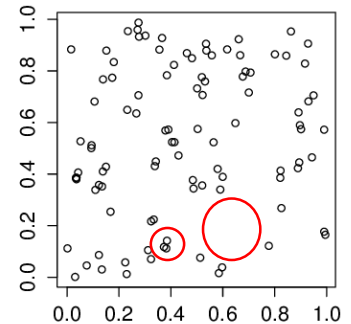
- ▶ einfache Zufallswerte (random values)
  - ▶ z.B. in Python über `numpy.random`

- ▶ „near-random“ Zufallswerte

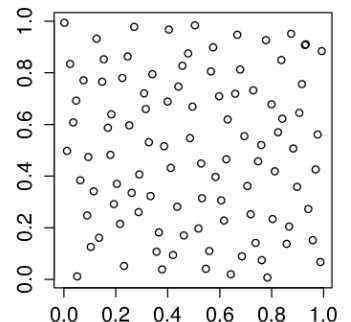
- ▶ Latin Hypercube sampling
- ▶ Orthogonal sampling
- ▶ Sobol Sequenzen
- ▶ ...



**Random Uniform**

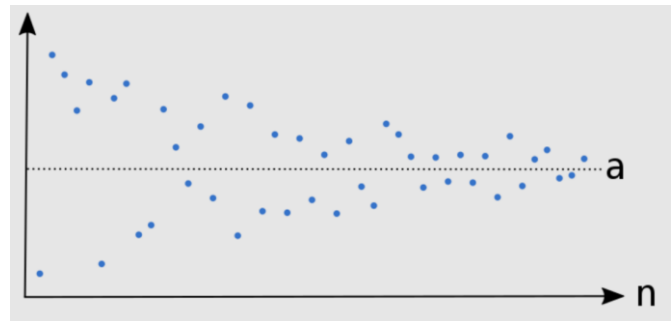


**LH Sampling**



# Wie groß muss $n$ sein?

- ▶ Konvergenz gegen unbekanntes  $\Phi(z)$
- ▶ 1. Option:  $n$  sehr hoch ansetzen (z.B.  $10^7$ )
  - ▶ lange Rechenlaufzeiten!
- ▶ 2. Option:  $n$  schrittweise erhöhen 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):

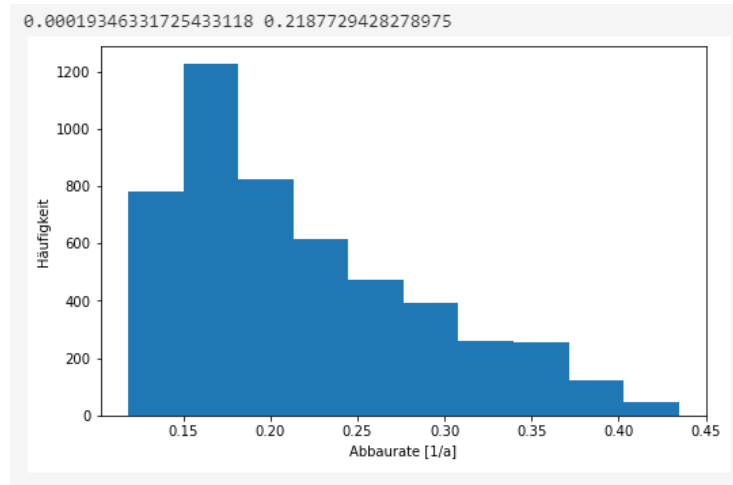
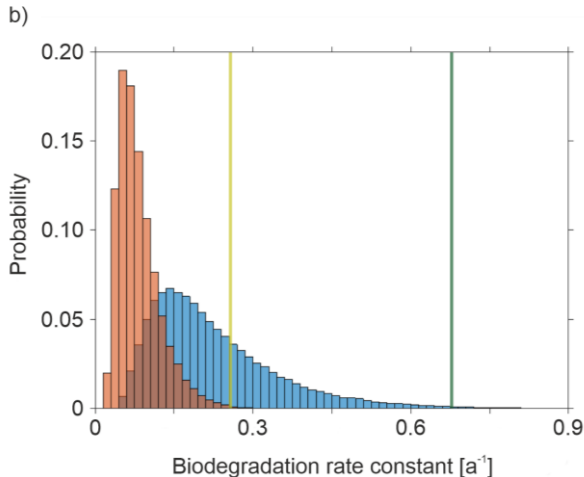
$$\lambda = \frac{\Delta\delta^{13}C \cdot k_f \cdot i}{\varepsilon \cdot S \cdot n_e}$$

- ▶ Angaben zu Parametern in Tab. 1
- ▶ Brunnen G10m – G30u
- ▶ Aufgaben in Jupyter Notebook:  
06\_Monte\_Carlo\_uebung



# 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



