

Monte Carlo: Zufall gegen Systematik

Hansruedi Künsch

Seminar für Statistik
Departement Mathematik, ETH Zürich

22. Mai 2023, Emeriten-Stamm, Winterthur

- Monte Carlo Algorithmen simulieren den Zufall, um Probleme zu lösen. Die Problemstellung selber kann auch rein deterministisch sein.
- “Zufällig” ist nicht das Gleiche wie “willkürlich”, “beliebig” oder “auf’s Geratewohl”.
- Monte Carlo ist oft einfacher, schneller, besser als ein deterministischer Algorithmus oder eine theoretische Analyse.
- Monte Carlo ermöglicht a posteriori probabilistische Fehlerangaben meist ohne grossen zusätzlichen Aufwand.
- Die einfachste Monte Carlo Methode ist die Zufallsstichprobe.
- Der erste Teil des Vortrags zeigt Beispiele für die Anwendung von Monte Carlo Methoden. Der zweite Teil behandelt die Simulation des Zufalls mit dem Computer.

Hertzprung und van de Hulst

Hertzprung 1942: “The simplest way to deal with exorbitant observations is to reject them. In order to avoid special rules for onesided rejection reject the largest deviations symmetrically to each side. ... How much is, in the case of Gaussian distribution of errors, the weight of the result diminished by symmetrical rejection of outstanding observations?”

Hertzprung benutzte Monte Carlo: “... mathematical treatment of this question appears to be laborious beyond the needs”. Er simulierte 1000 Messreihen der Länge 24 und berechnete jeweils 12 Mittelwerte (Weglassen von 0,1,...,11 Werten an beiden Enden). Damals war das ebenfalls aufwändig.

Van de Hulst entwickelte 1943 eine analytische Lösung im Grenzfall, wo die Länge n der Messreihe gegen unendlich geht und der Anteil weggelassener Werte fest bleibt.

Exkurs: Was bedeutet “weight of the result?”

Gewicht des gestutzten Mittels gibt an, wie viele Messungen man benötigt, um mit dem gewöhnlichen Mittel die gleiche Genauigkeit zu erreichen.

Werden in einer Messreihe der Länge 24 auf beiden Seiten je die 5 extremsten Werte weggelassen, dann ergab Monte Carlo dafür das Gewicht 21. Man bezahlt die Absicherung gegen grobe Fehler bei Gaussverteilung mit einem Informationsverlust von 3 Messungen (12.5%), nicht von 10, wie man naiverweise vermuten könnte.

Vergleich Monte Carlo und analytische Lösung

Fehler bei Monte Carlo:

Zufallsfehler (nur 1000 Stichproben) und Fehler bei Extrapolation auf Werte $n \neq 24$. Bei der analytischen Lösung: nur Fehler wegen Extrapolation rückwärts von $n = \infty$.

Verallgemeinerung auf nicht-Gaussverteilte Messwerte:

Monte Carlo anwendbar für beliebige Verteilungen. Analytische Lösung für beliebige stetige Verteilungen.

Verteilung der Messwerte unbekannt:

Monte Carlo benutzt empirische Verteilung der Messfehler (Bootstrap, Efron 1979). Analytische Lösung benutzt den Zentralen Grenzwertsatz mit empirischer Varianz der Messfehler. Fehler beim Bootstrap geht schneller gegen null für $n \rightarrow \infty$.

Mehrdimensionale Integration

Approximiere das Integral

$$I(f) = \int \dots \int_{[0,1]^d} f(x_1, \dots, x_d) dx_1 \dots dx_d$$

durch ein arithmetisches Mittel

$$I(f) \approx \frac{1}{N} \sum_{i=1}^N f(x_1^{(i)}, \dots, x_d^{(i)})$$

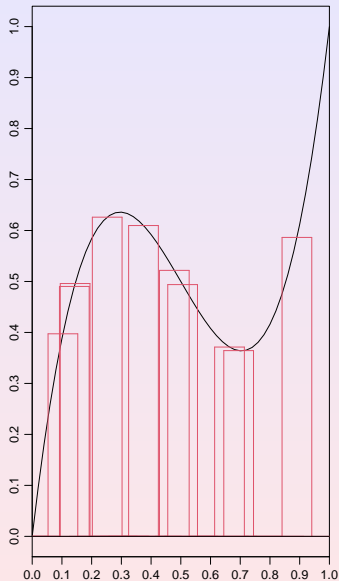
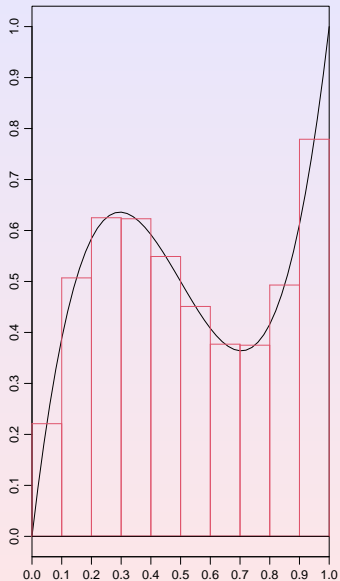
mit geeignet gewählten Stützstellen.

Systematisch: Wähle $N = n^d$ regelmässige Stützstellen

$$(x_1^{(i)}, \dots, x_d^{(i)}) \in \left\{ \frac{0.5}{n}, \dots, \frac{n-0.5}{n} \right\}^d.$$

Monte Carlo: Wähle die N Stützstellen unabhängig und uniform verteilt auf $[0, 1]^d$.

Illustration für $d = 1$



Vergleich der beiden Methoden

Systematisch:

Der Fehler geht für zweimal stetig differenzierbare Funktionen wie $N^{-2/d}$ gegen Null. Restriktion $N = n^d$. Fehlerabschätzung durch Vergleich der Approximationen für zwei verschiedene Werte von n .

Zufällig:

Der Fehler geht für beliebige quadrat-integrierbare Funktionen und für alle Dimensionen wie $N^{-1/2}$ gegen Null. Keine Restriktion für N . Fehlerabschätzung vom Zentralen Grenzwertsatz.

Verwandtes Problem: Wie platziert man Kugeln im d -dimensionalen Raum, so dass möglichst wenig Raum leer bleibt ?

Maryna Viazovska löste das Problem in Dimension 24 und erhielt dafür 2022 die Fields-Medaille: “ ... die beste Anordnung [in hohen Dimensionen ist] wahrscheinlich zufällig, nicht strukturiert wie in Räumen niedriger Dimensionen. Eine wichtige Frage: Siegt der Zufall über die Struktur ?” (Interview, Horizonte 136, 2023)

Das Ising Modell

Einfachstes Modell eines Ferromagnets: Interagierende Spins auf dem Gitter $L_n = \{-n, -n+1, \dots, n-1, n\}^d$.

Konfigurationen $\sigma = (\sigma_i; i \in L_n)$ mit $\sigma_i = \pm 1$ (up/down).
Wahrscheinlichkeit einer Konfiguration σ gegeben durch

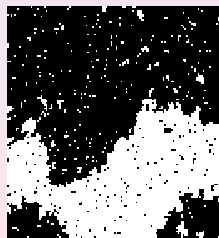
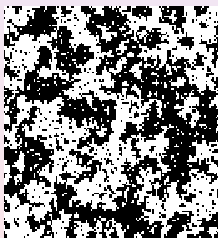
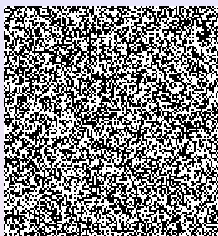
$$p_{n,b}(\sigma) = \frac{1}{Z_{n,b}} \exp(-bH_n(\sigma))$$

wobei $Z_{n,b}$ = Normierung, b proportional zu $1/\text{Temperatur}$,
 $H_n(\sigma)$ = Energie der Konfiguration σ :

$$H_n(\sigma) = \sum_{i,j \text{ benachbart}} \sigma_i \sigma_j.$$

Je tiefer die Temperatur, desto eher sind benachbarte Spins gleich. Dies illustriert eine Simulation sehr schön.

Simulation des Ising Modells



$n = 64$ und $b = 0, 0.3, 0.4, 0.5$

Magnetisierung im Ising Modell

Bei einem Magneten sollte das Verhalten temperaturabhängig sein: Lokale Interaktion führt zu globaler Abhängigkeit bei tiefer, aber nicht bei hoher Temperatur. Mathematisch:

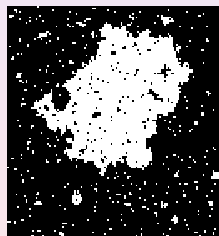
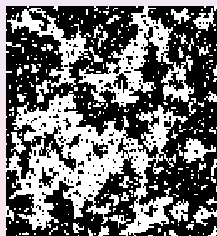
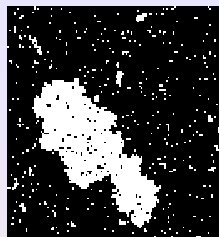
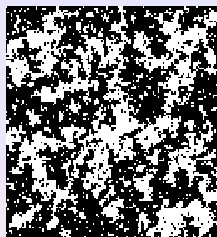
$$\lim_n \mathbb{E}_{n,b}(\sigma_{0,0} | \text{Spins am Rand} = 1) \begin{cases} = 0 & (b \text{ klein}) \\ > 0 & (b \text{ gross}) \end{cases}$$

(Phasenübergang).

Ising (1924): Für $d = 1$ und $d = 2$ hat das Modell keinen Phasenübergang. Für $d = 2$ ist sein "Beweis" leider falsch.

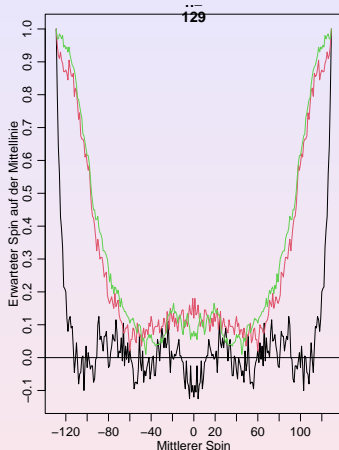
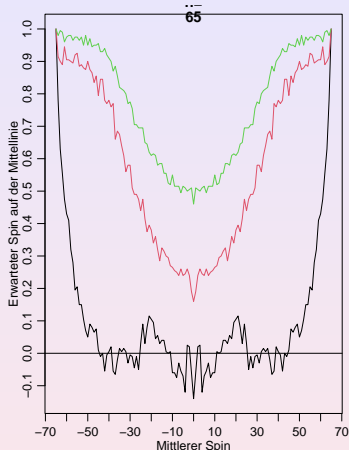
Peierls (1936): Im Wesentlichen korrekter Beweis für Phasenübergang bei $d = 2$. Onsager (1944): Exakte Lösung (Wert von b , an dem der Phasenübergang auftritt).

Ising Modell mit Randbedingung +1



Je 2 Simulationen mit $n = 65$ und $b = 0.4$ (links), $b = 0.5$ (rechts).

Simulation der Magnetisierung



$E_{n,b}(\sigma_{0,i} | \text{Spins am Rand} = 1)$ gegen i für $n = 65$ und $n = 129$.
Schwarz: $b = 0.3$, rot: $b = 0.4$, grün: $b = 0.5$

Ionenkanäle sind Proteine in der Zellmembran, welche gewisse Ionen in die Zelle hinein oder hinausleiten. Ziel ist, die Zeitpunkte, an denen sich der Kanal öffnet, bzw. schliesst, aus einer fehlerbehafteten Messreihe zu rekonstruieren.

de Gunst et al. (2001) entwickeln eine Monte Carlo Methode und vergleichen sie mit dem sogenannten Hinkley detector (1971), einem einfacheren deterministischen Verfahren. Die Abbildung zeigt das Hauptresultat. Mehr Details zum Verfahren später.

Für die Spezialisten: “The recorded samples are a cell-attached-patch measurement of a single potassium outward rectifier in a leaf protoplast of barley.”

Abbildung aus de Gunst et al. (2001)

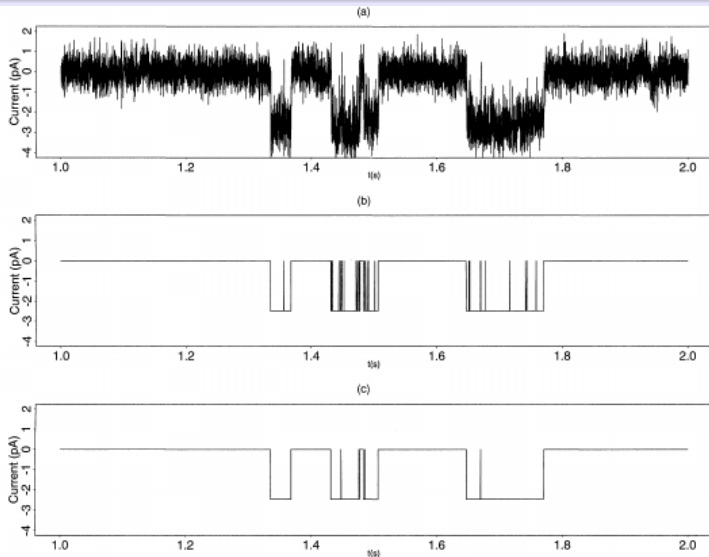


Figure 3. (a) Part of Recording 1 (1.0 sec); (b) the Outcomes of the Gibbs Sampler for indicator Loss; and (c) the Reconstruction From the Hinkley Detector.

Die Anfänge von Monte Carlo

Zunächst wurden einfache physikalische Experimente benutzt. Ein frühes Beispiel ist die Bestimmung von π durch die Zählung der Schnittpunkte von zufällig gestreuten Nadeln mit parallelen Linien (Buffon, 1733, 1777; Laplace, 1812).

Hertzprung verwendete zufällige Ziehungen aus einer Schachtel mit 12'534 Losen. Auf jedem Los stand eine Zahl, die Zahlenwerte folgten einer auf 2 Kommastellen diskretisierten Gaussverteilung.

Stanislaw Ulam schlug als erster vor, Zufallszahlen mit Computern zu erzeugen. Zusammen mit Nicholas Metropolis publizierte er 1949 das erste Paper dazu. Auch die Bezeichnung "Monte Carlo" stammt von dort.

Zufallszahlen auf dem Computer

In einem ersten Schritt werden Zufallszahlen generiert, welche auf $[0, 1]$ gleichverteilt sind. Mit geeigneten Transformationen kann man daraus Zufallszahlen mit anderen Verteilungen erzeugen.

Vom Computer produzierte Zufallszahlen sind natürlich nicht wirklich zufällig. Es genügt jedoch, wenn sie nicht von echten Zufallszahlen zu unterscheiden sind.

Bis gute computergenerierte Zufallszahlen einfach verfügbar waren, dauerte es: 1955 publizierte die RAND corporation das Buch “A Million Random Digits with 100,000 Normal Deviates”. Der Generator RANDU, der bis in die 70er Jahre verwendet wurde, ist gemäss Donald Knuth “truly horrible”.

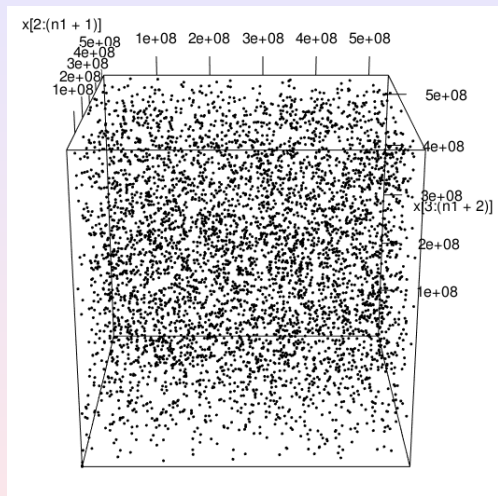
Der Weg zu guten Zufallsgeneratoren

Es genügt nicht, dass das Histogramm von n Zufallszahlen (für n gross) im wesentlichen konstant ist. Es sollten auch auch n d -tupel “gleichmässig” in $[0, 1]^d$ verteilt sein für $d = 2, 3, \dots$

RANDU versagt bereits bei $d = 3$, siehe nächste 2 Abbildungen.

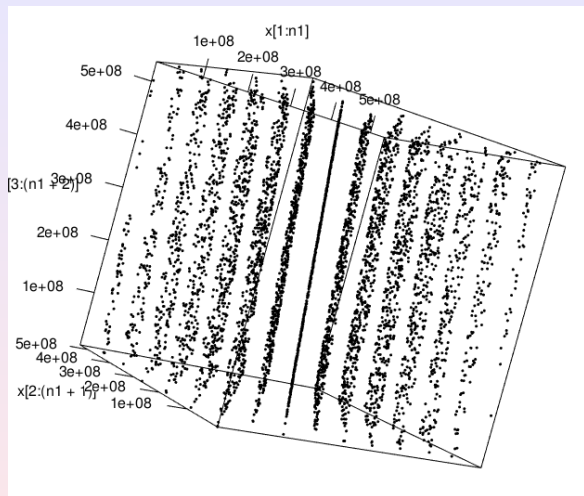
Heute gibt es zuverlässige Zufallszahlgeneratoren in verschiedenen Softwarepaketen. Diese benutzen meist eine Kombination verschiedener Methoden.

5000 Trippel von RANDU



Perspektivische Darstellung des Einheitswürfels.

5000 Trippel von RANDU, Fts.



Gleiche 5000 Trippel wie in vorangehenden Abbildung nach Drehung des Einheitswürfels.

Lineare Kongruenzgeneratoren

Dies sind die einfachsten Generatoren. Sie sind definiert als

$$x_j = \frac{z_j}{m}, \quad z_j = (az_{j-1} + c) \pmod{m}.$$

Dabei sind $m, a, c \in \mathbb{N}$ die Parameter des Generators. RANDU verwendet $m = 2^{29}$, $a = 2^{16} + 3$ und $c = 0$.

Die Qualität eines solchen Generators kann wie folgt beschrieben werden

- m kontrolliert die Periode des Generators.
- Alle d -tupel $(x_{i+1}, \dots, x_{i+d})$ liegen auf Scharen von parallelen Hyperebenen ("random numbers fall mainly in the planes"). Wie gross der maximale Abstand zweier benachbarter Hyperebenen ist, hängt von m , a und d ab. Es gibt gute und schlechte Wahlen von a für gegebenes m und d .

Zufallszahlen mit beliebiger Verteilung

Gegeben eine Zielverteilung p auf \mathbb{R}^d und eine unendliche Folge von Zufallszahlen x_1, x_2, \dots so dass die k -Tupel für alle k gleichverteilt sind auf $[0, 1]^k$.

Gesucht: Eine Transformation von k -Tupeln der x_i so dass die Werte

$$y_i = F(x_{(i-1) \cdot k + 1}, \dots, x_{i \cdot k})$$

die Verteilung p haben. Für kleines d meist einfach.

Für grösseres d muss man oft die Ansprüche reduzieren.
Markovketten Monte Carlo (MCMC) bestimmt eine rekursive Transformation

$$y_i = F(y_{i-1}, x_{(i-1) \cdot k + 1}, \dots, x_{i \cdot k})$$

so dass die y_i asymptotisch p -verteilt sind.

Zielverteilung im Beispiel der Ionenkanäle

Variablen:

\mathbf{x} = (latenter) Zustand des Kanals zu den Zeiten $\Delta, 2\Delta, \dots$

\mathbf{y} = Messwerte zu den gleichen Zeiten.

θ = Vektor der unbekannt Parameter (Übergangsraten des Kanals, Mittelwerte des Kanals im offenen/geschlossenen Zustand, Varianzen und Korrelationen der Messfehler).

Die folgenden Verteilungen müssen spezifiziert werden:

$p(\theta)$ = a priori Verteilung für θ , basierend auf Vorinformation.

$p(\mathbf{x}|\theta)$ = bedingte Verteilung von \mathbf{x} bei bekanntem θ , basierend auf einem Modell für den Ionenkanal.

$p(\mathbf{y}|\mathbf{x}, \theta)$ = bedingte Verteilung von \mathbf{y} bei bekanntem Zustand \mathbf{x} und bekanntem θ , basierend auf einem Modell für den Messvorgang.

Zielverteilung im Beispiel der Ionenkanäle, Fs.

Daraus ergibt sich die gemeinsame Verteilung

$$p(\mathbf{y}, \mathbf{x}, \theta) = p(\mathbf{y}|\mathbf{x}, \theta)p(\mathbf{x}|\theta)p(\theta).$$

Werte gemäss dieser Verteilung zu ziehen ist nicht schwierig.
Vorgehen in der Reihenfolge $\theta \rightarrow \mathbf{x} \rightarrow \mathbf{y}$.

Die Rekonstruktion von \mathbf{x} und die Schätzung von θ auf Grund der Messungen \mathbf{y} ist jedoch ein inverses Problem: Man benötigt Werte gemäss

$$p(\mathbf{x}, \theta|\mathbf{y}) = \frac{p(\mathbf{y}, \mathbf{x}, \theta)}{\int p(\mathbf{y}, \mathbf{x}, \theta) d\mathbf{x}d\theta}.$$

Dies ist wesentlich schwieriger \rightarrow Markovketten Monte Carlo.

Bei vielen Anwendungen hat man eine analoge Struktur.

Statt “Zufall oder Systematik ?” sollte es heissen “Zufall und Systematik !”

Jede Methode, die brauchbare Antworten auf eine interessante Fragestellung liefert, ist willkommen.

Monte Carlo kann die Intuition für eine systematische Behandlung liefern. Systematik kann helfen, Monte Carlo Algorithmen zu konstruieren und zu verstehen.

Ob der Zufall über die Struktur siegt, wird wohl offen bleiben.

VIELEN DANK FÜR'S ZUHÖREN !