Block 6 (Donnerstag 22.2.2024)

VIDEO: `video06a_grundlagen_r`

8 Datenanalyse und Zufallszahlen

Statistische Datenanalyse: für deterministische Simulation (Molekulardynamik, DGLs) und für stochastische Simulationen.
Für letzteres weiterhin: Anwendung von Zufallszahlen bei Computersimulationen:

- Systeme mit zufälligen Wechselwirkungen ("Spingläser")
- Simulationen bei endlichen Temperaturen mit Monte-Carlo Verfahren
- Randomisierte Algorithmen (aus deterministischen Algorithmen entstanden)

Zufallszahlen im Computer möglich (z.B. Schwankungen der Spannungen an einem Transistor durch thermisches Rauschen). Vorteil: Zufällig. Nachteil: Statistische Eigenschaften unbekannt und nicht kontrollierbar.
Daher: Pseudozufallszahlen = nicht zufällig, aber *möglichst* gleiche statische Eigenschaften (Verteilung, Korrelationen).

8.1 Grundlagen Wahrscheinlichkeitstheorie

Ω: Menge der Ausgänge eines Zufallsexperiments.

Bsp: $\Omega = \{\text{Kopf, Zahl}\}$ für Münzwurf.

Definition: Eine Wahrscheinlichkeitsfunktion P ist eine Funktion $P : 2^\Omega \longrightarrow [0,1]$ mit

$$P(\Omega) = 1 \tag{7}$$

und für jede Sequenz $A_1, A_2, A_3, \ldots$ disjunkter Ereignisse ($A_i \cap A_j = \emptyset$ für $i \neq j$):

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots \tag{8}$$

Eigenschaften (ohne Beweis), für $A, B \subset \Omega$, $A^c := \Omega \setminus A$:

$$P(A^c) = 1 - P(A)\,. \tag{9}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{10}$$

Für unabhängige Zufallsexperimente gilt:

$$P(A^{(1)}, A^{(2)}, \ldots, A^{(k)}) = P(A^{(1)})P(A^{(2)}) \ldots P(A^{(k)}) \tag{11}$$

Definition: Die bedingte Wahrscheinlichkeit A gegeben C ist

$$P(A|C) = \frac{P(A \cap C)}{P(C)}\,. \tag{12}$$

[Selbsttest]

Was ist P(Würfel 6|Würfel$>$ 3)?

Bayesche Regel

Da nach (12) $P(A|C)P(C) = P(A \cap C) = P(C \cap A) = P(C|A)P(A) \Rightarrow$

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)}\,. \tag{13}$$

VIDEO: `video06b_ZVen_r`

8.2 Zufallsvariablen

Zufallsvariable (ZV) X (unscharf): Zufallsexperiment mit $\Omega = \mathbb{R}$

Definition: Verteilungsfunktion (VF) einer ZV X ist eine Funktion F_X : $\mathbb{R} \longrightarrow [0,1]$ definiert über

$$F_X(x) = P(X \le x) \tag{14}$$

Bsp: Münze:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}\,. \tag{15}$$

x-Position eines Gasteilchens im Container $[0, L_x]$

$$F(x) = \begin{cases} 0 & x < 0 \\ x/L_x & 0 \le x < L_x \\ 1 & x \ge L_x \end{cases} . \tag{16}$$

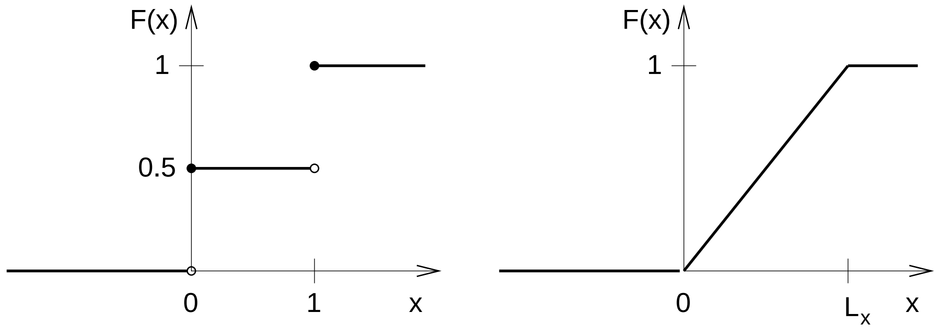

Figure 10: Verteilungsfunktionen für Münze und Gasteilchen.

Nach Zufallsexperiment gemäß X mit Ergebnis x, dann $y = g(x)$ berechnen: Transformation der ZV zu $Y = g(X)$, allgemein:

$$Y = \tilde{g}\left(X^{(1)}, X^{(2)}, \ldots, X^{(k)}\right) . \tag{17}$$

Verteilungsfunktion manchmal unhandlich → Wahrscheinlichkeits-/ Dichtefunktion:

VIDEO: `video06c_ZV_diskret_r`

8.2.1 Diskrete Zufallsvariablen

Definition: Die Wahrscheinlichkeitsfunktion (engl. probability mass function, pmf) $p_X : \mathbb{R} \to [0, 1]$ ist gegeben durch

$$p_X(x) = P(X = x) . \tag{18}$$

Verteilung für n fachen Münzwurf (0/1):

Definition: Die *binomial Verteilung* mit Parametern $n \in \mathbb{N}$ und p $(0 < p \le 1)$ beschreibt eine ZV X mit WF

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & (0 \le x \le n, x \in \mathbb{N}) \\ 0 & \text{sonst} \end{cases} \tag{19}$$

Notation $X \sim B(n, p)$.

Charakterisierung von ZVen:

$\{\tilde{x}_i\}$ Menge der Werte für die $p_X(\tilde{x}) > 0$.
Definition:

- Erwartungswert

$$\mu \equiv \mathrm{E}[X] = \sum_i \tilde{x}_i P(X = \tilde{x}_i) = \sum_i \tilde{x}_i p_X(\tilde{x}_i) \tag{20}$$

- Varianz

$$\sigma^2 \equiv \mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \sum_i (\tilde{x}_i - \mathrm{E}[X])^2 p_X(\tilde{x}_i) \tag{21}$$

- Standardabweichung

$$\sigma \equiv \sqrt{\mathrm{Var}[X]} \tag{22}$$

Eigenschaften:

$$\begin{aligned} \mathrm{E}[\alpha_1 X^{(1)} + \alpha_2 X^{(2)}] &= \alpha_1 \mathrm{E}[X^{(1)}] + \alpha_2 \mathrm{E}[X^{(2)}] && (23) \\ \sigma^2 = \mathrm{Var}[X] &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 && (24) \\ \Leftrightarrow \quad \mathrm{E}[X^2] &= \sigma^2 + \mu^2 && (25) \\ \mathrm{Var}[\alpha_1 X^{(1)} + \alpha_2 X^{(2)}] &= \alpha_1^2 \mathrm{Var}[X^{(1)}] + \alpha_2^2 \mathrm{Var}[X^{(2)}] && (26) \end{aligned}$$

$\mathrm{E}[X^n]$: n-tes Moment
Es gilt für die Binomialverteilung:

$$\begin{aligned} \mathrm{E}[X] &= np && (27) \\ \mathrm{Var}[X] &= np(1-p) && (28) \end{aligned}$$

VIDEO: `video06d_ZV_kont_r`

8.2.2 Kontinuierliche Zufallsvariablen

Definition: Für eine ZV X mit kontinuierlicher VF F_X, ist die Wahrscheinlichkeitsdichte (engl. probability density function, pdf) $p_X : \mathbb{R} \to \mathbb{R}$:

$$p_X(x) = \frac{dF_X(x)}{dx} \tag{29}$$

Damit:

$$F_X(x) = \int_{-\infty}^{x} p_X(\tilde{x})\, d\tilde{x} \tag{30}$$

Definition:

- Erwartungswert

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} dx\, x\, p_X(x) \tag{31}$$

- Varianz

$$\mathrm{Var}[X] = \mathrm{E}[(X-\mathrm{E}[X])^2] = \int_{-\infty}^{\infty} dx\, (x-\mathrm{E}[X])^2 p_X(x) \tag{32}$$

- Median $x_{\mathrm{med}} = \mathrm{Med}[X]$

$$F_X(x_{\mathrm{med}}) = 0.5 \tag{33}$$

Definition: Gleichverteilung mit Parametern $a < b$, beschreibt ZV X mit pdf

$$p_X(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \le x < b \\ 0 & x \ge b \end{cases} \tag{34}$$

Schreibweise: $X \sim U(a,b)$.
Mittels $g(X_{01}) = (b-a) * X_{01} + a$ erhält man $g(X_{01}) \sim U(a,b)$ falls $X_{01} \sim U(0,1)$.

[Selbsttest]

Berechnen sie $\mathrm{E}[X]$ und $\mathrm{Var}[X]$.

Am wichtigesten:
Definition: Die Gaußverteilung oder Normalverteilung mit Parametern μ und $\sigma > 0$, beschreibt die ZV X mit pdf

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{35}$$

Schreibweise: $X \sim N(\mu, \sigma^2)$.
Eigenschaften: $\mathrm{E}[X] = \mu$, $\mathrm{Var}[X] = \sigma^2$.
Mittels $g(X) = \sigma X_0 + \mu$ erhält man $g(X_0) \sim N(\mu, \sigma^2)$ falls $X_0 \sim N(0,1)$.

Zentraler Grenzwertsatz:
Für unabhängige ZVen $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ mit jeweils $E[X^{(i)}] = \mu$ und $\mathrm{Var}[X^{(i)}] = \sigma^2$ gilt:

$$X = \sum_{i=1}^{n} X^{(i)} \tag{36}$$

ist für große n $X \sim N(n\mu, n\sigma^2)$.
Dichten weiterer wichtiger Verteilungen:
Definition:

- Exponentialverteilung (für $x \geq 0$)

$$p_X(x) = \frac{1}{\mu} \exp\left(-x/\mu\right) \tag{37}$$

[Selbsttest]

Berechnen Sie die Verteilungsfunktion für die Exponentialverteilung

---

Definition:

- Potenz-Gesetz Verteilung oder Pareto Verteilung

$$p_X(x) = \begin{cases} 0 & x < 1 \\ \frac{\gamma}{\kappa}(x/\kappa)^{-\gamma-1} & x \geq 1 \end{cases} \tag{38}$$

  Für $\gamma > 1$ existiert Erwartungswert $\mathrm{E}[X] = \gamma\kappa/(\gamma-1)$, für $\gamma > 2$ $\mathrm{Var}[X] = \frac{\kappa^2\gamma}{(\gamma-1)^2(\gamma-2)}$

$$F_X(x) = 1 - (x/\kappa)^{-\gamma} \quad (x \geq 1) \tag{39}$$

- Fisher-Tippett Verteilung

$$p_X(x) = \lambda e^{-\lambda x} e^{-e^{-\lambda x}} \tag{40}$$

  (auch Gumbel Verteilung für $\lambda = 1$) $\mathrm{E}[X] = \nu/\lambda$, $\nu \equiv 0.57721\ldots$, Maximum bei $x = 0$, Verschiebung durch $x \to (x - \mu)$

[Selbsttest]

Können Sie die Verteilungsfunktion ablesen ?
