Block 6 (Donnerstag 22.2.2024)

VIDEO: `video06a_grundlagen_r`

8 Data Analysis and Random Numbers

Statistical data analysis: for deterministic simulations (Molecular Dynamics, DEs) and also for stochastic simulations.
For the latter ones also: Application of random numbers in computer simulations:

- Systems with random interactions (e.g. "Spin glasses")
- simulations at finite temperature with Monte Carlo algorithms
- randomised algorithms (modfied deterministic algorithms).

Generation of true random numbers in computer is possible (e.g. thermally-induced fluctuations of voltage measured at transistor etc). Advantage: random. Disadvantage: Statistical properties unknown and cannot be controlled.
Thus: pseudo random numbers = not random, but *as far as possible* equal statistical properties (distribution, correlations).

8.1 Fundamentals of Probability Theory

Ω: Set of outcomes of random experiment.

Ex: $\Omega = \{\text{head, tail}\}$ for coin toss.

Definition: A probability function P is a function $P : 2^\Omega \longrightarrow [0, 1]$ with

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

and for any sequence $A_1, A_2, A_3, \ldots$ of disjoint events ($A_i \cap A_j = \emptyset$ for $i \neq j$) holds:

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

Properties (without proof), for $A, B \subset \Omega$, $A^c := \Omega \setminus A$ holds:

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

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

Forr independent random experiments holds:

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

Definition: The conditional probability of outcome A conditioned to C ist

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

[Activator]

What is $P(\text{dice } 6|\text{dice} > 3)$?

= P(dice6 and dice>3) / P(dice>3)

P(dice6)=1/6
P(dice>3)=1/2
P(AnB)=P(A)=1/6 because A is subset of B.
P(A|B)= (1/6)/(1/2) = 1/3

Bayes' rule
Due to (12) holds $P(A|C)P(C) = P(A \cap C) = P(C \cap A) = P(C|A)P(A) \Rightarrow$

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

VIDEO: `video06b_ZVen_r`

8.2 Random Variables

Random variable (RV) X (sloppy): random experiment with $\Omega = \mathbb{R}$

Definition: Distribution function (DF) of a RV X is a function $F_X : \mathbb{R} \longrightarrow [0, 1]$ defined by

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

Ex: coin:

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

x position of gas particle in a container $[0, L_x]$

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

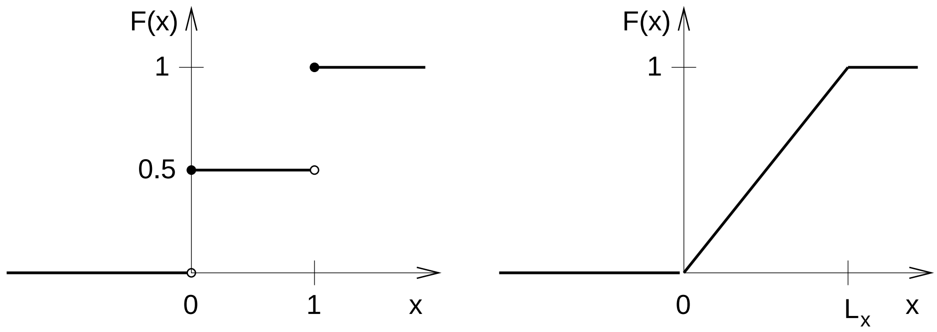

Figure 10: Distribution functions for coin and gas particle.

After random experiment corresponding to RV X with outcome x, next obtain $y = g(x)$:
Transformation of RV to $Y = g(X)$, in general:

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

Distribution function sometimes inconvenient → probability/density function:

VIDEO: `video06c_ZV_diskret_r`

8.2.1 Discrete Random Variables

Definition: The probability mass function (pmf) $p_X : \mathbb{R} \to [0, 1]$ is given by

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

Distribution for n coin tosses (0/1):
Definition: The *Binomial Distribution* with parameters $n \in \mathbb{N}$ and p $(0 < p \leq 1)$ describes a RV X with pmf

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

Notation $X \sim B(n, p)$.
Characterization of RVs:

$\{\tilde{x}_i\}$ set of possible values for which $p_X(\tilde{x}) > 0$ holds.
Definition:

- Expectation value

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

- Variance

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

- Standard deviation

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

Properties:

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

$\mathrm{E}[X^n]$: n'th moment
For the Binomial Distribution :

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

VIDEO: `video06d_ZV_kont_r`

8.2.2 Continuous Random Variables

Definition: For a RV X with continuous DF F_X, the probability density function (pdf) $p_X : \mathbb{R} \to \mathbb{R}$ is given by

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

Therefore:

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

Definition:

- Expectation value

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

- Variance

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

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

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

Definition: Uniform distribution with parameters $a < b$, describes a RV X with pdf

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

Notation: $X \sim U(a,b)$.
By using $g(X_{01}) = (b-a) * X_{01} + a$ one obtains $g(X_{01}) \sim U(a,b)$ if $X_{01} \sim U(0,1)$.

[Activator]

Calculate $\mathrm{E}[X]$ and $\mathrm{Var}[X]$.

Most important:
Definition: The Gaussian distribution or Normal distribution with parameters μ and $\sigma > 0$, describes the RV X having pdf

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

Notation: $X \sim N(\mu, \sigma^2)$.
Properties: $\mathrm{E}[X] = \mu$, $\mathrm{Var}[X] = \sigma^2$.
By using $g(X) = \sigma X_0 + \mu$ one obtains $g(X_0) \sim N(\mu, \sigma^2)$ falls $X_0 \sim N(0,1)$.

Central limit theorem:
For independent RVs $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ with $E[X^{(i)}] = \mu$ and $\mathrm{Var}[X^{(i)}] = \sigma^2$, the RV

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

is for large n distributed as $X \sim N(n\mu, n\sigma^2)$.

Densities of other important distributions:
Definition:

- Exponential distribution (for $x \geq 0$)

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

[Activator]

Calculate the distribution function for the Exponential distribution

---

Definition:

- Power-law distribution or Pareto distribution

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

For $\gamma > 1$ the expectation value is finite $\mathrm{E}[X] = \gamma\kappa/(\gamma-1)$, for $\gamma > 2$ $\mathrm{Var}[X] = \frac{\kappa^2\gamma}{(\gamma-1)^2(\gamma-2)}$

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

- Fisher-Tippett distribution

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

(also called Gumbel distribution for $\lambda = 1$) $\mathrm{E}[X] = \nu/\lambda$, $\nu \equiv 0.57721\ldots$, Maximum at $x = 0$, shift by $x \to (x - \mu)$

[Activator]

Can you read off the distribution function ?
