Masterseminar Summersemester 2023
Large Deviations and Concentration of Measure

Target audience: M.Sc. Mathematik/Technomathematik/Wirtschaftsmathematik/Data Science
Requirements: Stochastische Modellbildung (Stochastic Modelling), Wahrscheinlichkeitstheorie (Probability Theory), knowledge about functional analysis is helpful but not required

Outline: The students prepare a talk (≈ 60 min. + time for questions) and a handout (several pages)
Language: English or German, but all literature will be in English

Instructor: Dr. Markus Ebke (with supervision from Prof. Torben Krüger)
Time and place (preliminary): Weekly on Monday 14–16 in Übung 3 (Cauerstraße 11)
First meeting: February 27, 2023 at 14:00 (in Übung 3, attending via Zoom is also possible)

If you are interested: Please contact me via email (markus.ebke@fau.de) as soon as possible and also let me know if the suggested time is suitable for you (we may be able to change the time if free rooms are available).

Content

The **concentration of measure phenomenon** refers to the idea that most of the probability mass of a high-dimensional random variable is concentrated near its mean. This behaviour allows us to make robust statements about complex high-dimensional systems based just on the mean of random variables.
Large deviation theory is a mathematical framework for analysing the behaviour of random variables in the tail, or the far-off regions of the distribution where the probability is small. Here we study the decay rate of the probability of rare events as the sample size increases.
Together, they provide a complete picture for high-dimensional random variables, and are useful for analysing a wide range of physical and mathematical systems.
For example, take a fair coin and flip it N times. Let $S_N \sim \mathrm{Bin}(N, \frac{1}{2})$ be the number of heads. Then S_N may take any value in $\{0, 1, \ldots, N\}$, however values around $\mathbb{E}(S_N) = \frac{N}{2}$ are most likely. More precisely, we consider the probability $\mathbb{P}(|S_N - \mathbb{E}(S_N)|/N \geq \delta)$ with $\delta > 0$. Then Fig. 1 shows the concentration of measure phenomenon (for large N the probability decays quickly as δ increases) and Fig. 2 shows the decay rate of rare events, i.e. $\mathbb{P}(|S_N - \mathbb{E}(S_N)|/N \geq \delta) \sim \exp(-rN)$ as $N \to \infty$ for some rate $r > 0$ depending on δ.

In the seminar we will study selected chapters of the two books referenced below.

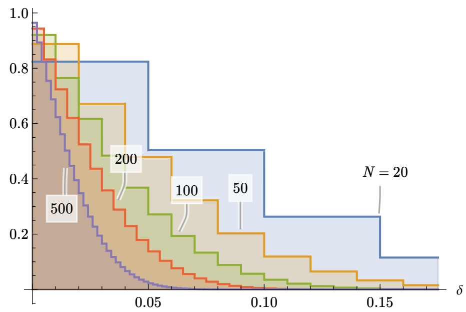

Fig. 1: $\delta \mapsto \mathbb{P}(|S_N - \mathbb{E}(S_N)|/N \geq \delta)$ for different N.

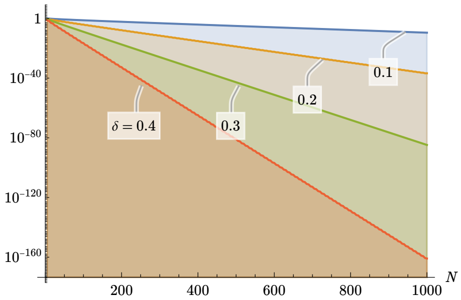

Fig. 2: $N \mapsto \mathbb{P}(|S_N - \mathbb{E}(S_N)|/N \geq \delta)$ for different δ, plot in log-scale.

Literature

- Amir Dembo, Ofer Zeitouni, **Large Deviations Techniques and Applications**, Springer (1998)
- Roman Vershynin, **High-Dimensional Probability: An Introduction with Applications in Data Science**, Cambridge University Press (2018)

(Electronically available via the university library)