Exercise Sheet 2

Exercise 1

Prove the union bound using Markov's inequality.

Exercise 2

Suppose a factory produces light bulbs, with each light bulb independently having a probability $p = 0.2$ of being defective. The factory tests $n = 20$ light bulbs each day.

Bernoulli- & Binomialverteilung sind miteinander verwandt
x

a) Use Markov's and Chebyshev's inequalities to bound the probability that at least 16 out of the 20 tested light bulbs are defective. Compare both bounds to the actual probability of observing 16 defective light bulbs.

b) Plot the bounds given by Markov's and Chebyshev's inequalities, as well as the actual probability that at least 16 out of n tested light bulbs are defective, as a function of the number n of tested light bulbs. Set bounds that exceed 1 to be equal to 1.

Exercise 3

In this exercise, you will apply Chebyshev's inequality to a binary classification problem. Notebook `exercise_3.ipynb` provides code to generate a ground-truth population of N training examples from a fixed distribution. Apply two different classifiers to the following experiment:

1. Sample a training set of size n from the ground-truth population.
2. Fit a classifier f on the training set.
3. Sample a test set of size n from the ground-truth population.
4. Compute the empirical risk $R_n(f)$ using the test set.
5. Compute the true risk $R(f)$ using the ground-truth population.
6. Compute Chebyshev's bound $\mathbb{V}[x]/t^2$.
7. Repeat steps 3-6 for multiple runs. Observe how often $R_n(f)$ deviates from $R(f)$ by more than $t = 0.05$. Compare the fraction of *large deviations* with the bound given by Chebyshev's inequality.
8. Repeat steps 2-7 for various training/test sizes n. Plot the estimated probability $\mathbb{P}(|R_n(f) - R(f)| \geq t)$ and the estimated Chebyshev bound as a function of n.