Winter term 2020/21
L. Rendsburg / U. von Luxburg

Assignment 9

Mathematics for Machine Learning

Submission due on **25.01.21, 8:00**

Justify all your claims.

Exercise 1 (**Limit theorems**, 2+3 points)**.**

a) A laundry bag contains one black and one white sock. Now Tom keeps throwing socks into the laundry bag. Every sock he throws is either black with probability $p \in [0, 1)$ or white with probability $1 - p$, independently of the previous socks. Let X_n be the fraction of black socks to total amount of socks and Y_n the fraction of black to white socks after $n \in \mathbb{N}$ throws. Prove that

   i) $X_n \to p$ almost surely as $n \to \infty$,

   ii) $Y_n \to \frac{p}{1-p}$ almost surely as $n \to \infty$.

b) Consider an i.i.d. sequence of real-valued random variables $(X_n)_{n\in\mathbb{N}}$ with $\mu = \mathbb{E}[X_1] \in \mathbb{R}$ and $\sigma^2 = \operatorname{Var}[X_1] < \infty$. Define $S_n := \sum_{k=1}^n X_k$ and let $a, b \in \mathbb{R}$ with $a < b$. Use the central limit theorem to prove

$$P(a \le S_n \le b) = \Phi\left(\frac{b - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a - n\mu}{\sqrt{n}\sigma}\right) + o(1)\,,$$

where Φ denotes the cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$ and $o(1)$ satisfies $o(1) \to 0$ as $n \to \infty$.
Hint: You may use the characterization

$$X_n \xrightarrow[n\to\infty]{} X \text{ in distribution} \quad \Leftrightarrow \quad F_n \xrightarrow[n\to\infty]{} F \text{ uniformly on } C_F\,,$$

where F_n, F denote the cumulative distribution functions of X_n, X, and $C_F = \{x \in \mathbb{R} \mid F \text{ is continuous at } x\}$.

Exercise 2 (**Borel-Cantelli**, 2+2+2 points)**.**

a) A monkey is furiously typing on a typewriter. The typewriter has 26 letters `A`-`Z`, and the monkey is hitting a letter X_n uniformly at random at every step $n \in \mathbb{N}$, independently of the previous letters. That is, $P(X_n = i) = \frac{1}{26}$ for $i \in \{\mathtt{A}, \ldots, \mathtt{Z}\}$. Prove that the sequence `M-O-N-K-E-Y` appears infinitely often with probability 1.
Hint: Use the Borel-Cantelli lemma to prove that $P(B_n \text{ i.o.}) = 1$, where

$$B_n = \{(X_{6n+1}, X_{6n+2}, X_{6n+3}, X_{6n+4}, X_{6n+5}, X_{6n+6}) = (\mathtt{M}, \mathtt{O}, \mathtt{N}, \mathtt{K}, \mathtt{E}, \mathtt{Y})\}, \quad n \in \mathbb{N}_0\,.$$

b) One direction of the Borel-Cantelli lemma requires the independence of the sequence $(A_n)_{n\in\mathbb{N}}$. To show that this assumption is necessary, give an example for a sequence $(A_n)_{n\in\mathbb{N}}$ which is not independent and satisfies

$$\sum_{n=1}^{\infty} P(A_n) = \infty \quad \text{and} \quad P(A_n \text{ i.o.}) \in (0, 1)\,.$$

c) Consider a sequence of random variables $X, X_1, X_2, \ldots$ that satisfies $X_n \to X$ in probability as $n \to \infty$. Prove that there exists a subsequene $(X_{n_k})_{k\in\mathbb{N}}$ for which $X_{n_k} \to X$ almost surely as $k \to \infty$.
Hint: Use the Borel-Cantelli lemma and the characterization

$$X_n \xrightarrow[n\to\infty]{} X \text{ almost surely} \quad \Leftrightarrow \quad \forall \varepsilon > 0: \; P(\{|X_n - X| > \varepsilon\} \text{ i.o.}) = 0 \,.$$

Exercise 3 (**Convergence of random variables**, 2+2+1 points)**.**

a) Consider the probability space $([0,1], \mathcal{B}, \lambda)$, where $\mathcal{B}$ denotes the Borel-σ algebra on $[0,1]$ and λ the Lebesgue measure. For every $n \in \mathbb{N}$ there exist unique $h, k \in \mathbb{N}_0$ with $0 \le k < 2^h$ such that $n = 2^h + k$. Then define the random variable X_n using these h, k as

$$X_n(\omega) = \mathbb{1}_{\left[\frac{k}{2^h}, \frac{k+1}{2^h}\right]}(\omega) = \begin{cases} 1, & \omega \in \left[\frac{k}{2^h}, \frac{k+1}{2^h}\right] \\ 0, & \text{otherwise} \end{cases} \qquad \forall \omega \in [0,1] \,.$$

A similar example was also given in the lecture. Prove that $X_n \to 0$ as $n \to \infty$ in probability and in L^1, but not almost surely.

b) As stated in Exercise 2c), convergence in probability implies the existence of a subsequence that converges almost surely. Find such a subsequence for the sequence given in a).

c) Consider real-valued random variables $X, X_1, X_2, \ldots$ with $X_n \to X$ almost surely as $n \to \infty$. Prove that for every continuous function $f\colon \mathbb{R} \to \mathbb{R}$ it also holds $f(X_n) \to f(X)$ almost surely as $n \to \infty$.

Exercise 4 (**Joint, marginal, conditional distribution**, 2+2 points)**.**

a) Consider two random variables X and Y, where $X \in \mathcal{X} = \{\text{Sun, Rain, Snow}\}$ describes the weather and $Y \in \mathcal{Y} = \{\text{Few, Many}\}$ describes how many pedestrians are taking a stroll. The joint distribution of X and Y is given by the following table, which contains the probabilities $P(X = x, Y = y)$:

X \ Y	Few	Many
Sun	0.1	0.4
Rain	0.27	0.03
Snow	0.06	0.14

Compute the probabilities of the following events:

i) Many pedestrians on a sunny day or few pedestrians on a rainy day.
ii) The sun shines.
iii) The sun shines given that many pedestrians take a stroll.

b) Now consider two random variables X, Y on $\mathcal{X} = \{1, \ldots, n\}, \mathcal{Y} = \{1, \ldots, m\}$ with $n, m \in \mathbb{N}$. Their joint distribution is described by a matrix $M \in \mathbb{R}^{n\times m}$ with $M_{x,y} = P(X = x, Y = y)$ for $x \in \mathcal{X}, y \in \mathcal{Y}$.

Let $J \in \mathbb{R}^{m\times n}$ denote the matrix of containing only 1s and $r_n \in \mathbb{R}^n$, $r_m \in \mathbb{R}^m$ denote the vectors containing increasing integers $1, \ldots, n$, respectively $1, \ldots, m$. Prove the following two statements:

i) X and Y are independent $\Leftrightarrow M = MJM$,
ii) X and Y are uncorrelated $\Leftrightarrow r_n^T M r_m = r_n^T MJM r_m$.