L. Rendsburg / U. von Luxburg

Assignment 8

Mathematics for Machine Learning

Submission due on **18.01.21, 8:00**

Justify all your claims.

Exercise 1 (**Independence**, 2+2+2 points)**.**

a) Let X describe the result of throwing a fair die. That is, X describes a random variable on a probability space $(\Omega, \mathcal{A}, P)$ such that $P(X = k) = \frac{1}{6}$ for $k \in \{1, \dots, 6\}$. For the following two pairs of events A_1, B_1 and A_2, B_2, prove whether they are independent:

$$A_1 = \{X \text{ is even}\} \quad \text{and} \quad B_1 = \{X \geq 5\}$$
$$A_2 = \{X \text{ is odd}\} \quad \text{and} \quad B_2 = \{X \geq 4\}$$

b) Consider two random variables $X, Y \colon (\Omega, \mathcal{A}, P) \to (\{1, 2, 3\}, \mathcal{P}(\{1, 2, 3\}))$, whose distribution is described by a probability matrix $M \in \mathbb{R}^{3\times 3}$ such that $P(X = x, Y = y) = M_{x,y}$ for $x, y \in \{1, 2, 3\}$. In the following two cases, prove whether X and Y are independent.

$$M_1 = \begin{pmatrix} 0.18 & 0.3 & 0.12 \\ 0.03 & 0.05 & 0.02 \\ 0.09 & 0.15 & 0.06 \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} 0.16 & 0.31 & 0.13 \\ 0.05 & 0.04 & 0.01 \\ 0.09 & 0.15 & 0.06 \end{pmatrix}$$

Hint: You may use that X and Y are independent if and only if $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all $x, y \in \{1, 2, 3\}$.

c) Consider the probability space $(\Omega, \mathcal{A}, P) = ([0, 2], \mathcal{B}, \frac{1}{2}\lambda_{[0,2]})$, where $\mathcal{B}$ is the Borel-σ-algebra restricted to $[0, 2]$ and $\frac{1}{2}\lambda_{[0,2]}$ is the normalized Lebesgue measure on $[0, 2]$, that is, $\frac{1}{2}\lambda_{[0,2]}([a, b]) = \frac{1}{2}(b - a)$. Define two random variables $X, Y \colon (\Omega, \mathcal{A}, P) \to ([0, 2], \mathcal{B})$ by $X(\omega) = \omega$ and $Y(\omega) = (\omega - 1)^2$ for $\omega \in \Omega$. Prove that X and Y are dependent.

Exercise 2 (**Surveillance**, 3 points)**.** A surveillance system is placed in the streets of London to detect criminals. It is equipped with a face recognition software and has access to a data base of criminal records. The software has a high true positive rate and low false positive rate when examining the face of a pedestrian: a criminal is correctly labeled as a criminal in 95% of the cases, while a righteous citizen is erroneously labeled as criminal in only 1% of the cases.
Aside from privacy concerns, discuss whether this surveillance system is as useful in detecting criminals as it seems. Invent some numbers to make your argument formal.

Exercise 3 (**Covariance and correlation**, 3+3+2 points)**.**

a) Consider two independent dice throws described by the random variables W_1 and W_2. That is, W_1 and W_2 are independent random variables with $P(W_i = k) = \frac{1}{6}$ for $k \in \{1, \dots, 6\}$ and $i \in \{1, 2\}$. Define the sum $X = W_1 + W_2$ and the product $Y = W_1 \cdot W_2$. Compute the correlation coefficient $\varrho_{X,Y}$.

b) Consider two binary random variables M, G on $\{0, 1\}$ that describe whether a student is motivated ($M = 1$) or not ($M = 0$) and whether her grades are good ($G = 1$) or not ($G = 0$). Assume additionally that they are independent, that is, $P((M, G) = (i, j)) = \frac{1}{4}$ for $i, j \in \{0, 1\}$. Compute the covariance $\mathrm{Cov}[M, G]$.
Now consider a scenario where a new course is offered, but students are only admitted if they are motivated or if their grades are high. What is the covariance between M and G of a student conditioned on the event that this student was admitted to the course?

c) Consider two random variables $X, Y \in L^2(\Omega, \mathcal{A}, P)$. Prove that the correlation coefficient satisfies $\varrho_{X,Y} \in [-1, 1]$.
Hint: Use the Cauchy-Schwarz inequality for random variables with the scalar product $\langle X, Y \rangle = \mathbb{E}[XY]$.

Exercise 4 (**Chebyshev's inequality**, 3 points)**.** Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables on $(\mathbb{R}, \mathcal{B})$ with $\mathbb{E}\left[X_1^2\right] < \infty$. Denote $\mu := \mathbb{E}[X_1]$ and consider the sum over the first n random variables denoted by $S_n := \sum_{k=1}^{n} X_k$. Prove that

$$P\left(\left|\frac{1}{n}S_n - \mu\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon > 0\,.$$

Hint: Use Chebyshev's inequality.

Bonus exercise (**Bernoulli trials**, 2+2+1 points)**.** A Bernoulli trial is a random experiment with only two outcomes, "success" and "failure". Assume that the probability of success is $p \in (0, 1)$. This Bernoulli trial is now repeated arbitrarily often, where the outcome of each trial is independent of the other trials.
For $n \in \mathbb{N}$, let K_n denote the number of successes within the first n trials, T_1 the number of failures until the first success, and T_2 the total number of failures until the second success.

a) Prove that $K_n \stackrel{d}{=} B(n, p)$ for any $n \in \mathbb{N}$. This means that K_n is distributed as the binomial distribution $B(n, p)$, which is defined by $B(n,p)(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, \ldots, n\}$.

b) Compute the distribution of T_1, T_2, and $X := T_2 - T_1$.

c) Are X and T_1 independent? (Prove your answer.)