FU Berlin
Stochastik I (Mono-Bachelor)
WiSe 2016–2017

Lectures: T. J. Sullivan [sullivan@zib.de]
TA 1: N. Gießing [noah.giessing@hushmail.com]
TA 2: I. Klebanov [klebanov@zib.de]
TA 3: H. C. Lie [hlie@zedat.fu-berlin.de]

Exercise Sheet 5

These exercises concern the lectures of Week 5, and possibly previous lectures and exercises. Please submit your solutions to these exercises at the end of the second lecture of Week 6, i.e. 10:00 on 24 November 2016. Solutions in German and in English are equally valid. The numbers in the margin indicate approximately how many points are available for each part.

Exercise 5.1 (Difficulties formalising causation). A friend suggests to you that probability could be used to assess whether one event causes another, in the sense that the effect should be more likely given the proposed cause than without it. After some dicussion, you and your friend formalise this idea by saying that, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with positive probability events $A, B \in \mathcal{F}$, B **causes** A if $\mathbb{P}(A|B) > \mathbb{P}(A)$. Show that this is not a good definition, because B causes A if and only if A causes B. **[1]**

Exercise 5.2 (Ordered sampling with and without replacement). Let $\Omega := \{1, \dots, n\}$, which you should think of as a collection of n distinguishable objects in a bag. Give appropriate probability models on Ω^k to model the following scenarios. As well as giving the general construction, explicitly write out the table of probability values for the points of Ω^2 in the case $n = 3$, $k = 2$.

(a) You record the ordered sequence resulting from fairly drawing one object from the bag, replacing it, reshuffling the objects, drawing the second, etc. **[1]**

(b) You record the ordered sequence resulting from fairly drawing one object from the bag, *not* replacing it, reshuffling the objects, drawing the second from the remaining objects, etc. **[1]**

Exercise 5.3 (Assessing a test for hepatitis). Consider the following probability model for the random testing of a population Ω of people for hepatitis. Assume that every person in the population is either free of hepatitis (event F), has hepatits A (event A), or has hepatitis B (event B). When applied to a person, the test only indicates whether or not the person has hepatitis or is hepatitis-free; the test does not distinguish between the A and B variants. Let D denote the event that the test result is a positive diagnosis, indicating that the person being tested has one of the two variants of hepatitis — however, the test is not perfectly accurate, and does not even have the same accuracy for the two variants of hepatitis!

Suppose, based e.g. on clinical trials and background information about hepatitis in the population, that the following probabilities hold for a member of the population selected uniformly (i.e. fairly) at random, and tested as described above:

$$\mathbb{P}(A) = 0.07, \qquad \mathbb{P}(B) = 0.03, \qquad \mathbb{P}(F) = 0.9,$$
$$\mathbb{P}(D|A) = 0.95, \qquad \mathbb{P}(D|B) = 1.0, \qquad \mathbb{P}(D|F) = 0.1.$$

(a) What proportion of the population will be diagnosed as having hepatitis? **[1]**

(b) If the test result is D, then what are the posterior probabilities for that person's hepatitis state? **[1]**

(c) How accurate is the test, as judged by its predictive errors $\mathbb{P}(F|D)$ (the probability of a **false positive**) and $\mathbb{P}(F^{\complement}|D^{\complement})$ (a **false negative**)? **[2]**

Exercise 5.4 (Practice with indicators). Let Ω be a set and let $\mathbb{I}_E \colon \Omega \to \{0, 1\}$ denote the indicator function of a set $E \subseteq \Omega$, i.e. $\mathbb{I}_E(\omega) := 1$ if $\omega \in E$, and $\mathbb{I}_E(\omega) := 0$ if $\omega \in \Omega \setminus E$.

(a) Let $E_i \subseteq \Omega$ for each i in some index set I. Show that (pointwise, i.e. for each $\omega \in \Omega$)

$$\begin{aligned}
\mathbb{I}_{E_i^{\complement}}(\omega) &= 1 - \mathbb{I}_{E_i}(\omega), \\
\mathbb{I}_{\bigcup_{i \in I} E_i}(\omega) &= \max_{i \in I} \mathbb{I}_{E_i}(\omega), \\
\mathbb{I}_{\biguplus_{i \in I} E_i}(\omega) &= \sum_{i \in I} \mathbb{I}_{E_i}(\omega), \\
\mathbb{I}_{\bigcap_{i \in I} E_i}(\omega) &= \min_{i \in I} \mathbb{I}_{E_i}(\omega) = \prod_{i \in I} \mathbb{I}_{E_i}(\omega), \\
\mathbb{I}_{E_i \triangle E_j}(\omega) &= |\mathbb{I}_{E_i}(\omega) - \mathbb{I}_{E_j}(\omega)|,
\end{aligned}$$

where $E_i \triangle E_j \coloneqq (E_i \setminus E_j) \cup (E_j \setminus E_i)$ denotes the symmetric difference of E_i and E_j. (Half marks for proving all of these claims in just the binary case $i = \{1, 2\}$.) **[3]**

(b) Hence give another proof of De Morgan's law $\left(\bigcup_{i \in I} E_i\right)^{\complement} = \bigcap_{i \in I} E_i^{\complement}$ by showing that the sets on either side have the same indicator. **[1]**

(c) Hence show that, when $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $E, F, G \in \mathcal{F}$, **[1]**

$$\mathbb{P}(E \triangle F) \leq \mathbb{P}(E \triangle G) + \mathbb{P}(G \triangle F).$$