TEACHERS Christian Schaffner and Philip Schulz TA Bas Cornelissen

Homework problem set 4

Your homework must be handed in **electronically via Moodle before Wednesday September 28th, 22:00h**. This deadline is strict and late submissions are graded with a 0. At the end of the course, the lowest of your 7 weekly homework grades will be dropped. You are strongly encouraged to work together on the exercises, including the homework. However, after this discussion phase, you have to write down and submit your own individual solution. Numbers alone are never sufficient, always motivate your answers.

Problem 1: Spam filters (1pt)

Suppose that 20% of the incoming emails are junk mail (in fact, it's more). A company is developing a spam filter that already manages to correctly label all junk mail as spam. But in doing so, it also wrongly labels some innocent emails as spam. Let's call the probability of wrongly labelling an innocent email as spam the *inaccuracy* α. The spam filter's inaccuracy can be decreased in such a way that all actual spam will still be intercepted. To what value should α at least be decreased if the company wants to be at least 99.9% sure that an email labelled as spam is actually junk mail?

Problem 2: A joint probability vase (4pt)

Consider a vase with 3 red balls, 2 white balls and 5 blue balls. You draw 5 balls uniformly at random without replacement and ignore their order. Let X be the number of red balls and Y be the number of white balls.

(a) **1pt** Let $C(x, y)$ be number of outcomes in which $X = x$ and $Y = y$. Fill in the following table with all values C can take.

$C(X,Y)$	$Y=0$	$Y=1$	$Y=2$
$X=0$			
$X=1$			
$X=2$			
$X=3$			

Justify that $P_{X,Y}(x,y) = \binom{10}{5}^{-1} C(x,y)$.

(b) **1pt** Provide a table of the cdf $F_{XY}(x,y) = P((X,Y) \le (x,y))$. (You might want to keep things simple and use C in some way.)

(c) **0.5pt** Let A be the event that you have drawn 1, 2 or 3 red balls and at most 1 white ball. Calculate $P_{XY}(A)$ from the table.

(d) **0.5pt** Calculate $P_{XY}(A)$ again, and indicate how you use F_{XY}

(e) **0.5pt** Find the marginal distributions P_X and P_Y.

(f) **0.5pt** Are X and Y independent?

Problem 3: 3pt

Let P_{XY} be the joint distribution of X and Y whose marginal distributions are $X \sim \text{Binomial}(2, 0.5)$ and $Y \sim \text{Bernoulli}(0.8)$. The joint distribution satisfies $P_{XY}(2,0) = 0$ and $P_{XY}(1,1) = c$ for $c \in \mathbb{R}$.

(a) **1pt** Find the joint distribution of X and Y and calculate $\text{Cov}(X,Y)$.

(b) **1pt** Calculate $\text{Cor}(X,Y)$.

(c) **1pt** Can we choose c in such a way that X and Y become independent? Give c if it exists or otherwise show it cannot exist.

Problem 4: Deterministic random variables (2pt)

A (discrete) random variable Y is called *constant*, *degenerate* or *deterministic* if there exists an outcome a such that $P_X(a) = 1$. Let Y be such a random variable.

(a) **1pt** Show that X is constant if and only if $\text{Var}[X] = 0$.

(b) **1pt** Again, let Y be constant and X not and further suppose that $E[X]$ is finite. Show that $\text{Cov}(X,Y) = 0$.