Exercise Sheet 2

The exercise sheet is subdivided into three parts:

- Part A contains warm-up questions you should do in your own time.
- You should solve all questions of Part B and put your solutions in the locker in the undergraduate common room in the Department of Statistical Science (Room number 117) assigned to your tutor group by the deadline stated on the exercise sheet. Don't forget to mark your solutions with your name and student number.

  Reminder: If you miss the deadline for reasons outside your control, for example illness or bereavement, you ***must*** submit a claim for extenuating circumstances, normally within a week of the deadline. Your home department will advise you of the appropriate procedures. For Statistical Science students, the relevant information is on the DOSSSH Moodle page.

  Your tutor will then mark one question selected at random after you handed in your solutions (mark counts towards ICA) and will return your marked solutions to you at the next tutorial at which the solutions to the part B questions will be discussed. Marks available for each part question are indicated in square brackets, e.g. [3]. Very succinct solutions to these questions will usually be made available on Moodle at 9:15am on Thursday mornings, so don't hand in late!
- Solutions to Part C questions will be made available in time for exam revision. These questions will not be discussed at the tutorials.

A: Warm-up questions

1. Let the random variable X take values $1, 2, 3$ with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ respectively.

   (a) What is the probability that $X \in A := \{1, 2\}$?

   (b) What is the conditional probability that $X = 1$ given that $X \in A$?

2. Add another random variable Y such that the joint distribution of the random variable X from above and the new random variable Y is given by the table

		X		
		1	2	3
	1	1/8	1/12	0
Y	2	1/8	1/6	1/12
	3	1/4	0	1/6

   (a) What is the probability of the event $\mathrm{P}(X = 2, Y = 1)$?

   (b) Confirm that you get back your result for question 1.(a) above for $\mathrm{P}(X \in A)$ from this table.

   (c) What is the conditional distribution of X given that $X \in A$? (Check that your answer agrees with your answer to question 1. (b) above!)

   (d) What is the probability that $Y = 1$?

   (e) What is the conditional distribution of X given that $Y = 1$?

3. Roll a fair die twice. Let X_1 be the number of times that face 1 shows, and let $X_2 = \lfloor \text{sum of faces}/4 \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x. Construct the joint probability table. Calculate the two marginal pmfs $p_1(x_1)$ and $p_2(x_2)$ and the conditional pmfs $p_{1|2}(x_1|x_2 = 1)$ and $p_{2|1}(x_2|x_1 = 1)$. Are X_1 and X_2 independent? Compute the means, μ_1 and μ_2, variances, σ_1^2 and σ_2^2, and covariance σ_{12}. Are X_1 and X_2 uncorrelated?

B: Answers to hand in by Thursday, 22 October, 9am

1. (a) Two factories produce widgets[1]. 20 percent of widgets produced by factory A are defective whereas only 5 percent of the widgets produced in factory B show a defect. Factory A produces twice as many widgets as factory B. Calculate the probability of obtaining a defective widget when randomly sampling one widget from the two factories' combined output. [4]

   (b) Assuming you purchase one widget and find out that it is defective, what is the probability that it was produced by factory B? [4]

   (c) Factories A and B both add new production lines that are faster but lead to defective widgets in 25% of cases. Factory A now produces 10% of its output on the new production line and factory B produces 20% of its output in this manner. The proportions of defectives on the old production lines in each factory are unchanged. However, the output rates are changed so that factory B produces twice as many widgets as factory A. Should you have delayed your purchase until after this change, to improve your chance of receiving an intact widget? [7]

2. Let $X \sim \mathrm{Bin}(50, 0.3)$ and let the conditional distribution of Y given X be given by $Y|X = x \sim \mathrm{Gam}(x + 1, 0.1)$.

   (a) Write down the conditional expectation for Y given $X = x$ in terms of x. [3]

   (b) Compute the marginal expectation of Y. [6]

   (c) Compute the expectation $\mathrm{E}[2X^3 + 2X^2 - \frac{1}{5}YX^2]$. [6]

3. A motor insurance company analyses historical data by considering the number of claims policyholders submitted in one calendar year. Each policyholder is classified into one of three categories: those who submitted $C = 0$ claims, those who submitted $C = 1$ claim and those who submitted $C \geq 2$ claims in that year. For each policyholder, the company also computes the amount of money paid out on claims made in the subsequent calendar year: the amounts are categorised as either no payout ($M = 0$), small payout ($M = 75$) or large payout ($M = 20000$) where M denotes the average amount paid out to each policyholder in the category.

   The motor insurance company estimates the following conditional probabilities of M given C from its analysis of the two years:

	$C = 0$	$C = 1$	$C \geq 2$
$M = 0$	0.9	0.85	0.7
$M = 75$	0.08	0.12	0.2
$M = 20000$	0.02	0.03	0.1

   (a) Explain why the entries in each column of this table have to sum to one but the entries in each row do not necessarily sum to one. [4]

   (b) Naively[2] assume that the insurance premium (i.e. the cost of taking out insurance) is equal to the average amount paid out to a policyholder. In this

[1] A 'widget' is a placeholder name for an unspecified or hypothetical manufactured product.

[2] Consider taking STAT3022 (Quantitative Modelling of Operational Risk and Insurance) next year if you want to find out more about how this is done in real life.

case, what is the annual premium for policyholders in each of the categories $C = 0$, $C = 1$ and $C \geq 2$? [6]

(c) Suppose now that the insurance company is not permitted to charge different premiums depending on policyholders' previous claims histories: Instead, every policyholder must be charged the same premium. What should this premium be? You may assume that 70% of policyholders have made no claims, 25% made one claim and 5% made two or more claims in the preceding year. [5]

C: Exam Practice Questions

1. Let Y indicate the health status of patients suffering from a specific disease, where $Y = 1$ corresponds to 'fairly well', $Y = 2$ to 'bad', and $Y = 3$ to 'very bad'. In a group of patients, 20% are doing 'fairly well', 50% are in category $Y = 2$, and 30% are doing 'very bad'. Let further X count the days until recovery after taking a newly developed drug. The pharmaceutical company claims that the conditional distribution of X given the different health status Y is given by

	x = days to recovery		
	4	5	6
$P(X = x \mid Y = 1)$	0.8	0.2	0
$P(X = x \mid Y = 2)$	0.2	0.6	0.2
$P(X = x \mid Y = 3)$	0	0.2	0.8

(a) Are X and Y independent? [5]

(b) On average, how long does it take overall for these patients to recover? [8]

(c) Derive the joint pmf of X and Y as well as the marginal distribution of X. [10]

Verify that your result in (b) is the same as the expectation of X computed using its marginal distribution. [2]

2. Suppose that the random variables X and Y can only take the values 0, 1, 2 and 0, 1 respectively. The joint probability mass function of X and Y, $p_{X,Y}$, and the marginal probability mass functions p_X and p_Y are given in the following incomplete table.

	X			p_Y
Y	0	1	2	
0				1/2
1		3/8	1/8	1/2
p_X	1/8	5/8	1/4	

(a) Compute the complete table. [4]

(b) Are X and Y independent? Give a reason for your answer. [3]

(c) Compute the marginal expectations $E[X]$ and $E[Y]$. [8]

(d) Compute the covariance of X and Y [10]