Problem 1

Let X and Y be i.i.d. $\mathsf{Geom}(p)$, and $N = X + Y$.

(a) Find the joint PMF of X, Y and N.

(b) Find the joint PMF of X and N.

(c) Find the conditional PMF of X given $N = n$.

Problem 2

Let X, Y and Z be random variables such that $X \sim \mathsf{N}(0, 1)$ and conditional on $X = x$, Y and Z are i.i.d. $\mathsf{N}(x, 1)$.

(a) Find the joint PDF of X, Y and Z.

(b) Find the joint PDF of Y and Z.

Problem 3

Let X and Y be continuous random variables with joint CDF $F(x, y)$. Show that the probability that (X, Y) falls in the rectangle $[a_1, a_2] \times [b_1, b_2]$ is

$$F(a_2, b_2) - F(a_1, b_2) + F(a_1, b_1) - F(a_2, b_1).$$

Problem 4

Let X and Y have joint PDF

$$f_{X,Y}(x, y) = x + y, \text{ for } 0 < x < 1 \text{ and } 0 < y < 1.$$

(a) Check that this is a valid joint PDF.

(b) Find the marginal PDFs of X and Y.

(c) Are X and Y independent?

(d) Find the conditional PDF of Y given $X = x$.

Problem 5

Let X and Y have joint PDF

$$f_{X,Y}(x, y) = cxy, \text{ for } 0 < x < y < 1.$$

(a) Find c to make this a valid joint PDF.

(b) Find the marginal PDFs of X and Y.

(c) Are X and Y independent?

(d) Find the conditional PDF of Y given $X = x$.

Problem 6

Let X and Y be i.i.d. $\mathsf{Unif}(0, 1)$.

(a) Use simulations in R (the statistical programming language) to numerically estimate the covariance of $X + Y$ and $X - Y$.

(b) Compute the covariance of $X + Y$ and $X - Y$.

(c) Are $X + Y$ and $X - Y$ independent?

Problem 7

Let X, Y and Z be i.i.d. $\mathsf{N}(0, 1)$. Find the joint MGF of $(X + 2Y, 3X + 4Z, 5Y + 6Z)$.

Problem 8

The social mobility data from Table 1 gives a joint probability distribution on

$$(Y_1, Y_2) = (\text{father's occupation, son's occupation})$$

		son's occupation				
		farm	operatives	craftsmen	sales	professional
	farm	0.018	0.035	0.031	0.008	0.018
	operatives	0.002	0.112	0.064	0.032	0.069
father's occupation	craftsmen	0.001	0.066	0.094	0.032	0.084
	sales	0.001	0.018	0.019	0.010	0.051
	professional	0.001	0.029	0.032	0.043	0.130

Table 1: Joint distribution of occupational categories of fathers and sons

These data can be inputed in R with the following commands (see below).

```r
y = matrix(c(0.018,0.035,0.031,0.008,0.018,
             0.002,0.112,0.064,0.032,0.069,
             0.001,0.066,0.094,0.032,0.084,
             0.001,0.018,0.019,0.010,0.051,
             0.001,0.029,0.032,0.043,0.130),nrow=5,byrow=TRUE)
colnames(y) = c("farm","operatives","craftsmen","sales","professional")
rownames(y) = colnames(y)
#make sure this is a joint distribution
sum(y)
#the returned value is indeed 1
```

Using this joint distribution, calculate the following distributions:

(a) the marginal probability distribution of a father's occupation

(b) the marginal probability distribution of a son's occupation

(c) [the conditional distribution of a son's occupation, given that the father is a farmer

(d) the conditional distribution of a father's occupation, given that the son is a farmer

Problem 9

You will analyze data from a study of the effects of aspirin on myocardial infarction – see Table 2. This is a contingency table that cross-classifies two binary variables: (i) Aspirin Use with levels Placebo and Aspirin, and (ii) Myocardial Infarction with levels Yes and No.

	Myocardial Infarction	
	Yes	No
Placebo	28	656
Aspirin	18	658

Table 2: Problem 9 – Study on Aspirin Use and Myocardial Infarction.

Please answer the following questions:

(a) Calculate the row and columns totals of this table. What is the grand total?

(b) Calculate the expected cell values under the hypothesis of interaction of Aspirin Use and Myocardial Infarction.

(c) Calculate the expected cell values under the hypothesis of independence of Aspirin Use and Myocardial Infarction.

(d) Perform an asymptotic test of independence vs. interaction of Aspirin Use and Myocardial Infarction based on Pearson's chi-square statistic:

$$X^2 = \sum_{\text{all cells}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}.$$

(e) Perform an asymptotic test of independence vs. interaction of Aspirin Use and Myocardial Infarction based on the likelihood ratio statistic G^2:

$$G^2 = 2 \sum_{\text{all cells}} (\text{Observed}) \log\left(\frac{\text{Observed}}{\text{Expected}}\right).$$

(f) Draw conclusions related to the effect of aspirin on the occurrence of myocardial infarction. Summarize your findings in a concise statement.