Neuroinformatics Exercise
Fall 2017

**Exercise E2**

Name: _____
Pascal Nieters

**Instructions:** (10 points) Solve the following problems. Write clearly and use same symbols as used in the lecture. Add comments, explanations or questions to your solution if necessary.
Solutions to exercises **1** and **2** are to be submitted as physical copies in groups of three to four. Please label each page you hand in **clearly** and **carefully** with the name of each group member and their student ID.
**Please staple all of your sheets.**
Deadline for this exercise sheet is: **24.11.2017**

(5 pts) **1.** In lecture 5, you were introduced to a graphical model relating students' letters of recommendation to their intelligence, performance in an exam, difficulty of the exam, and SAT scores. Use the numbers provided in the table in the slides (*i.e. L5, p.41*) to answer the following questions:

| |
|---|
| 5 pts |

(a) What's the probability of being intelligent given the grade is 3? (*i.e.* $p(i = 1|g = 3)$)

(b) What's the probability of being intelligent given the grade is 3, and the exam was difficult? (*i.e.* $p(i = 1|g = 3, d = 1)$)

(c) How has the additional information in (b) that the exam was difficult influenced the probability of the student being intelligent? Give a **brief** answer that relates to the relevant concept introduced in the lecture.

(5 pts) **2.** Tay-Sachs disease is a genetic disorder resulting in the destruction of nerve cells in the brain and spinal cord. The most common type, infantile Tay-Sachs disease, can occur in children with inherited genetic mutation in the HEXA genes on chromosome 15. Because the disease is an autosomal recessive genetic disorder, the probability of a child whos parents are both carriers to be affected is 25%.
Assume two carriers conceive and give birth to three children, the random variable $X$ represents the number of children that are affected.

| |
|---|
| 5 pts |

(a) List the different combinations of ill children that lead to a certain value of $X$. Give the probabilities of $X$ being zero, one, two or three.

(b) Given the problem statement above, what is the probability distribution over $X$? What assumptions are we making in order to be able to use this probability distribution?

(c) First, calculate the expected value and variance of $X$ manually, then check the formulas given for the probability distribution in (b) and confirm your calculations.

**3.** To calculate the variance of a random variable, you need to calculate $\mathbb{E}[X] - \mathbb{E}[X]^2$. $\mathbb{E}[X] \in \mathbb{R}$ is just a number, which is why $\mathbb{E}[X]^2$ is easy to calculate. $\mathbb{E}[X^2]$ is the expected value of a new random variable $Y = X^2$.

(a) Let $X \in \mathbb{R}^+$ be a positive random variable with cut-off standard normal distribution, i.e.

$$p_X(x) = \begin{cases} \frac{2}{\sqrt{2\pi}} exp(\frac{-x^2}{2}) & \text{if} x > 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the distribution of $Y = X^2$ (i.e. what is $p_Y(y)$)

(b) **ADVANCED** Suppose that $X \in \mathbb{R}$ is a random variable with standard normal distribution. What is the distribution of $X^2$?
*Hint*: $f(x) = x^2$ is no longer monotonic on the whole real axis. Instead, try to calculate the CDF $F_Y$ and derive the PDF $f_Y$ using the chain rule. How can you use your result to calculate the variance of the standard normal distribution?

**4.** A machine puts chocolate chips into cookies randomly. Lets call the average number of chocolate chips $\lambda$.

(a) Which probability distribution describes this process?

(b) Derive a general expression for the probability that a chocolate chip cookie contains two or less chocolate chips.

(c) Calculate the probability of three or less chocolate chips if $\lambda = 5$.