# Probability Theory and Statistics (BSc D-ITET)

**1. (10 Points)** A group of friends wants to go by train to the Jochpass region to do the "4 Lake Hike." It is known that the tour can be finished with probability 85% if the hiker is trained, with probability 60% if he is not. There are no direct connection from Zürich HB to Engelberg. Therefore, one has to change trains in Luzern. There is a train itinerary scheduled to depart at 09:04 and to arrive in Engelberg at 10:53, but, in order to make the connection in Luzern, the first train has to arrive on time in Luzern. The probability that the first train departs on time is 70% and the probability that it arrives on time in Luzern is 75%. Moreover, it is known that the probability that the first train departs on time *and* arrives on time is 60%.

**a)** On the day of the excursion, the train to Luzern departs with delay. Which is the probability that it arrives on time?

**b)** 20% of the members of the group are not trained for the hike. Suppose we select at random one member of the group. Calculate:

i) the probability that he finishes the tour.
ii) the probability that he has trained, knowing that he has finished the hike.

At the end of the hike, the group decides to stay in Engelberg for dinner. Let $N$ count the number of restaurants they have to visit before getting a table for 6. Suppose $N$ is Poisson distributed with random intensity $\Lambda$, with $\Lambda \sim \text{Exp}(1/2)$.

**c)** Find the distribution of $N$ (not conditioned on $\Lambda$).
HINT: The following formulas may be useful to solve the exercise

i) The joint distribution of $(N, \Lambda)$ is given by

$$P(N = n, \Lambda \in [\lambda, \lambda + d\lambda)) = e^{-\lambda}\frac{\lambda^n}{n!} \cdot \frac{1}{2}e^{-\lambda/2}d\lambda, \quad n = 0, 1, \ldots, \text{ and } \lambda > 0.$$

ii) For $n \in \mathbb{N}$, we have $\int_0^\infty e^{-y}y^n dy = n!$

**d)** What is the expected number of restaurants they have to visit before getting the table?
HINT: If you could not solve **c)**, suppose that $N$ has the following distribution

$$P(N = n) = p(1-p)^n, \quad p = \frac{1}{4}.$$

**Please Turn!**

**2. (10 Points)** Let $X$ be a random variable whose density is given by

$$f_X(x) = \begin{cases} 12(x^2 - x^3), & x \in (0,1) \\ 0, & \text{otherwise.} \end{cases}$$

Let $Y \sim \text{Ber}(p = X)$. Thus, the joint density function of $X$ and $Y$ is given by

$$f_{X,Y}(x,y) = \begin{cases} 12(x^2 - x^3)(1-x), & x \in (0,1),\ y = 0 \\ 12(x^3 - x^4), & x \in (0,1),\ y = 1 \\ 0, & \text{otherwise,} \end{cases}$$

**a)** Compute the distribution of $Y$.

**b)** Calculate the expectation of $X + Y$.

**c)** Calculate $\text{Cov}(X, Y)$. Are $X, Y$ independent?

**d)** Are the random variables $X^2$ and $\sin(2\pi Y)$ independent? Explain your answer.

**3. (10 Points)** As a researcher at ETH Zurich, you decide to buy a supercomputer with 100 computing chips. According to the seller, during initialization of the supercomputers he produces, each computing chip has probability $p = 0.01$ of failing, independently of the other ones, and won't be available for computations until the supercomputer is turned off. Let us define $X$ as the random variable counting the number of available chips after initialization.

**a)** Calculate $P(X \leq \text{Var}(X))$.

Suppose from now on that the value of the parameter $p$ is unknown, and that you have a sample $X_i$, with $1 \leq i \leq n$, of independent random variables that have the same distribution as $X$.

**b)** Calculate the expected value of $X$ (as a function of the unknown parameter $p$) and propose a natural estimator $\tilde{p}$ for $p$.

**c)** Calculate the maximum-likelihood estimator $\widehat{p}$ of the parameter $p$ as a function of the sample $X_i$, with $1 \leq i \leq n$.

**d)** After initializing the supercomputer a certain number of times, you collect the following data:

| # of failed chips after initialization | 0 | 1 | 2 | 3 | 4+ |
|---|---|---|---|---|---|
| # of occurrences | 1 | 64 | 20 | 15 | 0 |

Calculate the estimated value $p^*$ given by $\widehat{p}$ according to this data sample.

**e)** Assuming $p = 0.01$ and $n = 100$, calculate $\dfrac{p^* - E[\widehat{p}]}{\sqrt{\text{Var}(\widehat{p})}}$. Assume for simplicity of calculation that $\sqrt{0.99} \sim 1$.

Does this calculation support the vendor's claim that $p = 0.01$? Why?