SAMPLE exam NMAI059 Probability and Statistics 1 – 2020/21

1. (10 points) The following table describes the joint probability function $p_{X,Y}(x, y)$ of the random vector (X, Y). These random variables only take values indicated in the table. We know that $\mathbb{E}(X) = 1/2$.

x \ y	-1	0	1
0	a	$1/8$	$1/4$
1	$1/8$	b	$1/8$

(a) Determine a and b.
(b) Decide whether X and Y are independent.

Solution:

(a) We known that

$$\mathbb{E}(X) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p_{X,Y}(1, -1) + p_{X,Y}(1, 0) + p_{X,Y}(1, 1) = 1/8 + b + 1/8,$$

and so $b = 1/4$. All numbers in the table sum to 1, so $a = 1/8$.

(b) Even without solving (a) we know that $p_X(0) = p_X(1) = 1/2$. If X and Y were independent, both numbers in each column would be the same and this is not the same, e.g., in the last one.

---

2. (10 points) Peter repeatedly tries to defeat a stronger opponent in chess. If Peter wins, he gains 10 points, otherwise (even in the event of a draw) he loses 2 points. Peter wins with probability 1/4.

(a) In how many rounds does Peter win for the first time (on average)?

(b) If Peter has 6 points at the beginning, with what probability will he be at zero after at most five rounds?

(c) What is the distribution of S, Peter's score after five games, if he has zero points at the start (negative points are allowed)? Describe the probability mass function of S.

Solution:

(a) The number of round, when Peter first wins, follows $Geom(1/4)$ distribution, thus the desired expectation is $\frac{1}{1/4} = 4$.

(b) If Peter wins, he gets 16 points and not even four wins do not bring him to zero. Thus if he is at zero after ≤ 5 rounds, he had to lose three times in a row. This occurs with probability $(3/4)^3$.

(c) We can imagine that in each round Peter loses two points and then get a chance to win 12 points with probability 1/4. The total score is therefore $-10 + 12 \cdot Bin(5, 1/4)$. Thus Peter has $-10 + 12k$ points (for $k = 0, 1, \dots, 5$) with probability given by the binomial distribution, that is $\binom{5}{k}(1/4)^k(3/4)^{5-k}$.

(d) Expected gain in each round is $\mu = +10 \cdot \frac{1}{4} - 2 \cdot \frac{3}{4} = 1$. We could also calculate variance, but it is sufficient that it is some constant σ^2.

In ten rounds Peter will have ten points in expectation, in hudred rounds hundred points. Thus we ask for the probability of substantially better gain, even "linearly better than the average". This probability should decrease with increasing n, as for large n the average is closer to the expectation.

Now properly: Central limit theorem implies that the $(X_1 + \cdots + X_n - n\mu)/(\sigma\sqrt{n})$ approximately follows the standard normal distribution. When we move from $n = 10$ to $n = 100$, the numerator increases ten times, while the denominator only $\sqrt{10}$-times. Thus we ask, if the standard normal distribution generates much larger number, that is the probability for $n = 100$ will be (substantially) smaller.

(As a matter of fact, for $n = 10$ the probability is 0.224, for $n = 100$ just 0.027. You do not need to evaluate this during the test.)

---

3. (10 points) You're throwing a party for 100 guests and wondering how many sandwiches to order. You know from experience that the number of sandwiches eaten by a random guest follows a Poisson distribution with a mean of 3. Approximately how many sandwiches do you need to order so that with probability 0.95 no guest will go hungry?

(Use an appropriate limit theorem.)

Solution: Let X_i denote the number of sandwiches eaten by the i-th guest. We are told that $X_i \sim Pois(3)$. From properties of Poisson distribution it follows that $\mathbb{E}(X_i) = \mathrm{var}(X_i) = 3$. By Central Limit Theorem $\frac{X_1 + \cdots + X_{100} - 300}{\sqrt{3} \cdot \sqrt{100}}$ appropriately follows standard normal distribution $N(0, 1)$, thus

$$P(X_1 + \cdots + X_{100} \leq 300 + t\sqrt{300}) \doteq \Phi(t).$$

To make this 95 %, we have to buy $300 + \sqrt{300} \cdot \Phi^{-1}(0.95)$ sandwiches. (This equals 328.5, but you don't need to evaluate this in the test. In R we can ask also directly for **qnorm**(0.95,300,**sqrt**(300)).)

An alternative solution without the limit theorem: we know that the sum of Poisson distributions is again a Poisson distribution. Thus the number of sandwiches eaten by the 100 guests follows $Pois(300)$. We are interested in the quantile function of this distribution at point 0.95. There is no simplification of this, but R can evaluate it by **qpois**(0.95,300), we get 329.

---

4. (10 points) (a) Define the concept of probability density function of a random variable X.

(b) Describe how to use it to determine $\mathbb{E}(X)$.

Solution:

(a) We say that $f : \mathbb{R} \to [0, \infty)$ is a probability density function of a random variable X, if for every $x \in \mathbb{R}$ we have

$$P(X \le x) = \int_{-\infty}^{x} f(t)dt.$$

(b) By the LOTUS law and the formula to compute variance we get

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_{-\infty}^{\infty} t^2 f(t)dt - \Big(\int_{-\infty}^{\infty} tf(t)dt\Big)^2$$

---

5. (10 points) State the theorem – the universality of uniform distribution. Explain how it can be used.

Solution:

Let F be a function "of CDF-type": nondecreasing right-continuous function with $\lim_{x\to-\infty} F(x) = 0$ a $\lim_{x\to+\infty} F(x) = 1$. Let Q be the corresponding quantile function.

Let $U \sim U(0, 1)$ and $X = Q(U)$. Then X has CDF F.

The theorem is useful to sample random variables with a given distribution – we need to be able to sample a uniform random variable on $(0, 1)$ and to evalutate the quantile function. We also discussed another, related theorem:

Let X be a r.v. with CDF $F_X = F$, suppose F is increasing. Then $F(X) \sim U(0, 1)$.

We can use this one for testing: if we have a procedure for testing that a random variable follows the uniform distribution, we can use it to test any other continuous distribution.

---

6. (10 points) State the theorem – the weak law of large numbers. Prove it.

Let $X_1, \dots, X_n$ be i.i.d. with expectation μ and variance σ^2. Let $S_n = (X_1 + \cdots + X_n)/n$ be the sample mean. Then for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P(|S_n - \mu| > \varepsilon) = 0.$$

We say that sequence S_n converges to μ *in probability* and write $S_n \xrightarrow{P} \mu$).

We prove the result using Chebyshev inequality. We first calculate the expectation and variance of $\bar{X}_n$:

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n}(\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)) = \frac{n\mu}{n} = \mu$$
$$\mathrm{var}(\bar{X}_n) = \frac{1}{n^2}(\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)) = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

By Chebyshev inequality we have

$$P(|\bar{X}_n - \mu| > t \cdot \sigma/\sqrt{n}) < \frac{1}{t^2}.$$

For $t\sigma/\sqrt{n} = \varepsilon$ we get $1/t^2 = \frac{\sigma^2}{\varepsilon^2 n}$, which tend to 0 as $n \to \infty$.