Sheet 4

Please email your solution to your corresponding instructor before **Friday, March 20th** at **12h00** by an email with object "<First name> <Last name> Stochastik Sheet 4" and an attachment "<First name>_<Last name>.pdf".

Exercise 1. One could think that the day of the week, or the date, of birth of an individual picked randomly (uniformly) in the population follows a uniform distribution. Find factors explaining why it is not the case (and cite your sources!).

Solution of exercise 1. In Spain, the data collected between 2002 and 2011 in the SEN1500 network (Rodrigo et al., J Pediatr (Rio J). 2019;95(1):41–47) indicates the following repartition of birth by weekday:

Day	M	Tu	W	Th	F	Sa	Su
Proportion(%)	14.4	15.6	14.9	15.6	16.1	12.1	11.2

Why is there a decrease of the number of births during the weekends? A common explanation is the fact that a number of births are "elective", by induction of labour or cesarean sections. Since in this case the day of birth is chosen by the medical staff, these births occur mainly between Monday and Friday.

To check this hypothesis, in Denmark, in 2007, at the Hvidore Hospital (one of the largest in the country), the distribution of "non-elective" births by weekday was the following (Gam et al., BJM open 2013;3;2002920):

Day	M	Tu	W	Th	F	Sa	Su
Proportion(%)	14.5	15.1	13.9	14.7	14.2	14.1	13.5

As one can see, the occurence of "naturally occuring" births is more uniform among the weekdays than the occurence of all births. Following (Rodrigo et al., J Pediatr (Rio J). 2019;95(1):41–47), a further unfortunate effect is played by the higher morbidity of newborns in the week-ends, explained by the lesser amount of hospital staff present.

The effect of elective births is also present in the repartition of births by date. Indeed, following the British Office for National Statistics, the proportion of individuals born on January 1st, December 24th or December 25th is 13.3% smaller than on an average day. Apart from this, the repartition of births by date of the month is quite uniform once accounted for the fact that not all months have the same number of days. Following the same data, the repartition of births by months (normalised by the number of days in the month), however, is not uniform:

Month	J	F	M	A	M	J	J	A	S	O	N	D
Proportion(%)	8.15	8.21	8.18	8.16	8.31	8.43	8.56	8.41	8.73	8.44	8.28	8.12

There are slightly more babies born around September and slightly less around March. This difference cannot be explained by choosing the exact date of birth, since one cannot delay or speed up a pregnancy by several months! Several possible seasonal factors affecting the frequency or fertility rate of sexual intercourse are discussed in "Human birth seasonality" by Elison, Valeggia and Sherry: social factors (gatherings, holidays, religious festivities), weather (via a psychological or biological influence), alimentation (which affects female fecundity). □

Exercise 2. A new, deadly virus is spreading among the population. Its propagation is as follows: at an instant $t \in \mathbb{N}_0$, there is a certain number $N(t)$ of viruses alive. Between time t and $t+1$, each virus dies and creates new viruses following a probabilistic law: there is a probability measure $\mathbb{P}$ on $\mathbb{N}_0$ such that, in an independent manner, each virus has a probability $\mathbb{P}(\{k\})$ of making k new viruses.

1. Find how the expectation of the number of viruses evolves with time. This evolution will involve the quantity $E = \sum_{k=1}^{+\infty} k\mathbb{P}(\{k\})$.
2. Thanks to containment measures and thorough washing of hands, $E < 1$. Prove that, almost surely, the epidemic disappears in finite time.

(Remark: if $E > 1$ one can prove that, with non-zero probability, the virus never goes extinct, but this is much harder!)

Solution of exercise 2.

1. Each virus has E expected offspring, so that, if X_n denotes the random variable describing the number of viruses at time n,
$$\mathbb{E}[X_{n+1}] = E \times \mathbb{E}[X_n].$$
2. If $E < 1$ then the expected number of viruses decays exponentially with time. Since the number of viruses is a natural number, by the Markov inequality,
$$\mathbb{P}(X_n \geq 1) \leq \mathbb{E}[X_n].$$
In particular, $\mathbb{P}(X_n \neq 0)$ decays faster than a geometric sequence with ratio < 1, so that it is summable. By the Borel-Cantelli lemma (sheet 2, exercise 3), the probability that $X_n = 0$ for all n large enough is 1.

□

Exercise 3. A pharmacological lab has developed a new product for testing people against this virus. The rate of false positives (if you are not infected, what is the chance that the test says you are infected) is 2% and the rate of false negatives (if you are infected, what is the chance that the test says you are not infected) is 5%.

There are 8 million inhabitants in Switzerland and 200 of them are infected with the virus (and don't know about it). Everyone goes to their doctor to be tested. If you are tested positive, what is the chance that you are infected?

Solution of exercise 3. Let A be the event "You are infected" and B the event "You are tested positive". We know that $\mathbb{P}(A) = \frac{1}{40000}$. In particular,

$$\mathbb{P}(A \cap B) = \frac{1}{40000}\frac{19}{20} = \frac{19}{800000}$$

and

$$\mathbb{P}(A^c \cap B) = \frac{39999}{40000}\frac{1}{50} = \frac{39999}{2000000}.$$

We in fact know that B happened, so the probability that you are also in A is

$$\frac{\frac{19}{800000}}{\frac{19}{800000} + \frac{39999}{2000000}} \approx 0.00119.$$

If this test says that you are infected, you have in fact more than 99% chances not to be infected! □

Exercise 4. A student works from home because the courses have been cancelled. To be well-prepared for the exam, they solve every day a random number of exercises, following a fixed probability distribution on $\mathbb{N}_0$. We know that the expectation of the number of exercises they solve every day is 50.

1. Give an upper bound for the probability that, on a given day, they solve more than 75 exercises.

2. We now know that the standard deviation (square root of variance) of the number of exercises solved by day is 5 exercises. Improve the previous upper bound.

3. Can one give a lower bound for the probability to solve more than 75 exercises?

Solution of exercise 4.

1. We use the Markov inequality: if X denotes the number of exercises solved in one day, X is positive, so that

$$\mathbb{P}(X \geq 75) \leq \frac{\mathbb{E}(X)}{75} = \frac{2}{3}.$$

2. We now use the Markov inequality on $(X-50)^2$. First of all, $\{X \geq 75\}$ is included in $\{(X-50)^2 \geq 625\}$. Then,

$$\mathbb{P}(X \geq 75) \leq \frac{\mathbb{E}((X-50)^2)}{625} = \frac{1}{25}.$$

3. If the number of exercises solved every day is 45 with probability $\frac{1}{2}$ and 55 with probability $\frac{1}{2}$, then the average is 50 and the standard deviation is 5, but the probability that $X \geq 75$ is zero. Hence, there is no non-trivial lower bound on this probability.

□

Exercise 5. For the following examples, write down carefully the probability space, its measure, and the random variable(s) of interest. Then, solve the problem as much as possible.

1. Alice flips two coins and bets on the fact that the two coins will fall on different sides.
2. Bob tosses a fair dice, then flips as many coins as the number indicated by the dice; he counts the number of coins that fall on heads.
3. Caroline tosses a coin until it falls on heads, and counts the number of tries.
4. Davide takes k loto balls out of an urn containing n balls numbered from 1 to n. He counts the sum of the numbers indicated by the balls.
5. Eskaterina places n tickets numbered from 1 to n into n envelopes numbered from 1 to n (one ticket per envelope). She counts the number of tickets that share the same number as the envelope they are placed in.

Solution of exercise 5.

1. Here $\Omega = \{HH, HT, TH, TT\} = \{H,T\}^2$ with uniform probability. Then the event $\{HT, TH\}$ happens with probability $\frac{1}{2}$.
2. In this case, $\Omega = \{1,2,3,4,5,6\} \times \{H,T\}^6$. Indeed, the probability space cannot depend on a probability event, so we are forced to take into account 6 possible coin flips, then only considering some of them in the random variable. An element of this probability space is denoted $\omega = (d, f_1, f_2, f_3, f_4, f_5, f_6)$. The random variable of interest is

   $$X = \sum_{j=1}^{6} \mathbb{1}(f_j = H, d \geq j).$$

   The expectation of X is $\frac{7}{4}$; the distribution of X as a random variable on $\{0, \ldots, 6\}$ is

value	0	1	2	3	4	5	6
probability	$\frac{63}{384}$	$\frac{120}{384}$	$\frac{99}{384}$	$\frac{64}{384}$	$\frac{29}{384}$	$\frac{8}{384}$	$\frac{1}{384}$.

3. In this question, the number of possible tosses is not limited. Hence, our probability space must account for an infinite number of tosses. Let $\Omega = \{H,T\}^{\mathbb{N}}$, the set of infinite sequences of characters among H and T. Ω is not finite, it is not even enumerable! Rather than describing directly $\mathcal{F}$ and $\mathbb{P}$, we will build $\mathcal{F}$ as the σ-algebra generated by some events, and give the probabilities of these events.

   What should $\mathcal{F}$ contain? At least, one should be able to observe anything that concerns only the n first tosses, for all n. Consider the family of σ-algebras

   $$\mathbb{F}_n = \{A \times \{H,T\}^{\mathbb{N}}, A \subset \{H,T\}^n\},$$

and the family of probability measures $\mathbb{P}_n$ on $\mathbb{F}_n$ given by the uniform law on $\{H,T\}^n$. Then the family $\mathbb{F}_n$ is increasing, in the sense that $\mathbb{F}_n \subset \mathbb{F}_{n+1}$ for all $n \in \mathbb{N}$. Moreover the probability measures are compatible with this increasing property: if $A \in \mathcal{F}_n$, then $\mathbb{P}_{n+1}(A) = \mathbb{P}_n(A)$.

This suggests to define the $\mathcal{F}$ as the smallest σ-algebra containing all the $\mathcal{F}_n$. It is not simply equal to $\cup_{n\in\mathbb{N}}\mathcal{F}_n$; indeed the singleton $\{T,T,T,\ldots\}$ belongs to $\mathcal{F}$ (as a countable intersection of elements of $\mathcal{F}_n$) but does not belong to any of the $\mathcal{F}_n$'s.

One can easily define a function $\mathbb{P} : \cup_{n\in\mathbb{N}}\mathcal{F}_n \to [0,1]$: for any element A of $\cup_{n\in\mathbb{N}}\mathcal{F}_n$, there holds $A \in \mathcal{F}_{n_0}$ for some $n_0 \in \mathbb{N}$, and we let $\mathbb{P}(A) = \mathbb{P}_{n_0}(A)$. The compatibility of the family $(\mathbb{P}_n)$ ensures that this definition makes sense.

The fact that $\mathbb{P}$ extends from $\cup_{n\in\mathbb{N}}\mathcal{F}_n$ to $\mathcal{F}$ is non-trivial, here as for the construction of Lebesgue measure. First of all, $\cup_{n\in\mathbb{N}}\mathcal{F}_n$ is a ring, in the following sense: it contains the empty set (because the empty set belongs to any of the $\mathcal{F}_n$, which are σ-algebras), it is closed under finite union (because any finite union is a finite union in some $\mathcal{F}_{n_0}$ for n_0 large enough, which is a σ-algebra), and it is closed under substraction: for $A,B \in \cup_{n\in\mathbb{N}}\mathcal{F}_n$ one has $A \setminus B \in \cup_{n\in\mathbb{N}}\mathcal{F}_n$ (because A and B belong to $\mathcal{F}_{n_0}$ for some n_0 large enough, then $A \setminus B \in \mathcal{F}_{n_0}$ since it is a σ-field).

Then $\mathbb{P}$ is a "probability pre-measure" on $\cup_{n\in\mathbb{N}}\mathcal{F}_n$: one has $\mathbb{P}(\Omega) = 1$, and $\mathbb{P}$ is closed under countable disjoint union as long as the union belongs to $\cup_{n\in\mathbb{N}}\mathcal{F}_n$. A theorem by Carathéodory, which you may have used in the construction of the Lebesgue measure, ensures then that $\mathbb{P}$ extends to a probability measure on $\mathcal{F}$.

Here, the singleton $\{T,T,T\ldots\}$ has probability 0 (indeed, it is contained in $\{T\}^N \times \{H,T\}^{\mathbb{N}_0} \in \mathcal{F}_N$ for all N, and the probability of this event is 2^{-N}. In particular, with probability 1, the coin will indeed fall on heads after a finite number of tossings. To study the law of the number X of tossings before the coin falls on heads, we remark that, for any $n \in \mathbb{N}$, $\{X > n\} = \{T\}^n \times \{H,T\}^{\mathbb{N}_0} \in \mathcal{F}_n$ happens with probability 2^{-n}. In particular, for every $n \in \mathbb{N}$, $\{X = n\}$ with probability 2^{-n}.

4. The finite probability space here is $\Omega = \{f : \{1,\ldots,k\} \to \{1,\ldots n\} \text{ injective}\}$ with uniform probability measure. The random variable of interest is

$$X = \sum_{j=1}^{k} f(j).$$

Its expectation is $\frac{k(n+1)}{2}$. Indeed, for any fixed j the random variable $f(j)$ is a uniform random variable in $\{1,\ldots,n\}$; the fact that the $f(j)$'s are not independent from each other never affects the expectation.

5. Here, $\Omega = \mathfrak{S}_n$ is the set of permutations of $\{1, \dots, n\}$, with uniform probability measure. The random variable of interest is

$$X = \sum_{j=1}^{n} \mathbb{1}(j = \sigma(j)).$$

One has $\mathbb{E}(X) = 1$: on average, a permutation has 1 fixed point.

□