1st Semester, 2021-2022

3rd Year, Math & CompSci

# Mathematical Statistics

## Seminar Exercises: Week 2/Week 3

**Recap** (Basic notions). *(a) Let $X$ be a <u>discrete</u> random variable on a probability space $(S, \mathcal{K}, P)$, $S = \{x_i |\ i \in I\}$. Then:*

- $f_X : S \to \mathbb{R}$, $f_X(x_i) = p_i$ *is the* ***probability density function (pdf)****;*
- $F_X : \mathbb{R} \to \mathbb{R}$, $F_X(x) = P(X \le x) = \sum\limits_{x \le x_i} p_i$ *is the* ***cumulative density function (cdf)****;*
- $E(X) = \sum\limits_{i \in I} x_i \cdot P(X = x_i)$ *the* ***mean****;*
- $\nu_k = E(X^k) = \sum\limits_{i \in I} x_i^k \cdot P(X = x_i)$ *the* ***moment of order*** $k$.

*(b) Let $X$ be a <u>continuous</u> random variable on a probability space $(S, \mathcal{K}, P)$, $S = \{x_i |\ i \in I\}$. Then:*

- $f_X : S \to \mathbb{R}$ *is the* ***probability density function (pdf)****;*
- $F_X : \mathbb{R} \to \mathbb{R}$, $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)dt$ *is the* ***cumulative density function (cdf)****;*
- $E(X) = \int_{\mathbb{R}} x \cdot f_X(x)dx$ *the* ***mean****;*
- $\nu_k = E(X^k) = \int_{\mathbb{R}} x^{\cdot} \cdot f_X(x)dx$ *the* ***moment of order*** $k$.

*(c)* $V(X) = E(X^2) - E(X)^2$ *the* ***variance****;*

*(d)* $\sigma = \sigma(X) = \sqrt{V(X)}$ *the **standard deviation**.*

**Definition.** The trials of an experiment are called **Bernoulli trials** if they satisfy the conditions:

1. they are independent;
2. each trial has only two possible outcomes: **success** ($A$) and **failure** ($\overline{A}$);
3. the probability of success $p = P(A)$ is the same for each trial.

**Definition.** We will study the following probabilistic models:

- **The binomial model:** We have $n$ Bernoulli trials with probability of success $p$ and are interested in the probability of exactly $k$ successes occuring ($0 \leq k \leq n$).

  $$X \sim \mathcal{B}(n; p), \quad X \begin{pmatrix} k \\ C_n^k p^k (1-p)^{n-k} \end{pmatrix}_{k \in \overline{0,n}}$$

  $E(X) = np$ and $V(X) = np(1-p)$

- **The negative binomial (Pascal) model:** We have an infinite sequence of Bernoulli trials with probability of success $p$ (and probability of failure $q = 1-p$) and are interested in the probability of the $n$th success occuring after exactly $k$ failures (or, equivalently, the probability of $k$ failures occuring before the $n$th success).

  $$X \sim \mathcal{NB}(k, n, p), \quad X \begin{pmatrix} k \\ C_{n+k-1}^k p^n q^k \end{pmatrix}_{k \in \mathbb{N}}$$

  $E(X) = \dfrac{nq}{p}$ and $V(X) = \dfrac{nq}{p^2}$

- **The geometric model:** The negative binomial model for $n = 1$:

$$X \sim Geo(p), \quad X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k \in \mathbb{N}}$$

$E(X) = \dfrac{q}{p}$ and $V(X) = \dfrac{q}{p^2}$

**Exercise 1.** A computer virus has entered a system with a very large number of files. A computer manager identifies the type of the virus, thereby learning that each file is independently damaged with probability 0.2. Next, the manager runs a program to check the condition of each file. Find the probability that

(a) 2 of the first 10 scanned files are damaged;

(b) at least 3 of the first 20 scanned files are damaged;

(c) at least 19 files are scanned before 10 undamaged files are found.

**Recap.** *If $X \sim Gamma(a, b)$, $a, b > 0$, then its pdf is:*

$$f_X(x) = \frac{1}{\Gamma(a) b^a} \cdot x^{a-1} e^{\frac{-x}{b}}$$

*for $x > 0$ (and $0$ otherwise). We have:*

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx, \ a > 0$$

**Exercise 2.** A user spends $X$ minutes on a certain website, where

$$X \sim Gamma(a, b), \ a, b > 0$$

with a mean value of 12 minutes and a standard deviation of 6 minutes. Find the probability that the user spends at most 10 minutes on the website, given that he spends at least 1 minute.

**Recap.** *If* $X \sim Beta(a,b)$, $a,b>0$, *then its pdf is:*

$$f_X(x) = \frac{1}{B(a,b)} \cdot x^{a-1}(1-x)^{b-1}$$

*for* $x \in [0,1]$ *(and* $0$ *otherwise). We have:*

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx,\ \ a,b>0$$

**Exercise 3.** A darts player aims at the bullseye, which has a diameter of 1 cm. The distance from the center of the dartboard to the hit point of an arrow thrown by the player follows the $Beta(a,b)$ distribution, where $a,b>0$, with a mean value of $\frac{3}{4}$ cm and a standard deviation of $\frac{\sqrt{15}}{20}$ cm. We assume that the throws are independent.

(a) Find the probability that the player hits the bullseye on a throw.

(b) Find the mean value of the number of throws before the player hits the bullseye the first time.

(c) Find the probability that the player hits the bullseye 2 times out of 10 throws.

**Recap** (The law of total probability)**.** *If* $\{A_i|\ i=1,2,\ldots\}$ *is a finite or countably infinite partition of a sample space* $S$ *(which means that the events* $A_i$ *are pairwise disjoint and their union is the entire sample space), then for any event* $B$ *in the sample space* $S$ *we have:*

$$P(B) = \sum_{i\in I} P(B \cap A_i) = \sum_{i \in I} P(B|A_i) \cdot P(A_i)$$

**Exercise 4.** An electronics store purchases a certain type of motherboard in boxes of 100 pieces each. It is known that 40% of the purchased boxes have 3 defective motherboards each, while 60% of the purchased boxes have 2 defective motherboards each. An employee of the store tests 2 motherboards from a random box. Find the probability that at least one of the tested motherboards is defective.

**Recap.** *Let $a, b \in \mathbb{Z}$. We say that a discrete variable $X$ on a probability space $(S, \mathcal{K}, P)$ follows a **discrete uniform distribution**:*

$$X \sim \mathcal{U}\{a, b\}$$

*if $X(S) = \{a, a+1, \dots, b\}$ and:*

$$P(X = k) = \frac{1}{b - a + 1}$$

*for every $k \in \{a, a+1, \dots, b\}$.*

**Exercise 5.** We consider the following game: two dice are rolled $n$ times, where $n$ is a positive integer.

(a) Find the probability that the sum of the two dice is a prime number 3 times out of 9 throws.

(b) Find the probability of getting the first double after 5 throws.

**Recap** (The Leibniz integral rule for differentiation under the integral sign)**.**

$$\frac{d}{dx} \int_a^x f(x,t)dt = f(x,x) + \int_a^x \frac{\partial}{\partial x} f(x,t)dt$$

**Exercise 6.** A computer is connected to 2 printers: $P_1$ and $P_2$. A printing job is sent with probability 0.4 to $P_1$ and with probability 0.6 to $P_2$. $P_1$ prints an $A2$ poster in $T_1$ seconds, where $T_1$ has an exponential distribution with the mean value of 5 seconds, while $P_2$ handles the same job in $T_2$ seconds, where $T_2$ has a uniform distribution on the interval $[4, 6]$. An A2 poster is printed by using the computer. Find the expected value of the printing time.