1st Semester, 2021-2022

3rd Year, Math & CompSci

Mathematical Statistics

Seminar Exercises: Week 6

Recap. *For any random variables X, Y and $\alpha \in \mathbb{R}$ we have:*

- $E(X+Y) = E(X) + E(Y)$;
- $E(\alpha X) = \alpha E(X)$;
- $V(\alpha X) = \alpha^2 V(X)$;

If in addition X and Y are independent, then:

- $E(X \cdot Y) = E(X) \cdot E(Y)$;
- $V(X+Y) = V(X) + V(Y)$

If $X \sim \mathcal{N}(\mu, \sigma)$, then:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ x \in \mathbb{R}$$

If $X \sim \mathcal{N}(\mu, \sigma)$, then:

$$Y = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

$$Z = X - \alpha \sim \mathcal{N}(\mu', \sigma')$$

If X and Y are independent random variables with $X, Y \sim \mathcal{N}(\mu, \sigma)$, then for any $\alpha, \beta \in \mathbb{R}$:

$$\alpha X + \beta Y \sim \mathcal{N}(\mu', \sigma')$$

If $X_1, X_2, \ldots, X_n$ are i.i.d. random variables, $X_i \sim \mathcal{N}(\mu, \sigma)$, then:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Exercise 1. Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. (independent identically distributed) random variables that follow the normal distribution, $X \sim \mathcal{N}(\mu, \sigma)$. Find the constant k_n such that the sampling function

$$\overline{s} = k_n \sum_{j=1}^{n} |X_j - \overline{X}|$$

verifies $E(\overline{s}) = \sigma$.

Recap. • *If $X_1, \ldots, X_n$ are independent identically distributed random variables with distribution given by $X \sim \mathcal{N}(\mu, \sigma)$ and we define the statistic:*

$$V := \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{(n-1)s^2}{\sigma^2}$$

then

$$V \sim \chi^2(n-1)$$

- *If $X \sim \chi^2(n)$, then:*

$$f_X(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \ x > 0$$

where for every $\alpha > 0$ we have:

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

Exercise 2. (a) Let $X \sim \chi^2(n)$. Find the probability density function of the random variable $Y = \sqrt{\frac{X}{n}}$.

(b) Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. (independent identically distributed) random variables that follow the distribution $X \sim \mathcal{N}(\mu, \sigma)$. Find the constant k_n such that the sampling function

$$\overline{s} = k_n \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

verifies $E(\overline{s}) = \sigma$.

Recap. • *A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables converges* ***almost surely*** *(denoted by* ***a.s.*** *and written as $X_n \overset{a.s}{\to} X$) to a random variable X if:*

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

- ***The Strong Law of Large Numbers (SLLN):***

  If $(X_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables with $X_n \sim X$, then

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} E(X)$$

- *If $X \sim Unif[a,b]$, then:*

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, \text{ if } x \in [a,b] \\ 0, \text{ otherwise} \end{cases}$$

Exercise 3. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. (independent identically distributed) random variables that follow the distribution

$$X \sim Unif[a,b]$$

where $0 < a < b$, and consider the following statistics:

1. Arithmetic mean of selection:

$$a_n(X_1,\ldots,X_n) := \frac{1}{n}\sum_{i=1}^{n} X_i$$

2. Geometric mean of selection:

$$g_n(X_1,\ldots,X_n) := \sqrt[n]{\prod_{i=1}^{n} X_i}$$

3. Harmonic mean of selection:

$$h_n(X_1,\ldots,X_n) := \frac{n}{\displaystyle\sum_{i=1}^{n}\frac{1}{X_i}}$$

Prove that each of the above statistics converges almost surely to a constant, as $n \to \infty$ and find these constants.

Exercise 4 (Glivenko-Cantelli)**.** Let X be a characteristic with cumulative distribution function F and $X_1, \dots, X_n$ sample variables for a random sample of size n with the sample distribution function $\overline{F}_n$, given by:

$$\overline{F}_n(x) = \frac{\#\{i \in \{1, \dots, n\} : X_i \leq x\}}{n}$$

Prove that $\overline{F}_n(x) \xrightarrow{a.s.} F(x)$, as $n \to \infty$.

<u>Hint:</u> Use the SLLN.

Recap. • *If $X \sim Bern(p)$, then:*

$$X \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

- *If $X \sim Geo(p)$, then:*

$$X \begin{pmatrix} k \\ p(1-p)^k \end{pmatrix}_{k \in \mathbb{N}}$$

- *If $X \sim Exp(\lambda)$, then:*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, \textit{ for } x > 0 \\ 0, \textit{ for } x \leq 0 \end{cases}$$

Exercise 5. Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. (independent identically distributed) random variables that follow the distribution of a given characteristic X with finite mean and variance. Consider the statistic

$$s_n := s_n(X_1, \ldots, X_n) = X_1 + \ldots + X_n, \ n \in \mathbb{N}^*$$

Find the distribution of s_n in each of the cases below (using induction, if you so desire) and then find the mean and the variance of this statistic.

(a) $X \sim Bern(p)$, where $p \in (0, 1)$.

(b) $X \sim Geo(p)$, where $p \in (0, 1)$.

(c) $X \sim Exp(\lambda)$, where $\lambda > 0$.