Homework 8, due August 20

Problem 1. Consider the selection
13, 12, 13, 10, 11, 13, 12, 11, 13, 14.

(i) Find \( \overline{x} \) and \( s^2 \), the empirical mean and the empirical variance.
(ii) Find the confidence interval for the mean \( \mu \) corresponding to the confidence level 95% (assuming you can use normal approximation, which is actually questionable for this small selection).
(iii) Test the hypotheses \( \mu = 10 \) and \( \mu = 15 \) for significance level \( p = .05 \).

Problem 2. A hotly disputed Proposition X will be on the ballot in November elections. A poll of \( N_1 = 100 \) people shows that \( p_1 = 50\% \) (50 people) support Proposition X. A week later, a poll of \( N_2 = 400 \) people shows that \( p_2 = 55\% \) (220 people) support Proposition X. The local newspaper claims: “Support for Proposition X grows”.

(i) You are unconvinced and are trying to calculate the significance \( p \)-value for the null hypothesis: that the support stays on the level 50%. You use normal approximation. What is this \( p \)?
(ii) Can you reject the hypothesis with \( p = .05 \)? How about \( p = .01 \)?
(iii) What minimal \( N_2 \) (assuming \( N_1, p_1 \) and \( p_2 \) are unchanged) do you need to reject the null hypothesis with \( p = .05 \)?

Problem 3. For iid \( X_1, X_2, \ldots, X_N \) with \( \text{Var} X_k = \sigma^2 \), show that
\[
\mathbb{E}(X_k - \overline{X})^2 = \frac{n-1}{n} \sigma^2.
\]
Using this, show that
\[
\mathbb{E}s^2 = \sigma^2.
\]

Problem 4. Let \( X_1, X_2, \ldots \) be independent but not necessarily identically distributed random variables with \( \text{Var} X_k = \sigma_k^2 \), and let
\[
S_N = X_1 + \ldots + X_N.
\]

(i) Assume that
\[
\frac{1}{N^2} \sum_{k=1}^{N} \sigma_k^2 \to 0, \quad N \to \infty.
\]
Show that, as \( N \to \infty \),
\[
\frac{S_N - \mathbb{E}S_N}{N} \to 0,
\]
in other words, for every fixed \( \varepsilon > 0 \)
\[
\mathbb{P} \left( \left| \frac{S_N - \mathbb{E}S_N}{N} \right| \geq \varepsilon \right) \to 0.
\]

(ii) Show that the law of large numbers from the lecture, for iid random variables, is the consequence of this.
(iii) This law of large numbers is not always true: let

\[ X_n \sim \mathcal{N}(0, 2n + 1), \ n \geq 1, \]

be independent random variables. Show that

\[ \frac{S_N - \mathbb{E}S_N}{N} \sim \mathcal{N}(0, 1), \]

and therefore it does not tend to zero. Use the fact that if \( X \sim \mathcal{N}(0, \sigma_1^2) \) and \( Y \sim \mathcal{N}(0, \sigma_2^2) \) are independent, then \( X + Y \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2) \).

**Problem 5.** Estimate \( P(X \geq 10) \) using Markov’s inequality, for

(i) \( X \sim \text{Poi}(1) \);
(ii) \( X \sim \text{NB}(2, 1/2) \);
(iii) \( X \sim \text{Uni}(0, 10) \).

**Problem 6.** The average height of a person from a certain group is 1.7 meters, the standard deviation is 0.25 meters. Estimate the probability that the given person has height \( X \) such that \( |X - 1.7| \geq 0.3 \).

Use Chebychev’s inequality.

**Problem 7.** Candidates Smith and Johnson are running for US President. Peter the Pundit estimates the probability that Johnson wins in each of the 51 states (let us call Washington, DC also a state). For the sake of example, assume that each state sends only one person to Electoral College. The person who gets a simple majority wins.

There are 30 strongly pro-Smith states: Johnson wins in each of these states only with probability 20%. The other 21 states are strongly pro-Johnson: Johnson wins in each of them with probability 90%. All states vote for Smith or Johnson independently of each other.

A political weekly asks Peter: “Who will win in each state?” For pro-Smith states, Peter says ”Smith”. For pro-Johnson states, Peter says ”Johnson”.

(i) Approximate using Poisson distribution the number of states that Peter is going to get wrong. What is the probability that he gets all states right?
(ii) Estimate the probability that Johnson wins, using Normal approximation.