Exercise 1. The sample variance of random variables $X_1, \ldots, X_n$ is defined as

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

where $\bar{X}$ is the sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Suppose that the $X_i$’s all have the same expected value $\mu$ and variance $\sigma^2$. Prove that

$$\text{Ex}(S_n^2) = \sigma^2$$

Notice that this implies that the sample variance $S^2$ is an unbiased estimator of the true variance $\sigma^2$.

Exercise 2. We obtain 25 independent samples $X_1, X_2, \ldots, X_{25}$ from the Normal($\mu, \sigma^2$) distribution, where both $\mu$ and $\sigma^2$ are unknown. Consider the hypothesis test: $H_0 : \mu = 12, H_1 : \mu > 12$.

(a) Design and analyze a significance test for the hypothesis with confidence level $\alpha = 0.05$.

(b) Suppose that the observed sample mean is 14 and the observed sample variance is $(4.32)^2$. Can we reject the null hypothesis with significance level $\alpha = 0.05$?

Exercise 3. (Problem from an earlier exam) An online advertisement can convince a $\rho$-fraction of Internet users to click on the ad. The advertiser would like to estimate $\rho$. The search engine shows the ad to $n$ randomly and independently selected Internet users. Let $X_i = 1$ if the ith user clicks on the ad, and 0 otherwise. The search engine estimates $\rho$ using the sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- (2 points.) Determine $\text{Ex}(\bar{X}_n)$.

- (2 points.) Determine $\text{Var}(\bar{X}_n)$.

- (3 points.) True or False? If $n = 10^5$, then $\bar{X}_n$ estimates $\rho$ within $\pm 0.02$ with at least 95% confidence.

- To protect the privacy of search engine users, the advertiser is given a “noisy” estimate of the sample mean. That is, the advertiser is given $Y$ where

$$Y = \bar{X}_n + W$$
where the “noise” \( W \sim \text{Normal}(0, \sigma^2) \) is normally distributed with mean 0 and variance \( \sigma^2 = 10^{-5} \). The noise \( W \) is independent of \( \bar{X}_n \).

The advertiser sets up the following hypothesis test:

\[
H_0 : \rho = 0.1 \\
H_1 : \rho \neq 0.1
\]

The search engine tells the advertiser that \( Y = 0.12 \) and \( n = 10^3 \).

- (2 points.) What is \( \mathbf{E}X(Y) \) and \( \mathbf{Var}(Y) \)?

- (2 points.) Since \( n \) is so large, assume that \( \bar{X}_n \) is normally distributed.

Using the normal table\(^1\), what is the \( p \)-value for this hypothesis test? (Give the actual number; a response in closed form is not acceptable here.)

- (1 point.) True or False? The advertiser can reject the null hypothesis \( H_0 \) with 0.05 significance level.

\(^1\)The Normal table was attached to the exam.