Homework 5

Dec. 10, 2020

NOTE: Homework 5 is due next Thursday (Dec. 17, 2020). The questions started with * are for Exercise only, and you are not required to submit the answers.

1. Suppose that $X_1, \cdots, X_m$ i.i.d. $\sim N(a+c, \sigma_1^2)$, $Y_1, \cdots, Y_n$ i.i.d. $\sim N(a, \sigma_2^2)$, where $c, \sigma_1^2, \sigma_2^2$ are known, a is unknown. Suppose that X_i's and Y_j's are independent.
   - Derive the UMVUE of a;
   - Construct a confidence interval for a with confidence level $1 - \alpha$.
2. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from the population with p.d.f. $f(x, \theta) = e^{-(x-\theta)}, \ x > \theta$.
   - Show that the distribution of $X_{(1)} - \theta$ does not depend on θ;
   - Construct a confidence interval for θ with confidence level $1 - \alpha$.
3. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from a distribution with unknown (finite) mean μ and known (finite) variance σ^2, and suppose that n is large. Then:
   (i) Construct a confidence interval for μ with confidence level approximately $1 - \alpha$.
   (ii) Provide the form of the interval in part (i) for $n = 100$, $\sigma = 1$, and $\alpha = 0.05$.
   (iii) Refer to part (i) and suppose that $\sigma = 1$ and $\alpha = 0.05$. Then determine the sample size n, so that the length of the confidence interval is less than 0.1.
   (iv) Show that the length of the interval in part (i) tends to 0 in probability as $n \to \infty$.
4. Consider the data set "data.csv" in Homework 4. There are 1475 houses which have central air-conditioner (Central_Air = Y) and 123 houses which don't have air-conditioner (Central_Air = N). Compare the mean SalePrice and the variance of SalePrice of the houses with central air-conditioner (Central_Air = Y) and those without central air-conditioner (Central_Air = N). More specifically, let $X_1, \cdots, X_{1475}$ denote the SalePrice of the houses with central air-conditioner and suppose that they are i.i.d. r.v.'s. Let $Y_1, \cdots, Y_{123}$ denote the SalePrice of the houses without central air-conditioner and suppose that they are i.i.d. r.v.'s too. Suppose that X_i's and Y_j's are independent.
   (i) Suppose that X_i *i.i.d.* $\sim N(\mu_1, \sigma^2)$, Y_j *i.i.d.* $\sim N(\mu_2, \sigma^2)$, $\sigma = 79400$. Construct a 95% confidence interval for $\mu_1 - \mu_2$.
   (ii) Suppose that X_i *i.i.d.* $\sim N(\mu_1, \sigma^2)$, Y_j *i.i.d.* $\sim N(\mu_2, \sigma^2)$, σ is unknown. Construct a 95% confidence interval for $\mu_1 - \mu_2$.
   (iii) Suppose that X_i *i.i.d.* $\sim N(\mu_1, \sigma_1^2)$, Y_j *i.i.d.* $\sim N(\mu_2, \sigma_2^2)$, $\sigma_1^2 \neq \sigma_2^2$ are unknown. Construct a 95% confidence interval for σ_1/σ_2.
   (iv) Refer to part (iii), construct approximately 95% confidence intervals for $\mu_1 - \mu_2$ by using central limit theory.
   (v) Construct 95% confidence region for (μ_1, μ_2).
5. In the following examples, indicate which statements constitute a simple and which a composite hypothesis:

- When tossing a coin, let X be the r.v. taking value 1 if the head appears and 0 if the tail appears. The statement is: The coin is biased.
- X is a r.v. whose expectation is equal to 5.

6. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from a binomial distribution $B(1, p)$ where p is unknown and $n = 20$. To test: $H_0 : p = 0.1 \leftrightarrow H_1 : p \neq 0.1$, we use the following test function:

$$\varphi(\boldsymbol{X}) = \begin{cases} 1, & \sum_{i=1}^{20} X_i \geq 6 \text{ or } \sum_{i=1}^{20} X_i \leq 2, \\ 0, & \text{otherwise} . \end{cases}$$

   - Calculate the power function of the test at $p = 0, 0.1, 0.2, \cdots, 0.9, 1$.
   - Determine the level of significance α and the probability of type II error.

7. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from normal distribution $N(\mu, 4)$ where μ is unknown. Let $\bar{X}$ be the sample mean. To test: $H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0$, the rejection region is

$$D = \left\{ \boldsymbol{X} = (X_1, \cdots, X_n) : |\bar{X} - \mu_0| \geq c \right\} .$$

   - Determine c such that the level of significance $\alpha = 0.05$.
   - Determine the power function of the test.
   - How does the power function change with sample size n?

8. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from uniform distribution $U(0, \theta)$, where θ is unknown. Let $X_{(n)} = \max\{X_1, \cdots, X_n\}$. For the hypothesis testing problem $H_0 : \theta \leq \theta_0, \leftrightarrow \; H_1 : \theta > \theta_0$, consider the following test function:

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & X_{(n)} \geq c, \\ 0, & \text{otherwise} . \end{cases} \tag{1}$$

   (i) Calculate the power function of the test $\varphi(\boldsymbol{x})$, and show that it is an increasing function of θ.

   (ii) If $\theta_0 = 1/2$, determine the critical value c such that the level of significance $\alpha = 0.05$.

   (iii) Refer to (ii), determine the sample size n such that the power of the test at $\theta = 3/4$ is at least 0.98.

   (iv) Refer to (ii), determine the sample size n such that the type II error of the test is no more than 0.02 at $\theta = 3/4$.

9. A manufacturer claims that packages of certain goods contain 18 ounces. In order to check his claim, 100 packages are chosen at random from a large lot and it is found that $\sum_{i=1}^{100} x_i = 1570$ and $\sum_{i=1}^{100} x_i^2 = 32000$. Assume that the observations are Normally distributed, and formulate the manufacturer's claim as a testing hypothesis problem. Carry out the test at level of significance $\alpha = 0.01$ and compute the p-value.

*1. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from the Uniform distribution $U(\theta - 1/2, \theta + 1/2)$, construct a confidence interval for θ with confidence level $1 - \alpha$.

*2. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from Normal distribution $N(\mu, \sigma^2)$. To make $\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]^{1/2}/4$ be a 95% upper confidence limit of σ, how large sample size n should be?