DS3001: ADVANCED STATISTICS: ASSIGNMENT 3

HYPOTHESIS TESTING

PROBLEM 1

Let $X_1, \dots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ where σ^2 is unknown. We wish to test the hypothesis

$$H_0 : \mu = 5 \qquad \text{vs.} \qquad H_1 : \mu > 5.$$

Under H_0,

$$T^* = \frac{\bar{X} - 5}{S/\sqrt{n}} \sim t_{n-1}, \text{where } S \text{ is the sample standard deviation.}$$

Consider the following population models:

i. $N(\mu, 1)$,

ii. mixture of two normals: $0.9N(\mu, \sigma^2 = 1) + 0.1N(\mu, \sigma^2 = 25)$.

To obtain the power function, we will use a Monte Carlo method as described below:

i. Select a particular value of the parameter, say μ_1.

ii. For each replicate, indexed by $j = 1, \dots, 1000$:

  * Generate the $j-$th sample $x_1^{(j)}, \dots, x_n^{(j)}$ of size n from each of the population models described above with $\mu = \mu_1$.
  * Compute the test statistic T_j^* from the $j-$th sample.
  * Create an indicator variable to determine whether or not H_0 is rejected at $\alpha = 0.05$.

iii. Compute the proportion of significant tests.

iv. Repeat the process for different values of μ_1 ranging from 4.5 to 6.5 at intervals of 0.1.

Here's what you need to do:

(a) Plot the power curve for $n = 10, 20, 100$.

(b) What can you say about the power of the $t-$ test if the data are contaminated? What happens when the sample size increases?

PROBLEM 2

The dimension-free **measure of skewness** of a random variable X is given by

$$\beta_1 = \frac{E(X-\mu)^3}{\sigma^3},$$

where $\mu = E(X)$ and $\sigma^2 = Var(X)$. The index is zero when the distribution is symmetric about the mean. The sample measure of skewness denoted by b_1 is defined as

$$b_1 = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i-\bar{X})^3}{\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^{3/2}}$$

(a) Let $X_1, \dots, X_n$ be i.i.d. $N(0,1)$ random variables. Find the sampling distribution of b_1 for $n = 20, 30, 100$. Provide a graph of the sampling distributions and compute the mean and standard error of b_1.

(b) Theory suggests that b_1 is normally distributed with mean 0 and variance $\dfrac{6(n-2)}{(n+1)(n+3)}$. Are your results in part (a) consistent with the theory?

PROBLEM 3

Since a normal distribution is symmetric, we can define a test of normality based on the skewness. To test

$$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0,$$

we reject H_0 for large values of $|b_1|$.

Consider the class of contaminated normal distributions given by

$$(1-\epsilon)N(0,\sigma^2=1) + \epsilon N(0,\sigma^2=100), \qquad 0 \le \epsilon \le 1.$$

When ϵ is 0 or 1, the distribution is normal.

(a) Using the Monte Carlo method described in **Problem 1**, derive the power of the skewness test for a sequence of alternatives indexed by ϵ and plot the power curve. Use $\alpha = 0.01$ and $n = 30$.

(b) What can you infer from the plot?