MATH 756/856–Final Take-Home Exam

Out: May 8, 2023; Due: Monday, May 15, 2023

Name______________________________ Score______________

1. (12pts) The following 16 random samples; 5.33, 4.25, 3.15, 3.70, 1.61, 6.40, 3.12, 6.59, 3.53, 4.74, 0.11, 1.60, 5.49, 1.72, 4.15, 2.30; came from normal distribution with mean μ and variance σ^2, i.e., $X_1,\ X_2,\ \cdots, X_{16} \sim N(\mu,\ \sigma^2)$, with the density function

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

(a) (4pts) Find the maximum likelihood estimates of μ and σ^2, denoted with $\hat{\mu}$ and $\hat{\sigma}^2$.

(b) (4pts) Based on above $\hat{\mu}$ and $\hat{\sigma}^2$, construct 95% confidence intervals for μ and σ^2 separately.

(c) (4pts) Based on part (b) and the duality of confidence intervals and the hypothesis tests, carry out a hypothesis test of $H_0 : \mu = 4$ versus $H_1 : \mu \neq 4$ at $\alpha = 0.05$.

2. (10pts) Suppose that $X_1, X_2, \cdots, X_n$ are independent Poisson random variables with mean θ and that θ has a gamma prior distribution.

(a) (4pts) Show that the posterior distribution of θ is also gamma.

(b) (6pts) Determine the Bayes estimate of θ and show that it is a weighted average of the prior mean and the sample mean $\overline{X}$.

3. (16pts, Problem 46, page 322) The data of this exercise were gathered as part of a study to estimate the population size of the bowhead whale (Raftery and Zeh 1993). The statistical procedures for estimating the population size along with an assessment of the variability of the estimate were quite involved, and this problem deals with only one aspect of the problem–a study of the distribution of whale swimming speeds. Pairs of sightings and corresponding locations that could be reliably attributed to the same whale were collected, this providing an estimate of velocity for each whale. The velocities, $v_1, v_2, \cdots, v_{210}$ (km/h) were converted into times $t_1, t_2, \cdots, t_{210}$ to swim 1 km; $t_i = 1/v_i$. The distribution of the t_i was then fit by an exponential distribution and a gamma distribution. The times (or the data observations) are contained in the file *whales.txt* posted at Canvas.

(a) (6pts) Using R to make a histogram of the remaining 210 values of t_i. Does it appear that an exponential distribution would be a plausible model to fit? In specific, using R to construct a probability plot of the data against the quantiles of an exponential distribution to assess qualitatively whether the exponential is a reasonable model.

(b) (6pts) Fit the parameters of the exponential and gamma distribution by method of moments.

(d) (4pts) Using R to plot the exponential and gamma densities on top of the histogram. Do the fits look reasonable?

4. (12pts) Suppose that we are given a random sample $X_1, X_2, \cdots, X_n$ from the *p.d.f.*

$$f(x) = \theta x^{\theta-1}, \quad x \in (0, 1),$$

where $\theta > 0$ is an unknown parameter. The null hypothesis $H_0 : \theta = 1$ is to be tested against the alternative $H_1 : \theta > 1$.

(a) (4pts) Show that the family of uniformly most powerful tests have the rejection region of the form $\prod_{i=1}^{n} X_i > c_0$ for some constant c_0.

(b) (8pts) Let $X \sim U(0, 1)$ and $Y = -\log X$, show that Y follows exponential distribution with $\lambda = 1$, i.e., $Y \sim \exp(1)$. Use this result to find a uniformly most powerful test for above $H_0 : \theta = 1$ vs. $H_1 : \theta > 1$. (i.e., determining the exact value of that constant c_0 in part (a) in terms of a known distribution's quartile using exact distribution of the test statistics) at significance level $\alpha = 0.05$.

5. (12pts) Suppose that we are given a random sample $X_1, X_2, \cdots, X_n$ from $N(0, \sigma^2)$, where $\sigma > 0$ is an unknown parameter.

(a) (6pts) Find the most powerful test for: the null hypothesis $H_0 : \sigma = \sigma_0$ is to be tested against the alternative $H_1 : \sigma = \sigma_1$ at significance level $\alpha = 0.05$, where $\sigma_1 > \sigma_0$. The values of σ_0 and σ_1 are fixed.

(b) (6pts) Derive a likelihood ratio test of $H_0 : \sigma = \sigma_0$ verse $H_1 : \sigma \neq \sigma_0$ at significance level $\alpha = 0.05$. Provide the test statistic and rejection region as simple as possible and explain how to carry out above hypothesis test $H_0 : \sigma = \sigma_0$ verse $H_1 : \sigma \neq \sigma_0$.

6. (14pts) Suppose that n measurements are to be taken under a treatment condition and another n measurements are to be taken independently under a control condition. It is assumed that the measurements follow normal distribution and the standard deviation is about 10 under both conditions.

(a) (6pts) How large should n be so that the test of $H_0 : \mu_X = \mu_Y$ against the one-sided alternative $H_1 : \mu_X > \mu_Y$ has a power 0.8 if $\mu_X - \mu_Y = 2$ and $\alpha = 0.05$.

(b) (8pts) Consider conducting a two-sided test of the null hypothesis $H_0 : \mu_X = \mu_Y$ as described in part (a). Use R to plot power curves on the same plot for (1) $\alpha = 0.05, n = 20$; (2) $\alpha = 0.10, n = 20$; (3) $\alpha = 0.05, n = 40$; (4) $\alpha = 0.10, n = 40$. Comment on your result.

7. (12pts) An experiment was done to measure the effects of ozone, a component of smog. A group of 22 seventy-day-old rats were kept in an environment containing ozone for 7 days, and their weight gains were recorded. Another group of 23 rats of a similar age were kept in an ozone-free environment for a similar time, and their weight gains were recorded. The data (in grams) are given below. Analyze the data to determine the effect of ozone. Write a summary of your conclusions.

Controls: 41.0, 38.4, 24.9, 25.9, 21.9, 18.3, 13.1, 27.3, 28.5, -16.9, 17.4, 21.8, 15.4, 27.4, 19.2, 22.4, 17.7, 26.0, 29.4, 21.4, 22.7, 26.0, 26.6.
Ozone: 10.1, 6.1, 20.4, 7.3, 14.3, 15.5, -9.9, 6.8, 28.2, 17.9, -12.9, 14.0, 6.6, 12.1, 15.7, 39.9, -15.9, 54.6, -14.7, 44.1, -9.0, -9.0.

8. (12pts) An experiment was done to test a method for reducing faults on telephone lines (Welch 1987). Fourteen matched pairs of areas were used. The following table shows the fault rates for the control areas and for the test areas:

Test:	676	206	230	256	280	433	337	466	497	512	794	428	452	512
Control:	88	570	605	617	653	2913	924	286	1098	982	2346	321	615	519

Do you think it is more appropriate to use a t test or a nonparametric method to test whether the apparent difference between test and control could be due to chance? Why? Carry out both tests and compare.

Notes: Problems 1–3 come from Chapter 8, Problems 3(a), 4, and 5 come from Chapter 9, and Problems 6, 7 and 8 come from Chapter 11.