MIDTERM EXAM PRACTICE PROBLEMS

Problem 1:

Suppose that $X_1, X_2, \dots, X_n$ are i.i.d. random variables on the interval [0, 1] with the density function:

$$f(x;\alpha) = \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} x^{\alpha-1}(1-x)^{2\alpha-1}$$

where Γ is the gamma function and where $\alpha > 0$ is a parameter to be estimated from the sample. Given that

$$E(X) = \frac{1}{3}$$

$$Var(X) = \frac{2}{9(3\alpha+1)}$$

a) Find the method of moments estimator of α.
b) Write the equation for the maximum likelihood estimator of α.
c) Find a sufficient statistic for α.

Problem 2:

Suppose that *X* is a discrete random variable with $P(X=1) = \theta$ and $P(X=2) = 1-\theta$. Three independent observations of *X* are made: $x_1 = 1, x_2 = 2, x_3 = 2$.

a) Find the method of moments estimator of θ.
b) Write out and simplify the likelihood function.
c) Find the maximum likelihood estimator of θ.
d) If the prior distribution of θ is uniform on [0, 1], what is the posterior density function?
e) Find the Bayes Posterior Mean Estimate, the MAP estimate, the Posterior Median Estimate, and a 95% Bayesian Confidence Interval for the parameter θ.

Problem 3:

Given the following observations of a normal random variable.

5.3286	4.1563	3.2107	3.7684
1.8909	6.4002	3.2011	6.4485
3.5287	4.6993	0.1982	1.5976
5.4835	1.8539	4.2036	2.2840

What would you guess the mean and variance (μ and σ^2) of the normal distribution were?

a) Use the method of moments.
b) Use the maximum likelihood method.

Problem 4:

Let $X_1, X_2, \dots, X_n$ are i.i.d. random variables on the interval [0, 1] with the density function:

$$f(x; \theta) = (\theta + 1)x^{\theta}, \quad 0 \le x \le 1$$

a) Find the method of moments estimator of θ.
b) Find the maximum likelihood estimator of θ.
c) Find a sufficient statistic for θ.
d) Find the Fisher information for θ. Then use it to find a 95% confidence interval for θ.
e) Find the Cramer-Rao lower bound for θ.

Problem 5:

Let $Y_1, Y_2, \dots, Y_n$ denote a random sample from a normal distribution with mean μ and variance σ^2.

a) If μ is unknown and σ^2 is known, find a sufficient statistic for μ.
b) If σ^2 is unknown and μ is known, find a sufficient statistic for σ^2.

Problem 6:

Suppose that *X* is a discrete random variable with

$$P(X = 0) = \frac{2}{3}\theta$$
$$P(X = 1) = \frac{1}{3}\theta$$
$$P(X = 2) = \frac{2}{3}(1 - \theta)$$
$$P(X = 3) = \frac{1}{3}(1 - \theta)$$

where $0 \le \theta \le 1$ is a parameter. The following 10 independent observations were taken from such a distribution: (3, 0, 2, 1, 3, 2, 1, 0, 2, 1).

a) Find the method of moments estimate of θ.
b) What is the maximum likelihood estimate of θ?
c) If the prior distribution of θ is uniform on [0, 1], what is the posterior density function?
d) Find the Bayes Posterior Mean Estimate, the MAP estimate, the Posterior Median Estimate, and a 90% Bayesian Confidence Interval for the parameter θ.

Problem 7:

Let $X_1, X_2, \dots, X_n$ be an i.i.d. sample from a distribution with the density function

$$f(x|\theta) = \frac{\theta}{(1 + x)^{\theta + 1}}, \quad 0 < \theta < \infty \; and \; 0 \le x < \infty$$

Find a sufficient statistic for θ.

Problem 8:

Find sufficient statistics for the gamma distribution.

Problem 9:

Currently, 20% of potential customers buy soap of brand A. To increase sales, the company will conduct an extensive advertising campaign. At the end of the campaign, a sample of 400 potential customers will be interviewed to determine whether the campaign was successful.

a) State H_0 and H_a in terms of p, the probability that a customer prefers soap brand A.
b) The company decides to conclude that the advertising campaign was a success if at least 92 of the 400 customers interviewed prefer brand A. Find α. (Use the normal approximation to the binomial distribution to evaluate the desired probability.)

Problem 10:

Let X be a binomial random variable with n trials and probability p of success. What is the generalized likelihood ratio for testing H_0: p = 0.5 versus H_a: p ≠ 0.5 ?

Problem 11:

Suppose that X ~ Bin(100, p). Consider the test that rejects H_0: p = 0.5 in favor of H_a: p ≠ 0.5 for |X − 50| > 10. Use the normal approximation to the binomial distribution to find α.

Problem 12:

A coin is thrown independently 10 times to test the hypothesis that the probability of heads is 0.5 versus the alternative that the probability is not 0.5. The test rejects if either 0 or 10 heads are observed. What is the significance level of the test?

Problem 13:

Let $X_1, X_2, \dots, X_{25}$ be a sample from a normal distribution having a variance of 100. Find the rejection region for a test at level $\alpha = 0.1$ of $H_0: \mu = 0$ versus $H_a: \mu = 1.5$.

Problem 14:

Let $X_1, X_2, \dots, X_{10}$ be a random sample from an exponential distribution with the density function $f(x|\theta) = \theta\, exp[-\theta x]$. Find the rejection region for a test at level $\alpha = 0.05$ of $H_0: \theta = 1$ versus $H_a: \theta \neq 1$.

Problem 15:

For two factors—starchy or sugary, and green base leaf or white base leaf—the following counts for the progeny of self-fertilized heterozygotes were observed (Fisher 1958):

Type	Count
Starchy green	1997
Starchy white	906
Sugary green	904
Sugary white	32

According to genetic theory, the cell probabilities are $0.25(2+\theta)$, $0.25(1-\theta)$, $0.25(1-\theta)$ and 0.25θ, where θ $(0<\theta<1)$ is a parameter related to the linkage of the factors. Test the goodness of fit of the above data to the genetic model at level $\alpha = 0.05$.

Problem 16:

If gene frequencies are in equilibrium, the genotypes AA, Aa, and aa occur with probabilities $(1-\theta)^2$, $2\theta(1-\theta)$, and θ^2, respectively. Plato et al. (1964) published the following data on haptoglobin type in a sample of 190 people. Test the goodness of fit of the above data to the genetic model at level $\alpha = 0.05$.

Haptoglobin Type		
Hp1-1	Hp1-2	Hp2-2
10	68	112

Problem 17:

The National Center for Health Statistics (1970) gives the following data on distribution of suicides in the United States by month in 1970. At level $\alpha = 0.05$, is there any evidence that the suicide rate varies seasonally, or are the data consistent with the hypothesis that the rate is constant? (Hint: Under the latter hypothesis, model the number of suicides in each month as a multinomial random variable with the appropriate probabilities and conduct a goodness of fit test. Look at the signs of the deviations, $O_i - E_i$, and see if there is a pattern.)

Month	Number of Suicides	Days/Month
Jan.	1867	31
Feb.	1789	28
Mar.	1944	31
Apr.	2094	30
May	2097	31
June	1981	30
July	1887	31
Aug.	2024	31
Sept.	1928	30
Oct.	2032	31
Nov.	1978	30
Dec.	1859	31