COLUMBIA UNIVERSITY
Statistical Inference
Practice problems

Prof. Marco Avella
TA: George Chu, Joe Suk
June 24, 2020

Question 1

Suppose $X_1, \cdots, X_n$ is an i.i.d. random sample taken from an exponential distribution for which the value of the parameter θ is unknown, the prior distribution of θ is a specified gamma distribution, and θ must be estimated by using the squared error loss function.
Compute the Bayes estimator for the squared loss and show that the sequence of Bayes estimators for $n = 1, 2, \cdots$ is consistent for θ.

Question 2

Consider two distribution A and B with *unknown common mean* θ but with *known variance* σ_A^2 equal to four times σ_B^2. Let $X_1, \cdots, X_m$ be a sample of m random variables from distribution A, and $Y_1, \cdots, Y_n$ be a sample of n random variables from distribution B, assuming that all variables are independent.
Show that the estimator $\widehat{\theta} = \alpha \bar{X}_m + (1-\alpha)\bar{Y}_n$ is unbiased and, for fixed values of m and n, find the value of α that minimizes the variance of $\widehat{\theta}$?

Question 3

Suppose that $X_1, \cdots, X_n$ form an i.i.d. random sample from a normal distribution with mean 0 and unknown variance $\sigma^2 > 0$, and assume that one is interested in estimating $\theta = \log \sigma$.
Compute the maximum likelihood estimator for θ and the Fisher Information $I_n(\theta)$.

Question 4

Suppose that $X_1, \cdots, X_n$ form an i.i.d. random sample from a Gamma distribution where the value of β is known and the value of α is unknown.
Show that $T = \sum_{i=1}^{n} \log X_i$ is sufficient for α.

Question 5

Suppose that $X_1, \cdots, X_n$ form an i.i.d. random sample from the normal distribution with unknown mean and unknown variance, and let the random variable L denote the length of the shortest $1-\alpha$ confidence interval for the mean that can be constructed from the observed values in the sample.

Find an expression for $E(L^2)$ and compute its value for the sample size $n = 10$ and the confidence coefficient $\alpha = 0.05$. Furthermore, discuss the relationship between

1. the expected value of the squared length $E(L^2)$ and the sample size n (all other parameters being fixed),
2. the expected value of the squared length $E(L^2)$ and the confidence coefficient α (all other parameters being fixed).

Question 6

Suppose that $X_1, X_2, \ldots, X_n$ are independent random samples from a gamma distribution with parameter α unknown and β known.
Show that if n is large, the distribution of the MLE of α will be approximately a normal distribution with mean α and variance

$$\frac{\Gamma(\alpha)^2}{n\left(\Gamma(\alpha)\Gamma''(\alpha) - \Gamma'(\alpha)^2\right)}$$

Question 7

Suppose that $X_1, X_2, \ldots, X_8$ are independent random samples from a normal distribution with parameters μ and σ^2 unknown, and a t test at a given level of significance α_0 is to be carried out to test the following hypotheses:

$$\begin{aligned} H_0 : &\quad \mu \le \mu_0 \\ H_1 : &\quad \mu > \mu_0 \end{aligned}$$

Let $\pi(\mu, \sigma^2|\delta)$ denote the power function of this t test and assume that (μ_1, σ_1^2) and (μ_2, σ_2^2) are values of the parameters such that

$$\frac{\mu_1 - \mu_0}{\sigma_1} = \frac{\mu_2 - \mu_0}{\sigma_2}.$$

Suppose that the sample data are such that $\sum_{i=1}^{8} X_i = -11.2$ and $\sum_{i=1}^{8} X_i^2 = 43.7$. If $\mu_0 = 0$ and $\alpha_0 = 0.05$, should the null hypothesis be rejected or not? Furthermore, show that $\pi(\mu_1, \sigma_1^2|\delta) = \pi(\mu_2, \sigma_2^2|\delta)$.

Question 8

Let $(X_1, \ldots, X_m)$ be an independent random sample of m observations from a normal distribution for which both the mean μ_1 and the variance σ_1^2 are unknown, and $(Y_1, \ldots, Y_n)$ another independent random sample of n observations from another normal distribution for which both the mean μ_2 and the variance σ_2^2 are unknown. Suppose that we want to test the following pair of hypothesis at level α_0:

$$\begin{aligned} H_0 : &\quad \sigma_1^2 = \sigma_2^2 \\ H_1 : &\quad \sigma_1^2 \neq \sigma_2^2 \end{aligned}$$

Let V be test statistic

$$V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$$

and δ be the two-sided F test that rejects H_0 when either $V \leq c_1$ or $V \geq c_2$. Prove that the power function of δ is

$$\pi(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | \delta) = G_{m-1,n-1}\left(\frac{\sigma_2^2}{\sigma_1^1} c_1\right) + 1 - G_{m-1,n-1}\left(\frac{\sigma_2^2}{\sigma_1^1} c_2\right),$$

with G the c.d.f. of the F distribution with $m-1$ and $n-1$ degrees of freedom, and compute the values of c_1 and c_2 for a test at level $\alpha_0 = 0.05$ with $m = 11$ and $n = 21$.

Question 9

Consider a regression problem in which a patient's reaction Y to a new drug B is to be related to his reaction X to a standard drug A. Suppose that for each value x of X, the regression function is a polynomial of the form $E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2$. Suppose also that 20 pairs of observed values are as shown in the table below. Under the standard assumptions of the general linear model, determine the values of the M.L.E.'s $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\sigma}^2$ (round to 4 decimal places), and carry out a test of the following hypotheses at level $\alpha_0 = 0.05$:

$$\begin{aligned} H_0 : &\quad \beta_1 = 2 \\ H_1 : &\quad \beta_1 \neq 2 \end{aligned}$$

i	1	2	3	4	5	6	7	8	9	10
x	4.5	1.3	1.9	2.9	4.5	1	4.5	4.7	3.3	3.1
y	2.5	3	3.3	3.9	1.9	2.8	2.3	1.2	4	3.3
i	11	12	13	14	15	16	17	18	19	20
x_i	0.3	1	0.9	3.4	1.9	3.8	2.5	3.6	5	1.9
y_i	1.6	3.2	2.9	4.1	3.9	3.5	4.7	3.2	0.5	4

Table 1: Data for Question 9

Hints:

1. If

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

then

$$\left(X^\top X\right)^{-1} = \begin{bmatrix} 0.73237 & -0.56158 & 0.09019 \\ -0.56158 & 0.52066 & -0.09082 \\ 0.09019 & -0.09082 & 0.01663 \end{bmatrix} \quad \text{and} \quad X^\top Y = \begin{bmatrix} 59.800 \\ 157.810 \\ 505.533 \end{bmatrix}.$$

2. With X and Y as above, we also have that $\sum_{i=1}^{20} x_i = 56$, $\sum_{i=1}^{20} x_i^2 = 197.38$, $\sum_{i=1}^{20} y_i = 59.8$, $\sum_{i=1}^{20} y_i^2 = 200.48$, and $\left\|Y - X^\top \widehat{\beta}\right\|^2 = 2.33$.