Taras Bodnar
Department of Mathematics
Stockholm University
E-Mail: taras.bodnar@math.su.se

Theory of Statistical Inference

Re-exam, 2022/12/14

The only allowed aid is a pocket calculator provided by the department. The answers to the tasks should be clearly formulated and structured. All non-trivial steps need to be commented. The solutions should be given in English or Swedish.

The written exam is divided into two parts. The first part considers the most central of the course concepts and it is related to standard problems. The second part consists of problems that requires a higher level of understanding, the ability to generalize and to combine methods. Each part consists of three problems and will worth a maximum of 50 points. In order to receive grades A-E, a minimum of 35 points is required in the first part. The second part is only graded for students passing the first part. Given a minimum of 35 points in the first part, the final grade is determined by the sum of regular points in both parts of the exam and bonus points according to the following table:

Grade	A	B	C	D	E	F
Points	$\geq$90	(90-80]	(79-70]	(69-60]	< 60 and ≥ 35 in Part I	< 35 in Part I

Up to 10 bonus points (i.e., in addition to the ordinary 100 points) are given for the active participation in the problem sessions. A half of the bonus points will be used for the first part of the exam, while the second half of the bonus points will be used in the second part of the exam.

Part I:

Problem 1 [23P]

Suppose that we have an iid sample $X_{1:n} = (X_1, X_2, ..., X_n)$ from a Lomax distribution with density of X_i given by

$$f_{X_i}(x; \beta) = \beta(1+x)^{-(\beta+1)} \qquad \text{for} \qquad x \geq 0$$

where $\beta > 0$ is an unknown parameter.

(a) Derive the maximum likelihood estimate $\hat{\beta}_{ML}$ for β. **[4P]**

(b) Derive the ordinary Fisher information $I_{1:n}(\beta)$, the observed Fisher information $I_{1:n}(\hat{\beta}_{ML})$, and the expected Fisher information $J_{1:n}(\beta)$. **[3P]**

(c) Find a minimum sufficient statistic for β and explain your answer.**[3P]**

(d) Construct a 95% two-sided score confidence interval for β. Simplified the expression of the confidence interval as much as possible. **[3P]**

(e) Construct a 95% two-sided Wald confidence interval for β. Simplified the expression of the confidence interval as much as possible. **[3P]**

(f) Construct a 95% likelihood ratio confidence interval for β. Simplified the expression of the confidence interval as much as possible. **[3P]**

(g) Does the value of the parameter $\beta_0 = 2.5$ lie in each of the three constructed confidence intervals, when $n = 49$ and $\sum_{i=1}^{n} \log(1 + x_i) = 15.14$? **[4P]**

Hint: *Important quantiles of the standard normal distribution are:*

$z_{0.9}$	$z_{0.95}$	$z_{0.975}$	$z_{0.995}$
1.28	1.64	1.96	2.33

See Problem 4 for the important quantiles of the χ^2-distribution at various degrees of freedom.

Problem 2 [17P]

Let $X_{1:n} = (X_1, ..., X_n)$ denote a random sample from a Weibull distribution with density of X_i given by

$$f_{X_i}(x; \theta) = \theta k x^{k-1} \exp\left(-\theta x^k\right) \qquad \text{for} \qquad x \geq 0 \quad \text{and} \quad \theta > 0,$$

and known $k > 0$.

(a) Show that the conjugate prior for θ is given by the gamma distribution with hyperparameters $\alpha > 0$ and $\beta > 0$, that is

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta) \quad \text{for} \quad \theta > 0 \quad \text{and} \quad \alpha, \beta > 0$$

with prior mean $\mathbb{E}(\theta) = \frac{\alpha}{\beta}$. Determine the parameters in the corresponding posterior distribution. **[5P]**

(b) Compute the posterior mean of θ when the conjugate prior is used. **[2P]**

(c) Find the expression of the Jeffreys prior for θ and compute the corresponding posterior distribution. **[6P]**

(d) Calculate the posterior mean when the Jeffreys prior is used and compare it to the expression of the MLE estimator for θ obtained in the frequentist statistics. **[4P]**

Problem 3 [10P]

Define the score statistic in the case of a scalar parameter. Derive the asymptotic distribution of the score statistic.

Part II:

Problem 4 [17P]

Let $X_{1:n_1} = (X_1, ..., X_{n_1})$ denote a random sample from an inverse gamma distribution with density of X_i, $i = 1, ..., n_1$, given by

$$f_{X_i}(x;\theta_1) = \frac{\theta_1^k}{\Gamma(k)} x^{-(k+1)} \exp\left(-\frac{\theta_1}{x}\right) \quad \text{for} \quad x > 0 \quad \text{and} \quad k, \theta_1 > 0$$

and let $X_{n_1+1:n_1+n_2} = (X_{n_1+1}, ..., X_{n_1+n_2})$ denote a random sample from an inverse gamma distribution with density of X_j, $j = n_1+1, ..., n_1+n_2$, given by

$$f_{X_j}(x;\theta_2) = \frac{\theta_2^k}{\Gamma(k)} x^{-(k+1)} \exp\left(-\frac{\theta_2}{x}\right) \quad \text{for} \quad x > 0 \quad \text{and} \quad k, \theta_2 > 0$$

Assume that $X_{1:n_1}$ and $X_{n_1+1:n_1+n_2}$ are independent and let k be known.

The aim is to test the null hypothesis:

$$H_0: \ \theta_1 = \theta_2. \tag{1}$$

(a) Derive the generalized likelihood ratio statistic for testing H_0 in (1). Simplified the expression of the test statistic as much as possible. **[12P]**

(b) Determine the distribution of the test statistics derived in part (a). **[1P]**

(c) Perform the generalized likelihood ratio test at significance level of 5% when $k = 7$, $n_1 = 40$, $n_2 = 50$, $\sum_{i=1}^{n_1} x_i^{-1} = 6.15$ and $\sum_{j=n_1+1}^{n_1+n_2} x_j^{-1} = 9.25$. **[4P]**

Hint: *Important quantiles of the χ^2-distribution at various degrees of freedom are:*

d	1	2	3	4	5
$\chi^2_{0.9}(\text{df} = d)$	2.71	4.61	6.25	7.78	9.24
$\chi^2_{0.95}(\text{df} = d)$	3.84	5.99	7.81	9.49	11.07
$\chi^2_{0.975}(\text{df} = d)$	5.02	7.38	9.35	11.14	12.83

Problem 5 [18P]

Let $\mathbf{X} = (X_1, X_2, X_3, X_4)^\top$ be multinomially distributed with probability mass function expressed as

$$\mathbb{P}(\mathbf{X} = \mathbf{x}; \pi_1, \pi_2, \pi_3) = \frac{1}{x_1! x_2! x_3! x_4!} \pi_1^{x_1} \pi_2^{x_2} \pi_3^{x_3} (1 - \pi_1 - \pi_2 - \pi_3)^{x_4}$$

for $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$, $x_1, x_2, x_3, x_4 = 0, 1, 2, ...,$ with $x_1 + x_2 + x_3 + x_4 = n$, n is known. Furthermore, it holds that $\mathbb{E}(X_i) = n\pi_i$ for $i = 1, 2, 3, 4$.

(a) Derive the maximum likelihood estimator $\hat{\boldsymbol{\pi}}_{ML}$ for $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)^\top$. **[4P]**

(b) Compute the expected Fisher information $\mathbf{J}(\boldsymbol{\pi})$. **[4P]**

(c) Construct a 90% Wald confidence region for $\boldsymbol{\pi}$. Simplified the expression of the confidence region as much as possible.**[4P]**

(d) Construct a 90% likelihood ratio confidence region for $\boldsymbol{\pi}$. Simplified the expression of the confidence region as much as possible.**[2P]**

(e) Does the parameter vector $\boldsymbol{\pi}_0 = (0.1, 0.3, 0.4)^\top$ lie in each of the two constructed confidence regions, when $n = 36$, $x_1 = 4$, $x_2 = 10$, $x_3 = 10$ and $x_4 = 12$? **[4P]**

Hint: *Important quantiles of the χ^2-distribution at various degrees of freedom are:*

d	1	2	3	4	5
$\chi^2_{0.9}(\text{df} = d)$	2.71	4.61	6.25	7.78	9.24
$\chi^2_{0.95}(\text{df} = d)$	3.84	5.99	7.81	9.49	11.07
$\chi^2_{0.975}(\text{df} = d)$	5.02	7.38	9.35	11.14	12.83

Problem 6 [15P]

Formulate the statement of the Cramér-Rao lower bound and prove it.