UPPSALA UNIVERSITY
Department of Mathematics
Silvelyn Zwanzig

MATHEMATICAL STATISTICS
Inference II 1MS037
Dec 15, 2015

Tillåtna hjälpmedel: pocket calculator, two pages with handwritten notes, dictionary

Time: 5 hours. For a pass (mark 3) the requirement is at least 18 points. For the mark 4, 25-31 points are necessary. For an excellent test (mark 5) the requirement is at least 32 points. Every problem is worth 5 points. For the international ECTS the following main rules are valid: A: 36-40 points, B: 28-35 points, C: 23-27 points, D: 20-22 points, E: 18-19 points.

OBS: *Please explain your approach and write down your arguments. Solutions without any explanation will not be accepted!!!*

As Cecilia Linroth said Pareto distributions are important for modeling of insurance problems.

Suppose we have an i.i.d sample $X_1, \ldots, X_n$ from a r.v. X Pareto distributed with density

$$f(x;\theta) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & \text{for} \quad x \geq x_m \\ 0 & \text{else} \end{cases}, \ \theta = (\alpha, x_m) \in (0, \infty)^2 . \tag{1}$$

For $\alpha > 1$ it holds $E_\theta X = \frac{\alpha x_m}{\alpha - 1}$, for $\alpha > 2$ it holds $Var_\theta X = \left(\frac{x_m}{\alpha-1}\right)^2 \frac{\alpha}{\alpha-2}$.

Furthermore we know that

$$Y = \log\left(\frac{X}{x_m}\right) \sim Exp(\alpha), \text{ with } EY = \frac{1}{\alpha}, VarY = \frac{1}{\alpha^2}$$

and for independent random variables Y, Z it holds

$$Y \sim Exp(\alpha),\ Z \sim Exp(\mu) \text{ then } \min(Y, Z) \sim Exp(\alpha + \mu).$$

1. Consider an i.i.d sample $X_1, \ldots, X_n$ from a r.v. X Pareto distributed with $\theta = (\alpha, 1)$, $\alpha > 2$.
   (a) Is this family of Pareto distributions an exponential family?

(b) If it is an exponential family derive a natural parameter and a sufficient statistic.

(c) Find a parametrization $\gamma = g(\theta)$ and its efficient estimator $\widehat{\gamma}$.

(d) Calculate the Rao Cramer bound for γ .

2. Consider an i.i.d sample $X_1, \ldots, X_n$ from a r.v. X Pareto distributed with $\theta = (\alpha, 1),\ \alpha > 1$.

   (a) Derive the log likelihood function and the score function.

   (b) Calculate the Fisher Information.

   (c) Derive the maximum likelihood estimator for α.

   (d) Which properties asses the maximum likelihood estimator?

   (e) Compare it with the results of Problem 1.

3. Consider an i.i.d sample $X_1, \ldots, X_n$ from a r.v. X Pareto distributed with $\theta = (3, x_m),\ x_m > 1$.

   (a) Derive the log likelihood function.

   (b) Derive a sufficient statistic.

   (c) Is this family of Pareto distributions an exponential family?

   (d) Derive the maximum likelihood $\widehat{x}_m$ of x_m.

   (e) Is this estimator $\widehat{x}_m$ unbiased? If not, derive a corrected estimator $\widetilde{x}_m$.

   (f) It is reasonable to discuss the efficiency of the estimator $\widetilde{x}_m$? Why? or Why not?

4. Consider an i.i.d sample $X_1, \ldots, X_n$ from a r.v. X Pareto distributed with $\theta = (\alpha, x_m),\ \alpha > 2,\ x_m > 0$. We are interested in the following test problem:

$$H_0 : \alpha = 3 \text{ versus } H_1 : \alpha > 3$$

(a) Is the null hypothesis a simple hypothesis?

(b) Derive an optimal test under the additional assumption that $x_m = 1$. (Use results of Problem 1.)

(c) Which properties has this optimal test?

5. Consider an i.i.d sample $X_1, \ldots, X_n$ from a r.v. $X \sim U(a,b),\ \theta = (a,b),\ a < b$. Remind $EX = \frac{a+b}{2}$, $Var(X) = \frac{1}{12}(b-a)^2$ and $EX^k = m_k = \frac{1}{k+1}\sum_{i=1}^{k} a^i b^{k-i}$

   (a) Derive the sufficient statistics.

   (b) Find moment estimators for θ.

   (c) Are these estimators unbiased?

   (d) Is it possible to improve these estimators? Explain the method and formulate the equations which should be solved.

6. Let define two discrete densities:

x	-10	-5	-1	1	5	10
$p_0(x)$	0.1	0.02	0.33	0.05	0.3	0.2
$p_1(x)$	0	0.03	0.17	0.3	0.3	0.2

   (a) Give the Neyman Pearson test for $\alpha = 0.05$.

   (b) Give the Neyman Pearson test for $\alpha = 0.1$.

   (c) Calculate the probabilities for the errors of second type for both tests.

   (d) Give an alternative alpha test for $\alpha = 0.05$.

   (e) Compare your test in (d) with the Neyman Pearson test in (a).

7. Consider an i.i.d sample $X_1, \ldots, X_n$ from a r.v. $X \sim N(\mu, \mu),\ \mu > 0$.

   We are interested in the following testing problem

$$H_0 : \mu = 1 \quad \text{versus} \quad H_1 : \mu > 1.$$

Let us apply the approach of a test: assessing evidence against H_0. The following test statistics are proposed.

$$T_1 = \sum_{i=1}^{n} X_i, \;\; T_2 = \sum_{i=1}^{n} X_i^2, \; T_3 = \sum_{i=1}^{n} (X_i - 1)^2, \; T_4 = \# \{i, X_i < 1\}$$

(a) Determine the null distribution of each of these test statistics. (Hint: Let $Z_k \sim N(a_k, 1)$ *i.i.d.* then $\sum_{k=1}^{n} Z_k^2$ is noncentral chi squared distributed with n degrees of freedom and with noncentrality parameter $\lambda = \sum_{k=1}^{n} a_k^2$)

(b) Define the p-value $p-value_k$ for each of these test statistics $T_k, k = 1, \ldots, 4$.

(c) Having in mind the theory of testing as decision problem, which test statistic you would recommend?

8. Suppose an i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_n)$ from $X \sim N(0, \sigma^2)$ with $\sigma^2 \in \{1, 2\}$. Consider the test problem

$$H_0 : \;\; \sigma^2 = 1 \text{ versus } H_1 : \;\; \sigma^2 = 2.$$

(a) Explain, what is the error of first type ?

(b) Explain, what is the error of second type?

(c) Derive the class of Neyman Pearson tests for this test problem.

(d) Derive the power function of the Neyman Pearson tests. Explain the connection to the error of first and second type.

(e) Sign (roughly!) the $(\alpha, \beta)-$ presentation for this test problem.

Good Luck! Lycka till!! Viel Glück!!!