UPPSALA UNIVERSITY
Department of Mathematics
Silvelyn Zwanzig

MATHEMATICAL STATISTICS
Inference Theory II, 5c, 1MS037
January 12, 2017

Permitted aids: pocket calculator, one hand-written sheet of formulae (2 pages), 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!!!*

1. A manufacturer of ball point pens randomly samples with replacement $2n$ units per working day from the daily production. During the weekend (Saturday, Sunday) only n pens are sampled.
   (a) Assume that the probability of a defect p is constant during the whole week. Formulate a statistical model for one week.
   (b) Belongs this model to an exponential family? Determine the natural parameter and the sufficient statistic.
   (c) Which distribution has the sufficient statistic?
   (d) Determine the Fisher information.
   (e) Determine the Cramer Rao bound for estimating p.

2. Consider an i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_n)$ from $Geo(p)$, with probability function $p(x;p) = pq^x$, where $q = 1 - p, 0 < \ p < 1$ and with expected value $EX = \frac{q}{p}$ and variance $Var(X) = \frac{q}{p^2}$.
   (a) Derive the likelihood function of the sample.
   (b) Derive the score function.
   (c) Derive the maximum likelihood estimator $\widehat{p}_{MLE}$ for p.
   (d) Is the $\widehat{p}_{MLE}$ unbiased?

(e) Is the $\widehat{p}_{MLE}$ efficient?

3. A manufacturer of ball point pens randomly samples with replacement n units per day from the daily production.

   (a) Formulate a statistical model for one week, where the probability for defective pens during the weekend production is 2 times higher than during the working days production.

   (b) Belongs this model to an exponential family? Determine the natural parameters and the sufficient statistics.

   (c) Determine the covariance matrix of the sufficient statistics.

   (d) Is it a strictly k-parametric exponential family? In case of "yes", determine k.

4. Consider an i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_n)$ with the density (Epanechnikov quadratic kernel)

$$f_\theta(x) = \frac{3}{4h}(1 - (\frac{x-a}{h})^2) I_{[a-h,a+h]}(x), \theta = (a, h) \tag{1}$$

   with $Ex = a, Var(X) = \frac{1}{5}h^2$.

   (a) Does the distribution in (1) belong to an exponential family?

   (b) Derive moment estimators $\widehat{a}_{MME}, \widehat{h}_{MME}$ for a and h.

   (c) Is the moment estimator $\widehat{a}_{MME}$ for a unbiased?

   (d) Calculate the variance of $\widehat{a}_{MME}$.

   (e) Is the moment estimator $\widehat{a}_{MME}$ BUE?

5. Consider an i.i.d. sample $X = (X_1, ..., X_n)$ from $N(\mu, \mu)$, with density

$$f(x) = \frac{1}{\sqrt{2\pi\mu}} \exp(-\frac{1}{2\mu}(x-\mu)^2), \mu > 0.$$

   We are interested in the test problem

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

   Let us apply the approach of a test: assessing evidence.

(a) Is the null hypothesis simple? Give the model under H_0.

(b) Propose an appropriate test statistic.

(c) Determine the null distribution. Is the null distribution symmetric?

(d) Define the $p-value$.

(e) Which distribution has the $p-value$ under H_0? Why?

(f) Suppose the p-values takes the value 0.0001. What is your conclusion?

6. Consider the testing problem

$$H_0 : p_0(x) \text{ versus } H_1 : p_1(x)$$

where the distributions are given in the following table.

x	2	3	4	5	6	7	8
$p_0(x)$	0.1	0.02	0.33	0.05	0.2	0.1	0.2
$p_1(x)$	0	0.03	0.17	0.3	0.2	0.2	0.1

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

(b) Calculate the probability of the error of second type.

(c) Give an alternative alpha test for $\alpha = 0.1$.

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

7. Suppose an i.i.d. sample $\mathbf{X} = (X_1, .., X_n)$ from $X \sim N(\mu, 1)$ with $\mu \in \{-1, 1\}$.

Consider the test problem $H_0 : \ \mu = 1$ versus $H_1 : \ \mu = -1$

(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 relation between the error of first and second type.

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

(Hint:

x	0	0.67	1	1.64	1.67
$\Phi(x)$	0.5	0.75	0.8413	0.95	0.9525

)

8. Consider an i.i.d. sample $\mathbf{X} = (X_1, .., X_n)$ from $X \sim Beta(\alpha, \beta)$ with density

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha}(1-x)^{\beta-1} I_{[0,1]}(x), \ \alpha > 0, \beta > 0$$

with $EX = \frac{\alpha}{\alpha+\beta}$, mode= $\frac{\alpha-1}{\alpha+\beta-2}$, $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta-2)}$.

(a) Belongs the distribution of $\mathbf{X}$ to an exponential family? Derive the natural parameters and the sufficient statistics.

(b) When is the $Beta(\alpha, \beta)$ distribution symmetric?

(c) Consider the two sided test problem:

$$H_0: \ \alpha = 2, \beta = 2 \ versus \ H_0: \ \beta = 2, \ \alpha > 2 \qquad (2)$$

Derive the UMP $\alpha - test$.

(d) Give the properties of an UMP $\alpha - test$.