Politecnico di Milano
STATISTICS (079086), 2009/2010, Prof. A.Barchielli
Problem set n. 8

27th May 2010

Exercise 1 Let $X_1, \ldots, X_n$ be a random sample from a population with density

$$f(x, \theta) = 2\sqrt{\frac{\theta}{\pi}} e^{-\theta x^2} \mathbb{I}_{(0,\infty)}(x)$$

where θ is a positive unknown parameter.

1. Find the method of moments estimator $\overline{\theta}_n$ of θ. $[\overline{\theta}_n = \frac{1}{\pi \overline{X}^2}]$
2. Find the MLE estimator $\widehat{\theta}_n$ of θ. $[\widehat{\theta}_n = \frac{1}{2/n \sum_{i=1}^n X_i^2}]$

Exercise 2 In a quality control of a random sample of 100 mechanical pieces produced by a machine, 25 are found defective.

a) Exploiting the Central Limit Theorem, compute a bilateral confidence interval at level 90% for the proportion p of defective pieces produces by the machine [The asymptotic confidence interval based on the CLT is given by $\bar{X} \pm z_{\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}$; with the data above we have $0,25 \pm 0,071$]

b) Verify that $\hat{p} = \bar{X}$ is the MLE for p.

c) Compute the Fisher Information for X and hence shows that the confidence interval obtained from the asymptotic distribution of the MLE $\hat{p}$ coincides with the one based on the CLT [$I(p) = 1/(p(1-p))$; the confidence interval based on the asymptotic distribution of the MLE $\hat{p}$ is given by $\hat{p} \pm z_{\alpha/2}\sqrt{1/(nI(\hat{p}))} \equiv \bar{X} \pm z_{\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}$]

d) Can we reject the null hypothesis $H_0 : p = 0.2$ vs the alternative $H_1 : p > 0.2$ at level 5%? *[Thanks to the CLT we can approximate the distribution of $(\bar{X} - p)/\sqrt{p(1-p)/n}$ by a standard Gaussian distribution, where $p = p_0 = 0.2$ if H_0 is true. We thus reject H_0 if $\bar{X} > p_0 + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}$. With the data above, we do not reject H_0 at level 5%.]*

Exercise 3 To decide wether the running shoes produced by firm A are better than the running shoes produced by firm B, 10 runners are asked to run two 10km races, in two following weeks. Each runner competes one race wearing the shoes produced by firm A and the other wearing the shoes produced by firm B (but he does not know which firm produced the shoes he is wearing). Which shoes the runner wear in the first race is chosen randomly. The running times (in minutes) are the following:

Runner:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Shoes A	31.23	29.33	30.50	32.20	33.08	31.52	30.68	31.05	33.00	29.67
Shoes B	32.02	28.98	30.63	32.67	32.95	31.53	30.83	31.10	33.12	29.50

(a) Can you conclude that shoes A are better than shoes B, at significance level 5%? Specify if you need further hypotheses to perform the test. *[Let D_i be the difference of the running times obtained with shoes A and with shoes B for the i-th runner (-0.79, 0.35, -0.13, -0.47, 0.13, -0.01, -0.15, -0.05, -0.12, 0.17). We assume that $D_1, \ldots, D_{10}$ is a random sample from a normal distribution. We have $\overline{D} = -0.107$, $S_D = 0.326$. We want to carry out the test for $H_0 : \mu_D = 0$ vs $H_1 : \mu_D < 0$. We reject H_0 if $\overline{D} < -t_{0.05,9} \cdot (S_D/\sqrt{10})$. With the data above we do not reject H_0 at level 5%.]*

(b) Computer the p-value of the test. *[*$0.1 < p - value < 0.25$*]*

(c) Construct a confidence interval at level 95% for the mean difference of the running times using shoes A and B. *[*(-0.107 ± 0.233)*]*

Exercise 4** We want to make statistical inference on the probability p that a student, who is regularly registered to an exam session, do not actually come to sit the examination. For this reason, we look at the number of students registered at the exam session of 25/06/07 for the course of Statistica/Statistics and at the number of students who actually came to sit the examination: exactly 27 of the 128 students, who were regularly registered, did not come to sit the exam.

1. Verify the null hypothesis $H_0 : p \leq 0.25$ versus $H_1 : p > 0.25$, with an asymptotical test of level $\alpha = 3\%$.
2. Determine the power of the test at point 1 when the true value of p is 0.35.
3. Construct an (asymptotical) confidence upper bound for p with confidence level $\gamma = 95\%$.

The number of students registered to the next examination session for the course of Statistica/Statistics is 100. The professor thus prepares 100 copies of the examination paper.

4. On the basis of the results of point 3, with what level of (approximate) confidence we can assert that the mean number of unused copies will be less or equal to 27?

Solution.

1. We reject $H_0 : p \leq 0.25$ and accept $H_1 : p > 0.25$ if

$$\hat{p} \geq 0.25 + z_{97\%}\sqrt{\frac{0.25 \times (1 - 0.25)}{128}} \simeq 0.322,$$

where $\hat{p} = 27/128 \simeq 0.211$; since we have a large sample ($n = 128$) and $n\hat{p} > 5$, we can say that the approximate significance level of this test is $\alpha = 3\%$. Being $\hat{p} = 0.211 < 0.322$, we accept H_0, at approximate level 3%.

2. $\pi(0.35) = P_{0.35}(\hat{p} \geq 0.322) \simeq 1 - \Phi\left(\frac{0.322 - 0.35}{\sqrt{\frac{0.35\times(1-0.35)}{128}}}\right) \simeq 1 - \Phi(-0.66) \simeq 0.74537.$

3. An asymptotical confidence UB for p of level γ is given by

$$\mathrm{UB}(p) = \hat{p} + z_\gamma\sqrt{\frac{\hat{p} \times (1 - \hat{p})}{128}}.$$

With the above data and $\gamma = 95\%$ we get $\mathrm{UB}(p) \simeq 0.27$.

4. We expect that, in average, $100 \times p$ students will not come to sit the examination. We are here asked what is the confidence level of an UB, for $100 \times p$, which takes the value 27 with our data. Exploiting the result of the previous point we get that the confidence level of this UB is 95%.

Exercise 5** Let $X_1, \ldots, X_n$ be a random sample from a population with distribution

$$f(x, \theta) = \theta(1 - x)^{\theta-1}\mathbb{I}_{(0,1)}(x)$$

where θ is a positive parameter.

(a) Find the MLE for θ *[*$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} log(1-x_i)}$*]*

(b) Construct an asymptotic confidence interval for θ, using the asymptotic distribution of the MLE. Hint: $I(\theta) = 1/\theta^2$. *[The confidence interval based on the asymptotic distribution of the MLE* $\hat{\theta}$ *is given by* $\hat{\theta} \pm z_{\alpha/2}\sqrt{1/(nI(\hat{\theta}))} \equiv -\frac{n}{\sum_{i=1}^{n} log(1-x_i)}(1 \pm z_{\alpha/2}/\sqrt{n})$*]*