30 October 2008, 14:00 – 17:00

Rijksuniversiteit Groningen
Statistiek

Tentamen

1. **Maximum likelihood.** Let $X_1, ..., X_n$ be independently Poisson distributed with parameter λ, i.e.

$$p_{X_i}(x) = e^{-\lambda} \frac{x^\lambda}{x!}, \quad x = 0, 1, 2, \ldots$$

   (a) Derive the loglikelihood of λ. (**5 Marks**)
   (b) Find the Cramer-Rao lower-bound for an *unbiased* estimate of λ. (**10 Marks**)
   (c) Find the maximum likelihood estimate of λ? (**10 Marks**)
   (d) Is the MLE unbiased and does it attain the Cramer-Rao lowerbound? (**10 Marks**)
   (e) If $n = 100$ and $\bar{X} = 2$, then find the approximate 95% confidence interval for λ based on approximate normality of the maximum likelihood estimator. Use the fact that the 97.5% standard normal quantile is $z_{0.975} = 1.96$. (**5 Marks**)

2. **Rao-Blackwell.** The Rao-Blackwell theorem is given as follows: if $\hat{\theta}$ is an unbiased estimate of θ and T is a sufficient statistic of θ, then

$$\hat{\theta}^* = E(\hat{\theta}|T)$$

   is an unbiased estimate of θ with

$$V(\hat{\theta}^*) \leq V(\hat{\theta})$$

   (a) Prove that $\hat{\theta}^*$ is unbiased. (**5 Marks**)
   (b) Prove that $V(\hat{\theta}^*) \leq V(\hat{\theta})$. (**5 Marks**)
   (c) Let $X_1, \ldots, X_n$ be independently Bernoulli distributed with parameter θ. Show that

$$T = \sum_{i=1}^{n} X_i$$

   is a sufficient statistic for θ. (**5 Marks**)
   (d) Let $n = 3$. Consider the following estimator of θ,

$$\hat{\theta} = \frac{1}{6} X_1 + \frac{2}{6} X_2 + \frac{3}{6} X_3.$$

   Show that $\hat{\theta}$ is unbiased. (**5 Marks**)
   (e) Derive the estimator

$$\hat{\theta}^* = E(\hat{\theta}|T)$$

   for the above estimator $\hat{\theta}$ and the above sufficient statistic T. (**5 Marks**)

3. **Hypothesis testing.** Somebody wants to test whether or not there are an equal number of male and female students inclined to studying Mathematics in the US. In 100 universities he samples at random 4 mathematics students. The data are as follows:

Number of female maths students	Frequency
0	21
1	24
2	10
3	24
4	21

(a) Let X_i be the number of female students in university i, where $i = 1, \ldots, 100$. Assuming that $X_i \overset{\text{i.i.d.}}{\sim}$ Binomial$(4, p)$, test whether or not $p = 0.5$. Use $T = \sum_{i=1}^{100} X_i$ as test-statistic and the fact that $T \sim$ Binomial$(400, p)$. Write down the hypotheses, calculate the p-value and interpret the result using significance level $\alpha = 0.05$. **(10 Marks)**

(b) Somebody else claims that the test in the previous question is irrelevant, because it might be that the data do not support the Binomial model. Do a goodness-of-fit test to determine whether or not the Binomial model is correct. Write down the hypotheses, calculate the test-statistic, find the critical region using a significance level $\alpha = 0.05$ and interpret the results. Use the following table with some quantiles of chi-squared distribution in determining the Critical Region:

df	$\chi^2_{0.05}$	$\chi^2_{0.95}$
3	0.35	7.81
4	0.71	9.49
5	1.15	11.07
99	77.05	123.23
100	77.93	124.34

(10 Marks)

4. We try to fit the following model,

$$y_i = \alpha x_{i1} + \beta x_{i2} + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$. The data is given as follows

	y	x1	x2
1	15.08	3.70	6.00
2	16.19	5.40	6.50
3	16.82	3.30	6.20
4	16.96	8.20	4.40
5	21.68	5.70	8.00
6	15.94	3.40	5.80

(a) Use the data to estimate α and β. You can use the fact that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

(10 Marks)

(b) Using the fact that the estimated standard deviation $\hat{\sigma} = 1.3$, calculate the standard errors of $\hat{\alpha}$ and $\hat{\beta}$. **(5 Marks)**