LINKÖPINGS UNIVERSITET
Matematiska institutionen
Matematisk statistik

Kurskod: TAMS24
Provkod: TEN1
Datum: 2018-10-22
Institution: MAI

Exam in Statistics

TAMS24/TEN1 2018-10-22 8–12

You are permitted to bring:

- a calculator (no computer);
- Formel- och tabellsamling i matematisk statistik (from MAI);
- Formel- och tabellsamling i matematisk statistik, TAMS65;
- TAMS24: Notations and Formulas (by Xiangfeng Yang).

Grading: 8-11 points giving grade 3; 11.5-14.5 points giving grade 4; 15-18 points giving grade 5. Your solutions need to be complete, well motivated, carefully written and concluded by a clear answer. Be careful to show what is random and what is not. Assumptions you make need to be explicit. The exercises are in number order.

Solutions can be found on the homepage a couple of hours after the finished exam.

1. Let $f(x\,;\,\theta)$ be the density function for the Gamma distribution (with unknown parameter θ),
$$f(x\,;\,\theta) = \frac{\theta(\theta x)^{v-1}e^{-\theta x}}{\Gamma(v)},$$
where $x > 0$ and $v > 0$.

   (a) Find the ML-estimate $\widehat{\theta}$ for θ. (2p)

   (b) Show that the estimate $1/\widehat{\theta}$ is an unbiased estimate of $1/\theta$. *Hint: If X is Gamma distributed like above, then* $\mathrm{E}(X) = v/\theta$. (1p)

2. During one hundred days, Lena and Sture has been collecting used syringes at a central cemetery that's used by the local drug addicts. They've written down the number found each day and the corresponding frequencies can be found below.

Number found	0	1	2	3	4
Frequency	38	33	26	2	1

   Lena believes the data is Po(1)-distributed (Poisson distributed with expectation 1), but Sture disagrees. Use a suitable test to see if Sture is correct in rejecting the hypothesis at the 1%-level. (2p)

3. Belinda is a hobby chemist experimenting with organic peroxides. She's trying to synthesize hexamethylene triperoxide diamine (HMTD) using two slightly different methods. Method one uses citric acid and method two uses glacial acetic acid. Belinda is interested in if the yield is better when using citric acid, since this method produces more heat and requires more attention. She's done 10 experiments using each method and the yield (calculated as a fraction of the amount of hexamethylene diamine used) rounded off can be seen in the table below. We assume that the measurements are normally distributed and that different batches are independent. We also assume that the variance is the same for both methods.

	Yield										$\overline{x}$	s
Citric acid	55	36	55	64	53	58	55	45	51	40	51.2	8.5088
Acetic acid	50	38	39	40	27	54	47	40	53	35	42.3	8.5641

(a) Perform a test using at least one confidence interval to see if the method using citric acid produces a better yield with confidence level 90%. (2p)

(b) Perform a test to check if it was reasonable to assume that the variances were equal (with significance level 5%). (1p)

4. In MATLAB there is a command `randn(n)` for creating square matrices with random elements from a normal distribution. When testing to generate a growing sequence of matrices and timing the operation for each matrix (for example by using the command `cputime`), the following execution times were obtained.

x	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
y	0.03	0.09	0.21	0.37	0.58	0.84	1.15	1.49	1.90	2.32	2.83	3.36	3.93	4.57

The number x is the number of rows times divided by one thousand (so $x = 2$ means 2000 rows) and y is the execution time (the time it took to generate the matrix in question). Since the number of elements in the matrix grows quadraticly, the following model seems reasonable

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon,$$

where $\epsilon \sim N(0, \sigma)$ and different measurements are assumed to be independent.

(a) At the 1% level, can you reject the hypothesis that $\beta_1 = 0$? Interpretation? (1p)

(b) Find a prediction interval with degree of confidence 99% for the execution time if $x = 11.5$. (2p)

(c) Estimate the execution time if $x = 20$ using a reasonable estimate based on the model. Is there a problem with using the model for "large" x? (1p)

A helpful mathematician has done the following calculations for you.

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	$-0.2198 \cdot 10^{-3}$	$6.7413 \cdot 10^{-3}$
1	$0.5220 \cdot 10^{-3}$	$2.0676 \cdot 10^{-3}$
2	$23.2692 \cdot 10^{-3}$	$0.1341 \cdot 10^{-3}$

Analysis of variance

	Degrees of freedom	Square sum
REGR	2	29.3755
RES	11	$5.7582 \cdot 10^{-4}$
TOT	13	29.3761

$$(X^T X)^{-1} = \begin{pmatrix} 868.1319 & -239.0110 & 13.7363 \\ -239.0110 & 81.6621 & -5.1511 \\ 13.7363 & -5.1511 & 0.3434 \end{pmatrix} \cdot 10^{-3}.$$

5. Conny is ordering a bunch of seeds from the Netherlands, planning to do some "farming." The guy selling them claims that out of 10 seeds, at least 8 will grow pretty much no matter how much abuse you throw at them. Conny believes this and orders 15 seeds and plants them. After a while, he finds that only 10 has grown. Conny – who considers himself a decent enough statistician – forms the hypothesis test $H_0 : p = 0.8$ versus $H_1 : p < 0.8$ and assumes that the seeding of the seeds is independent.

   (a) Carry out the test at the significance level 5%. (1p)

   (b) What is the power of the test at $p = 0.65$? (1p)

   (c) How many seeds would Conny need to order to obtain a test with a power of 90% at $p = 0.65$ (using the same level of significane)? (2p)

6. Suppose that $\boldsymbol{Y} = (Y_1\ Y_2\ \cdots\ Y_k)^T$, where the components Y_i are normally distributed and independent with the same variance. If $A, B \in \mathbf{R}^{k\times k}$ are constant symmetric matrices such that $A^2 = A$ and $B^2 = B$, prove that $\boldsymbol{Y}^T A\boldsymbol{Y}$ and $\boldsymbol{Y}^T B\boldsymbol{Y}$ are independent if $AB = 0$. (2p)