TESTAT 2 Statistics and probability theory

SS 2006

Prof. Dr. M.H. Faber

ETH Zurich

Thursday, 6 July 2006 08:00 – 09:45

Surname:	
Name:	
Stud. Nr.:	
Course of studies:	

Date and duration:

Thursday, 6 July 2006 Start: 8:00 Duration: 90 minutes

Aids:

- Non-programmable pocket calculator
- No communication medium (e.g. cell phone)
- 8 pages (DIN A4 one-sided) original handwritten summary

Hints:

- Pease control first, if you have received all the materials (listed under: Contents).
- Please locate your Legi on your desk at the outside side.
- Please write your **name on every sheet of paper**, at the bottom left side.
- Use only the provided sheets of paper.
- **Do not open** the paper fastener.
- Place **all materials** back in the envelope after the examination and do not leave your seats until all examination papers are collected.
- Leave the examination room as quietly as possible in respect to those who are still writing.

Contents

- General information, exercices and tables (14 pages).
- 1 sheet of paper (checkered).
- Tables provided:
 - Standard normal distribution function,
 - Chi square distribution function,
 - Kolmogorov-Smirnov test statistic.

Part 1: 'Multiple Choice'

In answering the following multiple choice questions it should be noted that for some of the questions several answers may be correct. Tick the correct alternatives in every question as:

A question is considered as correctly answered if **ALL** and **ONLY** the correct alternatives are ticked.

1.1 According to the central limit theorem which of the following statement(s) holds?

The probability distribution function of the sum of a number of independent random variables approaches the normal distribution as the number of the variables increases.

The probability distribution function of the product of a number of independent random variables approaches the normal distribution as the number of the variables increases.

None of the above.

1.2 Consider a number of log-normal distributed and independent random variables. Which of the following statement(s) hold?

The probability distribution function of the sum of the random variables approaches the log-normal distribution as the number of the variables increases.

The probability distribution function of the sum of the random variables approaches the normal distribution as the number of the variables increases.

None of the above.

1.3 From past experience it is known that the shear strength of soil can be described by a log-normal distribution. 15 samples of soil are taken from a site and an engineer wants to use the data in order to estimate the parameters of the log-normal distribution. The engineer:

may use a probability paper to estimate the parameters of the log-normal distribution.

may use the maximum likelihood method to estimate the parameters of the lognormal distribution.

may use the method of moments to estimate the parameters of the log-normal distribution.

None of the above.

1.4 An engineer tests the null hypothesis that the mean value of the concrete cover depth of a concrete structure corresponds to design assumptions. In a preliminary assessment a limited number of measurements of the concrete cover depth are made, and after performing the hypothesis test the engineer accepts the null hypothesis. After a few years, a comprehensive survey of the concrete cover depth is carried out, i.e. many measurements are made. The survey shows that the mean value of the concrete cover depth does not fulfill the design assumptions. Which of the following statement(s) is(are) correct?

In the preliminary survey the engineer has performed a Type I error.	
In the preliminary survey the engineer has performed a Type II error.	
In the preliminary survey the engineer has performed a Type I and a Type II error.	
None of the above.	

1.5 It is given that the operational life (until breakdown) *T* of a diesel engine follows an exponential distribution, $F_T(t) = 1 - e^{-\lambda t}$, with parameter λ and mean value, $\mu_T = 1/\lambda$, equal to 10 years. The probability that the engine breaks down within 2 years, after placed in operation, is equal to:

 $P(T \le 2 y ears) = 0.181.$

 $P(T \le 2 y ears) = 0.819$.

 $P(T \le 2 y ears) = 0.0067$.

None of the above.

1.6 In a mediterranean city there are on average 5 snowfalls a year. Assume that the occurrence of snowfalls follows a Poisson process. The number of snowfalls in *t* years, *X*, is described by the discrete probability distribution function $P(X = k) = \frac{(vt)^{k}}{k!}e^{-vt}$ and with annual mean rate *v*. Which of the following statement(s) is(are) correct?

The probability of exactly 5 snowfalls in the next year is equal to 0.175.	
The probability of exactly 5 snowfalls in the next year is equal to 1.	
The probability of no snowfall in the next year is equal to 0.774.	
The probability of no snowfall in the next year is equal to 0.0067.	\square

1.7 A material property is described by a normal distributed random variable X with mean μ_X and known standard deviation σ_X . The sample characteristic \overline{X} has properties that in principle depend on the number of measurements n. Which of the following equation(s) can be used to estimate a one sided confidence interval for the mean μ_X , at the α significance level?

$P\left(-k_{\alpha} < \frac{\overline{X} - \mu_{X}}{\sigma_{X} \frac{1}{\sqrt{n}}}\right) = 1 - \alpha$	
$P\left(-k_{\alpha} < \frac{\overline{X} - \mu_{X}}{\sigma_{X} \frac{1}{\sqrt{n}}}\right) = \alpha$	
$P\left(-k_{\alpha/2} < \frac{\overline{X} - \mu_X}{\sigma_X \frac{1}{\sqrt{n}}}\right) = 1 - \frac{\alpha}{2}$	

1.8 Consider a timber beam subjected to an annual maximum bending moment *L*. The bending strength of the beam *R* is modeled by a normal distributed random variable with mean $\mu_R = 30kNm$ and standard deviation $\sigma_R = 5kNm$ and the annual maximum bending moment is modeled by a normal distributed random variable with mean $\mu_L = 9kNm$ and standard deviation $\sigma_L = 2kNm$. It is assumed that *R* and *L* are independent. The timber beam fails when the applied moment exceeds the bending strength. Which of the following statement(s) is(are) correct? (*HINT*: If *M* represents the safety margin, i.e., M = R - L then the probability of failure is given by: $P_F = P(M \le 0) = \Phi(-\beta)$. $\Phi(\cdot)$ is the probability distribution function of the standard normal distribution, and β is the so-called reliability index given as: $\beta = \frac{\mu_M}{\sigma_M}$, where μ_M and σ_M

are the mean and standard deviation of the safety margin M respectively.)

The reliability index of the timber beam corresponding to a one year reference period is equal to 3.9.

The annual probability of failure of the timber beam is equal to $1.0 \cdot 10^{-4}$.

The reliability index of the timber beam corresponding to a one year reference period is equal to 4.1.

The annual probability of failure of the timber beam is equal to $4.8 \cdot 10^{-5}$.

1.9 The Kolomogorov-Smirnov test is designed especially for:

discrete probability distribution functions.

continuous probability distribution functions.

None of the above.

1.10 An engineer wants to examine and compare the suitability of two distribution function model alternatives for a random material property. Measurements are taken of the material property. The engineer uses the two model alternatives to calculate the Chi-square sample statistics and the corresponding sample likelihoods. The results are given in the following table:

Model	Degrees of freedom	Chi-square sample statistic	Sample likelihood	
1	2	0.410	0.815	
2	1	0.407	0.524	

Which of the following statement(s) is(are) correct?

The engineer may accept model 1 at the 5% significance level.

The engineer may accept model 2 at the 5% significance level.

Model 1 is more suitable than model 2.

None of the above.

1.11 Based on experience it is known that the concrete compressive strength may be modeled by a normal random variable *X* with mean value $\mu_X = 30MPa$ and standard deviation $\sigma_X = 5MPa$. The compressive strengths of 20 concrete cylinders are measured. An engineer wants to test the null hypothesis H_o that *X* follows a normal distribution with the above given parameters. He/she carries out a Chi-square test by dividing the samples into 3 intervals. He/she calculates a Chi-square sample statistic equal to $\varepsilon_m^2 = 0.41$. Which of the following statement(s) is(are) correct?

The engineer can accept the null hypothesis H_o at the 5% significance level.	
The engineer can reject the null hypothesis H_o at the 5% significance level.	
The engineer can accept the null hypothesis H_o at the 10% significance level.	
None of the above.	

1.12 The Maximum Likelihood Method (MLM) enables engineers to estimate the distribution parameters of a random variable on the basis of data. Which of the following statement(s) is(are) correct?

The MLM provides point estimates of the distribution parameters.

The MLM provides information about the uncertainty associated with the estimated parameters.

The MLM provides no information about the uncertainty associated with the estimated parameters.

None of the above.

1.13 Consider *n* independent standard normal random variables $X_i, (i = 1, 2..., n)$.

For a random variable $Y = \sum_{i=1}^{n} X_i^2$ which of the following statement(s) is(are) correct?

If n = 3 then the random variable *Y* follows the Chi-square distribution with 3 degrees of freedom.

The random variable Y approaches the normal distribution as n increases.

If n = 3 then the random variable Y follows the Chi distribution with 3 degrees of freedom.

None of the above.

Part 2: 'Exercise'

Table 1 provides the annual maximum values of precipitation per hour (rainfall), x_i , i = 1, 2, ..., n, observed in the area of Zurich during the last n = 9 years. The data are given in ascending order.

i	Annual maximum precipitation per hour x_i (mm/hour)
1	73
2	85
3	93
4	98
5	100
6	108
7	115
8	121
9	130

Table	1. Data	of the	annual	maximum	preci	nitation	ner hour
I UDIO	T. Dulu		annaan	maximum	proor	pilation	por nour.

The modeling of the annual maximum precipitation per hour is of interest for the planning of safety measures. An environmental engineer believes that a Gumbel distributed random variable X is suitable for the modeling of the annual maximum precipitation per hour. Answer the following questions. (*HINT:* Question e. can be answered independently from questions a. to d.)

a. Accepting that the Gumbel distribution is suitable for the modeling of the random variable X, use the method of moments to calculate the distribution parameters α and u. Table 2 provides the expression of the Gumbel distribution function and its first two moments.

	Gumbel distribution		
Probability distribution function $F_x(x) = \exp(-\exp(-\alpha(x-u)))$			
Manager	First order: $u + \frac{0.5772}{\alpha}$		
Moments	Second order: $\left(u + \frac{0.5772}{\alpha}\right)^2 + \frac{\pi^2}{6\alpha^2}$		

Table 2: The Gumbel distribution function and its first two moments.

b. Plot the data given in Table 1 on the provided probability paper (Figure 1). You may use Table 3 for the necessary calculations. (*HINT:* The observed cumulative distribution function may be evaluated as $F_{x,o}(x_i) = \frac{i}{n+1}$, where *n* is the total number of observations).

Continued on the next page

Statistics and probability theory M.H.Faber, Swiss Federal Institute of Technology, ETH Zurich, Switzerland

i	Annual maximum precipitation per hour <i>x_i</i> (mm/hour)	$F_{X,o}(x_i) = \frac{i}{n+1}$	$-\ln(-\ln(F_{X,o}(x_i)))$		
1	73				
2	85				
3	93				
4	98				
5	100				
6	108				
7	115				
8	121				
9	130				





Figure 1: Probability paper for the Gumbel distribution.

- c. According to the resulting plot made in Figure 1, is the belief of the engineer, in regard to the choice of the Gumbel distribution function, correct?
- d. Assuming that the answer in c. is yes, obtain the Gumbel distribution parameters α and u from the probability paper. Compare the values with the ones calculated using the method of moments. What do you observe? Why is this occurring?

Continued on the next page

Carry out a Kolmogorov-Smirnov test, at the 5% significance level, of the e. hypothesis that the data in Table 1 are representative for a Gumbel distribution with parameters u=95 und $\alpha=0.05$. Use Table 4 for the necessary calculations. The sample statistic used in the Kolmogorov-Smirnov test is written as: $\varepsilon_{\max} = \max_{i=1}^{n} \varepsilon_i$ where $\varepsilon_i = |F_{X,o}(x_i) - F_{X,p}(x_i)|$. The observed cumulative distribution function is expressed as: $F_{X,o}(x_i) = \frac{i}{n}$ where *n* is the total number of

observations	5.
--------------	----

i	Annual maximum precipitation per hour	Observed distribution function $F_{-}(x)$	Postulated distribution function $E_{-}(x)$	Sample statistic
		$\Gamma_{X,o}(\lambda_i)$	$\Gamma_{X,p}(\lambda_i)$	$\boldsymbol{\mathcal{E}}_i$
1	73			
2	85			
3	93			
4	98			
5	100			
6	108			
7	115			
8	121			
9	130			
				$\varepsilon_{\rm max} =$

Table 4: Kolmogorov-Smirnov test.

Glossary

Accept	Annehmen
Average (in average)	Durchschnitt (im Durchschnitt / im Mittel)
Belief	Glaube
Breakdown	Betriebsunterbruch
Completion of a structure	Fertigstellung eines Tragwerkes
Comprehensive	Umfassend
Compressive strength	Druckfestigkeit
Concrete cover depth	Betonüberdeckung
Confidence interval	Konfidenzinterval
Continuous	Kontinuierlich
Design assumptions	Bemessungsannahmen
Deterioration	Alterung
Discrete	Diskret
Estimate	Schätzen
Experience	Erfahrung
Increase	Ansteigen
Independent	Unabhängig
Mean value	Mittelwert
Measurements	Messungen
Occurrence	Eintritt
Operational life	Betriebliche Lebensdauer
Plausibility	Plausibilität
Precipitation	Niederschlag
Preliminary survey	Vorstudie
Probability	Wahrscheinlichkeit
Probability density function	Wahrscheinlichkeitsdichtefunktion
Probability distribution function	Wahrscheinlichkeitsverteilungsfunktion
Rainfall	Niederschlag
Random variable	Zufallsvariable
Reject	Zurückweisen
Safety measures	Sicherheitsmassnahmen
Sample mean	Stichprobenmittelwert
Standard deviation	Standardabweichung
Suitable	Geeignet
To approach	Annähern (konvergieren)
Uncertainty	Unsicherheit
Concrete	Beton
Experience	Erfahrung
Soil	Boden
Estimate	Schätzen
Site	Bauplatz
Timber beam	Holzbalken
Point estimates	Punktschätzungen





Table 2. The quantile of the Chi-square distribution, χ^2 , with ν degrees of freedom. (*HINT: F* is equal to $1 - \alpha$ where α is the significance level used for a Chi-square test.)



V	F=0.01	F=0.05	F=0.10	F=0.25	F=0.50	F=0.75	F=0.90	F=0.95	F=0.99	F=0.995	F=0.999
1	0.0002	0.0039	0.0158	0.1015	0.4549	1.3233	2.7055	3.8415	6.6349	7.8794	10.8274
2	0.0201	0.1026	0.2107	0.5754	1.3863	2.7726	4.6052	5.9915	9.2104	10.5965	13.8150
3	0.1148	0.3518	0.5844	1.2125	2.3660	4.1083	6.2514	7.8147	11.3449	12.8381	16.2660
4	0.2971	0.7107	1.0636	1.9226	3.3567	5.3853	7.7794	9.4877	13.2767	14.8602	18.4662
5	0.5543	1.1455	1.6103	2.6746	4.3515	6.6257	9.2363	11.0705	15.0863	16.7496	20.5147
6	0.8721	1.6354	2.2041	3.4546	5.3481	7.8408	10.6446	12.5916	16.8119	18.5475	22.4575
7	1.2390	2.1673	2.8331	4.2549	6.3458	9.0371	12.0170	14.0671	18.4753	20.2777	24.3213
8	1.6465	2.7326	3.4895	5.0706	7.3441	10.2189	13.3616	15.5073	20.0902	21.9549	26.1239
9	2.0879	3.3251	4.1682	5.8988	8.3428	11.3887	14.6837	16.9190	21.6660	23.5893	27.8767
10	2.5582	3.9403	4.8652	6.7372	9.3418	12.5489	15.9872	18.3070	23.2093	25.1881	29.5879
11	3.0535	4.5748	5.5778	7.5841	10.3410	13.7007	17.2750	19.6752	24.7250	26.7569	31.2635
12	3.5706	5.2260	6.3038	8.4384	11.3403	14.8454	18.5493	21.0261	26.2170	28.2997	32.9092
13	4.1069	5.8919	7.0415	9.2991	12.3398	15.9839	19.8119	22.3620	27.6882	29.8193	34.5274
14	4.6604	6.5706	7.7895	10.1653	13.3393	17.1169	21.0641	23.6848	29.1412	31.3194	36.1239
15	5.2294	7.2609	8.5468	11.0365	14.3389	18.2451	22.3071	24.9958	30.5780	32.8015	37.6978
16	5.8122	7.9616	9.3122	11.9122	15.3385	19.3689	23.5418	26.2962	31.9999	34.2671	39.2518
17	6.4077	8.6718	10.0852	12.7919	16.3382	20.4887	24.7690	27.5871	33.4087	35.7184	40.7911
18	7.0149	9.3904	10.8649	13.6753	17.3379	21.6049	25.9894	28.8693	34.8052	37.1564	42.3119
19	7.6327	10.1170	11.6509	14.5620	18.3376	22.7178	27.2036	30.1435	36.1908	38.5821	43.8194
20	8.2604	10.8508	12.4426	15.4518	19.3374	23.8277	28.4120	31.4104	37.5663	39.9969	45.3142
21	8.8972	11.5913	13.2396	16.3444	20.3372	24.9348	29.6151	32.6706	38.9322	41.4009	46.7963
22	9.5425	12.3380	14.0415	17.2396	21.3370	26.0393	30.8133	33.9245	40.2894	42.7957	48.2676
23	10.1957	13.0905	14.8480	18.1373	22.3369	27.1413	32.0069	35.1725	41.6383	44.1814	49.7276
24	10.8563	13.8484	15.6587	19.0373	23.3367	28.2412	33.1962	36.4150	42.9798	45.5584	51.1790
25	11.5240	14.6114	16.4734	19.9393	24.3366	29.3388	34.3816	37.6525	44.3140	46.9280	52.6187
26	12.1982	15.3792	17.2919	20.8434	25.3365	30.4346	35.5632	38.8851	45.6416	48.2898	54.0511
27	12.8785	16.1514	18.1139	21.7494	26.3363	31.5284	36.7412	40.1133	46.9628	49.6450	55.4751
28	13.5647	16.9279	18.9392	22.6572	27.3362	32.6205	37.9159	41.3372	48.2782	50.9936	56.8918
29	14.2564	17.7084	19.7677	23.5666	28.3361	33.7109	39.0875	42.5569	49.5878	52.3355	58.3006
30	14.9535	18.4927	20.5992	24.4776	29.3360	34.7997	40.2560	43.7730	50.8922	53.6719	59.7022

Statistics and probability theory
M.H.Faber, Swiss Federal Institute of Technology, ETH Zurich, Switzerland

n	$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
1	0.995	0.990	0.975	0.950	0.900
2	0.929	0.900	0.842	0.776	0.684
3	0.829	0.785	0.708	0.636	0.565
4	0.734	0.689	0.624	0.565	0.493
5	0.669	0.627	0.563	0.509	0.447
6	0.617	0.577	0.519	0.468	0.410
7	0.576	0.538	0.483	0.436	0.381
8	0.542	0.507	0.454	0.410	0.358
9	0.513	0.480	0.430	0.387	0.339
10	0.489	0.457	0.409	0.369	0.323
11	0.468	0.437	0.391	0.352	0.308
12	0.449	0.419	0.375	0.338	0.296
13	0.432	0.404	0.361	0.325	0.285
14	0.418	0.390	0.349	0.314	0.275
15	0.404	0.377	0.338	0.304	0.266
16	0.392	0.366	0.327	0.295	0.258
17	0.381	0.355	0.318	0.286	0.250
18	0.371	0.346	0.309	0.279	0.244
19	0.361	0.337	0.301	0.271	0.237
20	0.352	0.329	0.294	0.265	0.232
21	0.344	0.321	0.287	0.259	0.226
22	0.337	0.314	0.281	0.253	0.221
23	0.330	0.307	0.275	0.247	0.216
24	0.323	0.301	0.269	0.242	0.212
25	0.317	0.295	0.264	0.238	0.208
26	0.311	0.290	0.259	0.233	0.204
27	0.305	0.284	0.254	0.229	0.200
28	0.300	0.279	0.250	0.225	0.197
29	0.295	0.275	0.246	0.221	0.193
30	0.290	0.270	0.242	0.218	0.190
31	0.285	0.266	0.238	0.214	0.187
32	0.281	0.262	0.234	0.211	0.184
33	0.277	0.258	0.231	0.208	0.182
34	0.273	0.254	0.227	0.205	0.179
35	0.269	0.251	0.224	0.202	0.177
36	0.265	0.247	0.221	0.199	0.174
37	0.262	0.244	0.218	0.196	0.172
38	0.255	0.241	0.215	0.194	0.170
39	0.252	0.238	0.213	0.191	0.168
40	0.249	0.235	0.210	0.189	0.165
n > 40	$1.63/\sqrt{n}$	$1.52/\sqrt{n}$	$1.36/\sqrt{n}$	$1.22/\sqrt{n}$	$1.07/\sqrt{n}$

Table 3. Kolmogorov-Smirnov-test statistic for a significance level α and sample size n.