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 log-normal 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:

$P(T \leq 2 years) = 0.181$. ☑

$P(T \leq 2 years) = 0.819$. ☐

$P(T \leq 2 years) = 0.0067$. ☐

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 $P(X = k) = \frac{(\nu t)^k}{k!} e^{-\nu t}$ and with annual mean rate ν . 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. ☑

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? (***<u>HINT</u>****: 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. ☐

DETAILED SOLUTION:

1.5 From the expression of the mean value, the parameter of the exponential distribution is calculated as:

$$\mu_T = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{\mu_T} = \frac{1}{10}$$

The probability that the engine breaks down within $t = 2 years$ after it was placed in operation can be calculated as:

$$P(T \leq 2) = F_T(t) = 1 - e^{-\lambda t} = 1 - e^{-\frac{1}{10}2} = 0.181$$

1.6 It is: $\nu = 5\ snowfalls / year$

The probability of exactly $k = 5\ snowfalls$ in the next year, i.e. $t = 1\ year$, is calculated as:

$$P(X=5)=\frac{(\nu t)^k}{k!}e^{-\nu t}=\frac{(5\cdot 1)^5}{5!}e^{-5\cdot 1}=0.175$$

The probability of no snowfall, i.e. $k=0\ snowfalls$, in the next year, i.e. $t=1\,year$, is calculated as:

$$P(X=0)=\frac{(\nu t)^k}{k!}e^{-\nu t}=\frac{(5\cdot 1)^0}{0!}e^{-5\cdot 1}=0.0067$$

1.8 The reliability index β can be calculated by the provided equation: $\beta=\frac{\mu_M}{\sigma_M}$.

The mean μ_M and standard deviation σ_M of the safety margin M can be calculated by applying the properties of the operators on the safety margin expression $M=R-L$. This gives:

$\mu_M=\mu_R-\mu_L=30-9=21kNm$ *and*

$$\sigma_M=\sqrt{\sigma_R^2+\sigma_L^2}=\sqrt{5^2+2^2}=5.39kNm$$

Hence the reliability index is equal to:

$$\beta=\frac{\mu_M}{\sigma_M}=\frac{21}{5.39}=3.9.$$

The annual probability of failure of the timber beam is:

$$P_F=P(M\leq 0)=\Phi(-\beta)=\Phi(-3.9)=4.8\cdot 10^{-5}$$

Where $\Phi(-3.9)=4.8\cdot 10^{-5}$, can be found from the table of the standard normal probability distribution function (Table 1 after Glossary).

Part 2: 'Exercise'

a. The first sample moment of the considered variable is calculated as:

$$m_1 = \frac{1}{n}\sum x_i = \frac{1}{9}923 = 102.56 mm/hour$$

The second sample moment is calculated as:

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{9}97237 = 10804 mm^2/hour^2$$

Using the expressions for the first two moments, provided in Table 2 of the "Questions" sheet of the Testat it is:

$$m_1 = u + \frac{0.5772}{\alpha} \Rightarrow 102.56 = u + \frac{0.5772}{\alpha} \Rightarrow u = 102.56 - \frac{0.5772}{\alpha} \qquad (1)$$

And

$$m_2 = \left(u + \frac{0.5772}{\alpha}\right)^2 + \frac{\pi^2}{6\alpha^2} \Rightarrow 10804 = \left(102.56 - \frac{0.5772}{\alpha} + \frac{0.5772}{\alpha}\right)^2 + \frac{\pi^2}{6\alpha^2} \Rightarrow \alpha = 0.076$$

And from Equation(1) it is:

$$u = 102.56 - \frac{0.5772}{\alpha} = 102.56 - \frac{0.5772}{0.076} = 94.94$$

b. Table 3 includes the necessary values for plotting the data on the probability paper provided. An example of calculation is given in the following:

$$F_{X,o}(x_1) = \frac{1}{9+1} = 0.1 \textit{ and } -\ln(-\ln(F_{X,o}(x_1))) = -\ln(-\ln(0.1)) = -0.834$$

Table 3: Calculations for the probability paper.

i	Annual maximum precipitation per hour x_i (mm/hour)	$F_{X,o}(x_i) = \frac{i}{n+1}$	$-\ln(-\ln(F_{X,o}(x_i)))$
1	73	0.1	-0.834
2	85	0.2	-0.476
3	93	0.3	-0.186
4	98	0.4	0.087
5	100	0.5	0.367
6	108	0.6	0.672
7	115	0.7	1.031
8	121	0.8	1.500
9	130	0.9	2.250

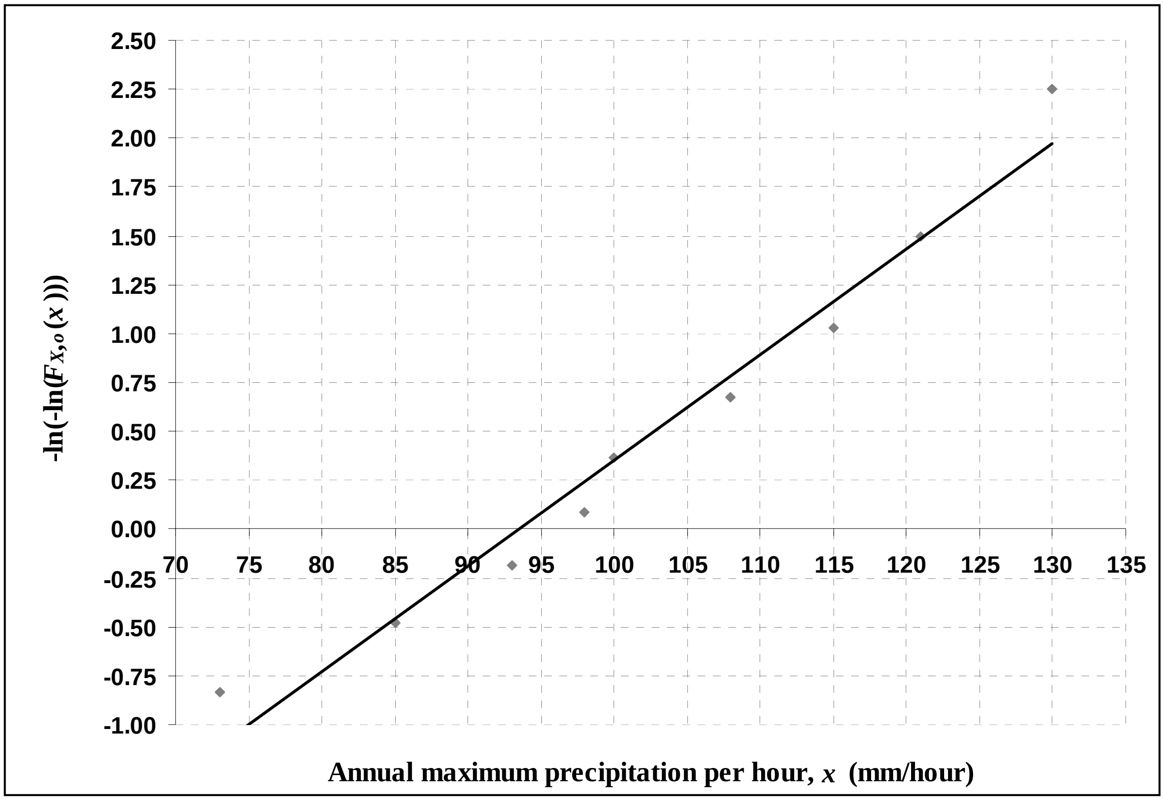

Figure 1: Probability paper for the Gumbel distribution.

c. Yes, the belief of the engineer is correct. Figure 1 shows the plotted data on the probability paper for the Gumbel distribution. A "best-fit" line is plotted. It is seen that the data fit quiet well to the straight "best-fit" line. Hence the belief of the engineer is correct.

d. From the probability distribution function of the Gumbel distribution, $F_X(x) = \exp\left(-\exp\left(-\alpha\left(x-u\right)\right)\right)$ it is seen that after a small transformation we get:

$-\ln(-\ln(F_X(x))) = \alpha x - \alpha u$ which is a linear relationship between:

$y = -\ln(-\ln(F_X(x)))$ and x with slope α and y-intercept αu .

Using the "best-fit" line two equations may be formed with two unknowns from where we can calculate the parameters of the distribution. It is:
Hence the two linear equations can be written using the general form: $y = \kappa x - \lambda$, where $\kappa = \alpha$ and $\lambda = \alpha u$.
For $x = 121$ it is $y = 1.5$, and for $x = 93$ it is $y = 0$.
It is:
$y = \kappa x - \lambda \Rightarrow 1.5 = \kappa 121 - \lambda$
and
$y = \kappa x - \lambda \Rightarrow 0 = \kappa 93 - \lambda$

Subtracting the above two equations we get:

$1.5 - 0 = \kappa 121 - \lambda - (\kappa 93 - \lambda) \Rightarrow 1.5 = 28\kappa \Rightarrow \kappa = 0.0536$

And then from one of the two equations it is: $y = \kappa x - \lambda \Rightarrow 0 = 0.0536 \cdot 93 - \lambda \Rightarrow \lambda = 4.9848$

Hence the parameters of the Gumbel distribution are:

$\alpha = \kappa = 0.0536$ and $\alpha u = \lambda \Rightarrow 0.0536u = 4.9848 \Rightarrow u = \frac{4.9848}{0.0536} = 93$.

It is observed that the values of the parameters calculated using the probability paper differ from the ones calculated with the method of moments. This can be explained from the fact that with the probability paper we use a "best-fit" line for the calculation of the parameters.

Note also that if you would use other combinations for the x and y values to form the set of the two linear relationships with the two unknowns the result may be slightly different.

e. For a Gumbel distribution with parameters $u = 95$ und $\alpha = 0.05$ Table 4 shows the calculation results for the Kolmogorov-Smirnov test.

Table 4: Kolmogorov-Smirnov test.

i	Annual maximum precipitation per hour x_i (mm/hour)	Observed distribution function $F_{X,o}(x_i)$	Postulated distribution function $F_{X,p}(x_i)$	Sample statistic ε_i
1	73	0.111	0.050	0.062
2	85	0.222	0.192	0.030
3	93	0.333	0.331	0.002
4	98	0.444	0.423	0.022
5	100	0.556	0.459	0.097
6	108	0.667	0.593	0.073
7	115	0.778	0.692	0.086
8	121	0.889	0.761	0.127
9	130	1.000	0.840	0.160
				$\varepsilon_{\max}$ =0.160

Example of calculation for $i = 1$:

$$F_{X,o}(x_1) = \frac{1}{n} = \frac{1}{9} = 0.111$$

$$F_{X,p}(x_1) = \exp\left(-\exp\left(-\alpha\left(x_1 - u\right)\right)\right) = \exp\left(-\exp\left(-0.05\left(x_1 - 95\right)\right)\right) = 0.066$$

$$\varepsilon_i = \left|F_{X,o}(x_i) - F_{X,p}(x_i)\right| = \left|0.111 - 0.066\right| = 0.045 .$$

The sample statistic is equal to $\varepsilon_{\max} = \max_{i=1}^{n} \varepsilon_i = 0.160$.

Using the corresponding table for the 5% significance level and a sample size equal to 9 the Kolmogorov-Smirnov statistic is found equal to **0.43**.

The sample statistic is smaller than the the Kolmogorov-Smirnov statistic. Hence the hypothesis of the Gumbel distribution with parameters $u = 95$ und $\alpha = 0.05$ cannot be rejected at the 5% significance level.