Sannolikhet, statistik och risk MVE300
Written examination - 25th May 2010, 14.00 - 18.00, V-huset.

Allowed aids: The course book "Probability and Risk Analysis - An Introduction for Engineers", copies of the e-book, earlier versions of the book (compendium). Mathematical or statistical tables. The tables of formulæ of the course. Any calculator (not PC). Dictionaries for translation. Jour: Igor Rychlik anknytning 3545 (examinator).

1 The random variable X describes the strength of a component. It is lognormally distributed with median 10 kN and the coefficient of variation R_X is 0.2. Find the characteristic strength of the component (the 0.9-quantile of X). (10 p)

2 Suppose that plankton growth in shallow lakes can be described by a linear function $g(X_1, X_2, X_3)$ of three variables: water temperature X_1, sun radiation X_2 and the concentration of nutrition X_3. Equilibrium is reached if $g(X_1, X_2, X_3) = 0$ while $g(X_1, X_2, X_3) < 0$ means an increase of plankton, indicating over-fertilization of lakes in the region. The "equilibrium" function g is given by

$$g(X_1, X_2, X_3) = a_0 - (a_1 X_1 + a_2 X_2 + a_3 X_3),$$

where $a_0 = 6.9$ [mg/m^3]; $a_1 = 0.08$ [mg/(m$^3\times$C)], $a_2 = 0.01$ [mg/(m$\times$W)] and $a_3 = 0.02$ (all other factors influencing the plankton growth are included in a_0).

Measurements taken from a number of lakes indicate that the three variables are normally distributed with the following expected values $m_1 = 16.0$ [C], $m_2 = 150$ [W/m^2] and $m_3 = 100$ [mg/m^3], respectively, and that the coefficients of variations are $R(X_1) = 0.5$, $R(X_2) = 0.3$ and $R(X_3) = 0.7$.

Compute the probability of plankton growth, i.e. that $g(X_1, X_2, X_3) < 0$. (10 p)

3 Suppose one is interested in the frequency λ of heavy rains in Göteborg. During a heavy rainy day, the measured rain precipitation (nederbörd) is exceeding 30 mm. Since these are extreme events one assumes that times between the heavy rains are independent exponentially distributed with average value $a = 1/\lambda$.

(a) Give the distribution for the number of heavy rain days in t years? (2 p)

(b) Suppose you need to estimate the probability of no rains in 2 years. You will employ the Bayessian approach to find the probability. First, based on yours personal experience, propose a "prior" distribution of frequency λ. Briefly give the reasons of your choice. (2 p)

(c) Suppose that you heard on radio that during the last 5 years there were 3 heavy rainy days. Use the information to compute the predictive probability of no heavy rains in the next two years. (6 p)

4 Suppose that at a location of a wind 2MW a wind turbine, intensity of the major storms is $\lambda = 2.5$ year^{-1}. A major storm is defined as storm with peak wind velocity exceeding 25 m/s. It is assumed that occurrence of the storms form a Poisson point process. The emergency stop of a wind turbine occurs when the wind speed exceeds 35 m/s. Suppose that for the last observed 10 major storms, the maximum wind speed exceeded the 25 [m/s] threshold by

$$1.4, \quad 11.8, \quad 4.2, \quad 5.0, \quad 1.2, \quad 4.0, \quad 13.4, \quad 2.3, \quad 5.3, \quad 2.5 \quad \text{[m/s]}.$$

Assume that the hight of the exceedences H are exponentially distributed, $H \in \exp(a)$.

(a) Estimate a and give the asymptotic confidence interval for a. (8 p)

(b) Estimate the probability of no emergency stops during one year. (12 p)

5 For a student lunch restaurant the following small survey of food preferences was carried out. 110 students were selected randomly and was asked whether they were vegetarian or not. The sex of the selected students were also registered. The result is given in the following table;

	Vegetarians	Non-vegetarians
Men	8	52
Women	12	38

Test the hypothesis that being vegetarian is independent of sex. (20 p)

6 Suppose that a number of small specimens of a material have been tested. The strength X of the specimens is well described by Weibull distribution with mean 100 MPa and coefficient of variation $R_X = 0.2$. Suppose that a structure is composed of 1000 of such specimens in series and has strength Y. Give the distribution of Y if it is equal to the strength of the weakest specimen, i.e. $Y = \min(X_1, \ldots, X_{1000})$. (Hint: use the following table to find the parameters of the X cdf.) (12 p)

Suppose that the load acting on the structure (pressure) varies in time and that its maximum value during one year is Gumbel distributed with parameters $a^* = 1.2$ MPa and $b^* = 1.5$ MPa. Estimate the safety index for the structure. (8 p)

c	2.00	2.10	2.70	3.00	3.68	4.00	5.00	5.73	8.00
R_X	0.523	0.500	0.400	0.363	0.302	0.281	0.220	0.200	0.118
$\Gamma(1+1/c)$	0.866	0.886	0.889	0.893	0.902	0.906	0.918	0.926	0.942

Good luck!

Solutions: Written examination - 25 May 2010

Sannolikhet, statistik och risk MVE300.

1 We search for $x_{0.9}$, i.e. 0.9-quantile of X. Since $Z = \ln X \in N(m, \sigma^2)$ where $m = \ln(10)$ while $\sigma^2 = \ln 1 + R_X^2 = 0.0392$, hence $z_{0.9} = m - \lambda_{0.1}\sigma$ and $x_{0.9} = \exp(z_{0.9})$. The answer: $x_{0.9} = 10\exp(-1.28 \cdot 0.198) = 7.76$.

2 The random variables $X_i \in N(m_i, \sigma_i^2)$, where $\sigma_i^2 = (m_i \cdot R_{X_i})^2$, giving the following variances 64, 2025 and 4900, respectively. By linearity the function g, $g(X_1, X_2, X_3) \in N(m, \sigma^2)$ and hence probability of plankton grow $P(g(X_1, X_2, X_3) < 0) = \Phi(-m/\sigma)$. The expected value $m = a_0 - \sum a_i\, m_i = 2.12$ and since X_i are uncorrelated (independent) then $\sigma^2 = \sum a_i^2\, \sigma_i^2 = 2.57$. The answer $P(g(X_1, X_2, X_3) < 0) = \Phi(-m/\sigma) = \Phi(-2.12/\sqrt{2.57}) = 0.093$.

3 **(a)** Let N be the number of heavy rains in t years then $N \in Po(m)$, $m = \lambda\, t$.

(b) Non-informative improper priors could be chosen, Gamma(0,0) (or if one knows that there were no heavy rains in the last 2 years then one could choose Gamma(0,2), etc.). There are other possible choices, but let assume that one has selected Gamma(a,b) as the prior density.

(c) If the answer in (b) were Gamma(a,b) then the posteriori density is Gamma(a+3,b+5). The event A = "no heavy rain in two years" is true if for $t = 2$ years "$N_t = 0$". Then conditionally that the intensity λ is known

$$P(N_2 = 0) = \exp(-2\lambda).$$

Hence the predictive probability

$$P^{pred}(N_2 = 0) = \int_0^\infty \exp(-2\lambda) f^{post}(\lambda)\, d\lambda = \left(\frac{b+5}{b+5+2}\right)^{a+3}.$$

For the improper priors, i.e. $a = 0$ and $b = 0$, $P^{pred}(N_2 = 0) = (5/7)^3 = 0.36$.

4 Denote by H the height of exceedences, $H \in \exp(a)$.

(a) The ML estimate of a is a^* equal to the average height 5.1 [m/s]. For $\alpha = 0.05$ the 95% asymptotic (approximative) conf. interval is $[a^* - \lambda_{\alpha/2} a^*/\sqrt{n}, a^* + \lambda_{\alpha/2} a^*/\sqrt{n}]$. The answer: $a^* = 5.1$, 95% conf. interval $[1.94, 8.26]$.

(b) The stream of emergency stops is also Poisson point process, with intensity $\lambda_e = \lambda \cdot P(H > 10)$ and hence the probability of no stops during one year $p = \exp(-\lambda_e \cdot 1)$. The answer: $\lambda_e^* = 2.5 \cdot \exp(-10/a^*) = 0.35$ year^{-1}. and $p^* = \exp(-0.35) = 0.7$.

5 The description of the survey was not exact. Two possible interpretations are: one planed to choose a fixed number of students, here 110 (this is the most natural assumption); one decided to ask all students that arrived during some fixed time period, e.g. 3 hours, then the number of interviewed students will be random (and if the survey will be performed in "non busy period" then the number would be Poisson distributed).

(a) Let A = "a person is a women" and B = "a person is a vegetarian". The hypothesis H_0 : sex and food preferences are independent. We will use χ^2-test to check whether the result of the survey contradicts the hypothesis. The level α is chosen to be 0.05. The Q statistics is

$$Q = \sum_{i=1}^{4} \frac{(n_i - n \cdot p_i^*)^2}{np_i^*},$$

where $n = 110$. If H_0 is true $p_1 = p_{A^c \cap B} = p_{A^c} p_B$, $p_2 = p_{A^c \cap B^c} = p_{A^c} p_{B^c}$, $p_3 = p_{A \cap B} = p_A p_B$ and $p_4 = p_{A \cap B^c} = p_A p_{B^c}$. From the survey we estimate $p_A^* = 5/11$ while $p_B^* = 2/11$ hence

$$p_1^* = \frac{6}{11}\frac{2}{11}, \quad p_2^* = \frac{6}{11}\frac{9}{11}, \quad p_3^* = \frac{5}{11}\frac{2}{11}, \quad p_4^* = \frac{5}{11}\frac{9}{11}.$$

Now $Q = 2.09$ and it should be compared with $\chi^2_{0.05}(4-2-1) = 3.84$. The answer is that the hypothesis can not be rejected.

(b) Here we shall use the notation introduced in a). From the description one could assume that one has observed four Poisson r.v. having expectations m_i corresponding to students of categories $A^c \cap B$, $A^c \cap B^c$, $A \cap B$ and $A \cap B^c$, respectively. The most complex model is that all m_i can have different values. The simpler model should represent the hypothesis that food preferences are independent of sex. These can be done as follows: suppose the survey was performed during $T = 1$ time unit and that the intensity of students were λ. Then the most general model is $m_i = \lambda \cdot p_i$ while the simpler is $p_1 = p_{A^c \cap B} = p_{A^c} p_B$, $p_2 = p_{A^c \cap B^c} = p_{A^c} p_{B^c}$, $p_3 = p_{A \cap B} = p_A p_B$ and $p_4 = p_{A \cap B^c} = p_A p_{B^c}$. From the survey we estimate $p_A^* = 5/11$ while $p_B^* = 2/11$. Clearly the complex model has 4 parameters while the simple only three. One can test whether data contradicts the simpler hypothesis by means of deviance

$$DEV = 2\sum_{i=1}^{4} n_i(\ln(n_i) - \ln(110\, p_i^*)) = -2.48 + 2.99 + 3.33 - 2.80 = 2.08.$$

Since $DEV = 2.08 < \chi^2_{0.05}(4-3) = 3.84$ the answer is that the hypothesis can not be rejected.

6 **(a)** Note that you have two equation $R_X = 0.2$ and $\mathsf{E}[X] = 100$ but three parameters a, c and b. Most often such problems do not have unique solution. This is the case here too. Simply the table is valid only when $b = 0$!

Since X has Weibull cdf with scale a and shape c then Y is also Weibull distributed with the same shape parameter c but another scale parameter $\tilde{a} = a/1000^{1/c}$. Hence in order to find cdf for Y one needs to estimate a and c. Using the table $R_X = 0.2$ corresponds to $c = 5.73$ and $a = 100/\Gamma(1+1/c) = 108.07$. The answer Y is Weibull distributed with $\tilde{a} = a/1000^{1/c} = 32.37$ and $c = 5.73$.

(b) Solution $Z = Y - S$, where S is the Gumbel distributed load and hence the index $\beta_C = (E[Y] - E[S])/\sqrt{V(Y) + V(S)}$. Computation of $E[Y]$, $V(Y)$: from the table $E[R] = 32.37 \cdot 0.926 = 30$. Since the coefficient of variation depends only on c hence $V(Y) = R_X^2 \cdot E[Y]^2 = 0.2^2 \cdot 30^2 = 36$.

Computation of $E[S]$, $V(S)$; from the table $E[S] = 1.5 + 1.2 \cdot 0.57772 = 2.19$, $V(S) = 1.2^2\pi^2/6 = 2.37$.

The answer $\beta_c = (30 - 2.19)/\sqrt{2.37 + 36} = 4.49$.