Sannolikhet, statistik och risk MVE300
Written examination - 23th May 2011, 8.30 - 12.30, 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) and copies of the 12 course-lectures. Mathematical or statistical tables. The tables of formulæ of the course. Any calculator (not PC). Dictionaries for translation. Jour: Igor Rychlik 0707405575.

1 A test consists of two steps; preparations and an experiment. Time the test takes varies. Suppose that in average each of steps take 30, 60, minutes with coefficients of variation 0.3, 0.2, respectively. One assumes that durations of the steps are normally distributed. Compute probability that the test can be completed in two hours assuming:

(a) Times needed to perform the steps are independent. (7p)

(b) The times are dependent, and the correlation coefficient is 0.9. (3p)

2 At a power plant, a certain type of failure, which leads to interrupt in the production, has its roots in presence of ice at important spots in a construction element. An unusually strong winter has made this problem evident to the security staff. One wants to investigate the probability p of an interupt during a production day, and a simple Bayesian approach is taken.

(a) As prior distribution, a Beta(1,1) distribution is proposed. Motivate why this would seem a natural choice. (2p)

(b) The following winter, in 100 days, there were 4 interrupts. Find the posterior distribution. (2p)

(c) In the next winter, in 100 days, there were 10 interrupts. Update the previous posterior distribution. (1p)

(d) Compute the predictive probability. (5p)

3 Often the uncertainty in the strength of components X is described by the coefficient of variation and the characteristic strength, which is the 0.9 quantile of X.

A producer estimated that in a series of tests 50% of the components failed under load of 200 kN and estimated the coefficient of variation of the strength to 0.2. Suppose that, by experience with similar data, he knows that the strength is log normally distributed. Then, what will be the estimate of the characteristic strength of his components? (10 p)

4 The yearly maximal water level in a river has been observed for 100 years. A preliminary analysis indicates that a Gumbel distribution is reasonable and that the observations can be considered independent. Computer software gives the ML estimates $a^* = 192$ and $b^* = 806$ of the scale and location parameter, respectively (unit: cm). Recall that for the ML estimators of a and b in a Gumbel distribution, the following asymptotic results hold:

$$V[A^*] \approx 0.61\frac{a^2}{n}, \quad V[B^*] \approx 1.11\frac{a^2}{n}, \quad \mathrm{Cov}[A^*, B^*] \approx 0.26\frac{a^2}{n}$$

(a) Estimate the 200-year return level, h_{200}. (5p)

(b) Find an approximate 95% confidence interval for h_{200}. (5p)

(c) Find the probability that the yearly maximum at least once during the following ten years will exceed h_{200}. (10p)

5 Even before the space shuttle Challenger exploded on January 20, 1986, NASA had collected data from 23 earlier launches. One part of these data was the number of O-rings that had been damaged at each launch;

2	0	0	0	0	1	0	0	0	0	0	2
0	1	0	0	1	0	0	0	0	1	0.	

Test whether N_i, numbers of damaged rings in ith launch, are independent Poisson distributed with the same expectation m. (20p)

6 Suppose that the strength of a welding (svetsning) is X with mean 15000 and coefficient of variation 0.1. The load is mainly due to wave acting on the structure and is equal to $Y = c_1 \cdot R + c_2 R^2$, where R is the wave amplitude with mean $\sqrt{\frac{\pi}{2}} \cdot \frac{h_s}{4}$ and variance $(2 - \frac{\pi}{2}) \cdot (\frac{h_s}{4})^2$.

The safety requirements are that the welding has to support a wave load due to 10000 years of storm which has $h_s = 20$ meters. Compute approximatively the safety index, i.e. $\frac{E(X-Y)}{D(X-Y)}$, for $c_1 = 20$ and $c_2 = 80$. (Hint: Use Gauss formulas) (20 p)

Good luck!

Solutions: Written examination - 23 May 2011
Sannolikhet, statistik och risk MVE300.

1 Let T be the time it takes to perform the experiment, while T_1, T_2 be the durations of steps 1 and 2, respectively, i.e. $T = T_1 + T_2$. We have $E[T] = E[T_1] + E[T_2] = 90$ minutes, while $D(T_1) = 30 \cdot 0.3 = 9$ and $D(T_2) = 60 \cdot 0.2 = 12$ minutes. We want to compute $P(T \leq 120) = \Phi((120 - E[T])/\sqrt{V(T)})$.

(a) Since T_1, T_2 are independent $V(T_1 + T_2) = V(T_1) + V(T_2) = 225$, and hence $P(T \leq 120) = \Phi(2) = 0.98$

(b) Since T_1, T_2 are correlated $V(T_1 + T_2) = V(T_1) + V(T_2) + 2D(T_1)D(T_2)\rho = 225 + 2 \cdot 9 \cdot 12 \cdot 0.9 = 419.4$, and hence $P(T \leq 120) = \Phi(1.465) = 0.93$

2 (a) The Beta(1,1) distribution is equivalent to a uniform distribution on $[0, 1]$. Without any specific prior knowledge about the unknown parameter p, this seems a reasonable choice.

(b) Updating with conjugated priors: $P^{\text{prior}} \in \text{Beta}(1, 1)$; $P^{\text{post}} \in \text{Beta}(1 + 4, 1 + 100 - 4) = \text{Beta}(5, 97)$.

(c) Again updating, now with $P^{\text{prior}} \in \text{Beta}(5, 97)$; $P^{\text{post}} \in \text{Beta}(5 + 10, 97 + 100 - 10) = \text{Beta}(15, 187)$.

(d) Predictive probability: $E[P^{\text{post}}] = 15/(15 + 187) \doteq 0.074$.

3 The 0.9-quantile, $x_{0.9}$, is defined as $P(X > x_{0.9}) = 0.9$. Since X is a lognormal distrubuted random variable, $\ln(X) \in N(m, \sigma^2)$. Thus the probability $P(\frac{\ln X - m}{\sigma} > \lambda_{0.9}) = 0.9$, where $\lambda_{0.9}$ is the 0.9-quantile in the standard Normal distribution. Furthermore, $\lambda_{0.9} = -\lambda_{0.1} = -1.28155$. So we have that $0.9 = P(\ln X > m + \sigma\lambda_{0.9}) = P(X > e^{m+\sigma\lambda_{0.9}})$. Thus, the characteristic strength, $x_{0.9} = e^{m+\sigma\lambda_{0.9}}$. Calculating the constants, $m = \ln(200)$ and $\sigma = \sqrt{\ln(R(X)^2 + 1)} = \sqrt{\ln(0.2^2 + 1)}$ gives $x_{0.9} = 200e^{\cdot -1.28155 \cdot \sqrt{\ln 1.04}} = 155$ kN.

4 (a) Gumbel distribution:

$$P(X \leq x) = \exp(\mathrm{e}^{-(x-b)/a}), \quad x \in \mathbb{R}.$$

The 200-year return value h_{200} is obtained by solving for h_{200} in $P(X > h_{200}) = 1/200$, which results in

$$h_{200} = b - a \ln(-\ln(1 - 1/200)).$$

An estimate h^*_{200} is then given by

$$h^*_{200} = b^* - a^* \ln(-\ln(1 - 1/200)) \doteq 1823 \text{ (cm)}$$

(b) Define $k = -\ln(-\ln(1 - 1/200))$. Then the variance of the estimate h^*_{200} is given by

$$\begin{aligned} V[H^*_{200}] &= V[B^* + k \cdot A^*] = V[B^*] + k^2 V[A^*] + 2k\mathrm{Cov}[A^*, B^*] \\ &\approx 1.11\frac{a^2}{n} + k^2 \cdot 0.61\frac{a^2}{n} + 2k \cdot 0.26 \cdot \frac{a^2}{n} \\ &= (1.11 + 0.61k^2 + 0.26 \cdot 2k)\frac{a^2}{n}. \end{aligned}$$

Plugging in a^* and k yields $V[H^*_{200}] \approx 7730.99$ cm^2. Hence an approximate 95% confidence interval is given by

$$[h^*_{200} \pm 1.96 \cdot \sqrt{V[H^*_{200}]}] = [1651,\ 1995].$$

(c) Introduce N = "The mumber of years with yearly maximum exceeding h_{200}". From (a), we know that $P(X \geq h_{200}) = 1/200$. Hence $N \in \mathrm{Bin}(10, 1/200)$ and we can compute

$$P(N \geq 1) = 1 - P(N = 0) = 1 - (1 - 1/200)^{10} \doteq 0.049.$$

(Alternative: use max stability of Gumbel cdf.)

5 The expected number of damaged rings during a launch m is estimated by $m^* = 8/23$. We shall employ χ^2 test. Since m^* is a small we introduce only 2 classes $A_1 =$ "$N_i = 0$", $A_2 =$ "$N_i = 1$" and $A_3 =$ "$N_i > 1$". Probabilities of A_i, p_i is estimated by

$$p^*_1 = \exp(-m^*) = 0.7, \qquad p^*_2 = m^* \exp(-m^*) = 0.2435, p^*_3 = 1 - 0.7 - 0.2435 = 0.0565.$$

There is $n = 23$ observations. The n_i be the number of observations in class A_i, $n_1 = 17$, $n_2 = 4$ and $n_3 = 2$, respectively. The Q statistics is

$$Q = \sum_{i=1}^{3} \frac{(n_i - n \cdot p^*_i)^2}{n \cdot p^*_i} = 0.89.$$

For $\alpha = 0.05$, the threshold $\chi^2_\alpha(f)$ with $f = 3 - 1 - 1 = 1$ is equal 3.84. Hence we can not reject hypothesis that the number of damaged rings during a lunch are independent Poisson distributed with a common expected value m. Note that $23 \cdot p_3 = 1.3 < 5$ but we neglect this.

6 The safety index is $\frac{E(X-Y)}{D(X-Y)} = \frac{E(X)-E(Y)}{\sqrt{V(X)+V(Y))}} = \frac{15000-E(Y)}{\sqrt{(0.1\cdot 15000)^2+V(Y)}}$. Now we use Gauss formulas to approximate $E(Y)$ and $V(Y)$. Let $Y = g(R) = 20R + 80R^2$, then $g'(R) = 20 + 160R$. According to Gauss formulas, with $E(R) = m_R = \sqrt{\frac{\pi}{2}} \cdot \frac{20}{4} = 6.2666$, and $V(R) = (2 - \pi/2) \cdot \frac{20^2}{4^2} = 10.73$

$$\begin{aligned} E(Y) &\approx = 20m_R + 160m_R^2 = 3267, \\ V(Y) &\approx (g'(E(R)))^2 \cdot V(R) = (20 + 160m_R)^2 \cdot V(R) = 1.1222 \cdot 10^7. \end{aligned}$$

So the safety index is,

$$\frac{15000 - 3267}{\sqrt{1500^2 + 1.1222 \cdot 10^7}} = 3.2$$