Sannolikhet, statistik och risk MVE240

Written examination - 8.30-12.30 Monday 9th March 2009 in M-building.

Allowed aids: The course book ”Probability and Risk Analysis - An Introduction for Engineers”, 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 Historical records of earthquakes in the San Francisco area show that during the period 1836-1961 there where 16 earthquakes of intensity VI or more. Suppose that occurrence of such high-energy earthquakes in this region can be modelled by means of a Poisson process with constant intensity.

(a) Estimate the intensity λ of the Poisson process. (3 p)

(b) What is the return period of earthquakes with energy VI or above? (2 p)

(c) Suppose you plan to study one semester in San Francisco (6 months). Estimate the probability that you will experience a strong earthquake during this period of time. (5 p)

2 In Finland year 2000 died 396 persons in traffic and about $46.6 \cdot 10^9$ km was driven. In Sweden the same year 580 persons died in traffic and about $69.6 \cdot 10^9$ km has been driven.

(a) Estimate rates of death per 10^9 km in both countries. (3 p)

(b) Test whether the rates are significantly different? (7 p)

3 Let go back to the year 2006. Anna-Carin Olofsson and Kati Wilhelm have been successful in women biathlon, season 2005/2006. Olofsson had 4 victories in total (3 World Cup/1 Olympic) and Wilhelm had 3 victories (2 World cup/1 Olympic). (Since Olofsson did not participate in Antholz, these races are not taken into account). In 8 other races another competitor has won. Define the events A = ”Olofsson wins race” and B = ”Wilhelm wins race”. Also, define the probabilities $p_A = \mathsf{P}(A)$ and $p_B = \mathsf{P}(B)$. Model uncertainty in p_A, p_B values using random variables P_A, P_B having a Dirichlet(1,1,1) pdf as prior. Now, use the information from the 15 races together with a Bayesian approach to update probabilities p_A and p_B.

(a) State the posterior density. (2 p)

(b) Compute the predictive probability, $\mathsf{E}[P_A]$, that Olofsson wins the next race. (4 p)

(c) Give a measure of the uncertainty of the value of p_A by computing the coefficient of variation $\mathsf{R}[P_A]$. (4 p)

4 In years 1900-2005 Sweden experienced 26 individual storm events with forest damage exceeding one million m^3 of forest. The hypothesis is that these events follow a Poisson process with an unknown intensity. The periods between the storm events are (days):

158	47	4033	13	979	2976	4378	655	705	7	33
730	3383	66	224	770	661	1132	1140	365	1035	1476
712	75	1075								

Test whether the hypothesis of a Poisson process can be rejected or not. (20 p)

5 In "Statistical Abstract of the United States" one can find the data for the number of crashes in the world during the years 1976-1985, which we denote as $k_1, \dots, k_{10}$ with values

$$24, \quad 25, \quad 31, \quad 31, \quad 22, \quad 21, \quad 26, \quad 20, \quad 16, \quad 22,$$

respectively. Suppose that the times for variability of observed flight crashes can be well described by means of the Poisson stream with intensity λ_A.

(a) Use data to estimate the intensity λ_A and give asymptotic 99 % confidence interval for λ_A. (6 p)

(b) Let T be the time to the next flight crash. Give the distribution $F_T(t)$ and estimate the probability that $T > t$, $t = 1$ week. (4 p)

(c) Adopt now the Bayesian approach and model the uncertainty of the value of the intensity λ_A by means of a random variable Λ having the probability density $f(\lambda)$. Propose a suitable prior density $f^{prior}(\lambda)$, then use the data to update the density, i.e. derive the posteriori density $f^{post}(\lambda)$. (4 p)

(d) Compute the predictive probability of $T > t$, $t = 1$ week. (6 p)

6 A Gumbel distribution is used to model the yearly maximum wind speed X in London (Ontario),

$$F_X(x) = \mathrm{e}^{-\mathrm{e}^{-\frac{x-b}{a}}}.$$

Based on the observations from year 1939 to 1961 the ML estimates of the model parameters are $a^* = 3$ and $b^* = 17$.

(a) Estimate the hundred year wind speed, i.e. x_{100}. (6 p)

(b) Use the asymptotic normality of the ML-estimates to compute an approximate 95% confidence interval for the hundred year wind speed x_{100}. (6 p)

(c) What is the expected maximum wind speed in 100 years? (8 p)

Good luck!

Solutions: Written examination - Monday 9th March 2009
Sannolikhet, statistik och risk MVE240.

1 **(a)** The estimate of the intensity for "strong" earthquakes is $\lambda^* = 16/126$, [year^{-1}].

(b) The estimate of return period is $T^* = 1/\lambda^* = 7.88$ years.

(c) Since exposure is one month, then N the number of earthquakes in the period is $\mathrm{Po}(m)$, $m = 6\lambda/12$. Answer $P = \mathsf{P}(N > 0) = 1 - \exp(-m)$ and $P^* \approx 0.062$.

2 **(a)** The rates of deaths are $\lambda_F^* = 396/46.6 = 8.5$ and $\lambda_S^* = 580/69.6 = 8.3$, for Finland, Sweden, respectively.

(b) Assume that the number of perished in traffic in the two countries are independent Poisson distributed variables with expected values m_F and m_S, respectively. Hypothesis is that there is no difference in the rates (although $m_F \neq m_S$) and hence the common rate is $\lambda^* = (580{+}396)/(46.6{+}69.6) = 8.3993$. Deviance can be used to test the hypothesis

$$DEV = 2 \cdot 580(\ln(580) - \ln(8.3993 \cdot 69.6)) + 2 \cdot 396(\ln(396) - \ln(8.3993 \cdot 46.6)) = 0.0925$$

which is below $\chi^2_{0.05}(1) = 3.84$. Answer: the hypothesis of equal rates year 2000 can not be rejected.

3 With the prior density $f^{pri}(p_A, p_B) \in \mathrm{Dirichlet}(1, 1, 1)$ we get:

(a) the posterior density $f^{post}(p_A, p_B) \in \mathrm{Dirichlet}(1{+}4, 1{+}3, 1{+}8) \sim \mathrm{Dirichlet}(5, 4, 9)$.

(b) the predictive probability that Olofsson wins next race is $\mathsf{E}[P_A] = \frac{5}{5+4+9} = \frac{5}{18}$.

(c) and the coefficient of variation $\mathsf{R}[P_A] = \frac{\mathsf{D}[P_A]}{\mathsf{E}[P_A]} = \frac{\sqrt{\frac{5(18-5)}{18^2(18+1)}}}{\frac{5}{18}} = 0.37$.

4 Poisson process $\Rightarrow$ time periods between events are Exponential.

(a) $H_0 : F_T(t) = 1 - \mathrm{e}^{-t/a^*}$, where T is the time period between two storm events and $a^* = \bar{t} = 1073.1$. We use a χ^2-goodness-of-fit test with five classes: $\{0{-}250{-}500{-}1000{-}1750{-}\infty\}$. The number of storm events in each of the five classes are $n_i = \{8, 1, 7, 5, 4\}$. We get $p_1^* = F_X(250) = 0.2078$, $p_2^* = F_X(500) - F_X(250) = 0.1646$, $p_3^* = F_X(1000) - F_X(500) = 0.2337$, $p_4^* = F_X(1750) - F_X(1000) = 0.1980$, $p_5^* = F_X(\infty) - F_X(1750) = 0.1958$. Since $np_i^* \approx 5$ five classes is the maximum number of classes we can use. Test quantity: $Q = \sum_{i=1}^{5} \frac{(n_i - np_i^*)^2}{np_i^*} = 4.2656 < \chi^2_{0.05}(5 - 1 - 1) = 7.8147$ and thus, it is not possible to reject H_0.

(b) Alternatively one could use the Barlow-Proschan's test. The quantity $Z = 13.58$ and it lies in the 95% confidence interval $[12 - 1.96\sqrt{2},\ 12 + 1.96\sqrt{2}]$.

5 **(a)** $\lambda^* = 23.8$ [year^{-1}], $(\sigma_\varepsilon^*)^2 = \lambda^*/n = 2.38$,

$$[\lambda^* - 2.58\sigma_\varepsilon^*, \lambda^* + 2.58\sigma_\varepsilon^*] = [19.8, 27.8]$$

(b) $T \in \exp(a)$, $a^* = 1/\lambda^*$, $P(T > t) = \exp(-23.8/52) = 0.63$.

(c) Two solutions are possible. One is to start with the gamma priors $f(\lambda) \in \text{Gamma}(a, b)$ then the posteriori density is $f(\lambda) \in \text{Gamma}(a + 238, b + 10)$, or to use the asymptotic result that $f(\lambda) \in N(23.8, 2.38)$.

(d) Let $P = \mathrm{e}^{-\lambda t}$ then $P^{\text{pred}}(T > t) = E[e^{-\lambda t}]$. For the first posteriori density $P^{\text{pred}}(T > t) = \left(\frac{b+10}{b+10+1/52}\right)^{a+238}$, which for $a = b = 0$ gives 0.633. The second posteriori density will give the following result: $-\Lambda\, t \in N(-23.8/52, 2.38/52^2)$. Since P is lognormally distributed

$$E[P] = e^{-23.8/52+0.5*2.38/52^2} = 0.633.$$

6 **(a)** The hundred year windspeed, x_{100}, is calculated by solving the equation $F(x_{100}) = 1 - 1/100 \Rightarrow x_{100} = b^* + a^* \cdot \ln(100) = 30.8$

(b) Since $x_{100} = b^* + a^* \cdot \ln(100)$, then $\sigma_\varepsilon^* = 3 \cdot \sqrt{\frac{1.11+0.61(\ln 100)^2+0.52\ln 100}{23}} = 2.54$ and with the asymptotic normality of the ML-estimates an approximate 95% confidence interval for x_{100} is
$[x_{100}^* - \lambda_{0.025} \cdot \sigma_\varepsilon^*,\ x_{100}^* + \lambda_{0.025} \cdot \sigma_\varepsilon^*] =$
$[30.8 - 1.96 \cdot 2.54,\ 30.8 + 1.96 \cdot 2.54] = [25.8,\ 35.8]$.

(c) By max stability of Gumbel cdf maximum wind speed in 100 years, M say, is also Gumbel distributed with parameters $\tilde{b}^* = b^* + a^* \ln(100) = 30.8$ and $\tilde{a}^* = a^* = 3$. Hence $\mathsf{E}[M] = \tilde{b}^* + \tilde{a}^* \cdot 0.5772 = 32.53$.