EXAMPLE OF POSSIBLE EXAMINATION PROBLEMS - III

1 A new test for a certain bacterial infection has a very high detection probability. This high sensitivity results in a number of false positive results. A medical evaluation of the test has shown that the probability of a positive result given a bacterial infection is 0.98. From large population studies it is also known that 15 persons in one thousand are infected with this bacteria. In a large screening test 932 (randomly chosen) persons were tested and 56 of these tested positive. How many of these 56 can be expected to be infected? The tests can be assumed to be independent. (10 p)

2 The concentration of a pollutant Y at a location is predicted using measurements from some monitoring stations and some uncertain parameters, all together forming an input vector $X_1, X_2, \ldots, X_n$. In other words, there is a function (program) that for inputs $x_1, ..., x_n$ gives the predictor of the pollutant.

$$y_{\text{pred}} = g(x_1, x_2, \ldots, x_n).$$

Obviously the true value y is unknown but can be measured.

After a few years of calibrations (training period), one has achieved that y_{pred} is an unbiased predictor, i.e. the fraction of days when the predicted concentration exceeds the true one is 50% (obviously in 50% cases the true value is underestimated). The relative error $k = y/y_{\text{pred}}$ is assumed to have a coefficient of variation equal to 0.2.

Probabilistic model: Let $X_1, X_2, \ldots, X_n$ be the unknown values of the inputs, Y the true value of the concentration of the pollutant and let

$$Y = K \cdot g(X_1, X_2, \ldots, X_n) = K \cdot Y_{\text{pred}}.$$

Suppose that K is a lognormal variable modelling the model uncertainty. We know that the median of K is one and the coefficient of variation $\mathrm{R}(K) = 0.2$. Compute the probability that the predictor Y_{pred} will underestimate the true value Y by more than 25%. Hint: compute $\mathsf{P}(Y_{\text{pred}}/Y < 0.75)$. (10 p)

3 At a determination of the yield Z (unit: %) when oxidating ammonia in a converter, the following quantities are measured: X = the amount NH_3 in the reactant gas; Y = the amount NO in the product gas. The yield Z was calculated as

$$Z = Y(100/X - 1.25).$$

The following standard deviations have been found:

$$\sigma_X = 7,7 \cdot 10^{-2}, \quad \sigma_Y = 9,6 \cdot 10^{-2}.$$

During one of the trials, one notes that $X = 12.0$ och $Y = 13.5$. Calculate approximately the standard deviation of the yield. Hint: Use the observed values 12.0 and 13.5 to estimate $\mathsf{E}[X], \mathsf{E}[Y]$, respectively, and suppose that X and Y are independent. (20 p)

4 The intensity λ (frequency) of ignition of a fire is approximately given by

$$\lambda = a \cdot A^b, \quad [\text{year}^{-1}]$$

where A is the total floor area of the building and a and b are constants related to occupancy, related to a particular type of building. It is estimated that for schools in UK the constants $a = 0.0002$ and $b = 0.75$. Suppose that the same constants are valid for schools in Mölndal.

(a) Suppose that a district contains 3 schools having total floor areas 20 000, 15 000 and 10 000 m^2, respectively. We will compare the probabilities of fire ignition during two years in these schools.
Calculate first, for each school, the probability of no fire during the two-year period. Use then your results to compute the probability of no fire ignition in the district (i.e. in any of the three schools) during a two-year period. (10p)

(b) It has been estimated that the sprinkler-system efficiency in avoiding fires is estimated to be 95% for newly installed systems and 75% for older systems. Recompute the probability of no fire in the district after installation of the sprinkler systems in the two larger schools. (The smallest one has no sprinkler system installed at all). (5 p)

(c) Compute the return period of school-fires in the district before and after installations of sprinklers. (5p)

5 At an emergency central, one has studied the time intervals between emergencies. A suitable probabilistic model for the time T (hours) between emergency calls is an exponential distribution:

$$F_T(t) = 1 - \mathrm{e}^{-\lambda t}, \quad t > 0.$$

The parameter λ is unknown and will here, using a Bayesian approach, be modelled as a random variable. Based on previous experience of stations of this size, one assumes a prior distribution for the intensity as $\text{Gamma}(1, 6)$.

(a) Give an interpretation of the prior distribution. (5 p)

(b) A collection of data is started and the following time intervals are found (hours): 8.5, 2, 3.5, 15, 7. Update the prior distribution for the intensity, i.e. calculate the posterior distribution. (6 p)

(c) Give the expected value of the posterior distribution, i.e. an estimate of the intensity. (4 p)

(d) Use the posterior distribution to calculate the probability that there will be a time period longer than 24 hours between two emergencies, i.e. $P(T > 24)$. (Hint. Law of total probability.) (5 p)

Solutions:

1 Events A = "Test positive", B = "A person is infected" From the text, we find $\mathsf{P}(A) = 56/932 = 0.06$, $\mathsf{P}(B) = 0.015$, $\mathsf{P}(A|B) = 0.98$. Thus

$$\begin{aligned} p = \mathsf{P}(B|A) &= \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(A)} = \frac{\mathsf{P}(B)\mathsf{P}(A|B)}{\mathsf{P}(A)} \\ &= \frac{0.015 \cdot 0.98}{0.06} = 0.245. \end{aligned}$$

Now let N be the number of infected between the 66 positively tested. Then $N \in Bin(56, p)$ and $\mathsf{E}[N] = 56 \cdot 0.25 = 14$.

2 Find $\mathsf{P}(\frac{1}{K} < 0.75)$, i.e. $\mathsf{P}(K > \frac{4}{3})$. Since K is log-normal, $\ln K \in \mathrm{N}(m, \sigma^2)$. The median of K equals one, hence $m = 0$. Now $\mathrm{R}(K) = \sqrt{\exp(\sigma^2) - 1}$, $\sigma^2 = \ln(1 + \mathrm{R}(K)^2)$ and hence $\sigma = 0.198$. Thus $\mathsf{P}(K > 4/3) = 1 - \Phi(\ln(4/3)/0.198) = 0.073$.

Answer: The underestimation by 25% happens in average seven days out of hundred.

3 We will use Gauss' approximation. Introduce $g(x, y) = y(100/x - 1.25)$. Then

$$\frac{\partial g}{\partial x} = -\frac{100y}{x^2}, \quad \frac{\partial g}{\partial y} = \frac{100}{x} - 1.25$$

and Gauss' approximation gives

$$\mathsf{V}(Z) = \sigma_X^2 \left(\frac{100}{\mathsf{E}(X)} - 1.25 \right)^2 + \sigma_Y^2 \left(-\frac{100\mathsf{E}(Y)}{(\mathsf{E}(X))^2} \right)^2 = 0.98.$$

It follows that $\mathsf{D}(Z) = 0.99$.

4 **(a)** The number of fires during two years in a school having area A is $\mathrm{Po}(m)$ with $m = 0.0002 \cdot A^{0.75} \cdot 2$ which gives for the three schools $m_1 = 0.673$, $m_2 = 0.542$ and $m_2 = 0.4$, respectively. Hence the probabilities of no fire ignition in schools $P_i = \exp(-m_i)$ are 0.51, 0.58, and 0.67 for the smallest school. The probability of no fire in the district is $P = P_1 \, P_2 \, P_3$ or, more directly, $P = \exp(-(m_1 + m_2 + m_3)) = 0.2$.

(b) Number of fires in a school having area A and a new sprinkler system during two years is $\mathrm{Po}(m)$ with $m = 0.0002 \cdot A^{0.75} \cdot 2 \cdot (1 - 0.95)$. This gives for the three schools $m_1 = 0.034$, $m_2 = 0.027$, and $m_3 = 0.4$, respectively. Hence $P = \exp(-(m_1 + m_2 + m_3)) = \exp(-0.461) = 0.63$.

(c) The intensity of fires before installation of sprinklers

$$\lambda = 0.0002 \cdot 20000^{0.75} + 0.0002 \cdot 15000^{0.75} + 0.0002 \cdot 20000^{0.75} = 0.8$$

and hence $T = 1/0.8$, i.e. one year and three months. After installation of sprinklers

$$\begin{aligned} \lambda &= 0.0002 \cdot 20000^{0.75} \cdot (1 - 0.95) + 0.0002 \cdot 15000^{0.75} \cdot (1 - 0.95) \\ &+ 0.0002 \cdot 20000^{0.75} = 0.23. \end{aligned}$$

Consequently, the return period is $T = 1/0.23$ i.e. 4 years, 4 months and one week.

5 (a) Interpretation: The expected value of the prior distribution is the intensity $1/6\ \mathrm{h}^{-1}$, which is equivalent to an average time interval of 6 hours between emergency calls.

(b) From the text, the prior distribution is $\mathrm{Gamma}(1, 6)$. Thus the posterior distribution is also Gamma: $\mathrm{Gamma}(1 + 5, 6 + 36)$, i.e. $\mathrm{Gamma}(6, 42)$.

(c) The table of formulæ immediately gives the expectation $6/42 = 1/7 = 0.142$ $[\mathrm{h}^{-1}]$.

(d) A version of the law of total probability (cf. Eq. (6.23)), gives

$$\begin{aligned} \mathsf{P}(T > t) &= \int_0^\infty \mathsf{P}(T > t | \Lambda = \lambda) f^{\text{ post}}(\lambda)\, \mathrm{d}\lambda = \int_0^\infty \frac{42^6}{\Gamma(6)} \mathrm{e}^{-\lambda t} \lambda^5 \mathrm{e}^{-42\lambda}\, \mathrm{d}\lambda \\ &= \left(\frac{42}{42 + t} \right)^6 . \end{aligned}$$

Plugging in values yield the probability asked for: $\left(\frac{42}{42+24} \right)^6 = 0.066$.