STA500 Introduction to Probability and Statistics 2, autumn 2018.

Exercise set 7

Note on Bayesian statistics	Exercises 1,2,3,4
Note on maximum likelihood:	Exercise 2

Exercise 1:

The times between failures for a type of security valves are assumed to be independent and exponentially distributed with parameter (hazard rate) λ.

Assume first that it is known that $\lambda = 0.25$ per year.

a) If one such valve has been functioning without failures for one year, what is the probability that it fails during the next year?
Which type of process do we have for the failure times under the given assumptions?
What is the probability of more than one failure in such a valve during two years?

In practise λ is unknown, but can be estimated from data. Let $T_1, \ldots, T_n$ be times between failures.

b) Show that the maximum likelihood estimator for λ becomes $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} T_i}$.
Calculate the estimate when we have $n = 4$ observations where $\sum_{i=1}^{4} t_i = 36.4$.

In addition to observed data there are information available from other sources which is expressed by a gamma prior distribution for λ:

$$p(\lambda) = \frac{1}{b^a \Gamma(a)} \lambda^{a-1} e^{-\lambda/b}$$

Expectation and variance in the gamma distribution are respectively ab and ab^2.

c) What is the interpretation of a posterior distribution?
Show that the posterior distribution for λ becomes a gamma distribution with parameters $n + a$ and $1/(\sum_{i=1}^{n} t_i + 1/b)$.
Explain that a reasonable estimator for λ will be $\hat{\lambda}_{\text{Bayes}} = \frac{n+a}{\sum_{i=1}^{n} t_i + 1/b}$.

In a gas system valves of the considered type are used as security valves that should open if the pressure gets too high. Based on information reported in a reliability database from usage in other installations a prior distribution is specified as a gamma distribution with expectation 0.2 and variance 0.004.

d) Calculate the Bayes estimate for λ. How is the prior information versus the information from the data weighted? Explain and comment.

(Wait with point e) below until we have learned about Bayes intervals.)

e) Find a 90% Bayes interval for λ. (Hint: A transformation result can be helpful.) What is the interpretation of this interval?

Exercise 2:

The mean concentration μ of a pollutant in a lake is to be determined. The concentration is assumed to be normally distributed with mean μ and known standard deviation $\sigma = 5$. In addition to measurement data $X_1, \ldots, X_n$ the experts also have knowledge about μ from previous measurements and insight in the development of the concentration level over time. This priori knowledge of μ is expressed by a normal distribution with mean μ_p and standard deviation σ_p. It can be shown that the posterior distribution of μ then is a normal distribution with expectation $\frac{(\sum_{i=1}^n x_i)\sigma_p^2 + \mu_p\sigma^2}{n\sigma_p^2+\sigma^2}$ and variance $\frac{\sigma_p^2\sigma^2}{n\sigma_p^2+\sigma^2}$.

a) Set up the expression we need to start with to calculate the posterior distribution in this case. (If you like to you can also try to show that we from this get the resulting posterior distribution given above, but do not spend too much time on this.)

b) Show that the Bayes estimate for μ (the posterior mean) can be written as

$$\hat{\mu}_{\text{Bayes}} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma^2/n}\bar{x} + \frac{\sigma^2/n}{\sigma_p^2 + \sigma^2/n}\mu_p$$

and comment what this shows about how the information in data and the prior information is weighted in different situtations.
Calculate the Bayes estimate when $\mu_p = 11$, $\sigma_p = 1.5$ and $n = 5$ measurements gave an average of 13.6.

c) Use the posterior distribution to find an interval which μ is in with 95% probability (often called a Bayes interval or a credibility interval). Compare this interval with the standard confidence interval for μ in this situation and comment.

d) Discuss briefly pros and cons by using a Bayesian approach in this situation.

Some answers:

Bayesian note 3 6.86 and 6.76;

MLE Note 2 c) $(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/\sum_{i=1}^n X_i}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/\sum_{i=1}^n X_i})$; d) $\hat{\mu} = \bar{X}$, unbiased;
e) $\hat{p} = 0.85$, $\hat{\mu} = 7.07$, [0.78,0.92]

1 a) 0.22, 0.09, Poisson process ; b) 0.11; d) 0.16

2 b) 11.8 ; c) (9.4, 14.2) and (9.2, 18.0);