STA500 Introduction to Probability and Statistics 2, autumn 2018.

Exercise set 8

Note on Bayesian statistics	Exercise 5
Textbook (Walpole), 9. edition:	Exercise 6.51
Textbook (Walpole), 8. edition:	Exercise 6.49

Exercise 1:

The size of certain types of insurance claims received by an insurance company is assumed to be exponentially distributed with expectation β. The company needs to update their estimate of β and will collected data for the size of n recent claims $X_1, \ldots, X_n$.

a) Show that the likelihood function for β becomes $L(\beta) = (1/\beta^n)\exp(-\sum_{i=1}^n x_i/\beta)$ and show that the maximum likelihood estimator becomes $\hat{\beta} = \bar{X}$.
Calculate the estimate when the observations are (in million kr) 2.1, 3.3, 5.6, 8.7, 4.4, 1.9.

b) Use results on transformations and linear combinations to explain that $\sum_{i=1}^n (2X_i/\beta)$ is χ^2_{2n}-distributed

c) Derive a 95% confidence interval for β using the information from point b).
What is the reason for not using a Wald confidence interval in this case?

In practice the insurance company has lots of relevant information about β from other sources than just data from the current situation (e.g. from previous data combined with knowlegde about the development over time etc). To combine this information with the information from the current data to get a better estimate of β they use a Bayesian approach.

As prior distribution for β they use an inverse gamma distribution. The inverse gamma distribution has probability density function

$$f(y) = \frac{1}{b^a \Gamma(a)} y^{-a-1} e^{-1/(yb)} \qquad y > 0$$

and $\mathrm{E}(Y) = 1/(b(a-1))$ and $\mathrm{Var}(Y) = 1/(b^2(a-1)^2(a-2))$. (The relation to the ordinary gamma distribution is that if Z is having a gamma distribution then $Y = 1/Z$ is having an inverse gamma distribution.)

d) Show that the posterior distribution also becomes an inverse gamma distribution and find an expression for the Bayes estimate for β. Calculate the estimate when $a = 4$, $b = 0.1$ and the data are as given in a).

It can be shown by the transformation formula that if Y is having an inverse gamma distribution with parameters a and b, then $Z = 2/(Yb)$ is having a χ^2_{2a}-distribution. (You do need to show this, but this results can be useful in the next point.)

e) Find a 95% Bayes interval for β. Compare this interval to the confidence interval in c) and comment.
Do you see any potential dangers in using a Bayesian approach in this setting?

Exercise 2:

In a pipeline systems a problem which needs to be monitored is rust attacks on the pipeline. Experience has shown that rust attacks occur as a Poisson process along the pipeline (Notice that "time" in this Poisson process is the distance along the pipeline.)

First assume that it is known that the expected number of rust attacks per km is 5, i.e. $\lambda = 5$.

a) Calculate the probability of finding more than 8 rust attacks in 1 km of pipeline.
Calculate the probability of finding more than 4 rust attacks in 0.5 km of pipeline.
If 10 independent 0.5 km pieces of pipeline are examined, what is the probability that at least one of these pieces has more than 4 rust attacks?

b) What is the probability that there is less than 100 meters between two consecutive rust attacks?
When 100 meters of pipeline has been examined since the last rust attack without finding any new rust attacks, what is the probability that it will take more than another 200 meters before the next rust attack is found?
When the inspection starts, what is the probability that the distance until rust attack number 8 is less than 1km?

I practice λ, the expected number of rust attacks per km of pipeline, is unknown. Based on independent measurements $Y_1, \ldots, Y_n$ of the number of rust attacks in n pieces of pipeline of length respectively $t_1, \ldots, t_n$ km we want to estimate λ.

c) Show that the maximum likelihood estimator (MLE) for λ is:

$$\hat{\lambda} = \frac{Y_1 + \cdots + Y_n}{t_1 + \cdots + t_n} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n t_i}$$

Calculate the expectation and variance of the estimator.

When the expected number of rust attacks per km exceeds 15, i.e. $\lambda > 15$, the pipeline is shut down and all rust attacks are repaired. Based on a series of measurements we want to decide whether the pipeline should be shut down for rust repair.

d) Derive and calculate the 90% Wald confidence interval for λ when $n = 8$ independent pieces of the pipeline system gave the data reported in the table below.

measurement i	1	2	3	4	5	6	7	8
y_i	0	4	9	2	2	3	4	6
t_i	0.1	0.1	0.3	0.15	0.2	0.15	0.2	0.3

Here $\sum_{i=1}^8 y_i = 30$ and $\sum_{i=1}^8 t_i = 1.5$.

Some answers:

6.51/ 6.49 0.0183;

1 a) 4.33; b) χ_2^2; c) (2.2, 11.8); d) 4; e) (2.1, 7.5);

2 a) 0.0681, 0.1088, 0.684; b) 0.393, 0.368, 0.133; c) $E(\hat{\lambda}) = \lambda$ and $Var(\hat{\lambda}) = \lambda / \sum_{i=1}^n t_i$; d) [14, 26]