STA500 Introduction to Probability and Statistics 2, autumn 2018.

Solution exercise set 8

Note on Bayesian statistics, Exercise 5

First we write down the likelihood function

$$L(p) = \prod_{i=1}^{N} \binom{n}{y_i} \theta^{y_i}(1-\theta)^{n-y_i} \propto \theta^{\sum_i y_i}(1-\theta)^{Nn-\sum_i y_i}.$$

Thus the posterior kernel may be written as

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto \theta^{a-1}(1-\theta)^{b-1}\theta^{\sum_i y_i}(1-\theta)^{Nn-\sum_i y_i} \\ &= \theta^{(a+\sum_i y_i)-1}(1-\theta)^{(b+Nn-\sum_i y_i)-1}, \end{aligned}$$

which is easily recognized as a beta($a + \sum_{i=1}^N y_i, b + Nn - \sum_{i=1}^N y_i$)-distribution.

6.51/6.49

In the Weibull distribution the density is:

$$f(t) = \alpha\beta t^{\beta-1}e^{-\alpha t^\beta} \ , \quad t \geq 0,$$

By substitution we get (observing that if we set $u = \alpha x^\beta$ we get $du = \alpha\beta x^{\beta-1}dx$):

$$F(t) \ = \ \int_0^t \alpha\beta x^{\beta-1}e^{-\alpha x^\beta}dx = \int_0^{\alpha t^\beta} e^{-u}du = [-e^{-u}]_0^{\alpha t^\beta} = -e^{-\alpha t^\beta} - (-1) = 1 - e^{-\alpha t^\beta}$$

The failure rate (also called hazard rate) then becomes

$$z(t) = \frac{f(t)}{1-F(t)} = \frac{\alpha\beta t^{\beta-1}e^{-\alpha t^\beta}}{e^{-\alpha t^\beta}} = \alpha\beta t^{\beta-1}$$

With a failure rate of $z(t) = 1/\sqrt{t} = t^{-1/2} = t^{1/2-1}$ we see that we have $\alpha = 2$ and $\beta = 0.5$. Then we get:

$$P(T > 4) = 1 - P(T \leq 4) = 1 - F(4) = 1 - (1 - e^{-2 \cdot 4^{0.5}}) = \underline{\underline{0.018}}$$

Exercise 1:

a) The density for the exponential distribution parameterised with the expectation β is

$f(x) = \frac{1}{\beta} e^{-x/\beta}$. Then:

$$
\begin{aligned}
L(\beta) &= = \prod_{i=1}^{n} \frac{1}{\beta} e^{-x_i/\beta} = \underline{\underline{\frac{1}{\beta^n} e^{-\sum_{i=1}^{n} x_i/\beta}}} \\
l(\beta) &= \ln L(\beta) = \ln(1) - \ln(\beta)^n - \sum_{i=1}^{n} x_i/\beta = -n\ln(\beta) - \frac{1}{\beta}\sum_{i=1}^{n} x_i \\
\frac{\partial l(\beta)}{\partial \beta} &= -\frac{n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} x_i = 0 \\
& n\beta = \sum_{i=1}^{n} x_i \qquad \Rightarrow \quad \beta = \frac{1}{n}\sum_{i=1}^{n} x_i
\end{aligned}
$$

I.e MLE becomes $\hat{\beta} = \underline{\underline{\frac{1}{n}\sum_{i=1}^{n} X_i}}$.

The estimate becomes: $\hat{\beta} = \bar{x} = (2.1 + 3.3 + 5.6 + 8.7 + 4.4 + 1.9)/6 = 26/6 = \underline{\underline{4.33}}$.

b) From the result on transformations (see collection of formulas) we have for the exponential distribution that $2X/\beta$ has a χ^2_2-distribution if X is exponentially distributed with parameter β.

Further we have that since a sum of independent χ^2-distributed variables is χ^2-distributed with parameter ("degrees of freedom") equal to the sum of the parameters in the distribution of each variable (collection of formulas), we will have that $Z = \sum_{i=1}^{n}(2X_i/\beta)$ is having a $\underline{\underline{\chi^2_{2n}\text{-distribution}}}$ (by being a sum of n indep. χ^2_2-distributed variables).

c)

$$
Z = \sum_{i=1}^{n}(2X_i/\beta) = (2/\beta)\sum_{i=1}^{n} X_i = (2n/\beta)\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{2n}{\beta}\hat{\beta} \sim \chi^2_{2n}
$$

$$
\begin{aligned}
\Rightarrow P(\chi^2_{1-\alpha/2,2n} \le \frac{2n}{\beta}\hat{\beta} \le \chi^2_{\alpha/2,2n}) &= 1-\alpha \\
P\left(\frac{\chi^2_{1-\alpha/2,2n}}{2n\hat{\beta}} \le \frac{1}{\beta} \le \frac{\chi^2_{\alpha/2,2n}}{2n\hat{\beta}}\right) &= 1-\alpha \\
P\left(\frac{2n\hat{\beta}}{\chi^2_{\alpha/2,2n}} \le \beta \le \frac{2n\hat{\beta}}{\chi^2_{1-\alpha/2,2n}}\right) &= 1-\alpha
\end{aligned}
$$

Observed data: $n = 6$, $\hat{\beta} = 4.33$. With $\alpha = 0.05 \quad \Rightarrow \quad \chi^2_{0.025,12} = 23.337$ and $\chi^2_{0.975,12} = 4.404$. Give 95% conf. int. for β:

$$
\left[\frac{2\cdot 6\cdot 4.33}{23.337}, \frac{2\cdot 6\cdot 4.33}{4.404}\right] = \underline{\underline{[2.2,\ 11.8]}}
$$

The Wald confidence interval is an approximate confidence interval, where the approximation is better the more data we have. With only 6 observations we can not trust the approximation to be good.

d) With prior distribution $p(\beta) = \frac{1}{b^a\Gamma(a)}\beta^{-a-1}e^{-1/(\beta b)}$ and the likelihood in a) we get the posterior:

$$\begin{aligned} p(\beta|\text{data}) &= c \cdot L(\beta) \cdot p(\beta) \\ &= c\frac{1}{\beta^n}e^{-\sum_{i=1}^n x_i/\beta}\frac{1}{b^a\Gamma(a)}\beta^{-a-1}e^{-1/(\beta b)} \\ &= c_2\beta^{-n-a-1}e^{-(\sum_{i=1}^n x_i+1/b)/\beta} \end{aligned}$$

We see that this (as a function of β) is on the same form as an inverse gamma distribution with parameters $a^* = n + a$ and $b^* = 1/(\sum_{i=1}^n x_i + 1/b)$. I.e. the distribution is an $\underline{\underline{\text{inverse gamma distribution with parameters } n + a \text{ and } 1/(\sum_{i=1}^n x_i + 1/b).}}$

The standard Bayes estimate is the expectation in the posterior distribution. Using the formula for the expectation in the inverse gamma distribution given in the text we get:

$$\hat{\beta}_{\text{Bayes}} = \frac{1}{b^*(a^*-1)} = \underline{\underline{\frac{\sum_{i=1}^n x_i + 1/b}{n+a-1}}} = \frac{26 + 1/0.1}{6+4-1} = \underline{\underline{4}}$$

e) Since the posterior distribution for β is an inverse gamma distribution with parameters $a^* = n + a = 6 + 4 = 10$ and $b^* = 1/(\sum_{i=1}^n x_i + 1/b)) = 1/(26 + 1/0.1) = 0.0278$ we can use the result given in the text to find the Bayes interval. We would like to find the quantiles in the inverse gamma distribution which are such that

$$P(\xi_{1-\alpha/2,a^*,b^*} < \beta < \xi_{\alpha/2,a^*,b^*}) = 1 - \alpha$$

By the transformation result in the text:

$$\begin{aligned} P(\frac{1}{\xi_{1-\alpha/2,a^*,b^*}} > \frac{1}{\beta} > \frac{1}{\xi_{\alpha/2,a^*,b^*}}) &= 1-\alpha \\ P(\frac{2}{b^*\xi_{1-\alpha/2,a^*,b^*}} > \frac{2}{b^*\beta} > \frac{2}{b^*\xi_{\alpha/2,a^*,b^*}}) &= 1-\alpha \\ P(\frac{2}{b^*\xi_{\alpha/2,a^*,b^*}} < Z < \frac{2}{b^*\xi_{1-\alpha/2,a^*,b^*}}) &= 1-\alpha \end{aligned}$$

where Z is $\chi^2_{2a^*}$. I.e. $\frac{2}{b^*\xi_{\alpha/2,a^*,b^*}} = \chi^2_{1-\alpha/2,2a^*}$ which implies $\xi_{\alpha/2,a^*,b^*} = \frac{2}{b^*\chi^2_{1-\alpha/2,2a^*}}$ and similar $\xi_{1-\alpha/2,a^*,b^*} = \frac{2}{b^*\chi^2_{\alpha/2,2a^*}}$. Thus the 95% Bayes interval becomes:

$$\begin{aligned} \left(\frac{2}{b^*\chi^2_{\alpha/2,2a^*}}, \frac{2}{b^*\chi^2_{1-\alpha/2,2a^*}}\right) &= (\frac{2}{0.0278\chi^2_{0.025,20}}, \frac{2}{0.0278\chi^2_{0.975,20}}) \\ &= (\frac{2}{0.0278 \cdot 34.170}, \frac{2}{0.0278 \cdot 9.591}) = \underline{\underline{(2.1,\ 7.5)}} \end{aligned}$$

Exercise 2:

a) Y = number of rust attacks in t km $\sim$ Poisson$(5t)$

$$t = 1: \qquad P(Y > 8) = 1 - P(Y \leq 8) \overset{table}{=} 1 - 0.9319 = \underline{\underline{0.068}}$$
$$t = 0.5: \qquad P(Y > 4) = 1 - P(Y \leq 4) \overset{table}{=} 1 - 0.8912 = \underline{\underline{0.109}}$$

Let X be the number of pieces with more than 4 rust attacks. Then $X \sim B(10, 0.109)$ and:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{10}{0} 0.109^0 (1 - 0.109)^{10} = \underline{\underline{0.684}}$$

b) The distance between points in a Poisson process is exponentially distributed with parameter λ. Then we get (to keep things in the same scale we use that 100 meters is 0.1km):

$$\begin{aligned} P(T < t) &= \int_0^t \lambda e^{-\lambda u} du = [-e^{-\lambda u}]_0^t = 1 - e^{-\lambda t} \\ P(T < 0.1) &= 1 - e^{-5 \cdot 0.1} = 1 - e^{-0.5} = \underline{\underline{0.393}} \\ P(T > 0.3 | T > 0.1) &= P(T > 0.2) = 1 - P(T \leq 0.2) = 1 - (1 - e^{-5 \cdot 0.2}) = e^{-1} = \underline{\underline{0.368}} \end{aligned}$$

In the last calculation we have used the memory-less property of the exponential distribution which implies that $P(T > 0.3 | T > 0.1) = P(T > 0.2)$ (notice this is only the case for the exponential distribution and the geometric distribution!) Alternatively we could do this using the definition of conditional probability e.g. as follows (which is what we have to do for this type of calculations for other distributions):

$$P(T > 0.3 | T > 0.1) = \frac{P(T > 0.3 \cap T > 0.1)}{P(T > 0.1)} = \frac{P(T > 0.3)}{P(T > 0.1)} = \frac{e^{-5 \cdot 0.3}}{e^{-5 \cdot 0.1}} = \underline{\underline{0.368}}$$

The last question in this point is most easily solved by defining Y as the number of events in the interval $[0, 1]$ and calculate $P(Y \geq 8)$ (the event that it takes less than 1 km until attack number 8 is the same as the event that at least 8 attacks occur during the 1 km). We get:

$$P(Y \geq 8) = 1 - P(Y \leq 7) \overset{table}{=} 1 - 0.8666 = \underline{\underline{0.133}}$$

The other (and more challenging) way to solve this is to define S_8 as the distance until rust attack number 8. We know that the distribution of the distance until event number 8 in a Poisson process with $\lambda = 5$ is having a gamma distribution with parameters $\alpha = 8$ and $\beta = 1/5$ (being the sum of 8 exponentially distributed variables with expectation 1/5). We then get that:

$$\begin{aligned} P(S_8 < 1) &= \int_0^1 \frac{1}{\beta^\alpha \Gamma(\alpha)} s^{\alpha - 1} e^{-s/\beta} ds = \int_0^1 \frac{1}{(\frac{1}{5})^8 \Gamma(8)} s^{8-1} e^{-5s} ds \\ &\overset{u=5s}{=} \int_0^5 \frac{5^8}{\Gamma(8)} (\frac{u}{5})^{8-1} e^{-u} du/5 = \int_0^5 \frac{u^{8-1}}{\Gamma(8)} e^{-u} du \overset{tableA.23/A.24}{=} \underline{\underline{0.133}} \end{aligned}$$

Table A.23 (A.24 in 8.edition) referred to above is on page 767 (791 in 8.ed) in Walpole.

c) $Y_i \sim \text{Poisson}(\lambda t_i)$

$$\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n} f(y_i;\lambda) = \prod_{i=1}^{n} \frac{(\lambda t_i)^{y_i}}{y_i!} e^{-\lambda t_i} = \frac{\prod_{i=1}^{n}(\lambda t_i)^{y_i}}{\prod_{i=1}^{n} y_i!} e^{-\lambda \sum_{i=1}^{n} t_i} \\
\ln(L(\lambda)) &= \sum_{i=1}^{n} y_i \ln(\lambda t_i) - \ln(\prod_{i=1}^{n} y_i!) - \lambda \sum_{i=1}^{n} t_i \\
\frac{\partial \ln L(\lambda)}{\partial \lambda} &= \sum_{i=1}^{n} \frac{y_i}{\lambda} - \sum_{i=1}^{n} t_i = \frac{1}{\lambda}\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} t_i = 0 \\
\Rightarrow \quad & \hat{\lambda} = \underline{\underline{\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}}}
\end{aligned}$$

$$\begin{aligned}
\text{E}(\hat{\lambda}) &= \frac{1}{\sum_{i=1}^{n} t_i}\sum_{i=1}^{n} \text{E}(Y_i) = \frac{1}{\sum_{i=1}^{n} t_i}\sum_{i=1}^{n} \lambda t_i = \lambda \\
\text{Var}(\hat{\lambda}) &= (\frac{1}{\sum_{i=1}^{n} t_i})^2 \text{Var}(\sum_{i=1}^{n} Y_i) \overset{\text{indep.}}{=} (\frac{1}{\sum_{i=1}^{n} t_i})^2 \sum_{i=1}^{n} \text{Var}(Y_i) \\
&= (\frac{1}{\sum_{i=1}^{n} t_i})^2 \cdot \sum_{i=1}^{n} \lambda t_i = \frac{\lambda}{\sum_{i=1}^{n} t_i}
\end{aligned}$$

d) We start by finding the second derivative of the log-likelihood at $\hat{\lambda}$:

$$\begin{aligned}
J(\lambda) &= \frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} Y_i \\
J(\hat{\lambda}) &= -\frac{1}{\hat{\lambda}^2}\sum_{i=1}^{n} Y_i = -\frac{\sum_{i=1}^{n} Y_i}{\left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}\right)^2} = \frac{1}{\frac{\sum_{i=1}^{n} Y_i}{(\sum_{i=1}^{n} t_i)^2}} = -\frac{\sum_{i=1}^{n} t_i}{\hat{\lambda}}
\end{aligned}$$

The Wald confidence interval then becomes:

$$[\hat{\lambda} - z_{\alpha/2}\sqrt{-1/J(\hat{\lambda})}, \hat{\lambda} + z_{\alpha/2}\sqrt{-1/J(\hat{\lambda})}] = \underline{\underline{[\hat{\lambda} - z_{\alpha/2}\sqrt{\hat{\lambda}/\sum_{i=1}^{n} t_i}, \hat{\lambda} + z_{\alpha/2}\sqrt{\hat{\lambda}/\sum_{i=1}^{n} t_i}]}}$$

With $\sum_{i=1}^{8} y_i = 30$ and $\sum_{i=1}^{8} t_i = 1.5$ we get $\hat{\lambda} = 30/1.5 = 20$. Also $z_{\alpha/2} = z_{0.05} = 1.645$ and we get the interval:

$$[\hat{\lambda} - z_{\alpha/2}\sqrt{\hat{\lambda}/\sum_{i=1}^{n} t_i}, \hat{\lambda} + z_{\alpha/2}\sqrt{\hat{\lambda}/\sum_{i=1}^{n} t_i}] = [20 - 1.645\sqrt{20/1.5},\ \ 20 + 1.645\sqrt{20/1.5}] = \underline{\underline{[14,\ \ 26]}}$$

Since the lower limit of this interval is less than 15 we can not from this interval claim that $\lambda > 15$. However, the interval is quite wide and most of the values in the interval are above 15 so it is probably wise to gather more data to get a more precise interval. (Or, if we have information from other sources than data we can combine this with the data information and calculate a Bayes interval. See the last problem on the exam fall 2010.)