STA500 Introduction to Probability and Statistics 2, autumn 2018.

Solution exercise set 7

Note on Bayesian statistics, exercise 1

Remember that the true density in this case is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-1)^2}{2}\right).$$

First expression: $h(x) = \exp\left(-\frac{(x-1)^2}{2}\right)$
Clearly $c = (2\pi)^{-1/2}$ will give us that $ch(x) = f(x)$

Second expression: $h(x) = 2\exp\left(-\frac{1}{2}x^2 + x\right)$
Note that

$$\begin{aligned} h(x) &= 2\exp\left(\underbrace{-\frac{1}{2}x^2 + x - \frac{1}{2}}_{-\frac{(x-1)^2}{2}} + \frac{1}{2}\right), \\ &= 2\exp(1/2)\exp\left(-\frac{(x-1)^2}{2}\right), \end{aligned}$$

thus $c = 1/(2\exp(1/2)\sqrt{2\pi})$ will make $f(x) = ch(x)$ true.

Third expression: $h(x) = \frac{1}{\pi}\exp\left(-\frac{(x-1)^2}{2} + 1\right)$
Clearly

$$h(x) = \frac{1}{\pi}\exp(1)\exp\left(-\frac{(x-1)^2}{2}\right)$$

which gives us the normalizing constant $c = \pi/(\exp(1)\sqrt{2\pi})$.

Note on Bayesian statistics, exercise 2

With the exponential density written as $f(t|\theta) = \theta e^{-\theta t}$ we get the likelihood

$$L(\theta) = \prod_{i=1}^{n} f(t_i|\theta) = \prod_{i=1}^{n} \theta e^{-\theta t_i} = \theta^n e^{-\theta \sum_{i=1}^{n} t_i}$$

The prior density is given as $p(\theta) = 2e^{-2\theta}$. We then get the following posterior distribution:

$$p(\theta|\text{data}) = c \cdot L(\theta) \cdot p(\theta) = c\theta^n e^{-\theta \sum_{i=1}^{n} t_i} 2e^{-2\theta} = c_2 \theta^{n+1-1} e^{-\theta(\sum_{i=1}^{n} t_i + 2)}$$

We see that this (as a function of θ) is on the same form as a gamma distribution with parameters $n+1$ and $1/(\sum_{i=1}^{n} t_i + 2)$. I.e. the distribution is a gamma distribution with parameters $n+1$ and $1/(\sum_{i=1}^{n} t_i + 2)$.

Note on Bayesian statistics, exercise 3

First we observe that $n = 10$ and $\bar{y} = 69/10 = 6.9$, thus

$$\hat{\lambda}_{Bayes} = \frac{a + n\bar{y}}{n + b^{-1}} = \frac{3 + 69}{10 + 1/2} \approx 6.86,$$

and

$$\hat{\lambda}_{MAP} = \frac{a + n\bar{y} - 1}{n + b^{-1}} = \frac{3 + 69 - 1}{10 + 1/2} \approx 6.76,$$

Note on Bayesian statistics, exercise 4

For the Bernulli(θ)-distribution we have that

$$L(\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{\sum_i y_i}(1-\theta)^{n-\sum_i y_i}.$$

Thus the posterior kernel can be written as

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

which we recognize as beta($a + \sum_{i=1}^{n} y_i, b + n - \sum_{i=1}^{n}$) distribution.

Exercise 2 in MLE-note:

a) The situations is characterised by:

- Independent trials - it is independent from plate to plate whether it is OK or not.
- We repeat the trials until success number k - we examine plates until k OK plates are found.
- We record "success"/not "success" - whether a plate is OK or not.
- The probability of "success" is the same in all trials - same probability of OK for each plate.

Then X = "number of plates we need to check to find k OK plates" is having a negative binomial distribution with parameters k and p.

b) For the negative binomial distribution we have $f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$, and we then get:

$$
\begin{aligned}
L(p) &= \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \binom{x_i - 1}{k-1} p^k (1-p)^{x_i - k} = \left(\prod_{i=1}^{n} \binom{x_i - 1}{k-1} \right) p^{nk} (1-p)^{\sum_{i=1}^{n} x_i - nk} \\
l(p) &= \ln L(p) = \ln \left(\prod_{i=1}^{n} \binom{x_i - 1}{k-1} \right) + \ln p^{nk} + \ln (1-p)^{\sum_{i=1}^{n} x_i - nk} \\
&= \ln \left(\prod_{i=1}^{n} \binom{x_i - 1}{k-1} \right) + nk \ln p + \left(\sum_{i=1}^{n} x_i - nk \right) \ln(1-p) \\
\frac{\partial l(p)}{\partial p} &= \frac{nk}{p} - \frac{\sum_{i=1}^{n} x_i - nk}{1-p} = 0 \\
& \frac{nk(1-p)}{p(1-p)} - \frac{p \sum_{i=1}^{n} x_i - pnk}{p(1-p)} = 0 \\
& nk(1-p) - p \sum_{i=1}^{n} x_i + pnk = 0 \\
& nk - p \sum_{i=1}^{n} x_i = 0 \\
& \quad \Rightarrow \quad p = \frac{nk}{\sum_{i=1}^{n} x_i}
\end{aligned}
$$

I.e MLE becomes $\hat{p} = \underline{\underline{\frac{nk}{\sum_{i=1}^{n} X_i} = \frac{k}{\bar{X}}}}$

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

$$\begin{aligned}
J(p) &= \frac{\partial^2 l(p)}{\partial p^2} = -\frac{nk}{p^2} - \frac{\sum_{i=1}^n x_i - nk}{(1-p)^2} \\
J(\hat{p}) &= -\frac{nk}{\hat{p}^2} - \frac{\sum_{i=1}^n x_i - nk}{(1-\hat{p})^2} = -\frac{nk(1-\hat{p})^2 + (\sum_{i=1}^n x_i - nk)\hat{p}^2}{\hat{p}^2(1-\hat{p})^2} \\
&= -\frac{nk - 2nk\hat{p} + nk\hat{p}^2 + \sum_{i=1}^n x_i\hat{p}^2 - nk\hat{p}^2}{\hat{p}^2(1-\hat{p})^2} = -\frac{nk - 2nk\hat{p} + \sum_{i=1}^n x_i(nk/\sum_{i=1}^n x_i)\hat{p}}{\hat{p}^2(1-\hat{p})^2} \\
&= -\frac{nk - nk\hat{p}}{\hat{p}^2(1-\hat{p})^2} = -\frac{nk}{\hat{p}^2(1-\hat{p})}
\end{aligned}$$

The Wald confidence interval then becomes:

$$[\hat{p} - z_{\alpha/2}\sqrt{-1/J(\hat{p})}, \hat{p} + z_{\alpha/2}\sqrt{-1/J(\hat{p})}] = \underline{\underline{[\hat{p} - z_{\alpha/2}\sqrt{\hat{p}^2(1-\hat{p})/nk}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}^2(1-\hat{p})/nk}]}}$$

d) For the negative binomial distribution we have that $\mu = k/p$ and by the invariance property of maximum likelihood estimators it follows that (since $\mu = k/p$ is a one-to-one transformation from μ to p and k is known)

$$\hat{\mu} = \frac{k}{\hat{p}} = \frac{k}{k/\bar{X}} = \underline{\underline{\bar{X}}}$$

$$\mathrm{E}(\hat{\mu}) = \mathrm{E}(\bar{X}) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \underline{\underline{\mu}}$$

I.e. the estimator is unbiased.

e) Since $\sum_{i=1}^{15} x_i = 6+6+7+6+8+6+9+6+7+8+6+10+6+8+7 = 106$ we get:

$$\begin{aligned}
\hat{p} &= \frac{nk}{\sum_{i=1}^{15} x_i} = \frac{15 \cdot 6}{106} = \underline{\underline{0.85}} \\
\hat{\mu} &= \bar{x} = 106/15 = \underline{\underline{7.07}}
\end{aligned}$$

With $\alpha = 0.05$ we have $z_{\alpha/2} = z_{0.025} = 1.96$ and we get

$$\begin{aligned}
&[\hat{p} - z_{\alpha/2}\sqrt{\hat{p}^2(1-\hat{p})/nk}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}^2(1-\hat{p})/nk}] \\
&= [0.85 - 1.96\sqrt{0.85^2(1-0.85)/15 \cdot 6},\ \ 0.85 + 1.96\sqrt{0.85^2(1-0.85)/15 \cdot 6}] = \underline{\underline{[0.78,\ \ 0.92]}}
\end{aligned}$$

Exercise 1:

a) First remember that for the exponential distribution we have

$$F(t) = P(T < t) = \int_0^t \lambda e^{-\lambda u} du = [-e^{-\lambda u}]_0^t = 1 - e^{-\lambda t}$$

Then:

$$\begin{aligned} P(T < 2|T > 1) &= \frac{P(T < 2 \cap T > 1)}{P(T > 1)} = \frac{P(1 < T < 2)}{1 - P(T < 1)} \\ &= \frac{F(2) - F(1)}{1 - F(1)} = \frac{1 - e^{-0.25\cdot 2} - (1 - e^{-0.25\cdot 1})}{1 - (1 - e^{-0.25\cdot 1})} = \underline{\underline{0.22}} \end{aligned}$$

Or we can do this simpler by using the fact that the exponential distribution is memoryless (notice this only holds for the exponential distribution):

$$P(T < 2|T > 1) \overset{memoryless}{=} P(T < 1) = 1 - e^{-0.25\cdot 1} = \underline{\underline{0.22}}$$

Since the times between failures are independent and exponentially distributed we have a homogeneous Poisson process. The number of failures during two years is then Poisson distributed with expectation $\lambda t = 0.25 \cdot 2 = 0.5$. I.e. if X=number of failures during two years:

$$P(X > 1) = 1 - P(X \le 1) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{0.5^0}{0!}e^{-0.5} - \frac{0.5^1}{1!}e^{-0.5} = \underline{\underline{0.09}}$$

b) Recall that the density of the exponential distribution with parameter λ is $f(t|\lambda) = \lambda e^{-\lambda t}$. The likelihood is then

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^n f(t_i|\lambda) = \prod_{i=1}^n \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^n t_i} \\ \ln(L(\lambda)) &= \ln(\lambda^n) + \ln(e^{-\lambda \sum_{i=1}^n t_i}) = n \ln(\lambda) - \lambda \sum_{i=1}^n t_i \\ \frac{\partial \ln(L(\lambda))}{\partial \lambda} &= \frac{n}{\lambda} - \sum_{i=1}^n t_i = 0 \qquad \Rightarrow \quad \hat{\lambda} = \underline{\underline{\frac{n}{\sum_{i=1}^n T_i}}} \end{aligned}$$

Estimate: $\hat{\lambda} = \frac{4}{36.4} = \underline{\underline{0.11}}$

c) The interpretation of a posterior distribution is that it is a (subjective) probability distribution which expresses our knowledge/uncertainty about a parameter (here λ) based on prior knowledge and information in data.

The posterior distribution is here given by:

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

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

A common estimator is the expectation in the posterior distribution. In the gamma distribution the expectation is the first parameter times the second parameter, i.e. we get:

$$\hat{\lambda}_{\text{Bayes}} = \frac{n + a}{\sum_{i=1}^n t_i + 1/b}$$

d) First we have to find the values of a and b. From the information in the text we have that the expectation in the prior distribution is $ab = 0.2$ and variance $ab^2 = 0.004$. Inserting the first in the latter we get $0.2b = 0.004$ which implies $b = 0.02$ and thus $a = 10$. We then get:

$$\hat{\lambda}_{\text{Bayes}} = \frac{4 + 10}{36.4 + 1/0.02} = 0.16$$

We see that the Bayes estimate falls approximately midways between the MLE estimate 0.11 and the prior estimate 0.20 (the prior expectation). This shows that in this case is the prior information and the information in data approximately equally weighted.

e) The posterior distribution for λ is a gamma distribution with parameters $a^* = n + a = 4 + 10 = 14$ and $b^* = 1/(\sum_{i=1}^n t_i + 1/b)) = 1/(36.4 + 1/0.02) = 0.0116$. To set up a Bayes interval we then just need to find the quantiles in this distribution, i.e. to find $\gamma_{1-\alpha/2,a^*,b^*}$ and $\gamma_{\alpha/2,a^*,b^*}$ such that

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

This can be done by using the relationship between the gamma distribution and the χ^2-distribution given in the list of transformation results in the collection of formulas. This result says that if X has a gamma(a, b)-distribution then $Y = (2/b)X$ has a χ^2_{2a}-distribution. Thus:

$$\begin{aligned} P(\frac{2}{b^*}\gamma_{1-\alpha/2,a^*,b^*} < \frac{2\lambda}{b^*} < \frac{2}{b^*}\gamma_{\alpha/2,a^*,b^*}) &= 1 - \alpha \\ P(\frac{2}{b^*}\gamma_{1-\alpha/2,a^*,b^*} < Y < \frac{2}{b^*}\gamma_{\alpha/2,a^*,b^*}) &= 1 - \alpha \end{aligned}$$

where Y is $\chi^2_{2a^*}$ I.e. $\gamma_{1-\alpha/2,a^*,b^*} = \frac{b^*}{2}\chi^2_{1-\alpha/2,2a^*}$ and $\gamma_{\alpha/2,a^*,b^*} = \frac{b^*}{2}\chi^2_{\alpha/2,2a^*}$. Thus the 90% Bayes interval becomes:

$$\begin{aligned} \left(\gamma_{1-\alpha/2,a^*,b^*}, \gamma_{\alpha/2,a^*,b^*}\right) &= \left(\frac{b^*}{2}\chi^2_{1-\alpha/2,2a^*}, \frac{b^*}{2}\chi^2_{\alpha/2,2a^*}\right) = (\frac{0.0116}{2}\chi^2_{0.95,28}, \frac{0.0116}{2}\chi^2_{0.05,28}) \\ &= (\frac{0.0116}{2} \cdot 16.928, \frac{0.0116}{2} \cdot 41.337) = \underline{\underline{(0.10,\ 0.24)}} \end{aligned}$$

Interpretation: There is a (subjective) 90% probability that the parameter λ is in this interval.

Exercise 2:

a) To calculate a posterior distribution we need the likelihood and the prior distribution. For observations $X_1, \ldots, X_n$ from a normal distribution with known variance σ^2 the likelihood becomes:

$$L(\mu) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

According to the information in the text, the prior density is:

$$p(\mu) = \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{(\mu-\mu_p)^2}{2\sigma_p^2}\right)$$

The posterior then becomes

$$\begin{aligned} p(\mu|\text{data}) &= c \cdot L(\mu) \cdot p(\mu) \\ &= \underline{\underline{c \cdot \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\mu-\mu_p)^2}{2\sigma_p^2}\right)}} \end{aligned}$$

Below it is also shown how we from the above starting point get the posterior distribution result given in the text (this was the optional part of the problem).

As a first step we notice the following property of the normal distribution density. The general normal distribution density is given as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2-2\mu x+\mu^2}{2\sigma^2}\right) = c\cdot\exp\left(-\frac{x^2-2\mu x}{2\sigma^2}\right)$$

where c is some constant. I.e. to identify the parameters of the normal distribution we only need to know the part $-\frac{x^2-2\mu x}{2\sigma^2}$ of the exponent. The mean is given after the 2 in the x-term and the variance after the 2 in the denominator.

Starting from the expression for the posterior given above (and note that this is a function of μ) we get:

$$\begin{aligned} p(\mu|\text{data}) &= c \cdot \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\mu-\mu_p)^2}{2\sigma_p^2}\right) \\ &= c_2 \cdot \exp\left(-\frac{\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + \sum_{i=1}^n \mu^2}{2\sigma^2}\right) \exp\left(-\frac{\mu^2-2\mu\mu_p+\mu_p^2}{2\sigma_p^2}\right) \\ &= c_2 \cdot \exp\left(-\frac{(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2)\sigma_p^2 + (\mu^2-2\mu\mu_p+\mu_p^2)\sigma^2}{2\sigma^2\sigma_p^2}\right) \\ &= c_2 \cdot \exp\left(-\frac{(n\sigma_p^2+\sigma^2)\mu^2 - 2(\sum_{i=1}^n x_i\cdot\sigma_p^2 + \mu_p\sigma^2)\mu + \sum_{i=1}^n x_i^2\cdot\sigma_p^2 + \mu_p^2\sigma^2}{2\sigma^2\sigma_p^2}\right) \\ &= c_3 \cdot \exp\left(-\frac{\mu^2 - 2[(\sum_{i=1}^n x_i\cdot\sigma_p^2 + \mu_p\sigma^2)/(n\sigma_p^2+\sigma^2)]\mu}{2\sigma^2\sigma_p^2/(n\sigma_p^2+\sigma^2)}\right) \end{aligned}$$

This is now as a function of μ on the form of a normal distribution density, and if we compare with our general normal density above we see that this posterior distribution is a normal distribution with mean $\frac{\sum_{i=1}^{n} x_i \cdot \sigma_p^2 + \mu_p \sigma^2}{n\sigma_p^2 + \sigma^2}$ and variance $\frac{\sigma^2 \sigma_p^2}{n\sigma_p^2 + \sigma^2}$ as it should be.

b) The standard Bayes estimate is the mean in the posterior distribution, and we here have:

$$\hat{\mu}_{\text{Bayes}} = \frac{\sum_{i=1}^{n} x_i \cdot \sigma_p^2 + \mu_p \sigma^2}{n\sigma_p^2 + \sigma^2} = \frac{\sum_{i=1}^{n} x_i \cdot \sigma_p^2}{n\sigma_p^2 + \sigma^2} + \frac{\mu_p \sigma^2}{n\sigma_p^2 + \sigma^2} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma^2/n}\bar{x} + \frac{\sigma^2/n}{\sigma_p^2 + \sigma^2/n}\mu_p$$

This shows that the Bayes estimate is a weighted average of the data average $\bar{x}$ and the prior mean μ_p. We see that the data is given most weight when $\sigma_p^2 > \sigma^2/n = \text{Var}(\bar{X})$ and that the prior is given most weight when $\sigma_p^2 < \sigma^2/n = \text{Var}(\bar{X})$. I.e. the part with smallest variance of the data average and the prior mean is given most weight.

With the given data and prior information the estimate becomes:

$$\hat{\mu}_{\text{Bayes}} = \frac{5 \cdot 13.6 \cdot 1.5^2 + 11 \cdot 5^2}{5 \cdot 1.5^2 + 5^2} = 11.8$$

Since $\sigma_p^2 = 1.5^2 = 2.25 < \sigma^2/5 = 5^2/5 = 5$ the prior information is given most weight in this case. We see that 11.8 is closer to $\mu_p = 11$ than to $\bar{x} = 13.6$.

c) The posterior distribution is a normal distribution with mean 11.8 and variance $\sigma_p^2\sigma^2/(n\sigma_p^2 + \sigma^2) = 1.5^2 \cdot 5^2/(5 \cdot 1.5^2 + 5^2) = 1.55$. Thus we have the posterior probability:

$$\begin{aligned} P(-z_{\alpha/2} < \frac{\mu - 11.8}{\sqrt{1.55}} < z_{\alpha/2}) &= 1 - \alpha \\ P(11.8 - z_{\alpha/2} \cdot \sqrt{1.55} < \mu < 11.8 + z_{\alpha/2} \cdot \sqrt{1.55}) &= 1 - \alpha \end{aligned}$$

With $\alpha = 0.05$ and thus $z_{\alpha/2} = 1.96$ we get the 95% Bayesian interval:

$$(11.8 - 1.96 \cdot \sqrt{1.55},\ 11.8 + 1.96 \cdot \sqrt{1.55}) = (9.4,\ 14.2)$$

The 95% confidence interval for μ in the normal distribution when σ is known is

$$(\bar{x} - 1.96 \cdot \sigma/\sqrt{n},\ \bar{x} + 1.96 \cdot \sigma/\sqrt{n}) = (13.6 - 1.96 \cdot 5/\sqrt{5},\ 13.6 + 1.96 \cdot 5/\sqrt{5}) = (9.2,\ 18.0)$$

We see that the confidence interval is much wider which is natural. When we use less information there is more uncertainty.

d) A nice thing by using a Bayesian approach is that we can use the knowledge from the experts based on other sources than data from the current situtation and combine this knowledge with information from data. This gives us more precise estimates.

A potential danger by using the Bayesian approach is if the expert knowledge is wrong. For instance if there has been some recent development in the pollution level which the experts are not aware of (for instance a recent unknown discharge) then the (wrong) prior information from the expert may drag the estimate in a wrong direction. This would in paricular be problematic if a strong prior distribution is used (i.e. a prior distribution with low variance reflecting that the experts think they have very precise knowledge).