Solution exam November 18, 2013

Problem 1:

a) Let X be the compressive strenght of a randomly selected module.

$$\begin{aligned} P(X > 45) &= 1 - P(X \leq 45) = 1 - P(Z < \frac{45-49}{3}) \\ &= 1 - P(Z < -1.33) = 1 - 0.0918 = 0.9082 \approx \underline{\underline{0.91}} \\ P(45 < X < 50) &= P(X < 50) - P(X < 45) = P(Z < \frac{50-49}{3}) - P(Z < \frac{45-49}{3}) \\ &= P(Z < 0.33) - P(Z < -1.33) = 0.6293 - 0.0918 = \underline{\underline{0.5375}} \\ P(\min(X_1, \ldots, X_{10}) > 45) &= P(X_1 > 45 \cap \ldots \cap X_{10} > 45) \\ &\overset{indep.}{=} P(X_1 > 45) \cdots P(X_{10} > 45) = 0.9082^{10} = \underline{\underline{0.38}} \end{aligned}$$

b) We have a situation where

- We check "success" or not "success" in each trial - whether the module pass the test or not.
- The probability of "success" is the same in all trials, $p = P(X > 45) = 0.91$.
- Independent trials - independent from module to module whether it passes the test or not.
- The trials continue until a specified number of "successes" are observed - we test modules until we have found $k = 10$ modules which pass the test.

Y is then having a negative binomial distribution with $k = 10$ and $p = 0.91$.

$$\begin{aligned} P(Y \geq 12) &= 1 - P(Y < 12) = 1 - (P(Y = 10) + P(Y = 11)) \\ &= 1 - \binom{10-1}{10-1} 0.91^{10}(1-0.91)^{10-10} - \binom{11-1}{10-1} 0.91^{10}(1-0.91)^{11-10} \\ &= 1 - 0.39 - 0.35 = \underline{\underline{0.26}} \end{aligned}$$

$$\mathrm{E}(X) = \frac{k}{p} = \frac{10}{0.91} = \underline{\underline{11}}$$

Problem 2:

a) The number of cars arriving during 6 minutes= $1/10 = 0.1$ hour, Y, is Poisson-distributed with expectation $\lambda t = 8 \cdot 0.1 = 0.8$.

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \frac{0.8^0}{0!} e^{-0.8} = \underline{\underline{0.55}}$$

The time between consecutive events in a Poisson process is exponentially distributed with parameter λ. We let T denote the time between two consequtive cars. Then we get:

$$P(T < 1/6) = \int_0^{1/6} 8e^{-8t} dt = [-e^{-8t}]_0^{1/6} = -e^{-8 \cdot (1/6)} - (-1) = \underline{\underline{0.74}}$$

If we let S_2 denote the time until the second car arrive we may solve the last question by finding $P(S_2 < 0.1)$ where S_2 has a gamma distribution with parameters $\alpha = 2$ and $\beta = 1/\lambda = 1/8$. Or, easier, we can use the observation that:

$$P(S_2 < 0.1) = P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - \frac{0.8^0}{0!} e^{-0.8} - \frac{0.8^1}{1!} e^{-0.8} = \underline{\underline{0.19}}$$

b) The value of X_{n+1} is given from X_n plus the number of new cars arriving during the wash of car $n+1$ (which is independent of the past since cars arrive as a Poisson process) - i.e. we do not need to know $X_{n-1}, X_{n-2}, \ldots$ and the process is thus Markov. The state space is $\{0, 1, 2, 3\}$ since these are the possible number of cars waiting in line.

The third row in the transition probability matrix shows the probabilities for the various number of cars in line after the next wash when there were 2 cars waiting at the end of a wash. It is impossible to have no cars in line after the next wash as only one car leaves the line, thus $p_{20} = 0$. The other possibilities depend on how many new cars arrive during the next car wash. If we let, as in point a), Y denote the number of new cars arriving during the 0.1 hour a wash take we get:

$$\begin{aligned} p_{21} &= P(Y = 0) = \frac{0.8^0}{0!} e^{-0.8} = \underline{\underline{0.45}}, \qquad p_{22} = P(Y = 1) = \frac{0.8^1}{1!} e^{-0.8} = \underline{\underline{0.36}} \qquad \text{and} \\ p_{23} &= P(Y \geq 1) = 1 - P(Y = 0) - P(Y = 1) = \underline{\underline{0.19}} \end{aligned}$$

The number of steps we stay in state 3 has a geometric distribution (independent from step to step if we go to a new state or not, same probability of going to a new state in each step, we record whether we go to a new state or not and we record the number of steps until first time we leave the state) with probability $p = 0.45$ of leaving the state. The expected number of steps we stay in the state is thus $1/p = 1/0.45 = \underline{\underline{2.2}}$

c)

$$\begin{aligned} P(X_2 = 2 | X_1 = 0) &= \underline{\underline{0.14}} \\ P(X_{12} = 2 | X_{10} = 1) &= \underline{\underline{0.19}} \\ P(X_5 = 3 | X_3 = 2, X_2 = 1, X_1 = 1) &\overset{\text{Markov}}{=} P(X_5 = 3 | X_3 = 2) = \underline{\underline{0.20}} \\ P(X_{n+2} = 1, X_{n+1} = 2 | X_n = 3) &= p_{32} p_{21} = 0.45 \cdot 0.45 = \underline{\underline{0.203}} \\ P(X_6 = 1, X_5 = 2, X_4 = 3 | X_2 = 2, X_1 = 1) &= P(X_6 = 1, X_5 = 2, X_4 = 3 | X_2 = 2) \\ &\overset{\text{Markov}}{=} p_{23}^2 p_{32} p_{21} = 0.20 \cdot 0.45 \cdot 0.45 = \underline{\underline{0.04}} \end{aligned}$$

d)

From
$$\begin{pmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} = \begin{pmatrix} 0.45 & 0.45 & 0 & 0 \\ 0.36 & 0.36 & 0.45 & 0 \\ 0.14 & 0.14 & 0.36 & 0.45 \\ 0.05 & 0.05 & 0.19 & 0.55 \end{pmatrix} \begin{pmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}$$

we get
$$\begin{aligned} \pi_0 &= 0.45\pi_0 + 0.45\pi_1 \\ \pi_1 &= 0.36\pi_0 + 0.36\pi_1 + 0.45\pi_2 + \\ \pi_2 &= 0.14\pi_0 + 0.14\pi_1 + 0.36\pi_2 + 0.45\pi_3 \\ \pi_3 &= 0.05\pi_0 + 0.05\pi_1 + 0.19\pi_2 + 0.55\pi_3 \end{aligned}$$

These and $\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$ are the steady state equations.

Expected number of cars waiting in steady state:

$0\pi_0 + 1\pi_1 + 2\pi_2 + 3\pi_3 = 1 \cdot 0.32 + 2 \cdot 0.25 + 3 \cdot 0.17 = \underline{\underline{1.33}}$

With three cars in line after a wash the expected number in line after the next wash is

$0p_{30} + 1p_{31} + 2p_{32} + 3p_{33} = 2 \cdot 0.45 + 3 \cdot 0.55 = \underline{\underline{2.55}}$

The last expectation is clearly largest which is reasonable since it is much more likely to also have a large number of cars in line after the next wash when there are 3 cars in line after a wash than in the steady state situation.

With a larger value of λ (12 instead of 8) more cars will arrive to the car wash per time unit. This implies that it is less likely to have no cars in line, and more likely to have many cars in line. In particular π_0 will decrease, π_3 will increase and both the expectations will increase.

Problem 3:

a)

$$\begin{aligned}
E(\hat\beta_1) &= \frac{1}{2}E(\bar T) + \frac{1}{2}E(\bar S) = \frac{1}{2}\beta + \frac{1}{2}\beta = \underline{\underline{\beta}} \\
E(\hat\beta_2) &= \frac{\sum_{i=1}^n E(T_i) + \sum_{i=1}^m E(S_i)}{n+m} = \frac{\sum_{i=1}^n \beta + \sum_{i=1}^m \beta}{n+m} = \frac{n\beta + m\beta}{n+m} = \underline{\underline{\beta}}
\end{aligned}$$

I.e. both estimators are unbiased. Further, since $\text{Var}(\bar T) = Var(T)/n = \beta^2/n$, we get (the calculation of the expression for $\text{Var}(\hat\beta_1)$ is just included for completeness):

$$\begin{aligned}
\text{Var}(\hat\beta_1) &= \text{Var}(\frac{1}{2}\bar T + \frac{1}{2}\bar S) \overset{\text{indep.}}{=} (\frac{1}{2})^2\text{Var}(\bar T) + (\frac{1}{2})^2\text{Var}(\bar S) \\
&= \frac{1}{4}\frac{\beta^2}{n} + \frac{1}{4}\frac{\beta^2}{m} = \frac{\beta^2}{4}(\frac{1}{n} + \frac{1}{m}) = \frac{\beta^2}{4}(\frac{1}{41} + \frac{1}{17}) = \underline{\underline{0.021\beta^2}} \\
\text{Var}(\hat\beta_2) &= \text{Var}(\frac{1}{n+m}[\sum_{i=1}^n T_i + \sum_{i=1}^m S_i]) = \frac{1}{(n+m)^2}\text{Var}[\sum_{i=1}^n T_i + \sum_{i=1}^m S_i] \\
&\overset{\text{indep.}}{=} \frac{1}{(n+m)^2}[\sum_{i=1}^n \text{Var}(T_i) + \sum_{i=1}^m \text{Var}(S_i)] = \frac{1}{(n+m)^2}[n\beta^2 + m\beta^2] \\
&= \frac{\beta^2}{n+m} = \frac{\beta^2}{41+17} = \underline{\underline{0.017\beta^2}}
\end{aligned}$$

Since both estimators are unbiased and $\text{Var}(\hat\beta_1) > \text{Var}(\hat\beta_2)$, $\underline{\underline{\hat\beta_2 \text{ is the best estimator}}}$.

b)

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

I.e the MLE becomes $\underline{\underline{\hat\beta = \frac{1}{n}\sum_{i=1}^n T_i = \bar T}}$.

Next we find the second derivative of the log-likelihood at $\hat\beta$:

$$\begin{aligned}
\frac{\partial l(\beta)}{\partial\beta} &= -\frac{n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^n t_i \\
J(\beta) &= \frac{\partial^2 l(\beta)}{\partial\beta^2} = \frac{n}{\beta^2} - \frac{2}{\beta^3}\sum_{i=1}^n t_i \\
J(\hat\beta) &= \frac{n}{\hat\beta^2} - \frac{2n}{\hat\beta^3}\frac{\sum_{i=1}^n t_i}{n} = \frac{n}{\hat\beta^2} - \frac{2n}{\hat\beta^2} = -\frac{n}{\hat\beta^2}
\end{aligned}$$

The Wald confidence interval then becomes:

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

With the given data we get $\hat\beta = 107/41 = \underline{\underline{2.61}}$. Further since $z_{\alpha/2} = z_{0.025} = 1.96$ we get

$$[\hat\beta - z_{\alpha/2}\sqrt{\hat\beta^2/n}, \hat\beta + z_{\alpha/2}\sqrt{\hat\beta^2/n}] = [2.61 - 1.96\sqrt{2.61^2/41},\ \ 2.61 + 1.96\sqrt{2.61^2/41}] = \underline{\underline{[1.81,\ \ 3.41]}}$$

The Wald confidence interval can safely be used since we have a relatively large amount of data.

c) With prior distribution given in the problem text and the likelihood in b) we get the posterior:

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

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 $b^* = 1/(\sum_{i=1}^n t_i + 1/b)$. I.e. the distribution is an $\underline{\underline{\text{inverse gamma distribution with parameters } n + a \text{ and } 1/(\sum_{i=1}^n t_i + 1/b).}}$

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:

$$\hat\beta_{\text{Bayes}} = \frac{1}{b^*(a^*-1)} = \underline{\underline{\frac{\sum_{i=1}^n t_i + 1/b}{n+a-1}}} = \frac{107 + 1/0.11}{41 + 4 - 1} = \underline{\underline{2.64}}$$

The 95% Bayes interval is the interval

$$P(\xi_{0.975,a^*,b^*} < \beta < \xi_{0.025,a^*,b^*}) = 0.95$$

in the posterior inverse gamme distribution. With $a^* = n + a = 41 + 4 = 45$ and $b^* = 1/(\sum_{i=1}^n t_i + 1/b)) = 1/(107 + 1/0.11) = 0.0086$ and we get by using the result given in the text that

$$\begin{aligned}
\xi_{0.975,45,0.0086} &= \frac{2}{0.0086\chi^2_{0.025,90}} = \frac{2}{0.0086 \cdot 118.136} = 1.97 \\
\xi_{0.025,45,0.0086} &= \frac{2}{0.0086\chi^2_{0.975,90}} = \frac{2}{0.0086 \cdot 65.647} = 3.54
\end{aligned}$$

I.e. the 95% Bayes interval is $\underline{\underline{(1.97,\ 3.54)}}$. This interval has approximately the same width as the confidence interval, and it is shifted slightly to the right. We do not gain much precision by adding the prior information in this case since we have quite a lot of data information.