Assignment 3 Solution

1. (a) We are interested in estimating $Y_1 + Y_2$. Note that

$$Y_2 = \rho Y_1 + Z_2 = \rho(\rho Y_0 + Z_1) + Z_2 = \rho^2 Y_0 + \rho Z_1 + Z_2$$

where $\rho = 7/10$ and $Y_0 = 2$ (million). Therefore,

$$Y_1 + Y_2 = (\rho Y_0 + Z_1) + (\rho^2 Y_0 + \rho Z_1 + Z_2) = (\rho + \rho^2)Y_0 + (\rho + 1)Z_1 + Z_2$$

One possible prediction for $Y_1 + Y_2$ is simply its mean, which is given by

$$E[Y_1 + Y_2] = (\rho + \rho^2)Y_0 + (\rho + 1)E[Z] + E[Z] = 5.08$$

(b) Let $S = Y_1 + \cdots + Y_{12}$, then S is the total profit over the next 12 months. To compute its mean and variance we should first note that

$$\begin{aligned}
Y_1 &= \rho Y_0 + Z_1 \\
Y_2 &= \rho^2 Y_0 + \rho Z_1 + Z_2 \\
Y_3 &= \rho^3 Y_0 + \rho^2 Z_1 + \rho Z_2 + Z_3 \\
&\vdots \\
Y_{12} &= \rho^{12} Y_0 + \rho^{11} Z_1 + \rho^{10} Z_2 + \cdots + \rho Z_{11} + Z_{12}
\end{aligned}$$

Adding them up gives:

$$\begin{aligned}
S &= \sum_{i=1}^{12} \rho^i Y_0 + \sum_{i=0}^{11} \rho^i Z_1 + \sum_{i=0}^{10} \rho^i Z_2 + \cdots + \sum_{i=0}^{1} \rho^i Z_{11} + Z_{12} \\
&= \frac{\rho(1-\rho^{12})}{1-\rho} Y_0 + \frac{(1-\rho^{12})}{1-\rho} Z_1 + \frac{(1-\rho^{11})}{1-\rho} Z_2 + \cdots + (1+\rho) Z_{11} + Z_{12}
\end{aligned}$$

Computing the mean and the variance follows easily, since $Y_0 = 2$ and the Z_i's are iid log-normal rv's with mean 1 and variance 2.

$$\begin{aligned}
E[S] &= \frac{2\rho(1-\rho^{12})}{1-\rho} + \frac{(1-\rho^{12})}{1-\rho} E[Z] + \frac{(1-\rho^{11})}{1-\rho} E[Z] + \cdots + (1+\rho)E[Z] + E[Z] \\
&= \frac{1}{1-\rho} \left[2\rho(1-\rho^{12}) + (1-\rho^{12}) + (1-\rho^{11}) + \cdots + (1-\rho^2) + (1-\rho) \right] \\
&= 36.93195067129199 \\
\text{Var}(S) &= \frac{(1-\rho^{12})^2}{(1-\rho)^2} \text{Var}(Z) + \frac{(1-\rho^{11})^2}{(1-\rho)^2} \text{Var}(Z) + \cdots + (1+\rho)^2 \text{Var}(Z) + \text{Var}(Z) \\
&= \frac{2}{(1-\rho)^2} \left[(1-\rho^{12})^2 + (1-\rho^{11})^2 + \cdots + (1-\rho^2)^2 + (1-\rho)^2 \right] \\
&= 185.7450278315905
\end{aligned}$$

(c) To approximate $S = Y_1 + \cdots + Y_{12}$ we use a special version of the CLT (for dependent rv's), that tells us that

$$\frac{S - E[S]}{\sqrt{\text{Var}(S)}} \overset{D}{\approx} N(0,1)$$

so

$$\begin{aligned} P(S \leq 10) &= P\left(\frac{S - E[S]}{\sqrt{\text{Var}(S)}} \leq \frac{10 - E[S]}{\sqrt{\text{Var}(S)}}\right) \\ &\approx P\left(N(0,1) \leq \frac{10 - E[S]}{\sqrt{\text{Var}(S)}}\right) \\ &= P(N(0,1) \leq -1.9761) \\ &= 0.0244 \end{aligned}$$

2. Given the historical data for 81 weeks, various forecasting models should be fitted and tested for providing the best 5 week forecast (i.e. for the 86th week). The models to be fitted are Autoregressive of order 1,2 and 3. The following procedures are taken:

1) For the first 41 data points (that is oil prices O_i, $i = 1, 2, \ldots, 41$) find parameters for Autoregressive models of order 1.
2) For $i = 46, 47, \ldots, 81$, predict the oil price P_i (the expected value) using parameters from Step 1 and data points from 1 till $(i - 5)$.
3) Find sum of square of error between predicted and actual values i.e. $SE = \sum_{i=46}^{81}(O_i - P_i)^2$
4) Repeat Steps 1-3 for each of the models to be tested
5) Choose the model with minimum SE and re-evaluate the parameters based on the entire data set of 81 weeks.
6) Using the chosen model, predict the value in 86th week i.e. 5 weeks from the last data point.

Matlab 7.0 is used for the various fits. Results are displayed below.

Autoregressive Order	Fit Parameters	Square Error
1 (red)	$b = 4.0539$, $a = 0.9664$	1.6739e+004
2 (green)	$b = 2.6740$, $a_1 = 0.8265$, $a_2 = 0.1568$	1.6148e+004
3 (blue)	$b = 2.5551$, $a_1 = 0.8183$, $a_2 = 0.1171$, $a_3 = 0.0499$	1.6818e+004

So Autoregressive of order 2 is the best-fit model among the three candidates. We then use the entire 81 points to re-calibrate the parameters and get $b = -0.2877$, $a_1 = 0.9456$, $a_2 = 0.0528$. Observe that a_2 is very small in comparison with a_1 and that the difference in Square Errors of fits of order 1 and is only around 4%. This indicates that a fit of order 1 might well be used. The parameters for this fit are: $b = 0.1018$, $a = 0.9949$. The predicted oil price is 41.98 if Autoregressive of order 1 is applied and 40.54 for order 2.

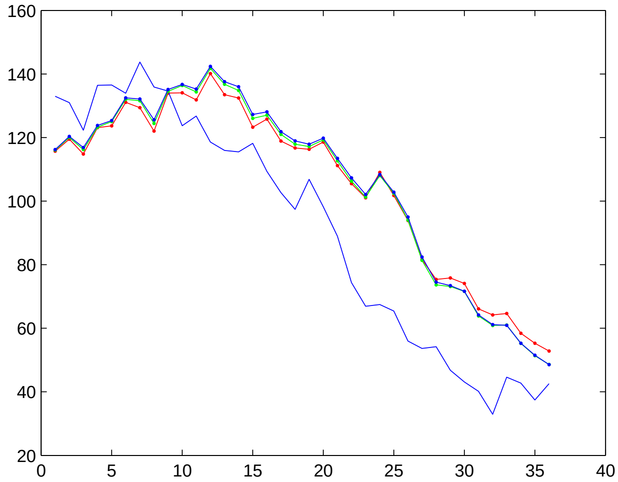

Figure 1: Comparison of Autoregressive of Order 1 (red), 2 (green), and 3 (blue).

3. (a) Let X_n be the number of umbrellas the man has at hand on the nth commute (a commute can be either from his house to the office or from the office to his house). The state space of this chain is then $S = \{0, 1, 2, \ldots, r\}$. Note that if it is raining when he leaves, he will take an umbrella with him (if there's any available). For example, if $_n = 3$ and it's raining when he leaves, $X_{n+1} = r - 3 + 1$, and if it's not raining, $X_{n+1} = r - 3$. When $X_n = 0$, then the following transition is to $X_{n+1} = r$, regardless of the weather. Therefore, the corresponding transition matrix is given by

$$P = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 & 1 \\ 0 & 0 & \ldots & 0 & q & p \\ 0 & 0 & \ldots & q & p & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & q & \ldots & 0 & 0 & 0 \\ q & p & \ldots & 0 & 0 & 0 \end{bmatrix}$$

(b) The steady state probability vector $\pi = (\pi_0, \pi_1, \ldots, \pi_r)$ satisfies the following system of equations:

$$\begin{aligned} q\pi_r &= \pi_0 \\ q\pi_{r-1} + p\pi_r &= \pi_1 \\ q\pi_{r-2} + p\pi_{r-1} &= \pi_2 \\ &\vdots \\ q\pi_1 + p\pi_2 &= \pi_{r-1} \\ \pi_0 + p\pi_1 &= \pi_r \\ \pi_0 + \pi_1 + \cdots + \pi_r &= 1 \end{aligned}$$

Using the first equation and substituting it on the previous to last we get

$$\pi_r = \pi_1 = \frac{1}{q}\pi_0,$$

and substituting these two values on the second equation we get

$$\pi_{r-1} = \frac{1}{q}\pi_0.$$

At this point our guess is that $pi_i = \frac{1}{q}\pi_0$ for all $i = 1, 2, \ldots, r$, and by substituting these values on the last equation we obtain

$$1 = \sum_{r=0}^{r} \pi_i = \pi_0 + \sum_{r=1}^{r} \frac{1}{q}\pi_0 = \frac{r+q}{q}\pi_0.$$

With these we conclude that the solution should be

$$\pi_0 = \frac{q}{r+q} \quad \text{and} \quad \pi_i = \frac{1}{r+q}.$$

To formally prove that this is the steady state probability vector it suffices to verify that it satisfies the system of equations defined above.

(c) The man gets wet if the chain is currently in state 0 and it rains, so

$$\text{Fraction of time he gets wet } = p\pi_0 = \frac{pq}{r+q}$$

(d) We want to maximize

$$f(p) = p\pi_0 = \frac{pq}{r+q} = \frac{p(1-p)}{r+1-p} = \frac{p(1-p)}{4-p}.$$

Note that

$$f'(p) = \frac{p^2 - 8p + 4}{(4-p)^2},$$

which is zero when $p = 4 \pm \sqrt{12}$. We discard $p = 4 + \sqrt{12}$, since only $p = 4 - \sqrt{12}$ is a probability; we just need to check whether its a maximum or minimum. To do this we note that

$$f''(p) = \frac{24}{(p-4)^3},$$

which is negative for $p = 4 - \sqrt{12}$, and therefore proves that it is a maximum.

NOTE: The following Markov Chain could have been used instead of the one suggested in the solution:

$$X_n = \text{ number of umbrellas at home (office) on the } n\text{th day.}$$

Its corresponding transition matrix has the same steady-state probabilities as the ones given above, but the proportion of time the person gets wet in this case is given by

$$\frac{1}{2}(p\pi_0 + pq\pi_r) = \frac{pq}{r+q}$$

This expression reflects the fact that it is possible to get wet in only two ways: in the morning, if all umbrellas are at the office (which occurs with probability $\pi_0 p$), and in the afternoon, if all umbrellas are at home AND it did not rain in the morning (which occurs with probability $\pi_r p(1-p)$). Since it is not possible to get wet twice in a day, simply summing these probabilities will double-count the quantity of interest, which in this case is the long-run proportion of time the individual gets wet (either in the morning or in the afternoon). To obtain the correct solution, it is necessary to divide the sum of the probabilities by 2. Note that this works only because rain is equally likely to occur in the morning and in the afternoon.

4. (a) The steady state probabilities $\pi = (\pi_0, \pi_1, \pi_2, \pi_3)$ satisfy

$$\pi P = \pi, \quad \pi_0 + \pi_1 + \pi_2 + \pi_3 = 1,$$

which has solution

$$(\pi_0, \pi_1, \pi_2, \pi_3) = (\frac{2}{13}, \frac{7}{13}, \frac{2}{13}, \frac{2}{13})$$

(b) The long-run expected average cost per day is given by

$$\text{Expected average cost } = 1000\pi_1 + 3000\pi_2 + 6000\pi_3 = 1922.70$$

(c) For state 0,

$$\text{Expected recurrence time } = \frac{1}{\pi_0} = 6.5$$

5. (a) Let X_n be the state of the recorder at the end of the nth year since it was bought. The possible states for the recorder are 0 = "new", 1 = "did not fail on the first year", 2 = "warranty expired without being used", and 3 = "failed". For example, if $X_1 = 1$, it means that the recorder did not fail during its first year. Clearly, $X_0 = 0$. The transition matrix of this chain is

$$P = \begin{bmatrix} 0 & 0.99 & 0 & 0.01 \\ 0 & 0 & 0.95 & 0.05 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(b) In the notation of the book, f_{02} = probability that the manufacturer will not have to honor the warranty, and f_{03} = probability that he will have to honor the warranty, and they satisfy

$$\begin{aligned} f_{03} &= p_{01}f_{13} + p_{03}f_{33} = 0.99f_{13} + 0.01 \\ f_{13} &= p_{12}f_{23} + p_{13}f_{33} = 0.05 \\ f_{23} &= 0 \\ f_{33} &= 1 \end{aligned}$$

And the solution is given by $f_{03} = 0.0595$.