Discrete Stochastic Modelling May/June 2003
Solutions

Question 1

i) (a) A random walk is a process $\{X_n: n \geq 0\}$ such that the increments $J_n = X_n - X_{n-1}$ form a sequence of i.i.d. r.v.s.

(b) A simple RW is one where the distribution of the J_n is $p(+1) = p, p(-1) = q = 1 - p$.

(c) A SSRW is a SRW with $p = 0.5$.

ii) (a) $E(J_n) = 1/3$, $\text{Var}(J_n) = \sigma^2 < \infty$ (value not important). Therefore

$E(X_n) = X_0 + n/3,\ \text{Var}(X_n) = n\sigma^2$, so that

$$P(X_n > 0) \simeq P\left(\frac{X_n - X_0 - n/3}{\sigma\sqrt{n}} > \frac{-X_0 - n/3}{\sigma\sqrt{n}} \right) \simeq \Phi\left(\frac{X_0}{\sigma\sqrt{n}} + \frac{\sqrt{n}}{3\sigma} \right).$$

This tends to 1 as $n \to \infty$, so that $P(X_n > 0 \text{ for some } n) \geq \sup_n P(X_n > 0) = 1$.

(b) The equation is $h_x = \frac{1}{4}h_{x-1} + \frac{1}{3}h_x + \frac{1}{4}h_{x+1} + \frac{1}{6}h_{x+2}$.

Seek solutions of the form $h_x = \theta^x$: we have $3 - 8\theta + 3\theta^2 + 2\theta^3 = 0$, implying that $(1 - \theta)(3 - 5\theta - 2\theta^2) = 0$, which in turn reduces to $(1 - \theta)(1 - 2\theta)(3 + \theta) = 0$. The solutions are $\theta = 1$, 0.5 or -3.

The value of -3 is clearly no good, since probabilities must be between 0 and 1. The value of 1 does not violate this rule, but it would imply that the process is recurrent, a fact which seems highly unlikely in view of the positive drift. This therefore leaves the solution $h_x = 0.5^x$.

(iii) There are two parts to this: converting the Uniform variables into the desired distribution and using the J_n to generate the chain. Something like this would do for the second part:

1	Time	J	X
2	0		=X0
3	=1+A2	?	=C2+B3

The question now becomes one of generating the J_n. The easiest way is to use a lookup table, with the cumulative probabilities in the first column and the values in the second, e.g.

0	–1
=3/12	0
=7/12	1
=10/12	2

If this table is named LT, the call =VLOOKUP(Random,LT,2,TRUE) will generate a value from the required distribution.

There are other methods available, for example using IF statements, which are equally acceptable.

Question 2

a) (i) The assumption is that the future movements of the director depend only on the present location and not on wherever the director has been in the past.

(ii) Unlikely. If the director has only recently visited one particular country, one might assume that it would be less likely that another visit will occur soon after.

(iii) There are various aspects to test: for example, one might wish to test whether $p_{ijk} = p_{kjk}$ for $k \neq i$. A contingency table test might be appropriate, but it needs a large collection of data.

b) The estimated probabilities are given by $\hat{p}_{ij} = n_{ij} / n_{i+}$, where n_{ij} is the number of observed transitions directly from i to j.
In this case, the estimated matrix of transition probabilities is

$$\begin{pmatrix} 0.143 & 0.571 & 0.143 & 0.143 & 0 \\ 0.25 & 0 & 0.375 & 0.25 & 0.125 \\ 0.25 & 0 & 0.167 & 0.25 & 0.333 \\ 0.24 & 0.04 & 0.20 & 0.20 & 0.32 \\ 0.10 & 0.15 & 0.30 & 0.35 & 0.10 \end{pmatrix}$$

c) It is irreducible, since every state can be reached from every other state in at most 2 steps. It is also aperiodic, since there is at least one state i such that $p_{ii} > 0$.

d) From the Law of Total Probability, we have

$$P(X_{m+n} = j \mid X_0 = i) = \sum_k P(X_n = k \mid X_0 = i) P(X_{m+n} = j \mid X_n = k, X_0 = i).$$

The Markov property ensures that the final probability is $P(X_{m+n} = j \mid X_n = k)$.
Therefore, in matrix form, $P^{(m+n)} = P^{(n)} P^{(m)}$.
Since $P^{(1)} = P$, we deduce by induction that $P^{(n)} = P^n$.

e) From Italy this week, the most likely destination for next week is clearly France.
To see the probabilities for 2 weeks from now, we square the estimated transition matrix.

$$\hat{P}^2 = \begin{pmatrix} 0.233 & 0.087 & 0.287 & 0.228 & 0.165 \\ 0.202 & 0.172 & 0.186 & 0.223 & 0.218 \\ 0.171 & 0.203 & 0.213 & 0.244 & 0.169 \\ 0.174 & 0.193 & 0.219 & 0.246 & 0.168 \\ 0.221 & 0.086 & 0.221 & 0.232 & 0.241 \end{pmatrix}$$

We see that the most likely place to be in two weeks' time is the UK.
For 3 weeks' time we need the cube of the estimated transition matrix, but we don't need to go beyond the first row:

$$\hat{P}^3 = \begin{pmatrix} 0.198 & 0.167 & 0.209 & 0.230 & 0.196 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

The most likely country to be in 3 weeks' time is Germany.

Question 3

(i) (a) π is a limiting distribution for X if $\lim_{n\to\infty} P(X_n = j \mid X_0 = i) = p_j$ for all i, j.
π is a stationary distribution for X if, given that $P(X_0 = i) = \pi_i$ for all i, it follows that $P(X_1 = i) = \pi_i$ for all i. Equivalently, if $\pi^T P = \pi^T$.

(b) We know from the Chapman-Kolmogorov equations that

$$P(X_{n+1} = j \mid X_0 = i) = \sum_k P(X_n = k \mid X_0 = i)\, p_{kj}.$$

If π is a limiting distribution, then the LHS converges to π_j and the RHS to $\Sigma\pi_k p_{kj}$ as $n \to \infty$. These two quantities must therefore be equal, so π is stationary.

(ii)

(a) Diagram.

(b) State 1 forms communicating class on its own: it is possible to leave it, but not to return to it. States 3 and 5 form a single c.c.. States 2, 4 and 6 form a third. Only the last of these is recurrent since it is closed.

(c) Since X is not irreducible, it cannot be ergodic.

(d) For j = 1, 3, 5 the limit must be 0, as they are all transient. For j = 2, 4, 6 the limit is π_j, where π is the solution to $\pi^T P = \pi^T$ satisfying $\Sigma\pi_i = 1$. This is because X, restricted to {2,4,6}, is irreducible and aperiodic, therefore ergodic.
The equations to solve are $0.8\pi_2 + 0.2\pi_4 = \pi_2$, $0.2\pi_2 + 0.6\pi_4 + 0.5\pi_6 = \pi_4$ and $\pi_2 + \pi_4 + \pi_6 = 1$. The first implies that $\pi_4 = \pi_2$, the second that $\pi_6 = 0.4\pi_2$, and the third that the solution is $(p_2, p_4, p_6) = \left(\frac{5}{12}, \frac{5}{12}, \frac{2}{12}\right)$.

(e) $\lambda_1 = 1$, $\lambda_2 = 0.7$, $\lambda_3 = 0.2$.
Since the restricted transition matrix P satisfies $PV = VL$, it follows that

$$P^n = V\Lambda^n V^{-1} = \frac{1}{60}\begin{pmatrix} 1 & 4 & 1 \\ 1 & -2 & -3 \\ 1 & -5 & 5 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.7^n & 0 \\ 0 & 0 & 0.2^n \end{pmatrix}\begin{pmatrix} 25 & 25 & 10 \\ 8 & -4 & -4 \\ 3 & -9 & 6 \end{pmatrix}.$$

Question 4

a) (i) $\left(1-0.7B-0.3B^2\right)X_t = (1+0.7B-0.6B^2)e_t$.

(ii) $(1-B)(1+0.3B)X_t = (1+0.7B-0.6B^2)e_t$, so X is ARIMA(1,1,2).

(iii) It is non-stationary because $d > 0$.

(iv) By definition $\nabla = 1 - B$.
$(1+0.3B)Y_t = (1+0.7B-0.6B^2)e_t$.

(v) Y is stationary, since the root of $1 + 0.3z = 0$ lies outside the unit circle.
Y is not invertible, since the roots of $1 + 0.7z - 0.6z^2 = 0$ are $z = 2$ and $z = 5/6$: the second of these roots does not lie inside the unit circle.

b) (i) $\gamma_1 = \alpha_1\gamma_0 + \alpha_2\gamma_1$, implying that $\rho_1 = \frac{\alpha_1}{1-\alpha_2}$.

$\gamma_2 = \alpha_1\gamma_1 + \alpha_2\gamma_0$, implying that $\rho_2 = \frac{\alpha_1^2}{1-\alpha_2} + \alpha_2$.

$\gamma_3 = \alpha_1\gamma_2 + \alpha_2\gamma_1$, implying that $\rho_3 = \frac{\alpha_1^3 + \alpha_1\alpha_2}{1-\alpha_2} + \alpha_1\alpha_2$.

(ii) $\hat{\mu} = \bar{x} = 104$, $\frac{\hat{\alpha}_1}{1-\hat{\alpha}_2} = 0.8$ and $\frac{\hat{\alpha}_1^2}{1-\hat{\alpha}_2} + \hat{\alpha}_2 = 0.73$.
The last two equations amount to

$$\hat{\alpha}_1 + 0.8\hat{\alpha}_2 = 0.8 \text{ and } 0.8\hat{\alpha}_1 + \hat{\alpha}_2 = 0.73,$$

which have solution $\hat{\alpha}_1 = 0.6$, $\hat{\alpha}_2 = 0.25$.

(iii) Since $\rho_3 = \alpha_1\rho_2 + \alpha_2\rho_1$, we would expect the value of r_3 to be about $0.6*0.73 + 0.25*0.8 = 0.638$.
The fact that the observed value of r_3 is rather different from this does not of itself contradict the estimates, but it does indicate that a fitting procedure other than MoM would produce different estimates.

(iv)
$$\hat{x}_{50}(1) = \hat{\mu} + \hat{\alpha}_1(x_{50} - \hat{\mu}) + \hat{\alpha}_2(x_{49} - \hat{\mu})$$
$$= 104 + 0.6(3) + 0.25(-2) = 105.3,$$
$$\hat{x}_{50}(2) = \hat{\mu} + \hat{\alpha}_1(\hat{x}_{50}(1) - \hat{\mu}) + \hat{\alpha}_2(x_{50} - \hat{\mu})$$
$$= 104 + 0.6(1.3) + 0.25(3) = 105.53.$$

Question 5

a) (i) The AIC is a method for determining the best-fitting model, avoiding the difficulty that increasing the number of parameters always decreases the sum of squares of the residuals. For each model fitted, -2 times the log of the maximum likelihood (equivalently, $n\log(\text{SSE})$) is added to $2(p+q)$, and the model with the smallest value of the AIC is the one selected.

(ii) The Ljung-Box Q statistic is calculated in conjunction with the fitting of a time series model to a set of data. Q_k is the sum, up to lag k, of the squares of the sample ACF of the residuals. If the observations are really generated from the fitted model, Q_k should be chi-squared distributed with $k-(p+q)$ d.f. This suggests a test for goodness of fit.

b) The decomposition is an attempt to analyse the extent to which the data exhibits linear trend and seasonal variation.
An additive model has been used to fit the data (clear from the seasonal indices, which would be clustered around 1 for a multiplicative model).
The "seasonal index" for a given month is the difference between the average observed value for that month and the overall average value.
The box-and-whisker plots in Figure 2 seem to indicate that there are significant differences between months, and therefore that the seasonal variation is significant.
Looking at the component analysis, there is a noticeable trend, but that does not necessarily imply that it is significant.

c) The SACF indicates that a MA(1) might be a good first model to fit. The SPACF shows significant values at lags 1, 2 and 4, so an AR(1) is unlikely to fit well.

d) Model 1, the MA(1), has SSE equal to 399.856 and 1 parameter, so the AIC is 732.9;
Model 2, the AR(1), has SSE equal to 455.879 and 1 parameter, so the AIC is 748.9;
Model 3, the ARMA(1,1), has SSE equal to 399.807 and 2 parameters, so the AIC is 734.9.
This means that the MA(1) fits best by this measure.
Comment: The Ljung-Box Q statistic for the MA(1) model is not particularly encouraging. It is possible that additional parameters must be included.

e) The Box-Jenkins approach to reducing a time series to stationarity is to difference it repeatedly until stationarity is achieved. In the case of a time series exhibiting seasonal variation, seasonal differencing (i.e. $1-B^{12}$) should be used along with standard differencing ($1-B$). This produces rather different results from the methods used above.
Once the differencing has been carried out, the actions carried out on the residuals are pretty much the same as those outlined in (d).