

In this practical we will analyse data on the purity of heroin from street seizures in Sydney and possibly related deaths from heroin overdose in the same area.

The data are in the dataframe `heroin.dat` which you can download from MyUni. The variable `purity` contains the mean heroin purity from fortnightly street seizures over a two year period. The variable `odeath` contains the fortnightly numbers of overdose deaths for the same period.

Read the data into R as follows:

```
heroin <- read.table("heroin.dat",
  header=TRUE)
attach(heroin)
heroin
objects(2)
```

This create a dataframe called `heroin`.

Look at the data.

Now create a bivariate time series object:

```
purity.ts <- as.ts(purity)
odeath.ts <- as.ts(odeath)
b.heroin <- ts.union(purity.ts,
  odeath.ts)
```

Creates the two time series.

This creates the bivariate series.

You learned how to handle bivariate time series in Computing Practical 3. Adapt the commands you used there to analyse the heroin data as outlined below.

1. Plot the two time series using the single `plot.ts` command.
2. Comment on the shapes of the plots, in particular whether you believe the series are stationary or not. Give reasons for your judgement.
3. Find the correlograms for each series. Discuss what the correlograms indicate, and comment on whether they are appropriate to use to examine the dependence structure of the series.
4. Plot the periodograms for each series. Discuss the information given by each periodogram and indicate which cyclic components appear important in each series.
5. Plot the cumulative periodograms for each series and comment on what they indicate.
6. Suggest a plausible model(s) for each series.

In the next part of the practical you will simulate some ARMA processes and examine the correlogram and periodogram for each realization.

The basic R command to simulate an ARIMA process (an extension of an ARMA process) is

```
arima.sim
```

The general form of the command is

```
arima.sim(n, model=list(order=c(p,d,q), ar = c(a1, a2, ..., ap),
                        ma=c(b1, b2, ..., bq)))
```

where n is the number of observations to generate, p , d and q are the autoregressive order, the difference order and the moving average order respectively, and the a_i and b_i are the values of the autoregressive and moving average parameters respectively (assumed to be zero if not given). In this exercise, d is zero.

For each time series generated, you are required to

1. plot the series,
2. plot the correlogram,
3. plot the (raw) periodogram, and
4. plot the estimated spectrum (i.e., the smoothed periodogram) for different degrees of smoothing.

It may help to plot these on one plot window, for example, using `par(mfrow=c(3,2))`.

The periodogram estimates the (power) spectrum, but recall we briefly discussed in lectures that the periodogram is what we call an *inconsistent* estimator. That is, the periodogram is asymptotically unbiased for the spectrum, but its variance does not decrease as $n \rightarrow \infty$. (We will discuss this further in Chapter 4.)

We can obtain an improved estimate of a continuous spectrum by ‘smoothing’ the raw periodogram. The default plot gives the estimates of the spectrum on the log scale:

```
<simseries.spec> <- spectrum(<simulated series>, spans=c(x,y))
```

`spans=c(x,y)` controls the degree of smoothing and trial-and-error is needed to choose x and y , which should be odd integers. The smoothing is done by repeated smoothing with Daniell smoothers, which are moving averages giving half weight to the end values of the span. Increasing x and y increases the degree of smoothing. Try $(5,7)$ or $(7,9)$.

To obtain an estimate of the spectrum on the original ordinate scale, plot

```
> plot.ts(<simseries.spec>$spec)
```

Now generate the following time series:

1. AR(p) processes of length 100 for $p = 1, 2$.
 - (a) AR(1): $\alpha = -0.8, -0.2, 0.2, 0.8$.
 - (b) AR(2): $\alpha_1 = 1, \alpha_2 = -0.5; \alpha_1 = -1, \alpha_2 = -0.5; \alpha_1 = 0.25, \alpha_2 = 0.5; \alpha_1 = -0.25, \alpha_2 = 0.5$.

2. MA(q) processes of length 100 for $q = 1, 2$.

(a) MA(1): $\beta = -0.8, -0.2, 0.2, 0.8$.

(b) MA(2): $\beta_1 = 1, \beta_2 = 0.5; \beta_1 = -1, \beta_2 = 0.5; \beta_1 = 0.25, \beta_2 = -0.5; \beta_1 = -0.25, \beta_2 = -0.5$.

3. ARMA(1,1) processes of length 100 with pairs of values drawn from

$\alpha = -0.8, -0.2, 0.2, 0.8$ and $\beta = -0.8, -0.2, 0.2, 0.8$.

If you have time, repeat some of the simulations with 50 or 200 observations to see how the plots change.

Patty Solomon

September 2004