STAT 479 - Assignment 1 - due date is on course outline

NOTES: 1. See the comments on 'Technical Writing', on the course web site, before submitting your work. 2. Don't be afraid to see me, or the TA, for help on the assignments. 3. As an incentive for getting started with R right away, there will be a 2% bonus on the course mark given to those who complete part (b) of question 9 (this can be done independently of everything else) and get it to me by the deadline in the course outline. To get the bonus you need only write an R program that does what is required, and e-mail it to doug.wiens@ualberta.ca as an attachment with the name 'yourname.R'. If I can run it in R and get the required output – the matrix with columns 'lag', 'acf' and 'p-value' and the acf plot – then the points are yours.

1. 1.4 in the text.

2. 1.5 in the text.

3. 1.6 in the text.

4. 1.7 in the text.

5. 1.9 in the text. (Recall that $\cos(A-B) = \sin A \sin B + \cos A \cos B$, and review the facts about covariances from Lecture 2.) The series constructed in this problem will form the basis of the *frequency domain* methods considered later in this course.

6. 1.12 in the text.

7. Let $\{x_t\}$ be a stationary normal process with mean μ_x and autocovariance function $\gamma(h)$. Thus for each t, x_t is a Normal random variable (r.v.) with mean μ_x and variance $\gamma(0)$, and also $\mathrm{COV}[x_t, x_{t+h}] = \gamma(h)$. Define the nonlinear time series $y_t = e^{x_t}$.

   (a) Express the mean function $E[y_t]$ in terms of μ_x and $\gamma(0)$. (Hint: Recall that the moment generating function (m.g.f.) of a random variable X is defined to be $M_X(\lambda) = E\left[e^{\lambda X}\right]$, assuming that this function of λ exists for λ near 0. Look it up, if you don't remember it, in the case that X is normally distributed.)

   (b) Determine the autocovariance function of $\{y_t\}$. (Here you will need the distribution of the sum $x_t + x_{t+h}$ of jointly normally distributed – but not independent – r.v.s. What is it?)

8. This question is designed to reinforce a feeling for correlation as a measure of the strength of a linear relationship. Consider a straight line regression model, which in the notation of this course can be written $y_t = \beta_0 + \beta_1 x_t + w_t,\ t = 1, .., n$. Here $\{y_t\}$ and $\{x_t\}$ are numerical values of observations on random variables $\{Y_t\}$ and $\{X_t\}$, and

$\{w_t\}$ is white noise. The least squares estimates of (β_0, β_1) are the minimizers of the sum of squares function

$$S(\beta_0, \beta_1) = \sum_{t=1}^{n} \{y_t - (\beta_0 + \beta_1 x_t)\}^2 .$$

(a) Using the notation $S_{YX} = \sum (y_t - \bar{y})(x_t - \bar{x})$, show that the minimizers are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \text{ and } \hat{\beta}_1 = \frac{S_{YX}}{S_{XX}}.$$

(b) Show that, in the notation of this course, the slope estimate can be expressed as

$$\hat{\beta}_1 = \hat{\rho}_{YX}(0) \sqrt{\frac{\hat{\gamma}_Y(0)}{\hat{\gamma}_X(0)}}.$$

(c) Now consider the same problem, but <u>before</u> any observations are made. We plan to predict Y_t by $\beta_0 + \beta_1 X_t$, and want to do so in such a way as to minimize the mean squared error

$$MSE(\beta_0, \beta_1) = E\left[\{Y_t - (\beta_0 + \beta_1 X_t)\}^2\right].$$

Show that the minimizers are

$$\beta_0 = E[Y] - \beta_1 E[X] \text{ and } \beta_1 = \rho_{YX}(0) \sqrt{\frac{\gamma_Y(0)}{\gamma_X(0)}}.$$

Thus the least squares estimates can be viewed as numerical solutions to this 'minimum mse' problem, obtained by replacing the (unknown) variances and covariances by sample estimates. And the slope is a multiple of the correlation, which we thus see measures the strength of the linear relationship. (In fact if Y_t and X_t were each divided by their standard deviations before we did anything else, the end result would be $\beta_1 = \rho_{YX}(0)$, i.e. the correlations would be the slope.)

9. Let $\{w_t\}$ be a Normal white noise process (with variance $\sigma_w^2 = 1$), and consider the series $x_t = w_t w_{t-1}$.

(a) Determine the mean and autocorrelation function of x_t. Is the series weakly stationary? Why or why not? With reference to your answers in part (b), comment on the relationship between these points and the horizontal confidence bands which are also on the acf plot. Do these numerical results give you reason to believe, or doubt, the answer you just gave? Explain.

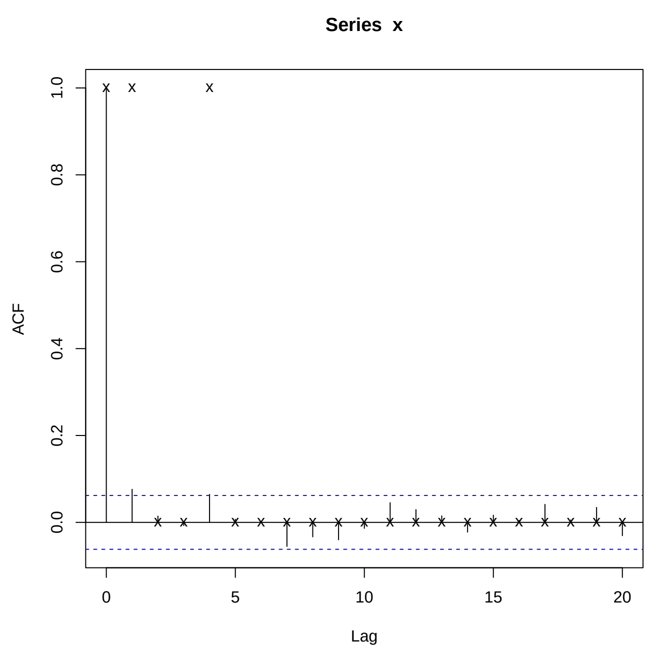
(b) To obtain numerical evidence for (or against) your answer in (a), carry out the following on R. First simulate $n = 1000$ values $\{x_t\}_{t=1}^{n}$, as follows: The command `w = ts(rnorm(1001,0,1))` gives you $\{w_t\}_{t=1}^{1001}$, then `lagw = lag(w, -1)` gives $\{w_{t+1}\}$. Enter the command `mat = cbind(w,lagw)`followed by (for example) `mat[1:5,]` to see the effect of all of this. Now x = w*lagw gives you $\{x_t\}_{t=1}^{n}$. Note that one point is lost in the lagging process. Plot the autocorrelation function at lags $m = 0, 1, ..., 20$. Then compute all 21 p-values associated with the tests of the null hypotheses that $\rho_x(m) = 0$. For this you can use the commands

`a = acf(x,20)` # Compute and plot the acf
`teststat = abs(sqrt(1000)*a$acf)` # two-sided alternative
`p.value = 2*(1-pnorm(teststat))` # Compute p-values: the command `pnorm(x)` computes $P(Z \leq x)$, when $Z \sim N(0,1)$).

Present your findings as a matrix with columns 'lag', 'acf' and 'p-value'; also place an '×' on the acf, at height 1 or 0, depending on whether or not the p-value was significant (< .05, say):

```
mat = cbind(0:20, round(a$acf,3), round(p.value,3))
colnames(mat) = c("lag", "acf", "p.value")
mat
points(0:20, (p.value < .05), pch="x")
```

Your acf plot should now look something like the following:



Sample acf. Points ('×') indicate whether or not the p-value is < .05.

10. 2.6 in the text.