MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.017/1.010 Computing and Data Analysis for Environmental Applications / Uncertainty in Engineering

---

Problem Set 5 (Solutions provided at end of each problem)
Due: Thursday, Oct. 23, 2003

Please turn in a hard copy of your MATLAB program as well as all printed outputs (tables, plots, etc.) required to solve the problem.

Problem 1 Computing joint probabilities and correlations of a time series

Suppose that the total streamflows (in m^3/month) Q_t and Q_{t+1} over two successive months t and t+1 are related by the following equation:

$$Q_{t+1} = 0.6Q_t + e_t \qquad t = 1....10$$

In deterministic applications such time series expressions are solved one step at a time, starting with a specified value of the initial value of the dependent (in this case, the initial streamflow Q_1) and using a specified series of e_t values (for t = 10). The resultis a series of Q_t values for $t > 1$.

When Q_1 and the e_t values are uncertain, we can use the time series equation to derive the mean, variance, and covariance of the streamflow for $t > 1$. These give a probabilistic description of the temporally variable streamflow.

Suppose that the random excitation values $e_1 \ldots e_{10}$ are independent normally distributed random variables, each with mean 4.0 and variance 0.64 (standard deviation 0,8). In addition, suppose that the initial streamflow Q_1 is normally distributed with mean 10.0 and variance 1.0 (standard deviation 1.0). Assume that Q_1 and the e_t values are independent.

Use the definitions of variance and correlation to evaluate the correlation between the two successive streamflows Q_2 and Q_3. First use the defining time series difference equation and the properties of the mean and variance of a linear function to find the means and variances of Q_2 and Q_3. Then use the same equation to derive an expression for the covariance and correlation between Q_2 and Q_3. Are Q_2 and Q_3 independent? Why? What do you think you would get for the means and variances of Q_9 and Q_{10}. Why?

Use MATLAB to determine the joint probability that Q_2 and Q_3 are both greater than 12.0. Also, plot a typical replicate of streamflow vs. time.

Some relevant MATLAB functions: normrnd, sum

Problem 1 Solution:

To find the mean and variance of Q_2 and Q_3:

$Q_2 = 0.6Q_1 + e_1$
$E[Q_2] = 0.6(10) + 4 = 10$

$Q_3 = .6(.6Q_1 + e_1)+e_2$
$E[Q_3] = .36(10)+.6(4)+4 = 10$

If the Q_i's are independent:

$$Var[a_1Q_1+a_2Q_2+\ldots+a_nQ_n] = a_1^2\sigma_{Q1}^2 + \ldots + a_n^2\,\sigma_{Qn}^2$$

Apply to the difference equations for Q_2 and Q_3 :

$Var[Q_2] = (.6)^2(1)^2+(.8)^2 = 1$
$Var[Q_3] = (.36)^2(1)^2+(.6)^2(.8)^2+(.8)^2 = 1$

Find the correlation between Q_2 and Q_3:

$Correl(Q_2,Q_3) = Cov(Q_2,Q_3)/(Var[Q_2]*Var[Q_3])$

$Cov(Q_2,Q_3) = E[(Q_2-\overline{Q}_2)(Q_3-\overline{Q}_3)]$

$Q_3-\overline{Q}_3 = 0.6(Q_2-\overline{Q}_2)+(e_2-\overline{e}_2)$
$E[(Q_2-\overline{Q}_2)(Q_3-\overline{Q}_3)] = 0.6*E[(Q_2-\overline{Q}_2)^2]+E[(e_2-\overline{e}_2)(Q_2-\overline{Q}_2)]$
$= 0.6*Var[Q_2] + Cov(Q_2,e_2)$

What is $Cov(Q_2,e_2)$?

Note: $Q_2-\overline{Q}_2 = 0.6(Q_1-\overline{Q}_1)+(e_1-\overline{e}_1)$
Then: $Cov(Q_2,e_2) = E[(Q_2-\overline{Q}_2)(e_2-\overline{e}_2)] = 0.6*E[(Q_1-\overline{Q}_1)(e_2-\overline{e}_2)] + E[(e_1-\overline{e}_1)(e_2-\overline{e}_2)]$

From the problem statement, Q_1 is independent of the e_i's, and the e_i's are independent of each other, Therefore:

$Cov(Q_2,e_2) = E[(Q_2-\overline{Q}_2)(e_2-\overline{e}_2)] = 0$

And:

$Cov(Q_2,Q_3) = 0.6*Var[Q_2]+Cov(Q_2,e_2) = 0.6*Var[Q_2] = 0.6$
$Correl(Q_2,Q_3) = 0.6/(1*1) = 0.6$

```
% Problem Set 5 -- Problem 1
clear all
close all
nrep=1000000;
emu=4;
esigma=.8;
qmu=10;
qsigma=1;
% create a matrix of random e's
emat=normrnd(emu,esigma,nrep,9);
% create the Q matrix
Qmat = zeros(nrep,10);
Qmat(:,1) = normrnd(qmu,qsigma,nrep,1);
for i=2:10
    Qmat(:,i) = 0.6*Qmat(:,i-1)+emat(:,i-1);
end
% to get the correlation, need mean of (Q2-Q2bar)*(Q3-Q3bar)
% divided by
% sqrt [var(Q2)*var(Q3)]
Q2bar=mean(Qmat(:,2));
Q3bar=mean(Qmat(:,3));
numerator = mean((Qmat(:,2)-Q2bar).*(Qmat(:,3)-Q3bar));
denominator = sqrt(var(Qmat(:,2))*var(Qmat(:,3)));
correl = numerator/denominator
covariance = numerator
vect = Qmat(:,2)>12 & Qmat(:,3)>12;
jointprob = sum(vect)/nrep
plot(1:10,Qmat(5,:),'-*')
xlabel('Time')
ylabel('Streamflow')
title('Sample Streamflow vs. Time')
%  One set of results:
%correl = 0.6007
% covariance =  0.6012
% jointprob = 0.0057
```

Problem 2 The Central Limit Theorem

The Central Limit Theorem states that the distribution of a linear function of independent random variables approaches the normal distribution as the number of variables increases. Use MATLAB to demonstrate this by carrying out the following virtual experiment:

Generate a random array of 1000 replicates (rows) each consisting of 32 independent identically distributed sample values (columns), using one of the non-normal random number generators provided in the MATLAB statistics toolbox (e.g. exponential, uniform, lognormal, etc.). Select your own values for all required distributional parameters. Plot histograms and CDFs (over all replicates) for the sample mean:

$$\boldsymbol{m}_{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i$$

For $N = 1$, 4, 16, and 64. Plot all 4 CDFs on the same axis using the MATLAB `normplot` function, which displays normally distributed data as a straight line. Your CDFs should more closely approach a straight line as N increases, indicating convergence to a normal probability distribution. The approach to normality should also be apparent in your 4 histograms (try plotting all four on one page, using the MATLAB `subplot` function)

Some relevant MATLAB functions: `exprnd, chi2rnd, lognrnd, hist, normplot, mean, subplot.`

Problem 2 Solution:

```
% Problem Set 5 -- Problem 2
clear all
close all
nrep=1000;
mat=exprnd(5,nrep,64);
m1=mat(:,1);
m4=sum(mat(:,1:4),2)/4;
m16=sum(mat(:,1:16),2)/16;
m64=sum(mat(:,1:64),2)/64;
figure
subplot(2,2,1), hist(m1)
subplot(2,2,2), hist(m4)
subplot(2,2,3), hist(m16)
subplot(2,2,4), hist(m64)
figure
normplot(m1)
hold
normplot(m4)
normplot(m16)
normplot(m64)
legend('N=1','N=4','N=16','N=32')
```