University of California at Berkeley
Department of Civil & Environmental Engineering

Structural Engineering Mechanics & Materials
Instructor: Armen Der Kiureghian

Student name:_____

CE93 -- Engineering Data Analysis
Final Examination
Tuesday, May 17, 2005

Work on all four problems. Write clearly and state any assumptions you make. The exam is closed books and notes, except for three sheets of paper and the table of Matlab® distributions.

The problems have the following weights:

Problem 1 (30 points) _____
Problem 2 (20 points) _____
Problem 3 (20 points) _____
Problem 4 (30 points) _____

Exam grade (100 points) _____

Student name:_____

Problem 1. (10+10+10 = 30 points)

Timely completion of a construction project depends on the weather and on the working conditions of machinery used on the job. The probabilities of delay for different combinations of weather condition (rain or no rain) and machinery breakdown are listed in the following table:

Weather	Machinery	Probability of delay in completion of the project
No rain	No breakdown	0.05
Rain	No breakdown	0.20
No rain	Breakdown	0.30
Rain	Breakdown	0.60

Weather forecast indicates 30% chance of rain. From previous experience, we know that the probability that the machinery will breakdown depends on the weather conditions (due to the added force required to move wet soil). If it rains the probability of breakdown is 0.25, whereas if it does not rain the probability of breakdown is only 0.10.

a) Determine the probability that the construction job will be completed on time.

b) If the job is known to have been completed on time, what is the probability that it rained?

c) If the job is not completed on time, what is the probability that machinery breakdown occurred?

Solution

Define R = rain, B = Breakdown, D = Delay.

a) $P(D) = P(D \mid RB)P(RB) + P(D \mid \overline{R}B)P(\overline{R}B) + P(D \mid R\overline{B})P(R\overline{B}) + P(D \mid \overline{R}\overline{B})P(\overline{R}\overline{B})$

$P(RB) = P(B \mid R)P(R) = 0.25 \times 0.30 = 0.075$

$P(\overline{R}B) = P(B \mid \overline{R})P(\overline{R}) = 0.10 \times (1 - 0.30) = 0.070$

$P(R\overline{B}) = P(\overline{B} \mid R)P(R) = 0.75 \times 0.30 = 0.225$

$P(\overline{R}\overline{B}) = P(\overline{B} \mid \overline{R})P(\overline{R}) = (1 - 0.10) \times (1 - 0.30) = 0.630$

$P(D) = 0.60 \times 0.075 + 0.30 \times 0.070 + 0.20 \times 0.225 + 0.05 \times 0.630 = 0.143$

$1 - P(D) = 1 - 0.143 = 0.857$ *Ans.*

b) $P(R \mid \overline{D}) = \dfrac{P(\overline{D} \mid R)}{P(\overline{D})} P(R)$

$$P(D \mid R) = P(D \mid RB)P(B \mid R) + P(D \mid R\overline{B})P(\overline{B} \mid R)$$
$$= 0.60 \times 0.25 + 0.20 \times (1 - 0.25)$$
$$= 0.30$$

$$P(R \mid \overline{D}) = \frac{1 - 0.30}{0.857} \times 0.30 = 0.245 \quad \textit{Ans.}$$

c) $$P(B \mid D) = \frac{P(DB)}{P(D)}$$

$$P(DB) = P(DB \mid R)P(R) + P(DB \mid \overline{R})P(\overline{R})$$
$$= P(D \mid BR)P(B \mid R)P(R) + P(D \mid B\overline{R})P(B \mid \overline{R})P(\overline{R})$$
$$= 0.6 \times 0.25 \times 0.30 + 0.3 \times 0.10 \times (1 - 0.30)$$
$$= 0.0660$$

$$P(B \mid D) = \frac{0.0660}{0.143} = 0.462 \quad \textit{Ans.}$$

Student name:_____

Problem 2. (10+10 = 20 points)

The following two equations describe the monthly water available to cities A and B. City A gets 40 million gallons through a contract with the water distribution system of a neighboring city, 40% of the water available from reservoir 1 and 70% of the water available from reservoir 2. City B gets 20 million gallons on contract basis, plus the remainder of the available water from reservoirs 1 and 2.

$W_A = 40 + 0.4X_1 + 0.7X_2$

$W_B = 20 + 0.6X_1 + 0.3X_2$

Let the available water from reservoir 1, X_1, be a random variable with mean 120 millions of gallon and a c.o.v. of 30%. Also, let the available water from reservoir 2, X_2, be a random variable with mean 80 million gallons and a standard deviation of 25 million gallons. Further, assume that the two variables have a correlation coefficient of 0.30.

a) Determine the means, the standard deviations and the correlation coefficient of the amounts of water available to cities A and B in one month.

b) Suppose the variables X_1 and X_2 are jointly normal. What is the probability that the amount of water available to city B in one month will be less than 80 million gallons?

Solution

a) $\mu_{W_A} = 40 + 0.4\times 120 + 0.7\times 80 = 144$ million gallons *Ans.*

$\mu_{W_B} = 20 + 0.6\times 120 + 0.3\times 80 = 116$ million gallons *Ans.*

$$\sigma_{W_A}^2 = 0.4^2\times(120\times 0.3)^2 + 0.7^2\times 25^2 + 2\times 0.4\times 0.7\times 0.30\times(120\times 0.3)\times 25$$
$$= 665$$

$\sigma_{W_A} = 25.8$ million gallons *Ans.*

$$\sigma_{W_B}^2 = 0.6^2\times(120\times 0.3)^2 + 0.3^2\times 25^2 + 2\times 0.6\times 0.3\times 0.30\times(120\times 0.3)\times 25$$
$$= 620$$

$\sigma_{W_B} = 24.9$ million gallons *Ans.*

$$\mathrm{Cov}[W_A, W_b] = 0.4\times 0.6\times(120\times 0.3)^2 + 0.7\times 0.3\times 25^2$$
$$+ (0.4\times 0.3 + 0.7\times 0.6)\times 0.30\times(120\times 0.3)\times 25$$
$$= 588$$

$$\rho_{w_a w_b} = \frac{588}{25.8\times 24.9} = 0.915$$

b) Since W_B is a linear function of jointly normal random variables, it is also normal.

$$P(W_B < 80) = \Phi\left(\frac{80-116}{24.9}\right) = \Phi(-1.446) = 0.074 \quad \textit{Ans.}$$

Student name:_____

Problem 3. (5+5+5+5 = 20 points)

The annual maximum water level in a river fits a lognormal distribution with mean 4m and a standard deviation of 1m.

a) What is the probability that the water level in the river will exceed 6m in a given year?

b) Suppose the annual maximum water levels in the river can be considered to be statistically independent. What is the probability that the water level in the river will exceed 6m during the next 5 years?

c) What is the probability that the first occurrence of a water level above 6m will occur on the fifth year?

d) What is the mean number of years that we must wait until the water level exceeds 6m?

Solution

a) $\zeta = \sqrt{\ln(1+0.25^2)} = 0.246$

$\lambda = \ln(4) - 0.5 \times 0.246^2 = 1.356$

$$P(\text{Level} > 6) = 1 - \Phi\left[\frac{\ln(6) - 1.356}{0.246}\right] = 1 - \Phi(1.771) = 1 - 0.962 = 0.0383 \quad \textit{Ans.}$$

b) $P(\text{water level exceed 6m in 5 tears}) = 1 - (1 - 0.0383)^5 = 0.177$ *Ans.*

c) $P(\text{first occurrence of water level above 6m in } 5^{\text{th}} \text{ year}) = (1 - 0.0383)^4 \times 0.0383$

$= 0.0328$ *Ans.*

d) 1/0.0383 = 26.1 years (average return period) *Ans.*

Student name:_____

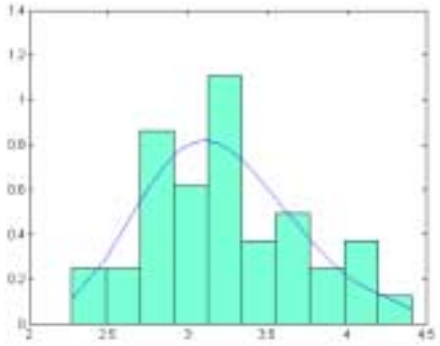
Problem 4. (10+10+10 = 30 points)

In the following Matlab® sessions, x is a 38x3 matrix. The first column contains the biochemical oxygen demand (BOD) and the second and third columns contain the corresponding values of nitrates (NO_3-N) and ammonia (NH_3-N) in a river in units of milligrams per liter. Carefully review the Matlab® commands and calculation results in each session and answer the questions listed below the session. Note that the results from each session are carried over to the next session.

Session 1

```
>> x = load('E6_2.dat');
>> size(x)
ans =
  38   3
mean(x)
ans =
    3.2176    8.5332    0.7811
>> std(x)
ans =
    0.5025    1.6012    0.2157
>> corrcoef(x)
ans =
    1.0000    0.6495    0.5153
    0.6495    1.0000    0.4154
    0.5153    0.4154    1.0000
>> [par, pari] = lognfit(x(:,1),0.05)
par =
    1.1569    0.1554
pari =
    1.1058    0.1267
    1.2080    0.2011
>> normfreq(x(:,1))
>> hold on
>> plot(sort(x(:,1)),lognpdf(sort(x(:,1)),par(1),par(2)))
>> hold off
```

Figure 1



Questions for Session 1

- What is the sample mean, sample standard deviation and sample c.o.v. of BOD data?

  Sample mean = 3.2176 milligrams per liter
  Sample stdev = 0.5025 milligrams per liter
  Sample c.o.v. = 0.5025/3.2175 = 0.156 or 15.6%

- What is the sample correlation coefficient between BOD and NO_3-N data?

  $\rho_{BOD,NO3-N} = 0.6495$

- What are the unbiased point estimates and 95% confidence intervals of the parameters λ and ζ of a lognormal distribution that is fitted to the BOD data?

  $\hat{\lambda} = 1.1569, \quad \hat{\zeta} = 0.1554$ unbiased estimates
  $<\lambda>_{0.95} = (1.1058, 1.2080), \quad <\zeta>_{0.95} = (0.1267, 0.2011)$ 95% confidence intervals

- Explain what is shown in Figure 1.

Student name:_____

Shown in Figure 1 are the normalized frequency diagram of the BOD data together with the PDF of the fitted lognormal distribution with the unbiased parameter estimates $\hat{\lambda}$ and $\hat{\zeta}$.

Session 2

```
>> sx = sort(x(:,1));
>> [h,p,kstat,cv] = kstest(sx,[sx,normcdf(sx,mean(sx),std(sx))])
h =
     0
p =
    0.6822
kstat =
    0.1138
cv =
    0.2155
>> [h,p,kstat,cv] = kstest(sx,[sx,logncdf(sx,par(1),par(2))])
h =
     0
p =
    0.9436
kstat =
    0.0837
cv =
    0.2155
```

Questions for Session 2

- Would you reject the hypothesis of a normal distribution fit to the BOD data? Explain your reasoning.

  I would not reject the hypothesis of a normal distribution fit to the BOD data because $h = 0$ after the first kstest command. Furthermore, the K-S statistic $D_n = 0.1138$ is smaller than the critical value $D_n^{\alpha} = 0.2155$ at the 5% significance level.

- Would you reject the hypothesis of a lognormal distribution fit to the BOD data? Explain your reasoning.

  I would not reject the hypothesis of a lognormal distribution fit to the BOD data because $h = 0$ after the second kstest command. Furthermore, the K-S statistic $D_n = 0.0837$ is smaller than the critical value $D_n^{\alpha} = 0.2155$ at the 5% significance level.

- Which of the two distributions (normal or lognormal) better fits the data? Explain your reasoning.

  The lognormal distribution fits better because the maximum deviation between the empirical and theoretical CDF's, the K-S statistic, is smaller for it $(D_n = 0.0837)$ than for the normal distribution .

Session 3

```
>> [b,bi,r,ri,stats] = regress(x(:,1),[ones(38,1),x(:,2)]);
>> b
    1.4784
```

Student name:_____

```
    0.2038
>> stats
    0.4219   26.2738    0.0000    0.1500
>> [b,bi,r,ri,stats] = regress(log(x(:,1)),[ones(38,1),x(:,2)]);
>> b
    0.5964
    0.0657
>> stats
    0.4579   30.4080    0.0000    0.0135
>> [b,bi,r,ri,stats] = regress(log(x(:,1)),[ones(38,1),x(:,2),x(:,3)]);
>> b
    0.5224
    0.0520
    0.2438
>> stats
    0.5527   21.6230    0.0000    0.0114
```

Questions for Session 3

- We have regressed BOD against NO_3-N and *ln*(BOD) against NO_3-N. Which model in your opinion is better? Explain your reasoning.

  The regression of *ln*(BOD) against NO_3-N is a better model than the regression of BOD against NO_3-N. This is because the R^2 value of the *ln*(BOD) regression (=0.4579) is greater than the R^2 value of the BOD regression (=0.4219).

- What fraction of the variability in *ln*(BOD) is being "explained" by NO_3-N?

  45.79 percent of the variance in *ln*(BOD) is being "explained" by NO_3-N.

- Sketch the mean and mean ± one standard deviation plots of the two regression models on the scatter diagram shown on the next page.

  For regression of BOD against NO_3-N:
  $E[BOD \mid NO_3\text{-}N] = 1.4784 + 0.2038(NO_3\text{-}N)$
  $\sigma_{BOD \mid NO3\text{-}N} = \sqrt{0.1500} = 0.387$

  For regression of *ln*(BOD) against NO_3-N:
  $E[ln(BOD) \mid NO_3\text{-}N] = 0.5964 + 0.0657(NO_3\text{-}N)$
  $\sigma_{ln(BOD) \mid NO3\text{-}N} = \sqrt{0.0135} = 0.116$

  The mean and ± one standard deviation curves are sketched in the figure. Note that the curves for *ln*(BOD) are obtained by taking exponential of the computed values and, therefore, are nonlinear.

- In regressing *ln*(BOD) against both NO_3-N and NH_3-N, how much of the variability in *ln*(BOD) is being explained by the two variables? Is this a better prediction model than the previous two? Explain your reasoning.

Student name:_____

When regressing ln(BOD) against both NO_3-N and NH_3-N, 55.27% of the variability in ln(BOD) is being explained, which is greater than the 45.79% mentioned above. Thus, the regression on both NO_3-N and NH_3-N is a better model.

- For NO_3-N = 10 milligram per liter and NH_3-N = 0.8 milligram per liter, what is the probability that BOD will exceed 4 milligrams per liter?

$$E[ln(BOD) \mid NO_3\text{-}N, NH_3\text{-}N] = 0.5224 + 0.0520(NO_3\text{-}N) + 0.2438(NH_3\text{-}N)$$
$$= 0.5224 + 0.0520\times 10 + 0.2438\times 0.8 = 1.237$$

$$\sigma_{\ln(BOD)|NO3\text{-}N,NH3\text{-}N} = \sqrt{0.0114} = 0.107 .$$

$$P[BOD>4] = P[\ln(BOD)>\ln(4)]$$

$$= 1-\Phi\left(\frac{\ln(4)-1.237}{0.107}\right) = 1-\Phi(1.391) = 1-0.918 = 0.082$$