Applied Statistics and Numerical Methods for Engineers
ChE 301, Spring 1999
Exam Number Four
Administered: 8:00-10:00 A.M. Thursday May 6, 1999

THE EXAM HAS 12*10=120 POINTS.
Reserve approximately 10 minutes for each of the 12 problems.

Problem 1. Random Variables

Given the joint PDF

	y				
x	1	2	3	4	5
0	0.01	0.02	0.03	0.04	0.05
0.25	0.02	0.03	0.04	0.05	0.11
0.5	0.03	0.04	0.05	0.06	0.02
0.75	0.04	0.05	0.06	0.02	0.03
1.0	0.05	0.06	0.02	0.03	0.04

Find $P(0 \le x \le 0.5, 2 < y \le 4)$

Solution:

$$P(0 \le x \le 0.5, 2 < y \le 4) = P(x=0, y=3) + P(x=0.25, y=3) + P(x=0.5, y=3)$$
$$+ P(x=0, y=4) + P(x=0.25, y=4) + P(x=0.5, y=4) = 0.27$$

Problem 2. Expectations

Given the PDF

$$f(x) = \frac{x^2}{9} \quad \text{for } 0 \le x \le 3$$

find the population variance of x.

Solution:

$$\sigma^2 = E\left[X^2\right] - E[X]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} x f(x)dx\right)^2$$

$$\sigma^2 = \int_0^3 x^2 f(x)dx - \left(\int_0^3 x f(x)dx\right)^2 = \left.\frac{x^5}{45}\right|_0^3 - \left(\left.\frac{x^4}{36}\right|_0^3\right)^2$$

$$\sigma^2 = 5.4 - 5.0625 = 0.3375$$

Problem 3. Discrete Distributions

In sampling a liquid product from the assembly line with a historical defect rate of 12%, what is the probability of finding your seventh sample turning out to be your second defect?

Solution:

Use Negative Binomial PDF:

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x\text{-}k} \quad \text{for } x = k, k+1, k+2...$$

x = # of (independent, repeated, with replacement, only 2 outcomes) Bernoulli trials = 7

p = probability of success on one trial = 0.12

$q = 1 - p$ = probability of failure on one trial = 0.88

k = # of successes = 2

$$b^*(7;2,0.12) = \binom{7-1}{2-1} 0.12^2 0.88^5 = 0.0456$$

Problem 4. Continuous Distributions

A steady temperature of an ink cartridge in a copier is essential for proper performance of the copier. If the temperature deviates more than 2 degrees Celsius from the desired temperature, the copying process fails. Consider a copier with a standard deviation of 0.8 degrees Celsius cartridge temperature. What percent of the time will the copier fail to function properly (excluding all other sources of copier mishaps)?

Solution:

We want the probability of failure, which is one less the probability of success:

$$1 - P(-2 \le x - \mu \le 2)$$

Use the normal distribution. Convert to standard normal variables.

$$z = \frac{x-\mu}{\sigma}$$

$$z_{low} = \frac{-2}{0.8} = -2.5 \text{ and } z_{high} = \frac{2}{0.8} = 2.5$$

$$1 - P(-2 \le x - \mu \le 2) = 1 - P(-2.5 \le z \le 2.5)$$

$$1 - P(-2 \le x - \mu \le 2) = 1 - [P(z \le 2.5) - P(z \le -2.5)]$$

Go to the table A.3 in the back of WMM to obtain the numerical values:

$$1 - P(-2 \le x - \mu \le 2) = 1 - [0.9938 - 0.0062] = 0.0124$$

Problem 5. Sampling and Estimation

Scientists are now saying that the recent outbreak of deformed frogs in Minnesota ponds is due to the increased presence of a certain type of bacteria, (possibly itself due to an increase in agriculture chemical run-off into the pond, which stimulates algae growth and increases the bacteria population). A scientist claims that 0.08 fraction of the total MN frog population has some sort of deformity.

The elementary school kids who made the initial discovery of the deformed frogs, found 12% of collected frogs to have deformities, with a sample standard deviation of 3% based on 16 samples.

Determine whether we can believe the scientist claim within a 95% confidence interval, based on the children's sampling results.

Solution:

We need to find a confidence interval for the mean, population variance unknown

Use t-distribution $T = \frac{\overline{X} - \mu}{s / \sqrt{n}}$, $t_{1-\alpha} = -t_{\alpha}$, $v = n - 1$

$$P(\overline{X} - t_{\alpha}\frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha}\frac{s}{\sqrt{n}}) = 1 - 2\alpha$$

$$T = \frac{0.12 - 0.08}{0.03 / \sqrt{16}} = 5.333 \text{, } v = n - 1 = 15 \text{, } \alpha = 0.025 \; t_{\alpha} = 2.131$$

$$P(0.12 - 2.131\frac{0.03}{4} < \mu < 0.12 + 2.131\frac{0.03}{4}) = 0.95$$

$$P(0.104 < \mu < 0.136) = 0.95$$

Since the scientists claimed population mean is not within this interval, his claim is found <u>not to be true</u> within 95% confidence.

Problem 6. Linear Algebra

Consider an nxn matrix, $\underline{\underline{J}}$, with rank = n-1. Indicate which of any of the following statements which are true.

(a) The inverse of $\underline{\underline{J}}$ exists.

(b) At least 2 rows of $\underline{\underline{J}}$ are linearly dependent.

(c) The determinant of $\underline{\underline{J}}$ is non-zero.

(d) There is a unique solution to the system of linear equations $\underline{\underline{J}}\underline{x} = \underline{R}$ for any real nx1 vector, $\underline{R}$.

(e) The reduced row echelon form of $\underline{\underline{J}}$ will not have any rows completely filled with zeroes.

Solution:

rank is less than n. So (a) is false, (b) is true, (c) is false, (d) is false, and (e) is false.

Problem 7. Solution of a Nonlinear Algebraic Equation

Consider the equation

$$3x - \sqrt{x} = 2$$

Using an initial guess of 3.0, complete one entire iteration of a Newton-Raphson step.

$$f(x) = 3x - \sqrt{x} - 2$$

$$f'(x) = 3 - \frac{1}{2\sqrt{x}}$$

$$f(x = 3) = 3x - \sqrt{x} - 2 = 5.27$$

$$f'(x = 3) = 3 - \frac{1}{2\sqrt{x}} = 2.71$$

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = 3.0 - \frac{5.27}{2.71} = 1.055$$

(Exact solution is 1.000)

Problem 8. Solution of a system of Nonlinear Algebraic Equations

Consider the following steady-state mass and energy balance for a non-isothermal continuous stirred tank reactor with an irreversible first order reaction:

$$0 = \frac{F_{in}C_{A,in} - F_{out}C_A - C_A k_o e^{-\frac{E_a}{RT}}}{V}$$

$$0 = \frac{F_{in}\tilde{H}_{in} - F_{out}\tilde{H}_{out}(T) + \Delta H_r C_A k_o e^{-\frac{E_a}{RT}} + \dot{Q}}{V}$$

Notation is the same as in the project.

A multivariate Newton Raphson has proved successful in giving us the steady-state values of C_A and T for slightly exothermic reactions (a value of ΔH_r). However, when we attempt the same solution technique for a more exothermic reaction with (a values of $10\Delta H_r$), we cannot find an initial guess close enough to the steady state values of C_A and T to converge to our solution. What do you suggest we do? Give an algorithm and equations if necessary.

Solution:

My suggestion is that we write a short code with a for-loop to run syseqn.m repeatedly.

(1) Our first cycle through the for-loop will use a value of ΔH_r and return the steady state values.

(2) We use the converged steady-state values of C_A and T for ΔH_r as the initial guess for the same equations with a heat of reaction of $2\Delta H_r$.

(3.a) If the code converges at $2\Delta H_r$, then use that converged solution as the initial guess for the same equations with a heat of reaction of $3\Delta H_r$.

(3.b) If the code cannot converges at $2\Delta H_r$, then lower the step size of our ΔH_r changes. For example, use the converged steady-state values of C_A and T from ΔH_r as initial guesses for the same equations with a heat of reaction of $1.1\Delta H_r$. If that works, then use that converged solution as the initial guess for the same equations with a heat of reaction of $1.2\Delta H_r$.

(4) Repeat with as small step-sizes as necessary until you reach $10\Delta H_r$.

Problem 9. Numerical Integration

Perform a numerical integration using Trapezoidal rule on the function, $f(x) = 500x^2$ over the range $-2 \le x \le 0$ using 2 intervals.

Solution:

$$\int_a^b f(x)dx \approx \frac{h}{2}\left[f(a) + f(b) + 2\sum_{i=2}^{n} f(x_i)\right]$$

$$\int_{-2}^{0} f(x)dx \approx \frac{1}{2}[2000 + 0 + 2 \cdot 500] = 1500$$

(Exact solution: 1333.333333333)

Problem 10. Solutions of Ordinary Differential Equations

Perform one Euler method step on the ordinary differential equation: $\frac{d^2y}{dx^2} = \frac{dy}{dx}\sin(x)\cos(y)$ subject to the initial conditions, $y(x = 1) = 2$ and $\frac{dy}{dx}(x = 1) = 0.5$, using a step size $\Delta x = 0.1$ x and y are given in radians. (Determine the values, $y(x = 1.1)$. and $\frac{dy}{dx}(x = 1.1)$.)

Solution:

first break into 2 first order ODEs.

$$\frac{dy_1}{dx} = y_2$$
$$\frac{dy_2}{dx} = y_2 \sin(x)\cos(y_1)$$

Then apply Euler method:

$$y_j(x_{i+1}) = y_j(x_i) + \Delta x * f_j(\{y(x_i)\}, x_i)$$

$$y_1(x = 1.1) = y_1(x = 1.0) + 0.1 * f_1(y_1 = 2.0, y_2 = 0.5, x = 1.0)$$

$$y_1(x = 1.1) = 2.0 + 0.1 * (0.5) = 2.05$$

$$y_2(x = 1.1) = y_2(x = 1.0) + 0.1 * f_2(y_1 = 2.0, y_2 = 0.5, x = 1.0)$$

$$y_2(x = 1.1) = 0.5 + 0.1 * (-0.1751) = 0.4825$$

$$y(x = 1.1) = y_1(x = 1.1) = 2.05$$

$$\frac{dy}{dx}(x = 1.1) = y_2(x = 1.1) = 0.4825$$

Problem 11. Regression

We are trying to fit some data to $\hat{y} = bx + a$

We are given the following information:

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 0.619 \qquad \sum_{i=1}^{n}(x_i - \overline{x})^2 = 6.32 \qquad n = 100$$

$$\sum_{i=1}^{n} y_i = 147.8 \qquad \sum_{i=1}^{n} x_i = 2509 \qquad SSE = 0.25$$

Calculate the (i) average slope, (ii) the standard deviation of the slope, and (iii) a 95% confidence interval on the slope.

Solution:

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{0.619}{6.32} = 0.098$$

$$\sigma^2 \approx s^2 = \frac{SSE}{n-2} = \frac{0.25}{98} = 0.002551$$

$$\sigma_b^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{0.002551}{6.32} = 0.0004036$$

$$\sigma_b = \sqrt{\sigma_b^2} = 0.0201$$

$$P(\hat{b} - t_{\alpha,n-2}\sqrt{\sigma_b^2} < b < \hat{b} + t_{\alpha,n-2}\sqrt{\sigma_b^2}) = 1 - 2\alpha$$

$$P(0.098 - 1.99 \cdot 0.0201 < b < 0.098 + 1.99 \cdot 0.0201 = 0.95$$

$$P(0.058 < b < 0.138) = 0.95$$

Problem 12. ANOVA

Consider the problem where we test three catalysts for their effect on moles per minute of product generated in a reactor.
We run the anova_1factor.m code and obtain this output:

```
» anova_1factor
  Sample Problem with    3 treatments and    6 replicates

y =

    1.0000    1.1000    1.2000    1.1000    1.2000    1.0000
    2.0000    2.2000    2.4000    2.2000    2.4000    2.0000
    1.0000    1.1000    1.2000    1.1000    1.2000    1.0000

  Ho: all treatments are equal
  Reject Ho if  151.25 >> f(    2,   15 )
  Hypothesis Rejected for 98 percent confidence interval (  151.25 >    6.36 )

  pvalue =   1.14e-010

  95 percent C.I. on the    1 treatment:  9.30e-001 <  1.10e+000 <  1.27e+000
  95 percent C.I. on the    2 treatment:  2.03e+000 <  2.20e+000 <  2.37e+000
  95 percent C.I. on the    3 treatment:  9.30e-001 <  1.10e+000 <  1.27e+000

  95 percent C.I. on the    1 -    2 treatment diff.: -1.34e+000 < -1.10e+000 < -8.60e-001
  95 percent C.I. on the    1 -    3 treatment diff.: -2.40e-001 <  0.00e+000 <  2.40e-001
  95 percent C.I. on the    2 -    3 treatment diff.:  8.60e-001 <  1.10e+000 <  1.34e+000
```

Answer these questions:

(i) What does the treatment represent?
(ii) What does y represent?
(iii) The null hypothesis was rejected. What does this mean--do the different catalysts make a difference?
(iv) Rank the catalysts in order of best performance.
(v) Does any one catalyst give a yield of at least 1.0 moles/minute more product than every other catalyst, within a 95% confidence interval? If so, which one?
(vi) What does the reported p-value mean?

Solution:

(i) What does the treatment represent?

The treatment represents different catalysts.

(ii) What does y represent?

y represents different yields of product in units of moles/minute.

(iii) The null hypothesis was rejected. What does this mean--do the different catalysts make a difference?

We rejected the hypothesis that the catalysts are equal. Therefore, we can say that one of them is making a difference, within the 98% confidence.

(iv) Rank the catalysts in order of best performance.

Catalyst 2 is best. The other two are equal.

(v) Does any one catalyst give a yield of at least 1.0 moles/minute more product than every other catalyst, within a 95% confidence interval? If so, which one?

Yes, catalyst 2 gives a 1.1 mole/minute higher yield than either other catalyst within a 95% confidence interval.

(vi) What does the p-value mean?

The p-value is the level-of-significance we can apply to the value of the F-statistic we obtained. So while we satisfied the 98% confidence, our results would also satisfy a 1-p confidence level, in this case a 99.99999999886 % confidence level.