Name: ______________________________ ID #: ____________________________

CENG114 – Probability and Statistical Methods for Engineers

Professor Shyue Ping Ong

Final

Winter 2016, March 15 2016

Instructions:

1. Write your name on the top right of this page.
2. Please answer **ALL** questions as far as possible.
3. You have a total of **3 hours** exactly.
4. Additional data are attached at the end of this exam. You may tear it out for easy reference and do not need submit it with your exam.
5. Write your answers in the space provided following the questions. If you need additional space, you may request for additional blank sheets of paper during the exam. If you use additional blank sheets, you should mark the top left of each sheet with your name and the question number and today's date, e.g., "John Smith, Q1, Jan 28 2016".
6. Indicative maximum points for each sub-question are provided only as a reference. Actual grading points may vary.
7. Write neatly and legibly. Illegible answers will be considered as wrong.

Question 1 (16 points): ____________

Question 2 (16 points): ____________

Question 3 (18 points): ____________

Question 4 (16 points): ____________

Question 5 (16 points): ____________

Question 6 (18 points): ____________

This exam has a total of 14 pages, including the cover page.

Qn 1. In statistical mechanics, the probability of a system being in a microstate *i* is given by:

$$P(i) = ce^{-\frac{E_i}{kT}}$$

where E_i is the energy of the microstate *i*, *k* is the Boltzmann constant, *T* is the temperature, and *c* is a constant.

Butane has three rotamers (each rotamer can be considered a microstate) *a*, *b*, and *c*, with normalized energies of $\frac{E_i}{k} = 0, 400, \text{ and } 400\text{ K}$ respectively. For the purposes of this question, we will assume a temperature of 300 K. Answer the following questions.

i. (2 points) What is the value of *c* that makes this is valid probability distribution?
ii. (2 points) What is the probability of a butane molecule being in the ground state, i.e., the state *a* with lowest energy?
iii. (4 points) What is the mean and standard deviation of the normalized energy of a butane molecule?
iv. (4 points) For a collection of 10 butane molecules, what is the probability that less than 7 molecules are in the ground state *a*?
v. (4 points) For a collection of 1 million butane molecules, estimate the probability that less than 655,000 molecules are in the ground state *a*.

Solution:

i. $P(a) + P(b) + P(c) = ce^0 + ce^{-\frac{4}{3}} + ce^{-\frac{4}{3}} = 1$, So value of c is $\frac{1}{1.5272} = 0.6548$
ii. The probability of butane molecule in the ground state is: $P(a) = c = 0.6548$
iii. $E(x) = \sum xP(x) = 0.6548 * 0 + 0.1726 * 400K + 0.1726 * 400K = 138.082\ K$

$$var(x) = \sum_x (x - E[X])^2 p(x)$$
$$= (138.082K)^2 * 0.6548 + (400K - 138.082K)^2 * 0.1726 + (400K - 138.082K)^2 * 0.1726 = 36165.91K^2$$

So standard deviation is $\sqrt{var(x)} = \sqrt{36165.91K^2} = 190.17K$

iv. The butane molecular state could be considered as a binomial with parameter p and n, we p = p(a) = 0.6548.
So the probability that less than 7 in ground state a is:

$$P(X < 7) = 1 - P(X = 7) - P(X = 8) - P(X = 9) - P(X = 10)$$
$$= 1 - \binom{10}{7} 0.6548^7 0.3452^3 - \binom{10}{8} 0.6548^8 0.3452^2 - \binom{10}{9} 0.6548^9 0.3452^1 - \binom{10}{10} 0.6548^{10}$$
$$= 1 - 0.2548 - 0.1812 - 0.07639 - 0.01449 = 0.47312 \approx 0.473$$

v. The distribution could be treated as a binomial distribution with parameter 1000,000 and p=0.6548. The mean and variance of the binomial distribution is:

$$\mu_x = np = 1000{,}000 * 0.6548 = 654800$$
$$\sigma_x^2 = np(1 - p) = 1000{,}000 * 0.6548 * 0.3452 = 226037$$

The standard deviation is: 475.43

The probability of less than 655000 molecules in ground state is:

$$P(X \leq 655000) = P(X \leq 655000 + 0.5) = P\left(Z < \frac{655000 + 0.5 - 654800}{475.43}\right)$$
$$\approx P(Z < 0.422) = 0.66276$$

Qn 2. Uranium-lead (U-Pb) dating is a method for dating rocks. There are actually two exponential decay chains in wide use: $^{238}U \rightarrow ^{206}Pb$ with a half-life of 4.468×10^9 years and $^{235}U \rightarrow ^{207}Pb$ with a half-life of 0.7038×10^9 years. The combination of the two methods provide a cross-validation of the estimated age of the rock. Answer the following questions:

i. (2 points) What is the rate constant of the radioactive decay process ^{238}U and ^{235}U for in year^{-1}?

ii. (5 points) A sample of rock containing zircon ($ZrSiO_4$) is dated. Zircon incorporates U, but strongly rejects Pb, i.e., any Pb found can be assumed to be from radioactive decay. The ratio of the atomic amount of ^{207}Pb to ^{235}U, N_{Pb207}/N_{U235}, in the sample is 5. Estimate the age of the rock.

iii. (5 points) The same rock in part ii is tested for ^{206}Pb and ^{238}U. Assuming that the rock age estimated in part ii is correct, what is the expected ratio of the atomic amount of ^{206}Pb to ^{238}U, N_{Pb206}/N_{U238} today?

iv. (4 points) Today, uranium occurs in the ratio of 0.711% ^{235}U to 99.289% ^{238}U (we can ignore other isotopes for simplicity). When the rock was first formed, what is the percentage of ^{235}U of total uranium content?

Solution:

i. $\lambda_{238U} = \frac{\ln(2)}{4.468\times10^9} = 0.1551\times10^{-9}\, year^{-1}$

$$\lambda_{235U} = \frac{\ln(2)}{0.7038\times10^9} = 0.9849\times10^{-9}\, year^{-1}$$

ii. $\frac{N_{Pb207}}{N_{U235}} = 5$, so 5/6= 0.833 N_{U235} has decayed.

$$1 - e^{-0.9849\times10^{-9}t} = 0.833$$

so $t = \frac{\ln(0.1667)}{-0.9849\times10^{-9}} = 1.819\times10^9\, year$

iii. Assuming the rock age is $1.819\times10^9\, year$, the amount of N_{U238} that have decayed is:

$$1 - e^{-0.1551\times10^{-9}*1.819\times10^9} = 0.2458$$

So the expected ratio of the atomic amount of ^{206}Pb to ^{238}U is: $\frac{0.2458}{1-0.2458} \approx 0.326$

iv. We the rock was first formed: $\frac{0.711*6}{0.711*6+99.289/(1-0.2458)} = \frac{4.266}{4.266+131.65} \approx 0.03139 =$ 3.14%

Qn 3. The drag force on a car F_D is related to its velocity by the following drag equation:

$$F_D = C_D A \frac{\rho v^2}{2} \quad (1)$$

where C_D is the drag coefficient, A is the frontal area, ρ is the density of air, and v is the velocity. We are going to use this equation to study the design of a Tesla Model S.

i. (8 points) You have measured the drag force on the Tesla Model S at various velocities, as given in the table below.

v (m/s)	F_D (N)							
3	3.99							
6	13.9							
9	29.8							
12	52.1							
15	80.8							
18	116.2							
21	158.3							
24	207.2							

Using performing a linear regression using the following slightly modified expression:

$$F_D = C_D A \frac{\rho v^2}{2} + k \quad (2)$$

determine the drag coefficient C_D. Note that k is a very small constant. Additional blank columns and rows are provided to make it easier for you to work out the solutions. You can choose whether you want to use them. If you use them, partial credit will be given for the <u>correct useful</u> numbers. The front area of the Tesla Model S is measured at $A = 2.34$ m^2, and the density of air $\rho = 1.225$ kg/m^3.

ii. (2 points) What is the estimated drag force on a Tesla Model S at its top rated speed of 210 km / h?

iii. (8 points) Rearranging the drag equation (1), you may also use the same table above to estimate the C_D at each speed to get a sample of 8 C_D values. Tesla Motors claims that the C_D of the Model S is 0.24. Perform a hypothesis test at the 5% level of significance to ascertain if the C_D is significantly greater than 0.24. *Show all workings and state all hypotheses clearly. Points will only be given if the steps and conclusions are properly outlined.*

Solution:

i.

v (m/s)	$Y=F_D$ (N)	$X=v^2$ (m/s)^2	Y^2	X^2	XY
3	3.99	9	15.9201	81	35.91
6	13.9	36	193.21	1296	500.4
9	29.8	81	888.04	6561	2413.8
12	52.1	144	2714.41	20736	7502.4
15	80.8	225	6528.64	50625	18180
18	116.2	324	13502.44	104976	37648.8
21	158.3	441	25058.89	194481	69810.3
24	207.2	576	42931.84	331776	119347.2
Sum	662.29	1836	91833.39	710532	255438.81

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{N} = 710532 - \frac{(1836)^2}{8} = 289170$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{N} = 91833.39 - \frac{(662.29)^2}{8} = 37004.88$$

$$SP = \sum xy - \frac{\sum x \sum y}{N} = 255438.81 - \frac{662.29 \times 1836}{8} = 103443.26$$

$$b = \frac{C_D A\rho}{2} = \frac{SP}{SS_{xx}} = \frac{103443.26}{289170} = 0.3577$$

$$C_D = \frac{2b}{A\rho} = \frac{2\times 0.3577}{2.34\times 1.225} = 0.25$$

$$k = \bar{Y} - b\bar{X} = 82.79 - 0.3577\times 229.5 = 0.688$$

ii. If v = 210 km/h = 58.33 m/s, then

$$F_D = 0.3577v^2 + 0.688 = 0.3577(58.33)^2 + 0.688 = 1217.72$$

iii. After rearrangement we will have:

$$C_D = \frac{2F_D}{A\rho v^2}$$

Then the data points for C_D are:

v (m/s)	*Y*=F_D (N)	*X*=v^2 (m/s)^2	C_d	$(C_d)^2$
3	3.99	9	0.309	0.0957
6	13.9	36	0.269	0.0726
9	29.8	81	0.257	0.0659
12	52.1	144	0.252	0.0637
15	80.8	225	0.251	0.0628
18	116.2	324	0.250	0.0626
21	158.3	441	0.250	0.0627
24	207.2	576	0.251	0.0630
Sum	662.29	1836	2.090	0.5490

$$H_0: C_D \leq 0.24, H_a: C_D > 0.24$$

$$\overline{C_D} = 0.26126;\ s = \sqrt{\frac{SS_{xx}}{n-1}} = \sqrt{\frac{0.002933}{7}} = 0.02047$$

$$t = \frac{\overline{C_D} - 0.4}{\frac{s}{\sqrt{n}}} = \frac{0.26126 - 0.24}{\frac{0.02047}{\sqrt{8}}} = 2.937$$

$$t_{crit} = 1.895$$

since $t > t_{crit}$, therefore reject H_o at 5% level of significance.

Q4. An article in the ACI Materials Journal (1987, Vol. 84, pp. 213–216) described several experiments investigating the rodding of concrete to remove entrapped air. A 3-inch × 6-inch cylinder was used, and the number of times this rod was used is the design variable. The resulting compressive strength of the concrete specimen is the response. The data are shown in the following table.

Rodding Level	Compressive Strength			
	1	2	3	
10	1530	1530	1440	
15	1610	1650	1500	
20	1560	1730	1530	
25	1800	1890	1710	

We wish to use an ANOVA analysis to determine if there are any substantive differences in compressive strength with different rodding levels.

i. (4 points) Construct the null hypothesis and alternative hypothesis for the test.

ii. (4 points) Calculate the mean square variability within and between samples.

iii. (8 points) Perform a hypothesis test at the 0.05 level and determine any significant differences in the compressive strength. In particular, identify the specific pairs of rodding levels with different compressive strength, if any, using Tukey's HSD test.

Solution:

i. $H_o: \mu_{10} = \mu_{15} = \mu_{20} = \mu_{25}$
$H_1: \mu_{10} \neq \mu_{15} \neq \mu_{20} \neq \mu_{25}$

ii.

Rodding Level	Compressive Strength			
	1	2	3	$T_{rodding}$
10	1530	1530	1440	4500
15	1610	1650	1500	4760
20	1560	1730	1530	4820
25	1800	1890	1710	5400
T	6500	6800	6180	19480

$$\sum \frac{T^2}{n} = \frac{1}{3}(4500^2 + 4760^2 + 4802^2 + 5400^2) = 31766666.7$$

$$\sum X^2 = 31823600$$
$$\frac{G^2}{N} = \frac{19480^2}{12} = 31622533.3$$
$$SS_{between} = \sum \frac{T^2}{n} - \frac{G^2}{N} = 144133.3$$
$$SS_{within} = \sum X^2 - \sum \frac{T^2}{n} = 56933.33$$
$$dof_{between} = \frac{12}{3} - 1 = 3$$
$$dof_{within} = 12 - \frac{12}{3} = 8$$
$$MS_{between} = \frac{SS_{between}}{dof_{between}} = 48044.44$$
$$MS_{within} = \frac{SS_{within}}{dof_{within}} = 7116.667$$

iii. $F = \frac{SS_{between}}{SS_{within}} = 6.751$

$F_{crit} = 4.0662$

Since F > F_{crit}, we reject the null hypothesis at 0.05 level.

$$HSD = q\sqrt{\frac{MS_{within}}{n}} = 4.529\sqrt{\frac{7116.667}{3}} = 220.587$$

$\mu_{10} = 1500, \mu_{15} = 1586.667, \mu_{20} = 1606.667, \mu_{25} = 1800$

Pair 1: $|\mu_{10} - \mu_{15}| = 86.667$

Pair 2: $|\mu_{10} - \mu_{20}| = 106.667$

Pair 3: $|\mu_{10} - \mu_{25}| = 300$

Pair 4: $|\mu_{15} - \mu_{20}| = 20$

Pair 5: $|\mu_{15} - \mu_{25}| = 213.33$

Pair 6: $|\mu_{20} - \mu_{25}| = 193.33$

Pair 3 has different compressive strength.

Q5. The table below shows the global mean sea level (GMSL) in mm obtained from NASA's website (Beckley et al., Marine Geodesy, Vol 33, Suppl 1, 2010, note that some data is simplified for easier analysis).

Time index	2014 GMSL, X_{2014} (mm)	2015 GMSL, X_{2015} (mm)	X_{2014}					
1	57.84	60.14						
2	55.8	62.49						
3	56.5	62.6						
4	55.11	64.75						
5	56.21	64.16						
6	58.74	67.53						

i. (4 points) Using the data, construct 95% confidence intervals for the GMSL in 2014 and 2015.

ii. (6 points) It is hypothesized that there has been an increase in the GMSL in 2015 over 2014. Assuming that the data in each time index pertains is measured at exactly the same day and time in each year and using exactly the same equipment, perform a test of the hypothesis at the 0.05 level of significance.

iii. (6 points) If instead, the data collected in 2014 and 2015 are independent, i.e., they do not necessarily correspond to the same time, perform a hypothesis test that the mean GMSL in 2015 is different from that in 2014 at the 0.05 level of significance. You may still assume that the data is collected using the same equipment and therefore, it is reasonable to assume that the standard deviation of the measurements in both 2014 and 2015 are the same.

Show all workings and state all hypotheses clearly. Points will only be given if the steps and conclusions are properly outlined.

Solution:

i. We construct confidence intervals for the population using the t distribution

Time index	2014 GMSL, X_{2014} (mm)	2015 GMSL, X_{2015} (mm)	GMSL X_{2014}^2	GMSL X_{2015}^2	D(2015-2014)	D^2		
1	57.84	60.14	3345.47	3616.82	2.3	5.29		
2	55.8	62.49	3113.64	3905.0	6.69	44.76		
3	56.5	62.6	3192.25	3918.76	6.1	37.21		
4	55.11	64.75	3037.11	4192.56	9.64	92.93		
5	56.21	64.16	3159.56	4116.51	7.95	63.2		
6	58.74	67.53	3450.39	4560.3	8.79	77.26		
sum	**340.2**	**381.67**	**19298.42**	**24309.95**	**41.47**	**320.65**		

From the t table with 0.05 and dof 5, t_{crit}=2.571

2014:

$$SS_X = \sum X^2 - \frac{(\sum X)^2}{6} = 19298.42 - \frac{(340.2)^2}{6} = 9.0794$$

$$S_X = \sqrt{\frac{SS_X}{6-1}} = \sqrt{\frac{9.0794}{5}} = 1.3475$$

$$S_{\bar{X}} = \frac{S_X}{\sqrt{6}} = \frac{1.3475}{\sqrt{6}} = 0.55$$

The confidence interval is: $\bar{X} \pm 2.571 * 0.55 = 56.7 \pm 1.414 = 55.286 - 58.114$.

2015:

$$SS_X = \sum X^2 - \frac{(\sum X)^2}{6} = 24309.95 - \frac{(381.67)^2}{6} = 31.28$$

$$S_X = \sqrt{\frac{SS_X}{6-1}} = \sqrt{\frac{31.28}{5}} = 2.5$$

$$S_{\bar{X}} = \frac{S_X}{\sqrt{6}} = \frac{2.5}{\sqrt{6}} = 1.02$$

The confidence interval is: $\bar{X} \pm 2.571 * 1.02 = 63.61 \pm 2.622 = 60.988 - 66.232$

ii. We work on the difference between pairs of scores.

$$H_0: \mu_D \leq 0$$
$$H_1: \mu_D > 0$$

$$SS_D = \sum D^2 - \frac{(\sum D)^2}{6} = 320.65 - \frac{41.47^2}{6} = 34.03$$

$$S_D = \sqrt{\frac{34.03}{6-1}} = 2.61$$

$$S_{\bar{D}} = \frac{S_D}{\sqrt{6}} = \frac{2.61}{\sqrt{6}} = 1.065$$

$$t = \frac{\frac{\sum D}{n}}{S_{\bar{D}}} = \frac{\frac{41.47}{6}}{1.065} = 6.49 > 2.015$$

Reject null hypothesis at 5% level, there has been an increase in GMSL in 2015 over 2014

iii.

$$H_0: \mu_{2015} - \mu_{2014} = 0$$

$$H_1: \mu_{2015} - \mu_{2014} \neq 0$$

$$\mu_{2015} = 63.61$$

$$\mu_{2014} = 56.7$$

$$SS_{2014} = 9.0794$$

$$SS_{2015} = 31.28$$

$$S_p^2 = \frac{31.28 + 9.08}{6+6-2} = 4.04$$

$$S_{\bar{X}_{2015}-\bar{X}_{2014}} = \sqrt{\frac{S_p^2}{6} + \frac{S_p^2}{6}} = \sqrt{\frac{2 * 4.04}{6}} = 1.16$$

$$t = \frac{63.61 - 56.7 - (0)}{1.16} = 5.96$$

Look up t-tables with dof = 6+6-2=10, we find t_{crit}=2.228

We reject the null hypothesis that the mean GMSL in 2015 is no difference from GMSL in 2014

Q6. There is an automated system for testing the purity of a reagent before it is shipped out to customers. A sample of 10 bottles were tested, and the following are the measured purities:

95.8%, 97.5%, 94.4%, 98.0%, 93.5%, 96.6%, 94.9%, 95.8%, 93.7%, 97.9%

i. (4 points) We want to estimate the overall (population) average and standard deviation of the purity. Use the sample data to calculate an estimate of the mean and standard deviation of the purity of the reagent.

ii. (2 points) We may assume that the purity is normally distributed with the mean and standard deviation derived in part i. What is the probability that the purity of a random bottle is < 94%?

iii. (3 points) A bottle with < 94% purity does not meet specifications. Tests are carried out on bottles coming off a production line. What is the probability that the 3rd faulty bottle is detected at the 15th bottle tested?

iv. (3 points) A production line is stopped and cleaned when more than 3 faulty bottles are found within the first 20 bottles tested. What is the probability of a production line being stopped, given that the 3rd faulty bottle is detected at the 15th bottle tested?

v. (4 points) A particular reaction succeeds 99% of the time when the purity of the reagent is > 96%, 80% of the time if the purity of the reagent is between 94% and 96%, and only 30% of the time when the purity of the reagent is < 94%. A CENG student carries out a successful reaction using a bottle of reagent. What is the probability that the bottle has > 96% purity?

Solution:

i. According to central limit theorem, population mean is close to sample mean.

$$population\ mean = \frac{958.1}{10} = 95.81\%$$

$$\sigma^2 = \frac{SS}{N-1} = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1} = \frac{91820.81 - \frac{(958.1)^2}{10}}{9} = 2.805$$

$$\sigma = \sqrt{2.805} = 1.675$$

ii.

$$P(X < 94\%) = P\left(Z < \frac{94 - 95.81}{1.675}\right) = P(Z < -1.08) = 0.14$$

iii. $P = \binom{14}{2}\big(P(X < 94\%)\big)^2\big(1 - P(X < 94\%)\big)^{12}\big(P(X < 94\%)\big)^1 = 91\times(0.14)^3\times(1-0.14)^{12} = 0.0409$

iv. $P_{last5pass} = \left(1 - P(X \leq 94\%)\right)^5 = 0.4704$

Then the probability of at least 1 bottom faulty found is = 1-0.4704 = 0.5296

v. $P(X > 96\%) = 1 - P(X \leq 96\%) = 1 - P\left(Z \leq \frac{96-95.81}{1.675}\right) = 1 - P(Z \leq 0.113) = 1 - 0.5438 = 0.4562$

$$P(94\% < X < 96\%) = P(X < 96\%) - P(X < 94\%)$$
$$= P\left(Z < \frac{96 - 95.81}{1.675}\right) - P\left(Z < \frac{94 - 95.81}{1.675}\right) = 0.5438 - 0.14$$
$$= 0.4038$$

$$P(X < 94\%) = P\left(Z < \frac{94 - 95.81}{1.675}\right) = P(Z < -1.08) = 0.14$$

E1: success | X > 96%, E2: success | 94%<X<96%, E3: success | X < 94%;

$$P = \frac{P(E1)P(X > 96\%)}{P(E1)P(X > 96\%) + P(E2)P(94\% < X < 96\%) + P(E3)P(X < 94\%)}$$

$$P = \frac{0.4562 * 0.99}{0.4562 * 0.99 + 0.4038 * 0.80 + 0.14 * 0.30} = 0.553$$