UNIVERSITY OF TORONTO

Faculty of Applied Science and Engineering

Division of Engineering Science

Final Examination, April 2018

Time allowed: 150 minutes

STA286H1S | Probability and Statistics

Exam Type C: A single, double-sided aid sheet is allowed.

Calculator Type 2: All non-programmable electronic calculators allowed

Instructors: Deepa Kundur and Mark Ebden

Surname: ______________________ **First name:** ______________________

Student Number: ______________

Instructions:

- Answer all questions, and note the value of each question. A total of 123 marks is available
- Answer each question directly on the examination paper. Indicate clearly where your work can be found
- Show all steps, and present results clearly. Use three significant figures unless indicated otherwise

Examiners' Report:

Question	Grade	Question	Grade	Question	Grade
1		7		13	
2		8		14	
3		9		15	
4		10		16	
5		11		17	
6		12		18	
		Total	/ 123		

This exam should have 15 pages including this page

Chapter 5 (26 marks)

1. (10 marks) A random committee of size 3 is selected from 2 professors and 4 students.

(a) Write a formula for the probability distribution of the random variable X representing the number of students on the committee.

(b) Find the probability that a majority of the members on the committee are students.

(c) Find the probability that a majority of the members on the committee are professors.

(d) Find the probability that there are no professors on the committee.

(e) Find the probability that there are no students on the committee.

2. (4 marks) A company purchased a large lot of a certain kind of electronic component. One percent of the lot is defective. Random samples of 100 units each are taken from within the large lot. Considering the samples of 100 units each,

(a) what is the mean number of defective units?

(b) what is the variance of the number of defective units?

3. (12 marks) A random variable X that assumes the values $x_1, x_2, \ldots, x_k$ is called a discrete uniform random variable if its probability mass function is $f(x) = 1/k$ for all $x_{-}, x_2, \ldots x_k$ and 0 otherwise.

(a) Find the mean and variance of X.

(b) Assuming that all values $x_1, x_2, \ldots, x_k$ must be distinct and real and that k is odd, find an assignment for $x_1, x_2, \ldots, x_k$ (i.e., find values for x_i for $i = 1, 2, \ldots, k$) such that the mean is zero and the variance can be represented as

$$\frac{k-1}{2}\left(\frac{k-1}{2}+1\right)\frac{1}{k}$$

For full marks, please show your work and discuss your reasoning.

Chapter 6 (20 marks)

4. (8 marks) The daily amount of coffee, in litres, dispensed by a machine located near the EngSci Common Room is a random variable X having a *uniform* distribution of the form $f(x;\, A, B)$ as given in the aid material.

(a) Given that $A = 4$ and $B = 20$, find the probability that on a given day the amount of coffee dispensed by this machine will be...

(i) at least 12.4 litres.

(ii) more than 10.4 litres but less than 18.4 litres.

(b) For general A and B, find the probability that on a given day the amount of coffee dispensed by this machine will be within one standard deviation of the mean of X.

5.(a) (4 marks) Discuss two scenarios similar to what was discussed in the lectures for which the exponential distribution would be an appropriate statistical model.

5.(b) (8 marks) The length of time for one individual to be served at a cafeteria is a random variable having an exponential distribution with a mean of 4 minutes. What is the probability that a person is served in less than 3 minutes on at least 4 of the next 6 days?

Chapter 7 (5 marks)

6. (5 marks) Let X be a random variable with a Poisson distribution as described in the aid material, in which $\lambda t = k$, a constant. Find the mean value of X by using the fact that the moment-generating function (mgf) of X is $M_X(t) = e^{k(e^t - 1)}$. Approaches not using the mgf will get zero marks.

Chapter 8 (16 marks)

7. (6 marks) Nine data points were recorded as follows: 5, 11, 4, 5, 10, 20, 6, 15, and 5. Find the following, explaining your work:

(a) the mean

(b) the median

(c) the mode

8. (7 marks) There are 48 independent random variables, X_i for $i \in \{1, 2, \ldots 48\}$, such that each is taken from a uniform distribution, $f(x; -1/2, 1/2)$. Approximately, what's the **probability** that their sum is greater than 3? Express your answer to three significant figures.

9. (3 marks) A random variable with a t-distribution can be thought of as the ratio of $X/\sqrt{Y}$, where X and Y are random variables whose distributions are:

A) χ^2
B) Normal
C) χ^2 and normal, respectively
D) Normal and χ^2, respectively

Chapter 9 (17 marks)

10. (6 marks) A device is manufactured such that its length varies slightly from device to device, following an approximately normal distribution with a standard deviation of 1.5 cm. A random sample of 75 devices has an average length of 310 cm. Find a 90% **confidence interval** for the mean of the lengths of all devices in the population. Answer with a minimum of 2 significant figures.

11. (6 marks) Students may choose between a course without labs and a course with labs. The final written examination is the same for each section. If 12 students in the section with labs made an average grade of 84 with a standard deviation of 4, and 18 students in the section without labs made an average grade of 77 with a standard deviation of 6, find a 95% **confidence interval** for the difference between the average grades for the two courses. Assume the populations to be approximately normally distributed with equal variances. Answer with a minimum of 2 significant figures.

12. (5 marks) Consider the Bernoulli distribution, given by $f(x;\mu) = \mu^x(1-\mu)^{1-x}$. Given an observed $x = 1$, calculate the **maximum likelihood estimator** for μ. (Don't forget to show all steps.)

Chapter 10 (19 marks)

13. (3 marks) A product engineer wants to test the (null) hypothesis that at least 10% of the public is allergic to her new product. Give an example of how the engineer could commit a **type I error**. (One sentence may be enough.)

14. (8 marks) Past experience indicates that the time required for students to complete a particular test is a normal random variable with a mean of 40 minutes. If a random sample of 20 students took an average of 38.1 minutes to complete this test with a standard deviation of 4.3 minutes, **test the hypothesis** that $\mu = 40$ minutes against the alternative that $\mu \neq 40$ minutes. Use $\alpha = 0.05$.

15. (8 marks) The grades in a first-year course were as follows:

Grade	A	B	C	D	F
Frequency, f	28	36	64	40	32

Test the hypothesis that the distribution of grades is uniform. (Begin by stating your hypotheses.) Use $\alpha = 0.05$.

Chapter 11 (23 marks)

16. (3 marks) In linear regression, the least squares procedure produces a line that **minimizes** the total of all...

A) Distances between the points and the line
B) Squared-distances between the points and the line
C) Vertical distances between the points and the line
D) Vertical squared-distances between the points and the line

17. (3 marks) In the linear regression model $Y = \beta_0 + \beta_1 x + \varepsilon$, the ε term represents the:

A) regression coefficient
B) random error
C) residual
D) true regression line

18. (17 marks) In a certain type of metal test specimen, the **normal stress** on a specimen is known to be functionally related to the **shear resistance**. The following is a set of experimental data:

Normal Stress, x	Shear Resistance, y
25	29
27	25
29	24

(a) (5 marks) Estimate the expression for the regression line, $\mu_{Y|x} = \beta_0 + \beta_1 x$. You do not need to plot anything.

(b) (5 marks) Calculate s^2, the mean squared error (MSE).

(c) (2 marks) Predict the shear resistance for a normal stress of 30.

(d) (5 marks) Provide a 90% confidence interval for the shear resistance for a normal stress of 30.

If you use this page for overflow, you must indicate so *at the original question.*

If you use this page for overflow, you must indicate so *at the original question.*

Aid material:

Number of ways to select k elements out of n elements: $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

Name	Probability distribution	Notes
Bernoulli	$b(x;\,1,p) = p^x(1-p)^{1-k}$	$x \in \{0,1\}$, mean $= p$, var $= p(1-p)$
Binomial	$b(x;\,n,p) = \binom{n}{x}\,p^x\,(1-p)^{n-x}$	$x \in \{0,1,2,\dots n\}$, mean $= np$, var $= np(1-p)$
Hypergeometric	$h(x;N,n,k) = \binom{k}{x}\binom{N-k}{n-x}/\binom{N}{n}$	$\max\{0, n-N+k\} \le x \le \min\{n,k\}$ mean $= nk/N$, var $= nk(N-n)(1-k/N)\,/\,[N(N-1)]$
Neg. binomial	$b^*(x;k,p) = \binom{x-1}{k-1}\,p^k\,(1-p)^{x-k}$	$x \in \{k, k+1, \dots\}$, mean $= k/p$, var $= k(1-p)/p^2$
Geometric	$g(x;p) = p\,(1-p)^{x-1}$	$x \in \{1,2,3,\dots\}$, mean $= 1/p$, var $= (1-p)/p^2$
Poisson	$p(x;\lambda t) = e^{-\lambda t}(\lambda t)^x\,/\,x!$	$x \in \{0,1,2,\dots\}$, mean $=$ var $= (\lambda t)$
Uniform	$f\,(x;A,B) = \frac{1}{B-A}$	$A \le x \le B$, mean $= (A+B)/2$, var $= (B-A)^2/12$
Normal	$n(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty < x < \infty$, mean $= \mu$, var $= \sigma^2$
Exponential	$f\,(x;\beta) = \frac{1}{\beta}e^{-x/\beta}$	$x > 0$, mean $= \beta$, var $= \beta^2$
Gamma	$f\,(x;\alpha,\beta) = \frac{1}{\beta^\alpha\,\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}$	$x > 0$, mean $= \alpha\beta$, var $= \alpha\beta^2$, $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}e^{-u}du$
χ^2	$f\,(x;v) = \frac{1}{2^{v/2}\,\Gamma(v/2)}\,x^{v/2-1}e^{-x/2}$	$x > 0$, mean $= v$, var $= 2v$
Weibull	$f\,(x;\alpha,\beta) = \alpha\beta\,x^{\beta-1}\exp\left(-\alpha x^\beta\right)$	$x > 0$, mean $= \alpha^{-1/\beta}\Gamma\left(1+\frac{1}{\beta}\right)$ var $= \alpha^{-2/\beta}\left\{\Gamma\left(1+\frac{2}{\beta}\right) - \left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2\right\}$
t	$h(t;v) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\,\sqrt{v\pi}}\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}}$	$-\infty < t < \infty$, mean $= 0$ for $v > 1$, var $= \frac{v}{v-2}$ for $v > 2$
F	$h(f;v_1,v_2) = \frac{\Gamma[(v_1+v_2)/2]}{\Gamma(v_1/2)\,\Gamma(v_2/2)}\left(\frac{v_1}{v_2}\right)^{v_1/2}\cdots$ $\cdots f^{v_1/2-1}\left(1+\frac{v_1}{v_2}f\right)^{-(v_1+v_2)/2}$	$f > 0$, mean $= v_2/(v_2-2)$ for $v_2 > 2$ var $= \frac{2\,v_2^2\,(v_1+v_2-2)}{v_1(v_2-2)^2(v_2-4)}$ for $v_2 > 4$

If X is approximately normal, or if $n \gtrsim 30$, then generally:

- $Z = (\bar{X} - \mu) \,/\, (\sigma / \sqrt{n})$. A CI for μ is $(\hat{\theta}_L, \hat{\theta}_U) = \left(\bar{x} - z_{\alpha/2}\,\sigma \,/\, \sqrt{n},\ \bar{x} + z_{\alpha/2}\,\sigma \,/\, \sqrt{n}\right)$
- $T = (\bar{X} - \mu) \,/\, (S / \sqrt{n})$ has a t-distribution with $n-1$ dof. A CI for μ is $\left(\bar{x} - t_{\alpha/2}\, s \,/\, \sqrt{n},\ \bar{x} + t_{\alpha/2}\, s \,/\, \sqrt{n}\right)$

If X is normal, a PI for the next observation is:

- $\left(\bar{x} - z_{\alpha/2}\,\sigma\sqrt{1 + 1/n},\ \bar{x} + z_{\alpha/2}\,\sigma\sqrt{1 + 1/n}\right)$, or
- $\left(\bar{x} - t_{\alpha/2}\, s\sqrt{1 + 1/n},\ \bar{x} + t_{\alpha/2}\, s\sqrt{1 + 1/n}\right)$ where the t-distribution has $n-1$ dof

If X_1 & X_2 are both approximately normal, or if $n_1, n_2 \gtrsim 30$, then generally:

- If σ_1 & σ_2 are both known, $Z = \left[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\right] \Big/ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ and
  a CI for $\mu_1 - \mu_2$ is $\left((\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
- If σ_1 & σ_2 are both unknown, and $\{\sigma_1^2 = \sigma_2^2$ or $\{n_1 = n_2$ and X_1 & X_2 are both approx'ly normal$\}$ $\}$:
  $T = \left[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\right] \Big/ s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ and
  a CI for $\mu_1 - \mu_2$ is $\left((\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$
  The t-distribution has $n_1 + n_2 - 2$ dof, and $s_p^2 = \left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right] / (n_1 + n_2 - 2)$

Chapter 11:

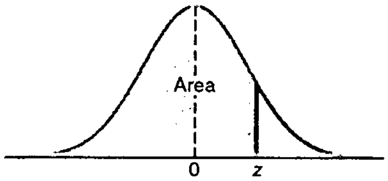
$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \qquad b_0 = \bar{y} - b_1\bar{x}, \qquad b_1 = \frac{S_{xy}}{S_{xx}}$$

From Table A.5:

Critical Values of the Chi-Squared Distribution

	α						
v	0.30	0.25	0.20	0.10	0.05	0.025	0.02
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622

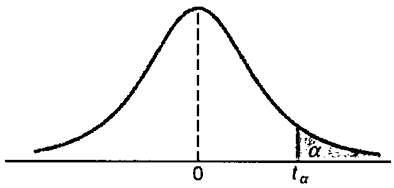
From Table A.3:



Areas under the standard normal curve, for positive z-values:

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

From Table A.4:



Critical values of the t-distribution, for degrees of freedom v and significance level α:

	α						
v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960