Midterm

SDS321

Spring 2015

You may use a single (2 sides) page of notes, and you may use a calculator.

This exam consists of five questions, containing multiple sub-questions. The assigned points are noted next to each question; the total number of points is 25. You have 75 minutes to answer the questions.

Please answer all problems in the space provided on the exam. Use extra pages if needed. Of course, please put your name on extra pages.

Read each question carefully, show your work and clearly present your answers. Note, the exam is printed two-sided - please don't forget the problems on the even pages!

Good Luck!

Name: ____

UTeid: ____

Question 1 (3 points)

(a) (1 point) How many solutions are there to the equation $x_1 + x_2 = 5$, where x_1 and x_2 are non-negative integers?

6. This can either be solved using a stars-and-bars type approach, or by direct enumeration of options ((0,5),(1,4),(2,3),(3,2),(4,1),(5,0)).

1 point for correct answer.

- If they used only positive numbers, give no marks but don't drop marks on the second part for only using positive numbers.
- If they only included half the pairs – e.g. counted (1,4) and (4,1) as the same – give no marks but don't drop marks on the second part if they do the same.

(b) (2 points) How many solutions are there to the equation $x_1 + x_2 + 2x_3 = 5$, where x_1, x_2 and x_3 are non-negative integers? *Hint: what values can x_3 take?*

12. x_3 can be 0,1 or 2. If $x_3 = 0$, there are 6 options as above. If $x_3 = 1$, then $x_1 + x_2 = 3$, so 4 options. If $x_3 = 2$, then $x_1 + x_2 = 1$, so 2 options.

2 points for correct answer. If the answer is incorrect but they showed working:

- If the answer is correct except they gave their answer from (a) for the options for $x_3 = 0$, give full marks for (b).
- If they excluded 0 values in part (a) and did the same here, but work is otherwise correct, give 2 marks for part (b) but no marks for part (a).
- If they excuded permutations in part (a) and did the same here, but the work is otherwise correct, give 2 marks for (b) but no marks for part (a).
- If the reasoning is correct but there is a minor math error, give one mark.
- **An incorrect answer with no working gets no marks**, even if it is consistent with one of the above errors.

Question 2 (3 points) Let $X \sim N(70, 100)$ be a normal random variable with expectation $\mu = 70$ and standard deviation $\sigma = 10$. For the following problems, please express your probability in terms of $\Phi(z)$ and then evaluate it using the standard normal table provided on the last page of this exam script. Recall that $\Phi(z) = P(Z \leq z)$, where $Z \sim N(0, 1)$.

(a) (1 point) Find $P(X < 60)$.

$$P(X < 60) = P\left(Z < \frac{60 - 70}{10}\right) = P(Z < -1) = 1 - P(Z \leq 1)$$
$$= 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

- 0.5 marks for translating into Z form (either $P(Z \leq 1)$ or $\Phi(1)$ notation are fine, it doesn't matter if they use "$<$" or "$\leq$ notation). 0.5 marks for right answer. More/fewer sig figs are fine.
- If they use an incorrect value for μ or σ in calculating Z, give 0 marks here but if they repeat the mistake in later parts, give marks.

(b) (1 point) Find $P(X \in [50, 60])$

$$P(X \in [50, 60]) = P(X \leq 60) - P(X \leq 50)$$
$$P(X \leq 50) = P\left(Z \leq \frac{50 - 70}{10}\right) = 1 - P(Z \leq 2) = 1 - 0.9772 = 0.0228$$
$$P(X \in [50, 60]) = 0.1587 - 0.0228 = 0.1359$$

- 0.5 marks for translating into Z form, 0.5 marks for right answer. They can either translate the whole thing into Z form, or just the $P(Z < 50)$ part. Again, it doesn't matter if they use "$<$" or "$\leq$ notation, and more/fewer sig figs are fine.
- If they made a mistake in part (a) and used that value here, but everything else is correct, give the mark.
- If they used an incorrect value for μ or σ here, and used the same incorrect value in part (a), give the mark assuming everything is consistent with the values used.

(c) (1 point) Find $P(X \in [50, 60] \cup [80, 90])$. *Hint: No further calculation necessary.* By symmetry, this is twice the answer in (b), i.e. 0.2718. 1 point for 2*answer in (b). Or, they can calculate it using tables – in which case, 0.5 marks for Z notation, 0.5 marks for correct answer (with the guidelines from part (b) for mistakes carried forward).

Question 3 (5 points) A continuous random variable X has probability density function.

$$f_X(x) = c\,e^{-|x|/2} = \begin{cases} c\,e^{-x/2} & x \geq 0 \\ c\,e^{x/2} & x < 0 \end{cases} \tag{1}$$

(a) (1 point) Find c.

$1 = \int_{-\infty}^{\infty} f(x)dx = 2 \cdot c \int_0^{\infty} e^{-x/2}dx = 2c[-2e^{-x/2}]_0^{\infty} = 4c \Rightarrow c = 1/4.$

1 point; 0.5 points for setting up correct integral relationship (either $\int_{-\infty}^{\infty} f(x)dx = 1$, or $\int_0^{\infty} f(x)dx = 0.5$); 0.5 points for solving integral correctly.

(b) (2 points) Find the CDF of X, i.e. $F_X(x) = \mathbf{P}(X \leq x)$ for $x \geq 0$ and $x < 0$. If you were unable to solve part a, you may give your answer in terms of c.

For $x < 0$, $F_X(x) = \int_{-\infty}^{x} f(t)dt = c\int_{-\infty}^{x} e^{t/2}dt = c[2e^{t/2}]_{-\infty}^{x} = 2ce^{x/2} = 0.5e^{x/2}$

For $x > 0$, $F_X(x) = \int_{-\infty}^{x} f(t)dt = c\int_{-\infty}^{0} e^{x/2}dt + c\int_0^{x} e^{-t/2}dt = 0.5 + c[-2e^{-t/2}]_0^{x} = 0.5 - 2ce^{-x/2} + 2c = 1 - 0.5e^{-x/2}$

- 1 mark for each case: 0.5 marks for setting up the integral, 0.5 marks for solving the integral.
- For the second case, it's OK to use a symmetry argument; if they use a symmetry argument but the answer for $x < 0$ was incorrect, give the mark if the answer for $x \geq 0$ is consistent with the answer for $x < 0$.
- If they got the wrong answer for c in part (a), don't drop marks for using that value. Also, don't drop marks if they don't use a numerical value for c.

(c) (2 points) Now consider two independent random variables X and Y both with the same pdf in Equation **??**, i.e. $f_X(x) = ce^{-|x|/2}$ and $f_Y(y) = ce^{-|y|/2}$. Find $E(XY)$. (*Hint: no integral required*). $E[XY] = 0$ - independent random variables mean that $E[XY] = E[X]E[Y]$, and by symmetry, $E[X] = E[Y] = 0$. 2 points for correct answer. It's fine if they solve it using $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x)f_Y(y)dxdy$, in which case, 1 mark for setting up integral (including appropriate functional forms for $f_X(x)$ and $f_Y(y)$), and 1 mark for solving integral.

Question 4 (6 points) Janet is concerned she might have a disease that affects 1% of the population. Luckily, a drug-store test is available. The test has a false-positive rate of 3% and a false-negative rate of 1% (remember, a false positive is when a person doesn't have the disease, but the test result is positive).

(a) (1 point) What is the probability of a positive test result?

Let T_1 be a positive test, D be have disease. $P(T_1|D) = 0.99$ (1-false negative rate); $P(T_1|D^c) = 0.03$ (false positive rate); $P(D) = 0.01$. Then $P(T_1) = P(T_1|D)P(D) + P(T_1|D^c)P(D^c) = 0.99 * 0.01 + .03 * .99 = 0.0396$

1 point. If they show working and have clearly mixed up the false positive and false negative numbers, give no points for this question but don't drop marks in later questions if they use these values.

(b) (2 points) Janet takes the test, and it returns positive. Given this information, what is the probability that she has the disease?

$$P(D|T_1) = \frac{P(T_1|D)P(D)}{P(T)} = \frac{0.99 * 0.01}{0.0396} = 0.25$$

2 marks;

- If the answer to part (a) was wrong and they use it here in the denominator, give 2 marks provided everything else is correct.
- If they flipped the false positive and the false negative rates in (a) and do the same here, give them 2 marks provided everything else is correce.
- Give 1 mark if they write down Bayes' Law/draw tree correctly but use incorrect numbers other than this.

(c) (3 points) Janet is aware that the test is imperfect, so she decides to take a second test. It also returns positive. Given the results of both tests, what is the probability that she has the disease? You may assume that the two test results are conditionally independent given her disease status. Let T_2 be the outcome of the second test.

- Method 1: $P(T_1 \cap T_2|D) = 0.99 * 0.99 = 0.9801$. $P(T_1 \cap T_2|D^c) = 0.03 * 0.03 = 0.0009$. So, $P(T_1 \cap T_2) = P(T_1 \cap T_2|D)P(D) + P(T_1 \cap T_2|D^c)P(D^c) = 0.9801 * 0.01 + 0.0009 * .99 = 0.0107$. So,

$$P(D|T_1 \cap T_2) = \frac{P(T_1 \cap T_2|D)P(D)}{P(T_1 \cap T_2)} = \frac{0.9801 \times 0.01}{0.0107} = 0.917$$

- Method 2: Treat answer from (b) as the new prior. $P(D|T_1) = 0.25$. $P(T_2|D, T_1) = P(T_2|D) = 0.99$. $P(D^c|T_1) = 0.75$. $P(T_2|D^c, T_1) =$

$P(T_2|D^c) = 0.03$. So,

$$P(D|T_1,T_2) = \frac{P(T_2|D,T_1)P(D|T_1)}{P(T_2|D,T_1)P(D|T_1) + P(T_2|D^c,T_1)P(D^c|T_1)}$$
$$= \frac{0.99 \times 0.25}{0.99 \times 0.25 + 0.03 \times 0.75} = 0.917$$

- Method 3: Draw a tree.

3 points; If incorrect give following working points: 1 point for a correct equation for $P(T_1 \cap T_2)$ or appropriate tree and 1 point for correctly identifying $P(T_1 \cap T_2|D) = P(T_1|D)P(T_2|D)$ OR 2 points for noticing you can use the posterior from (b) as the prior here..

Question 5 (8 points) A, B and C roll 3 fair dice. In the first round A starts by rolling a die, then B rolls a die, then C rolls a die. In the second round again A rolls a die, then B rolls a die and then C rolls a die and so on. The winner of the game is the first one to get a "6". In your solution please use $p = \frac{1}{6}$. Let X denote the number of rolls upto and including the first "6".

(a) (2 points) What is the probability that A wins in the k^{th} round, for some positive integer k. *Hint: what is the value of X if A wins in the k^{th} round?*

For A to win in the kth round, there need to have been $3(k-1)$ non-6 throws then 1 6 throw... so $P(\text{A wins in } k\text{th round}) = (1-p)^{3(k-1)}p = \left(\frac{5}{6}\right)^{3(k-1)} \frac{1}{6}$.

2 points (leaving in p/1-p form is fine). Lose 1 point if they do the prob that A first gets a 6 in the kth round, i.e. $(1-p)^{k-1}p$. Lose 1 point if they miss the p for throwing the 6.

(b) (2 points) What is $P(\{\text{A wins}\})$? *Hint: you can use $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ for $0 \leq x < 1$*

$\sum_{k=1}^{\infty} p(1-p)^{3(k-1)} = p\sum_{k=0}^{\infty}(1-p)^{3k} = p\sum_{k=0}^{\infty}((1-p)^3)^k = \frac{p}{1-(1-p)^3} = \frac{1/6}{1-125/216} = 0.396$

You can also do it using the memoryless property. $P(A|X \leq 3) = p$. $P(A|X > 3) = P(A)$.Now $P(A) = P(A|X \leq 3)P(X \leq 3) + P(A|X > 3)P(X > 3) = p + P(A)q^3$, since $P(A|X \leq 3)P(X \leq 3) = P(A, X \leq 3) = P(X = 1) = p$. Solving you get $P(A) = p/(1-q^3)$.

2 points; if they write the infinite summation out correctly but either leave unsimplified or make an error in simplification, do not drop points. Allow points carried forward for summing over the answer in (a). Do drop one point if the summation range is incorrect (check where it starts - 0 or 1). Leaving things in terms of p is fine.

If they do it by memoryless property, then full credit. Partial credit if

- Correct total probability rule–1 pt.
- Correct $P(A|X > 3) = P(A)$–1 pt.
- Correct $P(A|X < 3)P(X < 3)$–1 pt.
- Take a point off for each minor mistake, like wrong $P(X > 3)$, etc.

[Question continues on next page]

(c) (3 points) What is $P(\{\text{B wins}\})$? Now, the probability of B winning in the kth round is $p(1-p)^{3(k-1)+1}$, so $P(\{\text{B wins}\}) = \sum_{k=1}^{\infty} p(1-p)^{3(k-1)+1} = p(1-p)\sum_{k=0}^{\infty}(1-p)^{3k} = \frac{p(1-p)}{1-(1-p)^3} = \frac{1/6\times 5/6}{1-125/216} = 0.329$.

Or, can get from b by saying for every probability where A wins in the kth round, the probability that B wins in the kth round is (1-p) times that probability.

Memoryless: $P(B) = P(B|X \leq 3)P(X \leq 3) + P(B|X > 3)P(X > 3) = P(B, X \leq 3) + P(B)q^3 = P(X = 2) + P(B)q^3$. Solving, $P(B) = p(1-p)/(1-q^3)$. 3 points; if they write the infinite summation out correctly but either leave unsimplified or make an error in simplification, do not drop points. Partial credit:

- Full marks if they use 5/6 times the answer from par (b), even if part b is wrong.
- Give 1 point if they identify the probability of B winning in the kth round, but do not sum over all values of k.
- Leaving things in terms of p is fine.

Full points for memoryless property. Follow rubric on last part for memoryless property.

(d) (1 point) What is the probability of A winning conditioned on B not winning? $P(A|B^c) = \frac{P(A\cap B^c)}{P(B^c)} = \frac{P(A)}{P(B^c)} = \frac{0.4}{1-0.33} = 0.59$ 1 point; give marks if answer is consistent with parts (c) and (d).

[End of Exam]

Standard normal table

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.8413	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	0.9772	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	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990