Chris Piech
CS109

May 3rd, 2016

Midterm Examination

This is a closed calculator/computer exam. You are, however, allowed to use notes and your textbook in the exam. The last page of the exam is a Standard Normal Table, in case you need it.

You have 2 hours (120 minutes) to take the exam. The exam is 120 points, meant to roughly correspond to one point per minute of the exam. You may want to use the point allocation for each problem as an indicator for pacing yourself on the exam.

In the event of an incorrect answer, any explanation you provide of how you obtained your answer can potentially allow us to give you partial credit for a problem. For example, describe the distributions and parameter values you used, where appropriate. It is fine for your answers to include summations, products, factorials, exponentials, and combinations, unless the question specifically asks for a numeric quantity or closed form. Where numeric answers are required, the use of fractions is fine.

Problem	**Score**
1 (24 pts)	
2 (20 pts)	
3 (20 pts)	
4 (20 pts)	
5 (12 pts)	
6 (24 pts)	
Total (120 pts)	

THE STANFORD UNIVERSITY HONOR CODE

A. The Honor Code is an undertaking of the students, individually and collectively:

(1) That they will not give or receive aid in examinations; that they will not give or receive unpermitted aid in class work, in the preparation of reports, or in any other work that is to be used by the instructor as the basis of grading;

(2) That they will do their share and take an active part in seeing to it that others as well as themselves uphold the spirit and letter of the Honor Code.

B. The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid as far as practicable, academic procedures that create temptations to violate the Honor Code.

C. While the faculty alone has the right and obligation to set academic requirements, the students and faculty will work together to create optimal conditions for honorable academic work.

I acknowledge and accept the letter and spirit of the honor code:

Signature: ____

NAME (print): ____

1. (24 points) You are playing a card game that uses **four** standard decks of cards. There are 208 cards total.

   Each deck has 52 cards (13 values each with 4 suits). Cards are only distinguishable based on their suit and value, not which deck they came from.

   a. (5 Points) In how many distinct ways can the cards be ordered?

   b. (4 points) You will be dealt the first two cards from the four decks of cards. Cards with values 10, Jack, Queen, King and Ace are considered "good" cards. What is the probability of getting two "good" cards?

   c. (5 points) Over the course of several rounds you observe 100 cards played. Out of the cards played only 15 were "good" cards. You are dealt the next two cards. What is the probability of getting two "good" cards now? You may assume that previously seen cards are not re-dealt.

d. (10 Points) Consider a standard deck of 52 cards with the ace of spades on the bottom. Let the "step" be an algorithm where you take the card from the top of the deck and randomly place it in one of the 52 positions among the remaining deck including the top and bottom of the deck. If you repeat this step until the ace of spades is on the top (then execute the step one last time), the deck is perfectly shuffled. How many times do you expect to have to repeat the "step" until the deck is shuffled?

(Recall that it is fine for your answers to include summations, products, factorials, exponentials, or combinations.)

The first three parts are the basis of card counting in Black Jack. Part (d) is the starting point in Stanford professor Persis Diaconosis' numerical analysis on randomness of card shuffles.

2. (20 points) You randomly sample 11 numbers from a set of 101 **distinct** numbers.

   a. (6 Points) **How many ways** can you sample 11 numbers such that the median of the sample is equal to the median of the original set? A median is the number separating the higher half of a data-set from the lower half. For example, the median of {0, 1, 4, 5, 9, 10, 100} is 5.

   b. (4 points) What is the **probability** that the median of the sample is equal to the median of the original set?

c. (10 Points) Let us denote the probability you calculated from Part B as p. We draw 100 independent samples (with 11 numbers each) from the original set. What is the probability that less than 10 of those samples will have the same median as the original set? Leave your answer in terms of p.

3. (20 points) The probability that a Netflix user likes a movie T_i from the "Tearjerker" genre:
   Given that they **like** the Tearjerker genre is p_i.
   Given that they **do not like** the Tearjerker genre is q_i.
   60% of Netflix users like the Tearjerker genre.

   Netflix assumes that, given a user's preference for the genre, liking movie T_i and T_j are **conditionally independent** events. You may express all your answers in terms of qs and ps.

   a. (4 Points) What is the probability that a user likes movies T_1, T_2 **and** T_3 given that they like the Tearjerker genre?

   b. (6 Points) What is the probability that they like movies T_1, T_2 **or** T_3 given that they like the Tearjerker genre?

c. (10 points) What is the probability that they like the Tearjerker genre given that they like T_1, T_2 **and** T_3.

Three movies in the Tearjerker genre that the CS109 team likes are: Forest Gump, The Boy in the Striped Pajamas and Good Will Hunting. We're kinda old school like that. This is the formulation behind the machine learning algorithm called Naïve Bayes. You will learn the machine learning part soon!

4. (20 points) You are part of a group at Stanford researching whether or not it would be possible to have a clean energy grid that is generated exclusively by wind.

   a. (4 Points) Let X be the amount of energy generated when the wind is blowing for a particular wind turbine. $X \sim N(\mu = 2, \sigma^2 = 64)$ mWh. Given that the wind is blowing, what is the probability that the wind farm produces more than 4 mWh of energy? Provide a numerical answer.

   b. (6 Points) The population on the grid is 40,000 people. In a given hour each person has an independent probability of 0.5 of using electricity from the grid. What is the approximate probability that more than 20,300 people use electricity in a given hour?

c. (10 Points) The wind blows at two wind farms independently with probability p_1, and p_2 respectively. The amount of electricity produced, **given that the wind is blowing** is distributed differently for each wind farm:

Wind farm 1: Amount of energy when the wind blows $W_1 \sim N(\mu_1, \sigma_1^2)$ gWh
Wind farm 2: Amount of energy when the wind blows $W_2 \sim N(\mu_2, \sigma_2^2)$ gWh

If the wind doesn't blow at a wind farm, that farm produces zero gWh.

What is the probability that the sum of electricity produced by the two wind farms is less than θ gWh?

Arun Majumdar's lab at Stanford is working on modernizing the electric grid. The winds of freedom blow especially hard in that lab. Wind generates a ton of electricity but it is highly stochastic.

5. (12 points) An autonomous car has two different instruments for determining its direction:

   The first instrument reports a direction $D_1 = T + X_1$
   The second instrument reports a direction $D_2 = T + X_2$

   Where $X_1 \sim N(\mu = 0, \sigma^2 = 1)$ and $X_2 \sim N(\mu = 0, \sigma^2 = 4)$ are independent Gaussian noise and T is the true direction of the autonomous car. Before checking the instruments, the car believes that all directions in the range 50 to 60 degrees are equally likely.

   What is the **probability density function** for the true direction given that the first instrument reports 57 degrees ($D_1 = 57$) and the second instrument reports 59 degrees ($D_2 = 59$)? Use a constant K in your function.

Sebastian Thrun's Stanford team won the DARPA self driving car grand challenge.

6. (24 points) You are making a mobile phone ride sharing application where people travelling the same route can share a car. A particular route is requested on average 2 times in any 5-minute period. Each request is independent.

   a. (7 Points) A user requests the route and you commit a car to take her. What is the probability that at least one other user will request the same route in the next five minutes? *Hint: it should be the same as the probability of getting at least one user without the original request.*

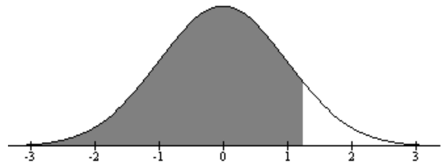
   b. (10 Points) A user requests the route and you commit a car to take her. All users who request the route in the next five minutes will be added to the car -- as long as the car has space. The car can fit up to three users. You will make \$6 for each user in the car (your revenue) minus \$7 (the operating cost). How much do you expect to make from this trip?

c. (7 points) What is the probability that it takes longer than 8 minutes to get one more request to the route?

This is the idea behind Lyft Line and Uber Pool.

Standard Normal Table

Note: An entry in the table is the area under the curve to the left of z, $P(Z \le z) = \Phi(z)$



z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
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	0.7703	0.7734	0.7764	0.7793	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	0.8906	0.8925	0.8943	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
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998