Chris Piech
CS109

Fall 2017
Nov 2$^{\text{nd}}$

CS109 Midterm Exam

This is a closed calculator/computer exam. You are, however, allowed to use notes 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.

LELAND STANFORD JUNIOR UNIVERSITY
DIE LUFT DER FREIHEIT WEHT

I acknowledge and accept the letter and spirit of the honor code. I pledge to write more neatly than I have in my entire life:

Signature: ____

Family Name (print): ____

Given Name (print): ____

Email (preferably your gradescope email): ____

1 Alpha TicTacToe Zero [18 points]

Consider a 3x3 TicTacToe board where each location can either have an empty space, an X, or an O. Each location is distinct (thus, even if two boards can be rotated to look the same, we consider them different). Here is an example board:

X		O
O	X	
		X

a. (5 points) How many unique ways are there of placing Xs and Os on a TicTacToe board such that there is at most one marker in each square? You do not have to follow the rules of the game and you do not have to fill each square.

$3^9 = 19683$

There are 3 possibilities (X, O, empty) for each of the 9 spaces on the board.

b. (6 points) The "Turn Rule" states that: players take turns and X always starts. After 5 moves how many unique TicTacToe boards are there that satisfy the Turn Rule (X has played three times and O has played twice)?

$$\binom{9}{3} \cdot \binom{6}{2} = \frac{9!}{3! \cdot 2! \cdot 4!}$$

From the 9 total spots on the board, choose 3 places to place an X. From the 6 remaining open spots, choose 2 to place an O.

c. (7 points) If both players play randomly, what is the probability that X will win after 5 moves (X has played three times and O has played twice)? X wins if their three pieces make a vertical, diagonal or horizontal rows. There are 8 such rows.

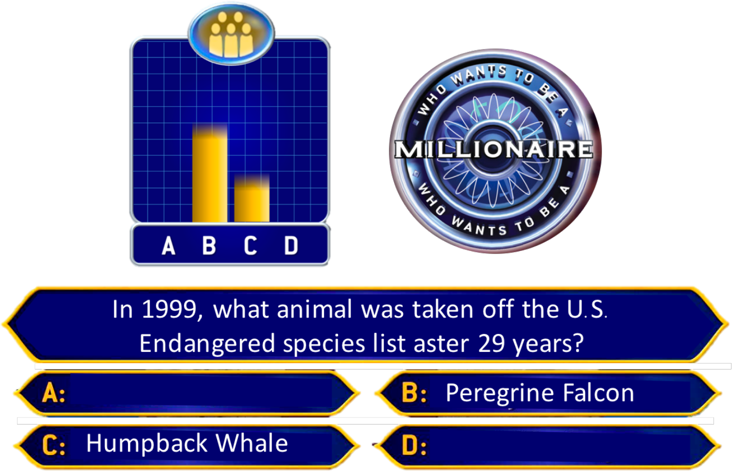
$$
\begin{aligned}
|S| &= \frac{9!}{3! \cdot 2! \cdot 4!} \\
|E| &= 8 \cdot \binom{6}{4} \\
p &= |E|/|S| = 0.095
\end{aligned}
$$

Every outcome in S is equally likely by symmetry. This answer was checked by simulation. p = 0.095384 after one million simulations.

2 Wisdom of the Crowds [24 points]

It doesn't take many experts for the collective to make a good decision! This phenomena is called the Wisdom of the Crowds and is heavily leveraged in crowd sourcing algorithms. Consider this example. In the game Who Wants to be a Millionaire, when contestants "ask the audience" each of the 200 members of the audience get to vote on the answer. Assume:

- There are two answers for each audience member to chose from, a Correct answer and an Incorrect answer.
- 10% of the audience are knowledgeable about the problem (call them experts). An expert votes for the Correct answer with a probability of 0.7, otherwise they vote for the Incorrect answer.
- 90% of the audience are not knowledgeable (call them non-experts). A non-expert votes randomly with equal likelihood between the Correct answer and the Incorrect answer.



In 1999, what animal was taken off the U.S. Endangered species list aster 29 years?

A:

B: Peregrine Falcon

C: Humpback Whale

D:

a. (6 points) What is the probability that exactly k of the experts vote for the Correct answer? You may assume that k is a number between 0 and 20 inclusive.

Answer: This is a binomial distribution with $n = 20$ and $p = 0.7$:

$$X \sim Bin(20, 0.7)$$

$$P(X = k) = \binom{20}{k}(0.7)^k(0.3)^{(20-k)}$$

b. (6 points) If exactly k of the experts vote for the Correct answer, what is the probability that the Correct answer will get at least 101 votes? (hint: the Correct answer needs at least 101 - k more votes from the non-experts).

Answer:

Now we have a binomial distribution with $n = 180$ and $p = 0.5$, and we need at least $101 - k$ votes from the non-experts:

$$Y \sim Bin(180, 0.5)$$

$$P(X + Y \geq 101 - k | X = k) = P(Y \geq 101 - k)$$

$$= \sum_{i=101-k}^{180} \binom{180}{i} (0.5)^i (0.5)^{(180-i)} = (0.5)^{180} \cdot \sum_{i=101-k}^{180} \binom{180}{i}$$

c. (6 points) Write an expression for the exact probability that the Correct answer will get at least 101 votes.

Answer: To get the probability of k experts correct and at least $101 - k$ others correct, we just multiply parts a and b. So to get the total probability, sum over all values of k:

$$\sum_{k=0}^{20} \left[\binom{20}{k} (0.7)^k (0.3)^{(20-k)} \sum_{i=101-k}^{180} \binom{180}{i} (0.5)^i (0.5)^{(180-i)} \right]$$

d. (6 points) Use an approximation to estimate the probability that the Correct answer gets at least 101 votes. You may leave your answer in terms of roots and/or values that could be looked up from the ϕ table. For full credit your approximation calculation should *not* include a summation or integral.

Answer: We will approximate both the expert distribution and the non-expert distributions as normals:

$$X \sim N(20 \cdot 0.7, 20 \cdot 0.7 \cdot 0.3) = N(14, 4.2)$$

$$Y \sim N(180 \cdot 0.5, 180 \cdot 0.5 \cdot 0.5) = N(90, 45)$$

Where we have used that the expectation of a binomial is np, and the variance is $np(1-p)$. The sum of two independent normal distributions is also a normal distribution:

$$X + Y \sim N(14 + 90, 45 + 4.2) = N(104, 49.2)$$

And then, making sure to apply continuity correction (since we are approximating a discrete distribution with a continuous one), we have

$$P(X+Y \geq 101) = 1 - P(X+Y < 101) = 1 - \Phi(\frac{100.5 - 104}{\sqrt{49.2}})$$

3 500 year flood planes [20 points]

The Huffmeister floodplane in Houston has historically been estimated to flood at an average rate of 1 flood every 500 years. A flood plane with that rate of flooding is called a "500 year" floodplane.

a. (4 points) What is the probability of observing at least 3 floods in 500 years?

Poisson RV with $\lambda = 1$ (flood per 500-year period)

$$P(X \geq 3) = 1 - \sum_{i=0}^{2} P(X = i)$$

$$= 1 - \sum_{i=0}^{2} \frac{\lambda^i}{i!} e^{-\lambda}$$

$$= 1 - \frac{5}{2e}$$

b. (5 points) What is the probability that a flood will occur within the next 100 years?

$$Y \sim Exp(\frac{1}{500}) \qquad F(Y) = 1 - e^{-\lambda Y}$$

$$F(100) = 1 - e^{-\frac{100}{500}} = 1 - e^{-\frac{1}{5}}$$

c. (5 points) What is the expected number of years until the next flood?

$$E[Y] = \frac{1}{\lambda} = 500$$

using $\lambda = \frac{1}{500}$ in terms of one year

d. (6 points) Say there are 10 independent 500 year floodplanes. What is the probability that more than 2 of them will have at least three floods in a given year?

First find the probability p^* that a single floodplane has at least 3 floods:

$X \sim Poisson(\frac{1}{500})$ = number of floods in a floodplane in one year

$$p^* = P(X \geq 3) = 1 - \sum_{i=0}^{2} \frac{\lambda^i}{i!} e^{-\lambda}$$

$$p^* = 1 - \sum_{i=0}^{2} \frac{(\frac{1}{500})^i}{i!} e^{-\frac{1}{500}}$$

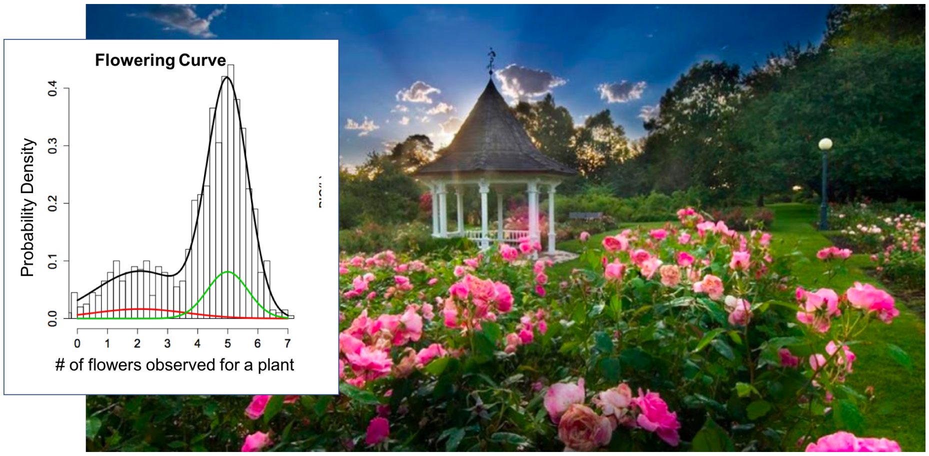
Then treat each floodplane as an independent trial of a binomial RV with $n = 10$ and $p = p^*$:

$Y \sim Bin(10, p^*)$ = number of floodplanes with $X \geq 3$

$$P(Y > 2) = \sum_{i=3}^{10} \binom{10}{i} (p^*)^i (1 - p^*)^{10-i}$$

4 Gaussian Mixture Model [22 points]

You are looking at the distribution of the number of flowers observed on rose bushes (a flowering curve) and you think you have made a discovery! Based on your flowering curve, you have reason to believe that there are two distinct species of roses. The species look the same but have different flowering patterns. The observed probability density of number of flowers is a “mixture”.



You believe that:
- Each flower is either from species A or species B.
- There are two times as many bushes of species A than there are of species B.
- The number of flowers on a bush for species A: $X_A \sim N(\mu = 4, \sigma^2 = 4)$.
- The number of flowers on a bush for species B: $X_B \sim N(\mu = 2, \sigma^2 = 9)$

a. (4 points) What is the probability that a rose bush from species A has more than 3 flowers?

Solution:
Note that for both a and b, you have to use continuity correction because you are using the Gaussian distribution to approximate a discrete distribution:

$$\begin{aligned} P(X_A > 3) &= 1 - P(X_A <= 3) \\ &= 1 - \phi\left(\frac{3.5 - 4}{2}\right) \\ &= 1 - \phi(-0.25) \\ &= \phi(0.25) \\ &= 0.5987 \end{aligned}$$

b. (4 points) What is the probability that a rose bush from species B has more than 3 flowers?

Solution:

$$\begin{aligned} P(X_B > 3) &= 1 - P(X_B <= 3) \\ &= 1 - \phi\left(\frac{3.5 - 2}{3}\right) \\ &= 1 - \phi(0.5) \\ &= 0.3085 \end{aligned}$$

c. (7 points) Without knowing which species a rose bush is from, what is the probability that a rose bush has more than 3 flowers? You can use p_a for your answer to part (a) and p_b for your answer to part (b).

Solution:

$$
\begin{aligned}
P(X > 3) &= P(X > 3, X = A) + P(X > 3, X = B) \\
&= P(X > 3|X = A)P(X = A) + P(X > 3|X = B)P(X = B) \\
&= P(X_A > 3)P(X = A) + P(X_B > 3)P(X = B) \\
&= p_a(\frac{2}{3}) + p_b(\frac{1}{3})
\end{aligned}
$$

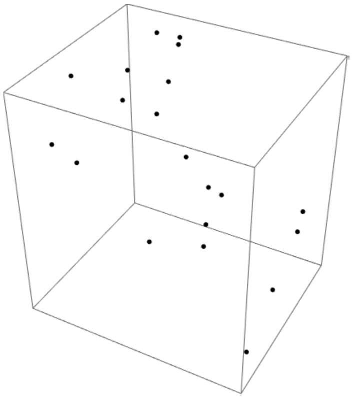
d. (7 points) A rosebush has 3 flowers. How many times more likely is the rosebush to be from species B?

$$
\begin{aligned}
\frac{P(X = B|X = 3)}{P(X = A|X = 3)} &= \frac{P(X = B, X = 3)}{P(X = A, X = 3)} \\
&= \frac{P(X = 3|X = B)P(X = B)}{P(X = 3|X = A)P(X = A)} \\
&= \frac{\varepsilon f(X = 3|X = B)P(X = B)}{\varepsilon f(X = 3|X = A)P(X = A)} \\
&= \frac{\frac{1}{\sqrt{2\pi 9}} e^{\frac{(3-2)^2}{2\cdot 9}} \cdot \frac{1}{3}}{\frac{1}{\sqrt{2\pi 4}} e^{\frac{(3-4)^2}{2\cdot 4}} \cdot \frac{2}{3}} \\
&= \frac{1}{3} e^{-\frac{5}{72}} \approx 0.357
\end{aligned}
$$

5 Curse of Dimensionality [20 points]

In machine learning projects we often work in high dimensions, and high dimension spaces have some surprising probabilistic properties.

A random value X_i is a Uni(0, 1).
A random point of dimension d is a list of d random values: $[X_1 \dots X_d]$.



a. (4 points) A random value X_i is close to an edge if X_i is less than 0.01 *or* X_i is greater than 0.99. What is the probability that a random value is close to an edge? **Solutions:**Let E be the event that a random value is close to an edge. $P(E) = P(X_i < 0.01) + P(X_i > 0.99) = 0.02$

b. (4 points) A random point $[X_1, X_2, X_3]$ of dimension 3 is close to an edge if *any* of it's values are close to an edge. What is the probability that a 3 dimensional point is close to an edge?

Solutions: The event is equivalent to the complement of none of the dimensions of the point is close to an edge, which is: $1 - (1 - P(E))^3 = 1 - 0.98^3$

c. (4 points) A random point $[X_1, \dots X_{100}]$ of dimension 100 is close to an edge if *any* of it's values are close to an edge. What is the probability that a 100 dimensional point is close to an edge?

Solutions: Similarly, it is: $1 - (1 - P(E))^{100} = 1 - 0.98^{100}$

6 Goodbye integral, my old friend [14 points]

When working with continuous random variables it can get hard to calculate probabilities by hand. For example consider a random variable X that can takes on values in the range $0 \leq x \leq 1$, and has probability density function:

$$f(x) = \frac{1}{K} \cdot g(x)$$

Where g(x) is some terribly nasty and non-integratable function. For your sanity I won't even write out g. It is that bad. In such a situation, we can turn to the power of computers to help us (you may assume that while g is impossible to integrate, it is straightforward to code g as a function in a programming language). The key idea that we are going to use is called Monte Carlo Integration:

Generate N values $(X_1, X_2, \ldots, X_N)$ uniformly sampled over a range (a, b). We can approximate the integral of a function h over (a, b) as:

$$\int_a^b h(x)\, dx \approx \frac{(b-a)}{N} \sum_{i=1}^{N} h(X_i)$$

Pretty amazing! Why did we bother with integrals at all? This question requires you to write pseudo code. Such code does not have to compile, but it should be specific enough that a knowledgable programmer could implement what you have described. You may use a function `random(a, b)` which returns a sample from $X \sim \text{Uni}(a, b)$.

a. (7 points) How could you use Monte Carlo Integration to find K? Give your answer as pseudo-code.

```
sum = 0
for i in range(N):
    sum += g(random(0, 1))
return sum / N
```

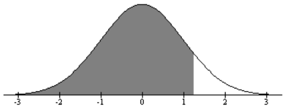
b. (7 points) Given that you know K, how could you find $P(X < 0.5)$? Again, provide pseudo-code.

```
sum = 0
for i in range(N):
        sum += g(random(0, .5))
return .5 * sum / (N * K)
```

That's the last question of the exam! We hope you had fun. Monte carlo sampling techniques (aka particle filters) are an increasingly prolific tool for solving more complex probability questions. Gaussian Mixture Models are a workhorse for unsupervised clustering. The Curse of Dimensionality and the Wisdom of the Crowds are real world phenomena that are worth understanding.

Standard Normal Table

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



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