Math 201: Introduction to Probability

Midterm 2

April 4, 2019

NAME (please print legibly): ________________________________

Your University ID Number: ________________________________

Instructions:

1. Indicate your instructor with a check in the appropriate box:

Krishnan	MW 14:00	
Tucker	MW 10:25	

2. Read the notes below:

   - **The presence of any electronic or calculating device at this exam is strictly forbidden, including (but not limited to) calculators, cell phones, and iPods.**
   - **Notes of any kind are strictly forbidden.**
   - **Show work and justify all answers. You may not receive full credit for a correct answer if insufficient work is shown or insufficient justification is given.**
   - **You do not need to simplify complicated numerical expressions such as $\binom{100}{30}$ and $50!$ to a number.**
   - **You are responsible for checking that this exam has all 16 pages.**

3. Read the following Academic Honesty Statement and sign:

   I affirm that I will not give or receive any unauthorized help on this exam, and that all work will be my own.

   Signature:________________________________

Formulae

If $X \sim \text{Binomial}(n, p)$, then

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, \ldots, n$$

We also have $E[X] = np, \text{Var}(X) = np(1-p)$.

If $X \sim \text{Geometric}(p)$, then

$$P(X = k) = (1-p)^{k-1} p \quad k = 1, 2, \ldots$$

The expectation of a Geometric is $E[X] = \frac{1}{p}$

The cdf of a random variable X is defined by

$$F(t) = P(X \leq t)$$

If X is continuous and has pdf $f(x)$,

$$F(t) = \int_{-\infty}^{t} f(x) dx.$$

QUESTION	VALUE	SCORE
1	5	
2	5	
3	10	
4	10	
5	10	
6	10	
7	15	
8	10	
9	10	
10	15	
TOTAL	100	

1. (5 points) Suppose a fair die is rolled 7 times. What is the mean value of the sum of the 7 rolls?

Solution: The expected value of one roll is 3.5. Let X_i represent the value of roll i. We must have $E[X_i] = 3.5$ for $i = 1, \ldots, 7$. Let $S = X_1 + \cdots + X_7$ be the sum of the seven rolls. By linearity of expectation,

$$E[S] = 7 \cdot 3.5$$

2. (5 points) Let X be a geometric random variable with parameter p (i.e. $X \sim Geom(p)$). If the average (mean) amount of time it takes to see the first success is 8, find the value of the parameter p.

Solution: The expectation of a geometric is $E[X] = 1/p$. This gives $p = 1/8$.

3. (10 points) Suppose $Z \sim N(\mu, \sigma^2)$ (that is, Z has the normal distribution with mean μ and variance σ^2), where μ and σ are unknown parameters.

(a) What is the probability that $Z \geq \mu + 2\sigma$, correct to 4 decimal places?

Solution: $P(Z \geq \mu + 2\sigma) = P((Z - \mu)/\sigma > 2)$. $(Z - \mu)/\sigma$ is a standard normal, and so we have

$$P((Z - \mu)/\sigma > 2) = 1 - \Phi(2) = 0.0228$$

(b) What is the variance of $3Z + 3$? You may express this in terms of μ and σ.

Solution:

$$\mathrm{Var}(3Z + 3) = 3^2 \,\mathrm{Var}(Z) = 9\sigma^2.$$

4. (10 points)

Suppose that we have a random variable X such that the probability density function f for X is given by

$$f(x) = \begin{cases} 0 & \text{if } x < -1 \\ 1/4 & \text{if } -1 \leq x < 0 \\ 3/4 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x > 1 \end{cases}$$

Find the cumulative distribution function (cdf) function for X.

Solution: Let F denote the cdf of X. Clearly $F(t)$ is 0 when $t < -1$. When t is between -1 and 0, we have $F(t) = \int_{-1}^{t} \frac{dx}{4} = \frac{t+1}{4}$. When t is 0 and 1, we have $F(t) = 1/4 + \int_0^1 \frac{3\,dx}{4} = \frac{1}{4} + \frac{3t}{4}$. When $t > 1$, we have $F(t) = 1$, so to summarize we hae

$$F(x) = \begin{cases} 0 & x < -1 \\ \frac{1}{4}(x+1) & -1 \leq x < 0 \\ \frac{1}{4} + \frac{3x}{4} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

5. (10 points)

Suppose that you are offered two games of chance. Game #1 is a game where a fair coin is flipped twice, and you lose 10 dollars if the number of tails that appear is zero , while you gain 3 dollars if the number of tails that appears is one or two. Let X be the random amount you get paid in Game #1. Game #2 is a game where you are paid Y, where Y is a random variable with probability density function

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2x/9 & \text{if } 0 \leq x \leq 3 \\ 0 & \text{if } x > 3 \end{cases}$$

You are only interested in a big pay off. Which game (#1 or #2) should you choose if you want to maximize your chance of winning at least two dollars? (Note: you may assume that any decimal amount of dollars is possible.)

Solution: In Game #1, you have a 3/4 chance of winning two or more dollars. In Game #2, we get a

$$\int_2^3 f(x)\,dx = \int_2^3 2x/9\,dx = \frac{3^2}{9} - \frac{2^2}{9} = 5/9$$

chance of winning at least two dollars. Since $5/9 < 3/4$, he should choose Game #1 to maximize his chance of winning at least two dollars.

6. (10 points)

Two numbers, X_1 and X_2, are chosen with replacement from the set $\{1, 2, 3\}$. We let $Y = |X_1 - X_2|$.

Find the probability mass function for Y.

Solution.

The number of ways of choosing X_1 and X_2 is $3 \cdot 3 = 9$. There are four ways that $|X_1 - X_2|$ can be 1 ($(1, 2)$, $(2, 1)$, $(2, 3)$, $(3, 2)$), three ways it can be zero (the obvious ones), and two ways it can be 2 ($(1, 3)$, and $(3, 1)$).

So the probablity mass function for $|X_1 - X_2|$ is given by

$$P(|X_1 - X_2| = i) = \begin{cases} 0 & i \neq 0, 1, 2 \\ 3/9 & i = 0 \\ 4/9 & i = 1 \\ 2/9 & i = 2 \end{cases}$$

7. (15 points) Let

$$F(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{3} & 0 \leq t < 1 \\ \frac{5}{6} & 1 \leq t < 3 \\ 1 & 3 \leq t \end{cases}$$

be the cdf of a random variable X.

(a) Is X continuous or discrete? If X is discrete, find the pmf (mass function); if it is continuous, find its pdf (density function).

Solution: X is discrete, so the pmf is

$$P(X = 0) = \frac{1}{3}$$
$$P(X = 1) = \frac{5}{6} - \frac{1}{3} = \frac{1}{2}$$
$$P(X = 3) = 1 - \frac{5}{6} = \frac{1}{6}$$

(b) As above, we have

$$F(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{3} & 0 \leq t < 1 \\ \frac{5}{6} & 1 \leq t < 3 \\ 1 & 3 \leq t \end{cases}$$

Find $E[X^4 e^{-X}]$. **Hint: You do not need to evaluate this sum explicitly, just write the correct expression down**

Solution: We use the formula

$$\mathbb{E}[g(X)] = \sum_{i \in ran(X)} g(i) P(X = i).$$

Here $\text{ran}(X) = \{0, 1, 3\}$. Then

$$E[X^4 e^{-X}] = 0\frac{1}{3} + 1^4 e^{-1}\frac{1}{2} + 3^4 e^{-3}\frac{1}{6}$$

8. (10 points) Sara finds that each of her math homework assignments are taking a random amount of time. She discovers that the probability that it takes her time t or more is equal to $1/(t+1)^2$ for $t > 0$. What is average (mean) amount of time it takes for her to finish her homework assignment? *Hint: You can find the pdf from the information given in the problem, or you can use a formula for expectation that we derived in class.*

Solution.

Her cumulative distribution function if $F(t) = 1 - \frac{1}{(t+1)^2}$, so her pdf $f(t) = F'(t) = \frac{2}{(t+1)^3}$. Thus the expectation is

$$\int_0^\infty t \cdot \frac{2}{(t+1)^3}\, dt.$$

Writing t as $(t+1) - 1$, we obtain

$$2\int_0^\infty \frac{1}{(t+1)^2} - \frac{1}{(t+1)^3}\, dt = 2(1 - 1/2) = 1.$$

Some of you may have seen (in office hours, study hall, or class) a formula similar to the one in Homework 8, number 2, which says that the expectation of a **nonnegative** random variable X is given by $\int_{-\infty}^{\infty} P(X \geq t)\, dt$, which in this case gives

$$\int_0^\infty \frac{1}{(t+1)^2}\, dt = 1.$$

(Note almost no one used this formula but it is correct.)

9. (10 points) A random sample of 1000 Americans are polled about their voting preferences. They vote either Democrat or Republican. Democrats make up 49% of the total number of voters. Since the sample of 1000 is taken from a very large pool of voters, you may assume that the sampling is done with replacement. Approximate the probability correct to two decimal places that at least 500 people in the sample of 1000 people will say they vote Democrat. *Hint: Use the approximation* $\sqrt{1000 \cdot 0.49 \cdot 0.51} \approx 16$.

Solution: Let $X \sim \text{Binomial}(1000, 0.49)$ represent the number of Democrats in a sample of 1000 voters. This has mean and standard deviation:

$$\mathbb{E}[X] = 1000 \cdot 0.49 = 490 \quad \sqrt{\text{Var}(X)} = \sqrt{1000 \cdot 0.49 \cdot 0.51} \approx 16$$

We have to approximate the probability that

$$\begin{aligned} P(X > 500) &= P(\frac{X - 490}{16} > \frac{500 - 490}{16}) \\ &\approx P(Z > \frac{5}{8}) \\ &= 1 - \Phi(0.625) = 0.2660 \end{aligned}$$

If you got 0.26 or 0.27 you got points for this problem. If you used the exact value of $\sqrt{1000 \cdot 0.49 \cdot 0.51}$, you would have gotten 0.2635 instead.

10. (15 points)

Consider the square in the plane consisting of all (x, y) such $0 \leq x \leq 2$ and $0 \leq y \leq 2$, and let $Q = (x, y)$ be a point chosen uniformly at random inside the square. Let $X = \max(x, y)$.

(a) **8 pts** Find the cumulative distribution function (cdf) for X. *Hint: Start with the definition for the cdf of X and draw a picture.*

Solution.

The set of points in the square $0 \leq x \leq 2$, $0 \leq y \leq 2$ such that $\max(x, y) < t$ (for $0 \leq t \leq 2$) forms a square with side length 2 and thus a set having area t^2. Hence the cumulative distribution function $F(x)$ is given by

$$F(t) = \begin{cases} 0 & t < 0 \\ t^2/4 & 0 \leq t \leq 2 \\ 1 & t > 2 \end{cases}$$

(where the division by 4 is because the area of the entire square $0 \leq x \leq 2$, $0 \leq y \leq 2$ is 4).

Some of you were also quite clever about this, and noticed that since Q was uniformly distributed in a square, the x and y coordinates of Q were independently distributed as Uniform$[0, 2]$ random variables. Then, if $Q = (A, B)$, then

$$\begin{aligned} P(\max(x, y) \leq t) &= P(A \leq t, B \leq t) \\ &= P(A \leq t)\, P(B \leq t) \\ &= \frac{t}{2}\frac{t}{2} \end{aligned}$$

when $0 \leq t \leq 2$.

(b) **7 pts** Find $E[X]$.

Solution.

Differentiating F from (a) we find that the p.d.f. is $f(t) = t/2$ for t between 0 and 2 and 0 otherwise, so the expectation is

$$\int_0^2 t \cdot t/2\, dt = \int_0^2 t^2/2\, dt = \frac{2^3}{6} = \frac{4}{3}.$$

If you realized that you had to use the cdf from part a to find the pdf, then you got 3 points. If you correctly computed an expectation using the pdf, then you got 4 more points. Even if you used an incorrectly computed cdf from part (a), you got points here.

Here is another clever solution due to Zachary Polansky. Take the interval $[0, 2]$, and drop points A and B on it. There are going to be three intervals $[0, \min(A, B)]$, $[\min(A, B), \max(A, B)]$ and $[\max(A, B), 2]$. Since

$$\begin{aligned} &|[0, \min(A, B)]| + |[\min(A, B), \max(A, B)]| + |[\max(A, B), 2]| \\ &\qquad = \min(A, B) - 0 + \max(A, B) - \min(A, B) + 2 - \max(A, B) \\ &\qquad = 2 \end{aligned}$$

By "uniformity" (one would have to prove this),

$$\begin{aligned} &\mathbb{E}[|[0, \min(A, B)]| + |[\min(A, B), \max(A, B)]| + |[\max(A, B), 2]|] = 2 \\ &\mathbb{E}[2 - \max(A, B)] = \mathbb{E}[\max(A, B) - \min(A, B)] = \mathbb{E}[\min(A, B) - 0] = \frac{2}{3} \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbb{E}[\min(A, B) - 0 + \max(A, B) - \min(A, B)] = \frac{2}{3} + \frac{2}{3} = \frac{4}{3} \\ &\Rightarrow \mathbb{E}[\max(A, B)] = \frac{4}{3} \end{aligned}$$

Appendix E

Table of values for $\Phi(x)$

	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	0.7704	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