NAME : Key

Math 20

Midterm 2
August 4, 2017

Prof. Pantone

INSTRUCTIONS: This is a closed book exam and no notes are allowed. You are not to provide or receive help from any outside source during the exam except that you may ask the instructor for clarification of a problem. You have 120 minutes and you should attempt all problems.

- Print your name in the space provided.
- **Calculators or other computing devices are not allowed.**
- Except when indicated, you must show all work and give justification for your answer. **A correct answer with incorrect work will be considered wrong.**

All work on this exam should be completed in accordance with the Dartmouth Academic Honor Principle.

If you use facts of a distribution, you must name the distribution and justify why it's appropriate.

TIPS:

- You don't have numerically expand all answers. For example, you can leave an answer in the form $10! \cdot \binom{5}{3}^2$, rather than 362880000.
- Use scratch paper to figure out your answers and proofs before writing them on your exam.
- Work cleanly and neatly; this makes it easier to give partial credit.

Problem	Points	Score
1	24	
2	20	
3	10	
4	16	
5	10	
6	8	
7	12	
8	0	
Total	100	

Section 1: True/False.

1. (24) Choose the correct answer. *No justification is required for your answers. No partial credit will be awarded.*

(a) The probability density function of a continuous random variable is a function $f(t)$ such that $P(a \leq X \leq b) = \int_a^b f(t)dt$.

This is the definition.

True   False

(b) The standard deviation of a Bernoulli random variable X with success probability p is $p(1-p)$.

$Var(X) = p(1-p)$

$\sigma_X = \sqrt{p(1-p)}$

True   False

(c) A Poisson random variable is memoryless.

Ex: Is $P(X>8 \mid X>5) = P(X>3)$?

$$\frac{\frac{\lambda^8 e^{-\lambda}}{8!}}{\frac{\lambda^5 e^{-\lambda}}{5!}} = \lambda^3 \cdot \frac{5!}{8!} \neq \frac{\lambda^3 e^{-\lambda}}{3!}$$

True   False

(d) An exponential random variable is memoryless.

True   False

(e) The probability density function of a random variable is the derivative of the cumulative distribution function.

True   False

(f) Markov's Inequality says that for any non-negative random variable X,

$$P(X \leq a) \geq \mathbb{E}[X]/a.$$

It says $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$

True   False

(g) If $\mathbb{E}[X] = 5$ and X is an independent random variable, then $\mathbb{E}[1/X] = 1/5$.

This is not a linearity property.

Ex: Let $\Omega = \{4, 6\}$
$m(4) = \frac{1}{2}$ $m(6) = \frac{1}{2}$.
$X(4) = 4$ $X(6) = 6$
Then, $\mathbb{E}[X] = 5$,
but
$\mathbb{E}[1/X] = \frac{1}{2}\left(\frac{1}{4}\right) + \frac{1}{2}\left(\frac{1}{6}\right)$
$= \frac{1}{8} + \frac{1}{12} = \frac{5}{24} \neq \frac{1}{5}$.

True   False

Also, the phrase "independent RV" makes no sense! Independent from what?

(h) A discrete random variable has only a finite number of non-zero outcomes.

True   False

Can be infinite.

Section 2: Short Answer.

2. (20) *No justification is required for your answers, unless otherwise stated. No partial credit will be awarded.*

(a) You flip a fair coin ten times. Is the probability that you get between 2 and 9 heads (inclusive) $\geq 90\%$ or $< 90\%$? Provide a short justification.

This is a binomial distribution.

Success = heads $\Rightarrow p = \frac{1}{2}$

10 trials $\Rightarrow n = 10$

X = # heads

$P(X=k) = \binom{10}{k} \cdot \frac{1}{2}^k \cdot \frac{1}{2}^{10-k}$

So, $P(2 \leq X \leq 9) = 1 - P(X=0) - P(X=1) - P(X=10)$

$= 1 - \binom{10}{0}\left(\frac{1}{2}\right)^{10} - \binom{10}{1}\cdot\left(\frac{1}{2}\right)^{10} - \binom{10}{10}\cdot\left(\frac{1}{2}\right)^{10}$

$= 1 - \frac{12}{1024}$ ← way less than $\frac{1}{10}$

so ≥ 0.9.

(b) Let X be a random variable with mean 0 and variance 2. Find the smallest r such that you can guarantee that $P(|X| \geq r) \leq \frac{1}{50}$.

Chebyshev's Inequality:

$$P(|X - \underbrace{E[X]}_{=0}| \geq r) \leq \frac{Var(X)}{r^2} \quad \leftarrow = 2$$

So, $P(|X| \geq r) \leq \frac{2}{r^2} \overset{\text{want}}{\leq} \frac{1}{50}$

What value of r makes $\frac{2}{r^2} \leq \frac{1}{50}$?

$\Rightarrow r^2 \geq 100 \Rightarrow r \geq 10$,

so, $\boxed{r = 10}$

(c) Assume that the probability that any given student sleeps through any given final exam is 1/500. If 2000 students take three final exams each in a quarter, what is the probability that exactly 10 students sleep through their exam? Use a Poisson approximation.

2000 students $\cdot$ 3 exams each = 6000 exams

Binomial would be $P(X=10) = \binom{6000}{10}\left(\frac{1}{500}\right)^{10}\left(\frac{499}{500}\right)^{5990}$.

Poisson Approximation: $\lambda = n \cdot p = 6000 \cdot \frac{1}{500} = 12$

Rate = $\dfrac{12 \text{ missed exams}}{6000 \text{ exams}}$

$$P(X=10) \approx \boxed{\frac{12^{10}e^{-12}}{10!}} \approx 10.48\%$$

(d) Suppose that X is an exponential random variable with $\tau = 3$. What is $P(1 \le X \le 4)$?

$\tau = 3 \Rightarrow \lambda = \frac{1}{\tau} = \frac{1}{3}$.

So, $f(t) = \frac{1}{3}e^{-\frac{1}{3}t}$.

Hence, $P(1 \le X \le 4) = \int_1^4 \frac{1}{3}e^{-1/3\, t}\,dt$

$$= \left[-e^{-t/3}\right]_{t=1}^{t=4} = -e^{-4/3} + e^{-1/3}$$

$$= e^{-1/3} - e^{-4/3}$$

$$= e^{-1/3}\left(1 - \frac{1}{e}\right).$$

Section 3: Free Response.

You must show all work to receive credit.

If you need more space you may use the back of the page. You must clearly indicate on the front of the page that there is more work on the back of the page. Please work neatly.

3. (10) All airlines overbook their flights, knowing that some people won't show up. Occasionally, too many people do show up and some passengers get bumped.

One of Delta's Boeing 777-200ER planes has 291 seats. Delta has sold 300 tickets.

7

(a) If each passenger shows up to the flight with probability 0.95, and if all passengers are mutually independent, what is the probability that someone will need to be bumped? (You may leave your answer as a summation.)

This is modeled by a binomial distribution, where success = missing a flight.

$p = 0.05$ $\quad X = \#$ of people who miss their flight

$n = 300.$

P(someone gets bumped) $= P(X < 9)$

$$\boxed{= \sum_{k=0}^{8} \binom{300}{k} (0.05)^k (0.95)^{300-k}.}$$

Could also set success = not missing a flight, $X = \#$ of people who show up. Then

$$P(X > 291) = \boxed{\sum_{k=292}^{300} \binom{300}{k} (0.95)^k (0.05)^{300-k}}$$

3

(b) Give at least one real-world reason why the assumption of independence isn't a good one.

Families traveling together

Weather delays

Long security lines

4. (16) Suppose that you have n identical pieces of radioactive material, and that each piece emits electrons at a rate of r per minute. Let X_i for $1 \leq i \leq n$ denote the random variable for the distribution of time between emission of electrons for the ith piece of radioactive material.

(a) Is X_1 a discrete random variable or a continuous random variable? Explain your answer.

2

Continuous — X_1 measures waiting time between emissions, which can take the value of any real # ≥ 0.

(b) More specifically, what is the name for the type of distribution that X_1 is? If you said X_1 is discrete, give its probability distribution function $P(X = k)$. If you said X_1 is continuous, give its probability density function.

2

Continuous waiting time $\Rightarrow$ exponential distribution

PDF $= f(t) = re^{-rt}$.

(c) Suppose you turn on your Geiger counter—which measures when electrons are emitted—and put it near the first sample. How long do you expect to wait until an electron is emitted? In other words, what is $\mathbb{E}[X_1]$?

2

$\mathbb{E}[X_1]$ = Expected waiting time $= \tau = \frac{1}{r}$.

We can verify this: $\mathbb{E}[X] = \int_0^\infty tre^{-rt}\,dt = \left[te^{-rt} + \frac{1}{r}e^{-rt}\right]_0^\infty$

$= \frac{1}{r}$.

integration by parts.

5

(d) Next, let's say that you want to check to make sure all your elements actually are radioactive, but you can only check one at a time. So you hold your Geiger counter in front of the first sample until you detect an electron emission, then instantly move it to the second sample until you detect an electron emission, etc, all the way up until you detect an emission on the nth sample. Let Y be the random variable for how long this process takes. Express Y in terms of $X_1, \ldots, X_n$ and find $\mathbb{E}[Y]$.

The length of this process =

[waiting time for sample 1] + [w.t. for 2] + ...

So, $Y = X_1 + X_2 + \cdots + X_n$.

Hence, $\mathbb{E}[Y] = \mathbb{E}[X_1 + \cdots + X_n]$

by linearity of expectation $\longrightarrow$ $= \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]$

$= \frac{1}{r} + \cdots + \frac{1}{r}$

$= \frac{n}{r}$

5

(e) Later, you take all the samples and bring them really close together, and you hold your Geiger counter up to the whole bunch. If any one sample emits an electron, you'll be able to detect it. Let Z be the random variable for the amount of time until any one of the samples emits an electron. Express Z in terms of $X_1, \ldots, X_n$ and find $\mathbb{E}[Z]$. What type of distribution does Z have?

Z is waiting time until any electron is emitted. So,

$$Z = \min(X_1, X_2, \ldots, X_n)$$

Let's find the distribution of Z, just like on the homework. Finding the CDF is easiest:

$$\underbrace{P(Z \le a)}_{\text{at least one is } \le a} = 1 - \underbrace{P(Z > a)}_{\text{all are } > a}$$

$$= 1 - P(X_1 > a)P(X_2 > a) \cdots P(X_n > a)$$

$$= 1 - e^{-ar} \cdot e^{-ar} \cdots e^{-ar}$$

$$= 1 - e^{-(nr)a}$$

Normally, we would take the derivative to find the PDF, then use the formula $\mathbb{E}[X] = \int_0^\infty x f(x)\,dx$, but, the fact that the CDF of Z is $1 - e^{-nra}$ tells us that Z is itself also an exp. dist. with rate nr. So, $\mathbb{E}[Z] = \boxed{\frac{1}{rn}}$.

5. (10)

5

(a) Prove that for all random variables X,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

$$\text{Var}(X) = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2]$$

$$= \mathbb{E}[X^2] - \underbrace{2\mu\,\mathbb{E}[X]}_{\mu = \mathbb{E}[X]} + \mu^2$$

by linearity of expectation

$$= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2$$

$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

5

(b) Prove that if X and Y are independent random variables then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).$$

Point out where in your proof you used independence.

Using (a) leads to a simpler proof! ☆

$$\text{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2$$

$$= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2$$

(by linearity) $$= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] + (\mathbb{E}[Y])^2$$

$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 + \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$$

independence for $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ $\longrightarrow$ $+ 2\mathbb{E}[X]\mathbb{E}[Y] - 2\mathbb{E}[X]\mathbb{E}[Y]$

$$= \text{Var}(X) + \text{Var}(Y). \quad \blacksquare$$

6. (8) A box contains k green marbles and ℓ yellow marbles. One at a time, you randomly draw a marble from the box. If it's green you put it back and draw again, if it's yellow you keep it and the experiment is over.

4

(a) What is the probability that the experiment takes exactly n draws to complete?

This is a geometric distribution.

X = waiting time until yellow is drawn

$p(\text{yellow}) = \frac{\ell}{k+\ell}$ $\quad$ $p(\text{green}) = \frac{k}{k+\ell}$

So, $P(X=n) = \boxed{\left(\frac{k}{k+\ell}\right)^{n-1}\left(\frac{\ell}{k+\ell}\right) = \frac{k^{n-1}\ell}{(k+\ell)^n}}$

4

(b) What is the expected number of draws that it takes to complete the experiment?

In a geometric distribution,

$\mathbb{E}[X] = \frac{1}{\text{prob. success}} = \frac{1}{\frac{\ell}{k+\ell}} = \boxed{\frac{k+\ell}{\ell} = 1 + \frac{k}{\ell}}$

7. (12) For each of the functions $f(x)$ below, check if $f(x)$ is a valid probability density function. If not, explain why not. If it is, find the cumulative density function $F(x)$.

4

(a) $f(x) = \begin{cases} 0, & x < 0 \\ x - \frac{1}{2}, & 0 \leq x \leq 2 \\ 0, & x > 2 \end{cases}$

$f(0) = -\frac{1}{2}$. Densities cannot take negative values.

4

(b) $f(x) = \begin{cases} 0, & x < 0 \\ \frac{3}{2}\sqrt{x}, & 0 \leq x \leq 1 \\ 0, & x > 1 \end{cases}$

$f(x) \geq 0$ for all x? ✓

$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 \frac{3}{2}\sqrt{x}\,dx = [x^{3/2}]_0^1 = 1$ ✓

CDF $\Rightarrow F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x f(t)\,dt = [t^{3/2}]_0^x = x^{3/2}$

if $0 \leq x \leq 1$

So, $F(x) = \begin{cases} 0, & x < 0 \\ x^{3/2}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$

4

(c) $f(x) = \begin{cases} 0, & x < 0 \\ 5(1-x)^4, & 0 \leq x \leq 1 \\ 0, & x > 1 \end{cases}$

$f(x) \geq 0$ for all x? ✓

$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 5(1-x)^4\,dx = [-(1-x)^5]_0^1$

$= (-0) - (-1) = 1$ ✓.

if $0 \leq x \leq 1$

CDF $\Rightarrow F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x 5(1-t)^4\,dt = [-(1-t)^5]_0^x$

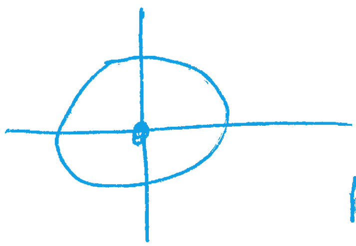
$= -(1-x)^5 + 1 = 1 - (1-x)^5$.

So, $F(x) = \begin{cases} 0, & x \leq 0 \\ 1-(1-x)^5, & 0 \leq x \leq 1 \\ 1, & x \geq 1 \end{cases}$

8. (0) **Bonus:** *(5 points)*

3

(a) Consider a circle of radius 1, centered at $(0,0)$. Let X be the random variable for the distance from the center of the circle to a point (x, y) chosen uniformly at random from all points within the circle. What is $\mathbb{E}[X]$?



Let $f(x)$ be the density function

$F(x)$ the CDF.

Note $F(X) = 0$ when $X < 0$

$F(X) = 1$ when $X > 1$.

Otherwise, $P(X \le d) =$ (area of subcircle of radius d)/(area of whole circle)

$= \pi d^2/\pi = d^2$.

Hence $F(x) = \begin{cases} 0, & x<0 \\ x^2, & 0 \le x \le 1 \\ 1, & x>1 \end{cases}$

$\Rightarrow f(x) = 2x$

$\Rightarrow \mathbb{E}[X] = \int_0^1 x(2x)\,dx = \left[\frac{2}{3}x^3\right]_0^1 = \boxed{\frac{2}{3}}$

2

(b) What if you choose a point not uniformly, but instead by first picking $r \in [0,1]$ uniformly at random, then an angle $\theta \in [0, 2\pi)$ uniformly at random, and make the point (r, θ) in polar coordinates. Now what is the expected distance from such a point to the origin?

Now r is the only thing that matters.

Hence $\mathbb{E}[X] = \mathbb{E}[r] = \boxed{\frac{1}{2}}$.

Lesson: You have to be very careful how you pick random points!