MATH 201 (SUMMER 2023, SESH A2)

LECTURE 19 : 06/20/23

ANURAG SAHAY
OFF HRS: BY APPT (VIA ZOOM)

LECTURES :
9:00 AM – 11:15 AM (ET)
M, T, W, R

email: anuragsahay@rochester.edu

Zoom ID :
979-4693-0650

COURSE WEB PAGE https://people.math.rochester.edu/grads/asahay/summer2023/math201/index.html

ALL PHOTOS TAKEN FROM TEXTBOOK

ANNOUNCEMENTS

1. UPCOMING DEADLINES : (i) WW 10 – TODAY (ii) HW 9, HW 10* – WED
   * : EXTRA CREDIT
2. FINAL EXAM ON THURSDAY, 9 :00 AM – 12 :00 PM } USUAL TIME.
3. CLASS ON WEDNESDAY / TOMORROW IS CANCELLED.
4. OFFICE HOURS : TW : 11:15 AM – 12:15 PM, W: 4:00 PM – 5:00 PM, W : 8:00 PM – 9:00 PM.
5. BE IN TOUCH IF YOU WANT AN EXTENSION.
6. PLEASE KEEP VIDEOS ON, IF POSSIBLE!

FINAL EXAM REVIEW

(A) **2. (15(A) points)**

Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} e^{-x} + 2cx & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

where c is a constant.

(a) Find c.

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

$$\therefore \int_0^1 (e^{-x} + 2cx)\,dx = 1$$

$$-e^{-x} + cx^2 \Big]_0^1 = -e^{-1} + c \cdot 1^2 - (-e^{-0})$$

$$= 1 - 1/e + c$$

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1 - \frac{1}{e} + c = 1$$

$$\Rightarrow \boxed{c = 1/e}$$

(b) Find $\mathbb{E}[X]$.

$$= \int_{-\infty}^{\infty} x\, f_X(x)\,dx = \int_0^1 x\left[e^{-x} + \frac{2x}{e}\right]dx$$

$$= \boxed{\int_0^1 x e^{-x}\,dx} + \boxed{\int_0^1 \frac{2x^2}{e}\,dx}$$

$$\frac{2}{3e}x^3\Big]_0^1 = \frac{2}{3e}$$

$$\int_0^1 \underbrace{x}_{u}\,\underbrace{e^{-x}\,dx}_{dv} \overset{\text{INT. BY PARTS}}{=} \left[x\left(-e^{-x}\right)\right]_0^1 - \int_0^1 1\cdot\left(-e^{-x}\right)dx$$

$du = dx$

$v = -e^{-x} dx$

$$= -e^{-1} + \int_0^1 e^{-x}\,dx$$

$$= -e^{-1} + \left[-e^{-x}\right]_0^1 = 1 - 2e^{-1}$$

f_X ?

(c) Find the cumulative distribution function (c.d.f.) of X.

$$F_X(t) = P(X \le t)$$

$$= \int_{-\infty}^{t} f_X(x)\, dx$$

① CASE 1 : $t < 0 \Rightarrow F_X(t) = 0$

② CASE 2 : $t > 1 \Rightarrow F_X(t) = \int_{-\infty}^{t} = \int_{0}^{1} = \underline{1}$

(3) CASE 3: $0 \le t \le 1$

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\,dx$$

$$= \int_{0}^{t} \left[e^{-x} + \frac{2x}{e} \right] dx$$

$$= -e^{-x} + \frac{x^2}{e} \Bigg]_{x=0}^{x=t} = 1 - e^{-t} + \frac{t^2}{e}$$

$$F_X(t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-t} + t^2/e & 0 \le t \le 1 \\ 1 & t > 1 \end{cases}$$

RECALL :

$G \sim \mathrm{Exp}(\lambda)$

$P(G \geq t) = e^{-\lambda t}$

$[\, t \geq 0 \,].$

(A) **3.** **(15(A) points)** Scientists are trying to do an experiment. They know the outcome of the experiment X is random variable with distribution Ber(1/2). However, due to inaccuracies in their measuring equipment, there is a noise parameter $G \sim \mathrm{Exp}(1/3)$, and what actually gets measured by the equipment is $M = X + G$. Assuming that X and G are independent, what is the probability that $X = 1$ if the measurement M is at least 4?

$$P(x=0) = P(X=1) = 1/2$$

ANS : $P(X = 1 \mid M \geq 4)$

BAYES : $= \dfrac{P(x=1, M \geq 4)}{P(M \geq 4)} = \dfrac{P(X=1, M \geq 4)}{P(X=1, M \geq 4) + P(X=0, M \geq 4)}$

FACT : $X=0$ & $X=1$ PARTITION Ω $\rightarrow$ $P(\overset{M \geq 4}{A}) = P(A \cap \overset{X=1}{B}) + P(A \cap B^c)$

$$\frac{P(X=1, M \geqslant 4)}{P(X=1, M \geqslant 4) + P(X=0, M \geqslant 4)}$$

NEXT STEP :

$$P(X=1, M \geqslant 4) = P(X=1, X+G \geqslant 4)$$

$$= P(X=1, G \geqslant 3)$$

(IND.) $= P(X=1)\, P(G \geqslant 3)$

$$= \frac{1}{2}\left[e^{-\left(\frac{1}{3}\right)\cdot 3}\right]$$

$$= 1/2e$$

$$P(X=0,\ M \geq 4) = P(X=0,\ X+G \geq 4)$$

$$= P(X=0,\ G \geq 4)$$

$$(\text{IND.}) \quad = P(X=0)\, P(G \geq 4)$$

$$= \frac{1}{2}\left[\ e^{-\frac{1}{3}\cdot 4}\ \right]$$

$$= \frac{1}{2e^{4/3}}$$

$$\text{ANS} = \frac{P(X=1, M \geq 4)}{P(X=1, M \geq 4) + P(X=0, M \geq 4)} = \frac{\frac{1}{2}e^{4/3}}{\left(\frac{1}{2e}\right) + \left(\frac{1}{2}e^{4/3}\right)}$$

CAN LEAVE LIKE THIS.

$$= \frac{1}{1 + e^{-1/3}}$$

(A) **4. (10(A) points)** In the city of Gotham, anyone who likes drinking tea doesn't like drinking anything else. If everyone likes at least one of the beverages tea, coffee, or soda, 10% of people like tea, 50% people like soda, and 60% of people like coffee, then how many people like both coffee and soda? $\rightarrow 20\%$

$T \rightarrow$ TEA $\qquad C \rightarrow$ COFFEE $\qquad S \rightarrow$ SODA

$T \cup C \cup S \rightarrow$ EVERYONE.

$\therefore \quad P(T \cup C \cup S) = 1$

$P(T) = 10\% = \frac{10}{100} = \frac{1}{10}$

$P(S) = \frac{50}{100} = \frac{1}{2}$

$P(C) = \frac{60}{100} = \frac{3}{5}$

$P(T \cap C) = P(T \cap S) = P(T \cap S \cap C) = 0$

Q: $P(S \cap C)$

ANS: INCLUSION - EXCLUSION

$$P(T \cup S \cup C) = P(T) + P(S) + P(C) - P(T \cap S) - P(T \cap C) - P(S \cap C) + P(T \cap S \cap C)$$

$$\Rightarrow \quad 1 = \frac{1}{10} + \frac{1}{2} + \frac{3}{5} - 0 - 0 - P(S \cap C) + 0$$

$$\Rightarrow \quad P(S \cap C) = \frac{1}{10} + \frac{1}{2} + \frac{3}{5} - 1$$

$$= \frac{1 + 5 + 6 - 10}{10} = \frac{12 - 10}{10} = \frac{2}{10} = \frac{1}{5} = 20\%$$

(A) **5.** **(20(A) points)**

$n = 90000 \Rightarrow \sqrt{n} = 300$

You want to find out how popular pineapple is on pizzas. You randomly call $\boxed{90,000}$ people around the US and among them 42,000 said pineapple on a pizza is unacceptable.

(a) Give a 95% confidence interval for the true proportion who find pineapple on a pizza unacceptable.

$p \to$ TRUE PROPORTION

$\hat{p} \to$ OBS. PROPORTION $\left(\hat{p} = \frac{42000}{90000} = \frac{7}{15} \right)$

$$P(|p - \hat{p}| < \underbrace{\epsilon}_{\text{ERROR}}) \geq 2\Phi(2\epsilon\sqrt{\underbrace{n}_{\text{SAMPLED}}}) - 1 \geq 95\%$$

GOAL: $2\,\Phi(2\epsilon\sqrt{n}) - 1 \geq 95\%$

$$\Rightarrow \quad 2\,\Phi(\;-\;) - 1 \geq \frac{95}{100}$$

$$\Rightarrow \quad \Phi(2\epsilon\sqrt{n}) \geq \frac{\frac{95}{100} + 1}{2} = \frac{1.95}{2} = 0.975$$

$$\therefore \quad \Phi\left(2\epsilon \cdot (300)\right) \geq 0.975$$

$$\Rightarrow \quad \Phi(600\epsilon) \geq 0.975 \approx \Phi(1.96)$$

$$\therefore \quad 600\epsilon \geq 1.96 \Rightarrow \epsilon = \frac{1.96}{600}$$

$$= \frac{0.0196}{6}$$

$$\approx 0.00326$$

$$\begin{array}{r}
0.466\ldots \\
15\,\overline{)\,70} \\
-\,60 \\
\hline
100 \\
-\,90 \\
\hline
10
\end{array}$$

$$\hat{p} = \frac{7}{15} = 0.46667$$

$$|p - \hat{p}| < \epsilon \Rightarrow \hat{p} - \epsilon \leq p \leq \hat{p} + \epsilon$$

$$p \in [0.46341, 0.46993]$$

(b) A national vote was held about pineapple on a pizza. The result says 40% of the population finds pineapple on a pizza unacceptable. Alfredo's sells pizzas with pineapple on them at \$9 per pizza, and pizzas without pineapple on them at \$10. Let X_n be the money Alfredo's makes from selling n pizzas. Express X_n in terms of a distribution you know.

ALFREDO'S SELLS Y_n (NON)-PINEAPPLE PIZZAS

$$10 Y_n + 9(n - Y_n)$$

$$= 9n + Y_n$$

$$Y_n \sim \mathrm{Bin}\left(n, \frac{40}{100}\right)$$

$$\boxed{X_n = 9n + Y_n} \quad \left(Y_n \sim \mathrm{Bin}(n, 0.4)\right)$$

(c) Find

$$\lim_{n\to\infty} \mathbb{P}(X_n > 9.5n). = 0 \quad (\text{BY L.L.N.})$$

$$P(9n + Y_n > 9.5n)$$

$$\Rightarrow P(Y_n > 0.5n)$$

$$\Rightarrow P\Big(\frac{Y_n}{n} - \underbrace{0.4}_{\downarrow\, p} > \underbrace{0.1}_{\epsilon}\Big) \leq 1 - P\Big(\Big|\frac{Y_n}{n} - p\Big| < \epsilon\Big)$$

$$\lim_{n\to\infty} P(X_n > 9.5n) \leq 1 - \lim_{n\to\infty} P\Big(\Big|\frac{Y_n}{n} - p\Big| < \epsilon\Big) = 1 - 1 = 0$$

(L.L.N.)

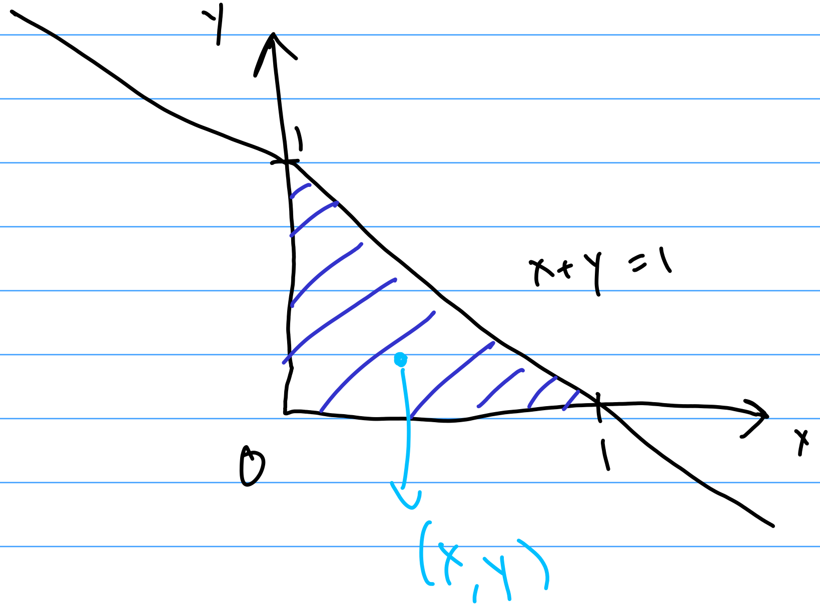
BE BACK AT
10 :20 AM

(B) **6. (30(B) points)**

Suppose (X, Y) are uniformly distributed on the triangular region

$$D = \{(x, y) : x + y \leq 1, x \geq 0, y \geq 0\}$$

(a) Find the marginal density functions f_X and f_Y.



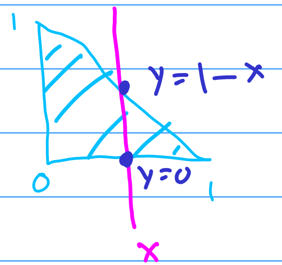
$$\text{AREA}(D) = \frac{1}{2} \cdot 1 \cdot 1 = \frac{1}{2}$$

$$f_{X,Y}(x,y) = \begin{cases} 2 & (x,y) \in D \\ 0 & \text{o.w.} \end{cases}$$

Q.

$$f_{X,Y}(x,y) = \begin{cases} 2 & (x,y) \in D \\ 0 & \text{o.w.} \end{cases}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$



CASE (I): IF $x < 0$ OR $x > 1$

$$f_X(x) = 0$$

CASE (II) IF $0 \le x \le 1$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^{1-x} 2\,dy = 2(1-x)$$

$$f_X(x) = \begin{cases} 2(1-x) & \text{IF } x \in [0,1] \\ 0 & \text{o.w.} \end{cases}$$

$$f_Y(y) = \begin{cases} 2(1-y) & \text{IF } y \in [0,1] \\ 0 & \text{o.w.} \end{cases}$$

SYMMETRY IN $X \leftrightarrow Y$

(b) Find $M_X(t)$. MOMENT GENERATING FUNCTION.

$$M_X(t) = \mathbb{E}\left[e^{tX} \right]$$

$$= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$$

$$= \int_{0}^{1} e^{tx}\, 2(1-x)\,dx$$

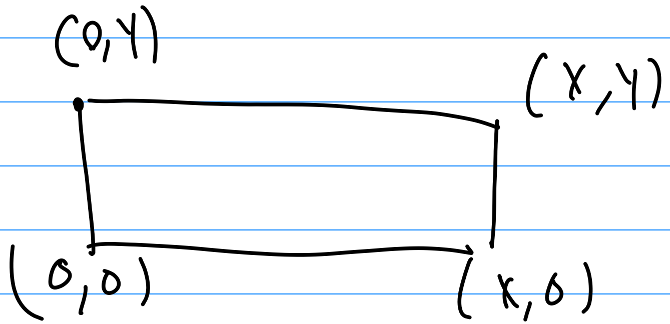
$$\int_0^1 e^{tx}\, 2(1-x)\, dx \qquad \begin{matrix} dv \\ u \end{matrix} \qquad \longrightarrow \quad t \neq 0$$

$$= \left[\frac{e^{tx}}{t}\, 2(1-x) \right]_{x=0}^{x=1} - \int_0^1 \frac{e^{tx}}{t} (-2)\, dx$$

$$= -\frac{2}{t} + 2 \left[\frac{e^{tx}}{t^2} \right]_0^1$$

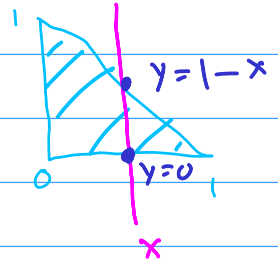
$$= \frac{-2}{t} + \frac{2e^t}{t^2} - \frac{2}{t^2} = \boxed{\frac{2}{t^2}\left[e^t - t - 1 \right]}$$

(c) Let A be the area of the ~~square~~ RECTANGLE bounded by the points $(0,0)$, $(X,0)$, $(0,Y)$, and (X,Y). Compute $\mathbb{E}[A]$.



$$f_{X,Y}(x,y) = \begin{cases} 2 & (x,y) \in D \\ 0 & \text{o.w.} \end{cases}$$

$$A = XY$$



$$\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \, f_{X,Y}(x,y) \, dx \, dy$$

$$= \int_{y=0}^{y=1} \int_{x=0}^{x=1-y} (xy)(2) \, dx \, dy$$

$$= \int_{y=0}^{y=1} \Big[x^2 y \Big]_{x=0}^{x=1-y} dy$$

$$= \int_{y=0}^{y=1} (1-y)^2 \, y \, dy$$

$$= \int_0^1 \left(y - 2y^2 + y^3 \right) dy$$

$$= \left[\frac{y^2}{2} - \frac{2y^3}{3} + \frac{y^4}{4} \right]_0^1 = \frac{1}{2} - \frac{2}{3} + \frac{1}{4} = \frac{6-8+3}{12}$$

$$\mathbb{E}[A] = \mathbb{E}[XY] = \frac{1}{12}$$

(d) Using your computations in (b) and (c) or otherwise, find the correlation coefficient $\mathrm{Corr}[X, Y]$.

$$\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}(X)} \cdot \sqrt{\mathrm{Var}\,Y}}$$

Q. $\mathrm{Cov}[X,Y] = \underbrace{\boxed{\mathbb{E}[XY]}}_{1/12} - \mathbb{E}[X] \cdot \mathbb{E}[Y]$

NEED TO FIND:

$\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$

DUE TO SYMMETRY, $\mathbb{E}[Y] = \mathbb{E}[X]$

$\mathrm{Var}(Y) = \mathrm{Var}(X)$

Now,

$$M_X(t) = \frac{2}{t^2}\left[e^t - 1 - t\right] \qquad e^t = \sum_{j=0}^{\infty} \frac{t^j}{j!}$$

$$= \frac{\cancel{2}}{\cancel{t^2}}\left[\cancel{1} + \cancel{t} + \frac{\cancelto{1}{t^2}}{\cancel{2}} + \frac{\cancelto{t}{t^3}}{\cancelto{3}{6}} + \frac{\cancelto{t^2}{t^4}}{\cancelto{12}{24}} + \cdots - \cancel{1} - \cancel{t}\right]$$

$$M_X(t) = 1 + \boxed{\frac{t}{3}} + \boxed{\frac{t^2}{12}} + \text{HIGHER ORDER TERMS.}$$

$$M_X(t) = \sum_{n=0}^{\infty} M_X^{(n)}(0) \cdot \frac{t^n}{n!}$$

$$= M_X(0) + M_X'(0)\,t + M_X''(0)\,\frac{t^2}{2} + \text{H.O.T.}$$

$$= 1 + \boxed{E[X] \cdot t} + \boxed{E[X^2] \cdot \frac{t^2}{2}} + \text{H.O.T.}$$

$$\left(E[X^k] = M^{(k)}(0) \right)$$

$$\therefore \mathbb{E}[X] = \frac{1}{3}, \quad \frac{\mathbb{E}[X^2]}{2} = \frac{1}{12} \Rightarrow \mathbb{E}[X^2] = \frac{1}{6}$$

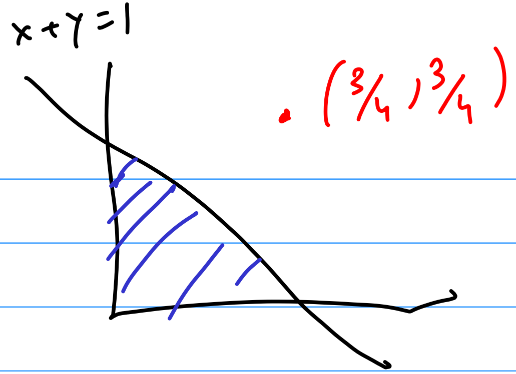
$$\therefore \mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$= \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{6} - \frac{1}{9}$$

$$= \frac{3-2}{18} = \frac{1}{18}$$

$\boxed{E}$

(e) Are X and Y independent? Remember to justify your answer.



NO! :

$$f_{X,Y}(x,y) \overset{\text{IND.}}{=} f_X(x)\, f_Y(y)$$

$$f_{X,Y}\left(\frac{3}{4},\frac{3}{4}\right) = 0 \neq \frac{1}{4} = f_X\left(\frac{3}{4}\right)\cdot f_Y\left(\frac{3}{4}\right)$$

$$2\left(1-\frac{3}{4}\right)\cdot 2\left(1-\frac{3}{4}\right)$$

$1/4 \cdot e^{0t}$

(B) **7. (30(B) points)**

Suppose X and Y are independent random variables with the following moment generating functions (m.g.f.),

$$M_X(t) = \boxed{\frac{1}{4}e^{-t}} + \boxed{\frac{1}{4}} + \boxed{\frac{1}{2}e^{t}},$$

$$M_Y(t) = \frac{2}{3}e^{-t} + \frac{1}{3}e^{2t}.$$

$X \in \{-1, 0, 1\}$

$Y \in \{-1, 2\}$

$M_X(t) = E[e^{tX}] = \sum_k e^{t \cdot k} P(X = k)$

(a) Find the probability mass functions (p.m.f.) p_X and p_Y.

$P(X = -1) = \frac{1}{4}$, $P(X = 0) = \frac{1}{4}$, $P(X = 1) = \frac{1}{2}$

$P(Y = -1) = \frac{2}{3}$, $P(Y = 2) = \frac{1}{3}$

(b) Find the joint p.m.f., $p_{X,Y}(x, y)$ and express it as a table.

$$p_{X,Y}(x,y) = p_X(x)\, p_Y(y)$$

$P(X=-1, Y=2) = \frac{1}{12}$

		Y	-1	2	
X	$p_X(x)$	$p_Y(y)$	$2/3$	$1/3$	
-1	$1/4$		$1/6$ [-2]	$\boxed{1/12}$	[1]
0	$1/4$		$1/6$ [-1]	$1/12$	[2]
1	$1/2$		$1/3$ [0]	$1/6$	[3]

[Z]

(c) Let $Z = X + Y$. What is $M_Z(t)$?

$Z = X + Y$ [IND.]

$$M_Z(t) = M_X(t)\, M_Y(t)$$

$$= \left(\frac{e^{-t}}{4} + \frac{1}{4} + \frac{e^{t}}{2}\right)\left(\frac{2}{3}e^{-t} + \frac{1}{3}e^{2t}\right)$$

$$= \frac{e^{-2t}}{6} + \frac{e^{-t}}{6} + \frac{e^{0}}{3} + \frac{e^{t}}{12} + \frac{e^{2t}}{12} + \frac{e^{3t}}{6}$$

$X \in \{-1, 0, 1\}$

<u>ALT.</u> CONVOLUTIONS

$Y \in \{-1, 2\}$

$X + Y \in \{-2, \ldots, 3\}$

$$p_Z(n) = \sum_k p_X(k)\, p_Y(n-k)$$

<u>FOR EG.</u>

$$p_Z(1) = \sum_k p_X(k)\, p_Y(1-k) = p_X(0)\, p_Y(1)$$

$k \in \{-1, 0, 1\}$

$1-k \in \{-1, 2\}$ $\rightarrow$ $k = 0$

(d) Compute the p.m.f. p_Z.

$$\boxed{\frac{e^{-2t}}{6}} + \boxed{\frac{e^{-t}}{6}} + \frac{e^{0}}{3} + \frac{e^{t}}{12} + \frac{e^{2t}}{12} + \boxed{\frac{e^{3t}}{6}}$$

$p_Z(-2) = \frac{1}{6}$

$p_Z(-1) = \frac{1}{6}$

$p_Z(3) = \frac{1}{6}$

(e) Compute $\mathrm{Var}[Z^2]$.

$$\mathrm{Var}[Z^2] = \mathbb{E}\left[(Z^2)^2\right] - \left(\mathbb{E}[Z^2]\right)^2$$

$$= \mathbb{E}\left[Z^4\right] - \left(\mathbb{E}[Z^2]\right)^2$$

$$= M_Z^{(4)}(0) - \left[M_Z^{(2)}(0)\right]^2$$

$$M_Z(t) = \frac{e^{-2t}}{6} + \frac{e^{-t}}{6} + \frac{e^{0}}{3} + \frac{e^{t}}{12} + \frac{e^{2t}}{12} + \frac{e^{3t}}{6}$$

$$M_Z''(t) = (-2)^2\frac{e^{-2t}}{6} + (-1)^2\frac{e^{-t}}{6} + 0^2\frac{e^{0}}{3} + 1^2\frac{e^{t}}{12} + 2^2\frac{e^{2t}}{12} + 3^2\frac{e^{3t}}{6}$$

$t=0,$

$$\frac{4}{6} + \frac{1}{6} + 0 + \frac{1}{12} + \frac{4}{12} + \frac{9}{6} = \square$$

$$M_Z^{(4)}(t) = (-2)^4\cdot\frac{e^{-2t}}{6} + \ldots + 3^4\frac{e^{3t}}{6}$$

$t=0$

$\square$.

BREAK TILL

10 : 20 AM

(B) **8. (20(B) points)**

Recall that NegBin(k,p) for $k \in \mathbb{N}$ and $0 < p < 1$ is the negative binomial distribution and it is defined as the number of independent Ber(p) trials before one sees k successes.

The p.m.f. of $X \sim \text{NegBin}(k,p)$ is given by

$$p_X(n) = P(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}.$$

$\text{NegBin}(1,p) = \text{Geom}(p)$

In the following parts, it may be useful to recall that if $Y \sim \text{Geom}(p)$, then

$$E[Y] = \frac{1}{p} \qquad \text{Var}[Y] = \frac{1-p}{p^2}.$$

(a) Find a formula for $\mathbb{E}[X]$ and $\text{Var}[X]$. (**Hint:** do not compute this directly! Instead, use the relationship between NegBin and Geom.)

RECALL: $X \sim \text{NegBin}(k,p) \Rightarrow X = X_1 + X_2 + \cdots + X_k$

$X_j \sim \text{Geom}(p)$ (i.i.d)

$$X = X_1 + X_2 + \cdots + X_k$$

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_k]$$

LoE.

$$= k\,\mathbb{E}[X_1] \qquad (\because X_j \text{ i.i.d.})$$

$$= k \cdot \left(\frac{1}{p}\right) = \frac{k}{p}$$

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_k) \qquad (\text{IND.})$$

$$= k\,\mathrm{Var}(X_1) = k(1-p)/p^2$$

(b) In class, we showed using the central limit theorem that when n is a large positive integer, then Poisson(n) behaves like a normal variable. Mimic this argument to show that when k is a large positive integer, then NegBin(k, p) behaves like a normal variable. What are its parameters?

$$X \sim \text{Pois}(n) \Rightarrow X = X_1 + X_2 + \cdots + X_n \quad (X_j \sim \text{Pois}(1) \text{ i.i.d.})$$

APPLY C.L.T.

$$X \sim \text{NegBin}(k,p) \Rightarrow X = X_1 + \cdots + X_k \quad (X_j \sim \text{i.i.d. Geom}(p))$$

$\therefore$ BY CLT

$$\tilde{X} = \frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \sim N(0,1) \quad \text{AS } k \longrightarrow \infty$$

$$\Rightarrow X \approx \sigma Z + \mu,$$

$\mu = E[X]$
$\sigma = \text{Var}[X]$
$Z \sim N(0,1)$

$$\mathbb{E}[X] = \frac{k}{p} \quad , \quad \mathrm{Var}[X] = k\left(\frac{1-p}{p^2}\right)$$

$$\therefore \quad X \approx N\left(\frac{k}{p}, \; k\left(\frac{1-p}{p^2}\right)\right)$$

(B) **9. (20(B) points)**

On average Watson and Holmes have to wait 20 days between visits by Lestrade. Lestrade just left after a visit, and suppose X is the number of days from now he will visit again.

(a) If you know nothing else about the distribution of X, what upper bound can you provide for the probability that $X > 50$? Clearly state any inequality you use, and explain why the hypotheses apply.

$X \geq 0$

$$\mathbb{E}[X] = 20.$$

$$P(X > 50) \leq \frac{\mathbb{E}[X]}{50} = \frac{20}{50} = \frac{2}{5}$$

(MARKOV'S INEQUALITY)

$$P(X > c) \leq \frac{\mathbb{E}[X]}{c}$$

(b) Suppose X is a continuous random variable. You can assume that X is memoryless – if Lestrade hasn't visited after t days, then probability he will take at least s more days is the same as the probability that he would have taken s days in the first place. That is,

$$P(X > s + t | X > t) = P(X > s).$$

Can you identify the distribution of X?

$\mathbb{E}[X] = 20$

$\Rightarrow \lambda = \frac{1}{20}$

$X \sim \mathrm{Exp}(\lambda)$

$\mathbb{E}[\mathrm{Exp}(\lambda)] = \frac{1}{\lambda}$

$\therefore X \sim \mathrm{Exp}\left(\frac{1}{20}\right)$

(c) Compute $P(X > 50)$ exactly.

$$P(X > 50) = \int_{50}^{\infty} \overbrace{\lambda e^{-\lambda t}}^{f_X(t)} \, dt$$

$$= e^{-\lambda (50)} = e^{-50/20}$$

(d) Compute an upper bound on $P(X > 50)$ using Chebyshev's inequality.

$$P(|X - \mu| > c) \leq \frac{\sigma^2}{c^2}$$

$$P(X > 50) = P(X - 20 > 30)$$

$$\leq P(|X - \underbrace{20}_{\mu = \mathbb{E}[X]}| > \underbrace{30}_{c}) \leq \frac{\sigma^2}{30^2}$$

$$\sigma^2 = \mathrm{Var}[X] = \frac{1}{\lambda^2} = \frac{1}{(1/20)^2} = 20^2$$

$$P(X > 50) \leq \frac{20^2}{30^2} = \frac{4}{9}$$

(B) **10. (20(B) points)**

Two-Face rolls a standard die repeatedly. Let A_j be the indicator random variable for the event that the jth roll is 1, and B_j be the indicator random variable for the event that the jth roll is 2.

(a) Find $\mathbb{E}[A_j B_k]$. (**Hint:** consider the cases $j = k$ and $j \neq k$ separately)

$j = k$. $\quad A_j B_j = \begin{cases} 1 & \text{~~IF A_j \& B_j BOTH OCCUR TOGETHER~~} \\ 0 & \text{O.W.} \end{cases}$

$A_j B_j = 0 \Rightarrow \mathbb{E}[A_j B_j] = 0$.

$j \neq k$.

A_j & B_k ARE IND. R.Vs IF $j \neq k$

$$\mathbb{E}[A_j B_k] = \mathbb{E}[A_j]\,\mathbb{E}[B_k]$$

$$= P(A_j)\;P(B_k)$$

$$= \left(\frac{1}{6}\right)\cdot\left(\frac{1}{6}\right) = \frac{1}{36}$$

$$\mathbb{E}[A_j B_k] = \begin{cases} 0 & j = k \\ 1/36 & j \neq k \end{cases}$$

(b) Let X be the number of 1s and Y be the number of 2s that show up in n rolls of the die. Find $\mathbb{E}[XY]$.

$$X = \sum_{j=1}^{n} A_j$$

$$Y = \sum_{k=1}^{n} B_k$$

$$XY = \sum_{j,k=1}^{n} A_j B_k \Rightarrow \mathbb{E}[XY] = \sum_{j,k=1}^{n} \mathbb{E}[A_j B_k]$$

$$\mathbb{E}[XY] = \frac{1}{36} \times \left[\# \quad \{ (j,k) : 1 \le j,k \le n \quad j \neq k \} \right]$$

$$= \frac{1}{36} \, n(n-1) = \frac{n(n-1)}{36}$$

(c) Compute Cov$[X, Y]$. $= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{j=1}^{n} A_j \right] = \sum_{j=1}^{n} \mathbb{E}[A_j] = \sum_{j=1}^{n} P(A_j) = \frac{n}{6}$$

III ly $\mathbb{E}[Y] = n/6$

$$\operatorname{Cov}[X,Y] = \underbrace{\frac{n(n-1)}{36}}_{\mathbb{E}[XY]} - \underbrace{\left(\frac{n}{6}\right)}_{\mathbb{E}[X]}\underbrace{\left(\frac{n}{6}\right)}_{\mathbb{E}[Y]}$$

$$= \frac{-n}{36}$$

$\text{Var}(X) \le \mathbb{E}[X^2] \le \mathbb{E}[25]$
$= 25$

(B) **11. (20(B) points)** No Hobbit in Middle-Earth is taller than 5 feet. You decide to go around randomly asking Hobbits their height to get a good estimate for their average height. Using the law of large numbers, estimate how many Hobbits you need to ask before there is a $\boxed{\text{99\% chance}}$ that the average height you sampled differs from the actual average by at most 1 inch? [**Note:** 12 inches is 1 feet.]

$\mathbb{E}[X]$.

$X_1, X_2, X_3, \ldots, X_n \sim \text{DISTR}(X)$ i.i.d.

$\hat{\mu}$ = OBSERVED MEAN $= \frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$.

$\mu \to$ ACTUAL AVG.

$|\hat{\mu} - \mu| < \frac{1}{12}$

MAXI. PROB.

$$P\left(|\mu - \hat{\mu}| < 1/12\right) > \frac{99}{100}$$

$$\therefore \quad P\left(|\mu - \hat{\mu}| \geq 1/12\right) \leq \frac{1}{100} \quad \left(= 1 - \frac{99}{100}\right)$$

$$P\left(\left|\frac{S_n}{n} - \boxed{\mu}\right| \geq \frac{1}{12}\right) \leq \frac{1}{100}$$

CHEBYSHEV

$$\downarrow$$

$$E[S_n/n]$$

$$Var[S_n/n] = Var\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n^2}\left[Var[X_1] + \cdots + Var[X_n]\right]$$

$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{n\,\mathrm{Var}[X]}{n^2} = \frac{\mathrm{Var}[X]}{n}$$

BY CHEBYSHEV

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \frac{1}{12}\right) \leq \frac{\mathrm{Var}[S_n/n]}{(1/12)^2}$$

$$= \frac{144\,\mathrm{Var}[X]}{n} \leq \frac{144 \cdot 25}{n}$$

$$= \frac{3600}{n} < \frac{1}{100}$$

$$\Rightarrow \quad n > (3600)(100) = 360{,}000$$