1. Response time for a mobile Web site is the speed of page downloads. Let X denote the number of bars of service and Y the response time (to the nearest second) for a particular user and site. The joint pmf of X and Y is shown below.

Y \ X	1	2	3
4	0.15	0.10	0.05
3	0.02	0.10	0.05
2	0.02	0.03	0.20
1	0.01	0.02	0.25
	0.2	0.25	0.55

(a) Find the marginal pmf of X.

X	1	2	3
P_X	0.2	0.25	0.55

(b) Given that a user has 2 bars of service, find the probability the response time will be 1 or 2 seconds.

$$P[Y=1 \text{ or } 2 \mid X=2] = \frac{P[Y=1 \text{ or } 2 \text{ and } X=2]}{P[X=2]} = \frac{0.03+0.02}{0.25}$$

$$= 0.2$$

(c) Given that a user has 2 bars of service, find the minimum mean square error estimate of Y, $\hat{y}_M(2)$.

$$P_{Y|X}(y|2) = \frac{P_{X,Y}(2,y)}{P_X(2)}$$

$Y \mid X=2$	1	2	3	4
$P_{Y\|X}(y\|2)$	0.08	0.12	0.4	0.4

$$E[Y \mid X=2] = 1\cdot 0.08 + 2\cdot 0.12 + 3\cdot 0.4 + 4\cdot 0.4$$

$$= 3.12$$

$$\hat{y}_M(2) = 3.12$$

(d) Given that a user has 2 bars of service, find the linear mean square error estimate of Y, $\hat{y}_L(2)$.

$E[X] = 1 \cdot 0.2 + 2 \cdot 0.25 + 3 \cdot 0.55 = 2.35$

$E[X^2] = 1^2 \cdot 0.2 + 2^2 \cdot 0.25 + 3^2 \cdot 0.55 = 6.15$

$Var[X] = E[X^2] - E[X]^2 = 0.6275$

$E[Y] = 1 \cdot 0.28 + 2 \cdot 0.25 + 3 \cdot 0.17 + 4 \cdot 0.3 = 2.49$

$E[Y^2] = 1^2 \cdot 0.28 + 2^2 \cdot 0.25 + 3^2 \cdot 0.17 + 4^2 \cdot 0.3 = 7.61$

$Var[Y] = E[Y^2] - E[Y]^2 = 1.4099$

$$\begin{aligned} E[XY] = &\ 1 \cdot 1 \cdot 0.01 + 1 \cdot 2 \cdot 0.02 + 1 \cdot 3 \cdot 0.02 + 1 \cdot 4 \cdot 0.15 \\ &+ 2 \cdot 1 \cdot 0.02 + 2 \cdot 2 \cdot 0.03 + 2 \cdot 3 \cdot 0.10 + 2 \cdot 4 \cdot 0.10 \\ &+ 3 \cdot 1 \cdot 0.25 + 3 \cdot 2 \cdot 0.20 + 3 \cdot 3 \cdot 0.05 + 3 \cdot 4 \cdot 0.05 = 5.27 \end{aligned}$$

$Cov[X,Y] = E[XY] - E[X]E[Y] = -0.5815$

$a = \frac{Cov[X,Y]}{Var[X]} = -0.9267 \qquad b = E[Y] - a E[X] = 4.6677$

$\hat{y}_L(x) = -0.9267x + 4.6677$

$\hat{y}_L(2) = 2.81$

(e) For each of the following statements, circle it if it is TRUE.

i. $\hat{y}_M(2)$ will always be more accurate than $\hat{y}_L(2)$

ii. $\hat{y}_L(2)$ will always be more accurate than $\hat{y}_M(2)$

iii. $\hat{y}_M(2)$ will sometimes be more accurate than $\hat{y}_L(2)$

iv. $\hat{y}_L(2)$ will sometimes be more accurate than $\hat{y}_M(2)$

v. $\hat{y}_M(2)$ will be more accurate than $\hat{y}_L(2)$ on average

vi. $\hat{y}_L(2)$ will be more accurate than $\hat{y}_M(2)$ on average

vii. $\hat{y}_M(2)$ and $\hat{y}_L(2)$ will always be equally accurate

2. The owner of a small gas station has his 1,500 gallon tank of 93-octane gas filled up once at the beginning of each week. The random variable X is the amount of 93-octane the station sells in one week (in thousands of gallons). The PDF of X is shown below.

$$f_X = \begin{cases} x, & 0 \leq x \leq 1 \\ 1, & 1 < x \leq 1.5 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find the unconditional ("blind") estimate of the amount of gasoline the owner will sell in one week.

$$\hat{X}_b = E[X] = \int_0^{1.5} x \cdot f_X(x)\, dx = \int_0^1 x^2\, dx + \int_1^{1.5} x\, dx$$

$$= \frac{1}{3} + \frac{5}{8} = \frac{23}{24} = 0.9583$$

(b) Find Var[X].

$$E[X^2] = \int_0^1 x^3\, dx + \int_1^{1.5} x^2\, dx = \frac{1}{4} + \frac{19}{24} = \frac{25}{24}$$

$$Var[X] = E[X^2] - E[X]^2 \approx 0.1233$$

(c) Assuming the amount of gasoline sold each week is independent of other weeks, estimate the probability that the total amount of gasoline sold in one year (52 weeks) is greater than 52,000 gallons.

$$W = X_1 + X_2 + \cdots + X_{52}$$

$$E[W] = 52 \cdot E[X] = 49.8\bar{3}$$

$$Var[W] = 52 \cdot Var[X] = 6.41$$

$$\sigma_W = 2.53$$

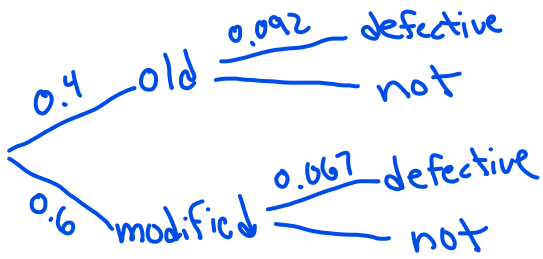
Central Limit Theorem:
W is gaussian$(49.8\bar{3}, 2.53)$, approximately.

$$P[W > 52] \approx 1 - \Phi\left(\frac{52 - 49.8\bar{3}}{2.53}\right) = 1 - \Phi(0.8558)$$

$$= 0.1949$$

3. **[21 pts]** In the manufacturing of crankshaft bearings, typically the proportion of defective bearings is .092. When a modification to the manufacturing process is made, the proportion of defective bearings is lowered to .067. Suppose 40% of bearings are manufactured using the old process and the remaining 60% are manufactured using the modified process.

(a) Consider a batch of thousands of bearings (where 60% of bearings are manufactured using the modified process). If a bearing is randomly selected from the batch and it is defective, find the probability it was manufactured using the modified process.



P[old and defective] = $0.4 \times 0.092 = 0.0368$

P[mod and defective] = $0.6 \times 0.067 = 0.0402$

$$P[\text{modified} \mid \text{defective}] = \frac{P[\text{mod and def}]}{P[\text{def}]} = \frac{0.0402}{0.0368 + 0.0402} = \boxed{0.5221}$$

(b) Suppose we have 10 bearings, 4 of which were manufactured using the modified process. If we randomly select 2 of these 10, find the probability that exactly 1 will have been manufactured with the modified process.

$$P[1 \text{ modified}] = \frac{\binom{4}{1}\binom{6}{1}}{\binom{10}{2}} = \frac{4 \cdot 6}{45} = \boxed{0.5333}$$

(d) Assume 10 bearings were all produced using the same process. Design a MAP hypothesis test that determines which manufacturing process was used based on the number of defective bearings. Be sure to clearly state your hypotheses in context of the problem.

H_0: old process used $\quad H_1$: modified process used

X = number of defective bearings out of 10

$$P[X=x|H_0]\cdot P[H_0] \geq P[X=x|H_1]\cdot P[H_1]$$

$$\cancel{\binom{10}{x}} 0.092^x \, 0.908^{10-x} \cdot 0.40 \geq \cancel{\binom{10}{x}} 0.067^x \, 0.933^{10-x} \, 0.60$$

$$\left(\frac{0.092}{0.908}\right)^x 0.908^{10} \, 0.40 \geq \left(\frac{0.067}{0.933}\right)^x 0.933^{10} \, 0.60$$

$$\left(\frac{0.092}{0.908}\cdot\frac{0.933}{0.067}\right)^x \geq \frac{0.933^{10}\cdot 0.60}{0.908^{10}\cdot 0.40}$$

$$x \ln\left(\frac{0.092}{0.908}\cdot\frac{0.933}{0.067}\right) \geq \ln\left(\frac{0.933^{10}\, 0.60}{0.908^{10}\, 0.40}\right)$$

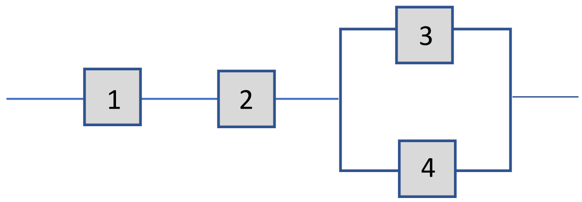
$$x \ln(1.4109) \geq \ln(1.9681)$$

$$x \geq \frac{\ln(1.9681)}{\ln(1.4109)} = 1.9667$$

If $x = 0, 1$, decide on H_0: old process used.

If $x = 2, 3, \ldots, 10$, decide on H_1: modified process used.

4. In the following system, each component's lifetime, X, is a Weibull random variable. (X is measured in months.) Specifically, $f_X(x) = 2xe^{-x^2}, x \geq 0$. The lifetime of each component is independent of the others. Components are replaced when the system fails to operate.



(a) Find the probability that any one of these components will last longer than 1 month.

$P[X>1] = \int_1^\infty 2xe^{-x^2}\,dx = \int_1^\infty e^{-u}\,du = -e^{-u}\Big|_1^\infty = 0-(-e^{-1}) = e^{-1}$

$= 0.368$

$u = x^2$
$du = 2x\,dx$

(b) Find the probability that this system will run for at least 1 month without having to replace any of the components.

3,4 parallel

$P[\text{3-4 combo works}] = 1-(1-0.368)^2 = 0.600$

$P[\text{entire system works}] = 0.368 \cdot 0.368 \cdot 0.600 = 0.0813$

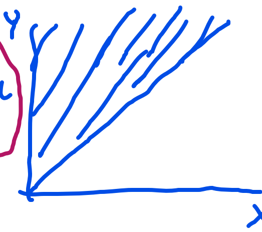
5. At a certain restaurant, a group's wait time to be seated after arrival is an exponential random variable, X, with $E[X] = 20$ minutes. Once seated, the time for the waiter to reach the table and take their order is an exponential random variable, Y, with $E[Y] = 5$ minutes. **<u>Assume X and Y are independent.</u>**

(a) Find the probability that a group will spend more time waiting for the waiter to take their order than to be seated.

$$f_{X,Y}(x,y) = f_X(x)\,f_Y(y) = \begin{cases} \frac{1}{20}e^{-x/20}\cdot\frac{1}{5}e^{-y/5} & x,y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

setup   setup

$$P[X<Y] = \boxed{\int_0^\infty \int_0^y \frac{1}{20}e^{-x/20}\,\frac{1}{5}e^{-y/5}\,dx\,dy} \text{ or } \boxed{\int_0^\infty \int_x^\infty \frac{1}{20}e^{-x/20}\cdot\frac{1}{5}e^{-y/5}\,dy\,dx}$$



$$= \int_0^\infty \frac{1}{5}e^{-y/5}\left(-e^{-x/20}\right)\Big|_0^y\,dy$$

$$= \int_0^\infty \frac{1}{5}e^{-y/5}\left(1-e^{-y/20}\right)dy$$

$$= \int_0^\infty \frac{1}{5}\left(e^{-y/5}-e^{-y/4}\right)dy = \frac{1}{5}\left(-5e^{-y/5}+4e^{-y/4}\right)\Big|_0^\infty = \frac{1}{5}(5-4) = \frac{1}{5}$$

(b) Find the expected value and the variance of the total wait time, $W = X + Y$, the time between arriving and ordering.

$$E[W] = E[X] + E[Y] = 20 + 5 = 25 \text{ minutes}$$

$$Var[W] = Var[X] + Var[Y] = 20^2 + 5^2 = 425 \text{ min}^2$$

(independence!)

(c) The restaurant claims that it is very unlikely that a group's total wait time will be longer than 45 minutes. Can you use Chebyshev's Inequality to support or refute this claim? Explain.

$$P[W>45] = P[W-25>20] \leq \frac{425}{20^2} = 1.0625$$

Chebyshev does not help.

(d) Set up (but do not compute) an integral or integrals that will give either the cdf or the pdf of the total wait time, W. It does not matter which you choose, but clearly indicate whether you have found the cdf or the pdf.

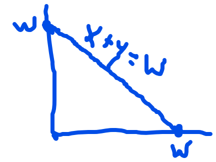
$$f_W(w) = f_X * f_Y(w) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(w-x)\, dx$$

PDF

$$= \int_0^w \frac{1}{20} e^{-x/20} \cdot \frac{1}{5} e^{-\frac{w-x}{5}}\, dx$$

$$= \frac{1}{100} e^{-w/5} \int_0^w e^{3x/20}\, dx \qquad (w \geq 0)$$

CDF $F_W(w) = P[X+Y \leq w] = \int_0^w \int_0^{w-y} \frac{1}{20} e^{-x/20} \frac{1}{5} e^{-y/5}\, dy\, dx$



(e) The cdf of W is $F_W(w) = \frac{1}{3}\left(e^{-\frac{w}{5}} - 4e^{-\frac{w}{20}} + 3\right), w \geq 0$

Find the probability that the total wait time will be more than 30 minutes.

$$P[W > 30] = 1 - F_W(30) = 1 - \frac{1}{3}\left(e^{-30/5} - 4e^{-30/20} + 3\right)$$

$$= 0.2967$$

(f) Set up (but do not compute) an integral that will give the expected total wait time for groups whose total wait time is at least 30 minutes.

$$f_W(w) = \frac{d}{dw} F_W(w) = \frac{1}{3}\left(-\frac{1}{5} e^{-w/5} + \frac{1}{5} e^{-w/20}\right) = \frac{1}{15}\left(e^{-w/20} - e^{-w/5}\right) \quad (w > 0)$$

$$f_{W|W>30}(w) = \frac{f_W(w)}{P[W>30]} = \frac{e^{-w/20} - e^{-w/5}}{15\,(0.2967)}$$

$$E[W \mid W > 30] = \int_0^{\infty} w\, \frac{e^{-w/20} - e^{-w/5}}{15\,(0.2967)}\, dw$$

6. A grocery store employee works between 0-50 hours each week. The pdf of X, the number of hours the employee works per week, is shown below.

$$f_X(x) = \frac{x+4}{1450}, \qquad 0 \le x \le 50$$

The employee currently makes \$12/hour if he works between 0-40 hours per week. If he works over 40 hours in a week, he gets paid overtime ("time and a half"); that is, he gets paid \$18/hour for those hours beyond 40. So the amount of his paycheck each week, W, in dollars, is

$$W = \begin{cases} 12X, & 0 \le X \le 40 \\ 480 + 18(X-40) & X > 40 \end{cases}$$

Find the cdf of W.

$0 \le W \le 660$

If $w \le 480$, then $F_W(w) = P[W \le w] = P[X \le \frac{w}{12}] = \int_0^{w/12} \frac{x+4}{1450}\,dx$

$= \frac{1}{1450}\left(x^2/2 + 4x\right)\Big|_0^{w/12} = \frac{1}{1450}\left(\frac{w^2}{288} + \frac{w}{3}\right)$

If $480 < w \le 660$, then $F_W(w) = P[W \le w] = P[W \le 480] + P[480 < W \le w]$

$= P[X \le 40] + P[480 < 480 + 18(X-40) \le w]$

$= P[X \le 40] + P[40 < X < \frac{w-480}{18} + 40]$ $\quad \frac{w}{18} + 13.\bar{3}$

$= \int_0^{40} \frac{x+4}{1450}\,dx + \int_{40}^{\frac{w-480}{18}+40} \frac{x+4}{1450}\,dx$

$= \frac{960}{1450} + \frac{1}{1450}\left(\frac{\left(\frac{w-480}{18}+40\right)^2}{2} + 4\left(\frac{w-480}{18}+40\right)\right)$

$$F_W(w) = \begin{cases} 0 & w < 0 \\ \frac{1}{1450}\left(\frac{w^2}{288} + \frac{w}{3}\right) & 0 \le w \le 480 \\ \frac{1}{1450}\left[\frac{\left(\frac{w-480}{18}+40\right)^2}{2} + 4\left(\frac{w-480}{18}+40\right)\right] & 480 \le w \le 660 \\ 1 & w > 660 \end{cases}$$

7. A town is trying to determine the rate of COVID-19 infections in their population. They have been told by state officials the infection rate could be as high as 5%, but they suspect it is much lower. They decide to randomly select and test individuals each day until they find one person who tests positive. They will do this for 30 days.

(a) If the infection rate is actually 5%, find the probability that on any given day, they will have to test no more than 5 people.

X = number of tests until first positive

X is geometric (0.05)

$$P[X \le 5] = 1 - (1-0.05)^5 = 0.2262$$

(b) Find the expected value and the variance of the average number of people that must be tested each day over the 30 day period.

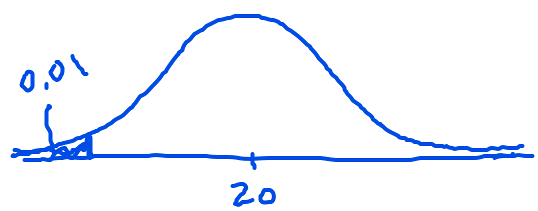
$$E[X] = \frac{1}{p} = \frac{1}{0.05} = 20 \qquad E[M_{30}] = 20$$

$$Var[X] = \frac{1-p}{p^2} = \frac{0.95}{0.05^2} = 380 \qquad Var[M_{30}] = \frac{380}{30} = 12.67$$

(c) Using the average number of people that must be tested each day over the 30 days, design a significance test with $\alpha = .01$ to determine whether to reject the null hypothesis that the infection rate is 5%. Design the test so they will reject the null hypothesis if the average number of people tested is too low. You should use the Central Limit Theorem.

H_0: $p = 0.05$

CLT: M_{30} is gaussian$(20, \sqrt{12.67})$



$$P[M_{30} < x] = 0.01$$

$$\Phi\left(\frac{x-20}{\sqrt{12.67}}\right) = 0.01$$

$$1 - \Phi\left(\frac{20-x}{\sqrt{12.67}}\right) = 0.01$$

$$\Phi\left(\frac{20-x}{\sqrt{12.67}}\right) = 0.99$$

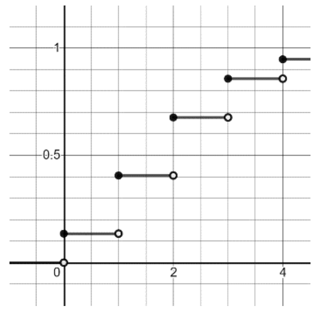
$$\frac{20-x}{\sqrt{12.67}} = 2.33$$

$$x = 11.814$$

$$\boxed{\text{rejection}: M_{30} < 11.814}$$

8. When a particular nuclear plant is operating normally, it sometimes releases a detectable amount of radioactive gases. The CDF of X, the number of radioactive gas emissions over a one month period, is given below. X is a Poisson random variable. Note the function/graph continues past the information shown.

$$F_X(x) = \begin{cases} 0 & x < 0 \\ e^{-2} & 0 \le x < 1 \\ 3e^{-2} & 1 \le x < 2 \\ 5e^{-2} & 2 \le x < 3 \\ \frac{19}{3}e^{-2} & 3 \le x < 4 \\ \vdots & \end{cases}$$



Poisson

$\alpha = 2$

(a) Find the expected value of X.

$E[X] = \alpha = 2$

(b) Find $P[X = 2]$.

$P[X=2] = 5e^{-2} - 3e^{-2} = 2e^{-2}$

(c) Find $P[X < 3]$.

$P[X<3] = P[X \le 2] = F_X(2) = 5e^{-2}$