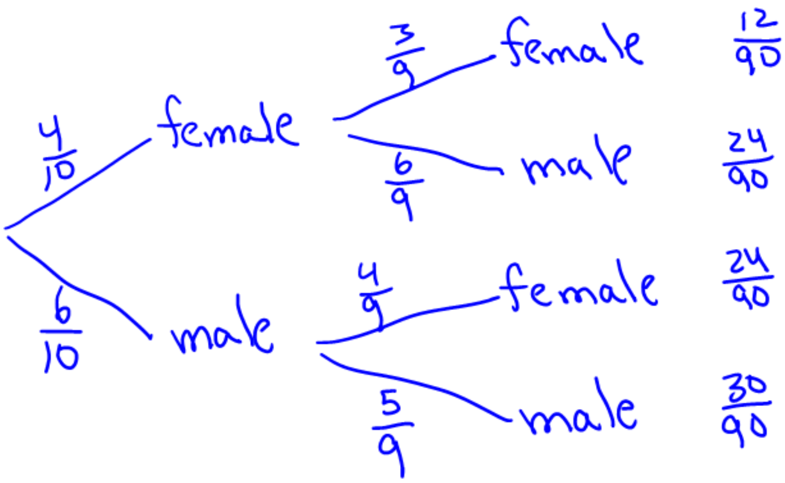
1. **[18 pts]** Two additional jurors are needed to complete a jury for a criminal trial. There are ten prospective jurors, four female and six male. Two jurors are randomly selected from the ten available.

(a) How many different pairs of 2 jurors could be selected from the 10?

$$\binom{10}{2} = 45$$

(b) Create a probability tree that shows the sequential choices (in terms of gender) of the first and second jurors. Your tree should show the outcomes and probabilities.



(c) Let X be the number of male jurors selected. Construct the CDF of X. (You do not have to graph the CDF.)

PMF

x	0	1	2
$P_X(x)$	$\frac{12}{90}$	$\frac{48}{90}$	$\frac{30}{90}$

CDF
$$F_X(x) = \begin{cases} 0 & x < 0 \\ 12/90 & 0 \le x < 1 \\ 60/90 & 1 \le x < 2 \\ 1 & x \geq 2 \end{cases}$$

$$\frac{12}{90} + \frac{48}{90} = \frac{60}{90}$$

2. **[17 pts]** The voltage V across a resistor is a continuous uniform random variable that ranges between −3 and 3 volts. The power Y dissipated by the resistor is dependent on the voltage, according to the formula $Y = V^2/10$ watts.

(a) Find the PDF of Y.

$$f_V(v) = \frac{1}{6} \quad (-3 \le v \le 3)$$

$$F_V(v) = \frac{v+3}{6} \quad (-3 \le v \le 3)$$

$$0 \le Y \le 9/10$$

$$F_Y(y) = P[Y \le y] = P\left[\frac{V^2}{10} \le y\right] = P\left[V^2 \le 10y\right] = P\left[-\sqrt{10y} \le V \le +\sqrt{10y}\right]$$

$$= F_V(+\sqrt{10y}) - F_V(-\sqrt{10y})$$

$$= \frac{\sqrt{10y}+3}{6} - \frac{-\sqrt{10y}+3}{6}$$

$$= \frac{\sqrt{10y}}{3}$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{\sqrt{10}}{3}\frac{d}{dy}(\sqrt{y}) = \frac{\sqrt{10}}{3}\cdot\frac{1}{2\sqrt{y}} = \frac{\sqrt{10}}{6\sqrt{y}}$$

$$f_Y(y) = \begin{cases} \frac{\sqrt{10}}{6\sqrt{y}} & 0 \le y \le \frac{9}{10} \\ 0 & \text{otherwise} \end{cases}$$

(b) If there are 35 such resistors (independent of each other), estimate the probability that the average power dissipated is less than 0.4 watts. (The expected value of Y is 0.3. The variance of Y is 0.072.)

$$M_{35} = \text{average power} = \text{gaussian}\left(0.3, \underbrace{\sqrt{\frac{0.072}{35}}}_{=0.04536}\right)$$

$$P[M_{35} < 0.4] = P\left[Z < \frac{0.4-0.3}{0.04536}\right] = P[Z < 2.20]$$

$$= \boxed{0.9861}$$

3. **[37 pts]** At a customer service center, the call rate is believed to be 2 calls per minute, and governed by a Poisson process.

(a) Find the probability the service center will receive more than 4 calls in a 1-minute period.

$$\underset{\text{in 1-minute}}{N = \#\text{ of calls}} = \text{poisson}(\alpha = 2) \qquad P_N(x) = \frac{2^x e^{-2}}{x!}$$

$$P[N > 4] = 1 - (P_N(0) + P_N(1) + P_N(2) + P_N(3) + P_N(4))$$

$$= 1 - \left(e^{-2} + 2e^{-2} + \frac{4e^{-2}}{2} + \frac{8e^{-2}}{6} + \frac{16e^{-2}}{24}\right)$$

$$= 1 - 0.9473 = \boxed{0.0527}$$

(b) The service center opens at 8:00 am. Find the probability the first call is received between 8:01 and 8:02 am.

$$X = \text{time until first call} = \text{exponential}(\lambda = 2) \qquad F_X(x) = 1 - e^{-2x}$$

$$P[1 \le X \le 2] = F_X(2) - F_X(1) = (1 - e^{-2\cdot 2}) - (1 - e^{-2\cdot 1})$$

$$= \boxed{0.1170}$$

(c) A service representative complains to her supervisor that they are receiving many more calls, on average, than 2 per minute. The supervisor designs a significance test (level 0.05) by counting the number of calls arriving during a 1-minute interval. If too many calls are received, she will reject the hypothesis of 2 calls per minute, on average. How many calls is too many?

H_0: N = number of calls = poisson$(\alpha = 2)$

Rejection Region: $\{N \geq n\}$

$P[N \geq n] \leq 0.05$

$P\{N \geq 5\} = 0.0527$

$P\{N \geq 6\} = 0.0166$ ★ first one below 0.05

Rejection Region: $\boxed{\{N \geq 6\}}$

Regardless of the number of calls received, 20% of all calls are complaints, and the remaining 80% are requests for assistance.

(d) If the center receives exactly 3 calls, find the probability that exactly 2 of them will be complaints.

Z = number of complaints (out of 3) = binomial$(n=3, p=0.2)$

$P\{Z=2\} = \binom{3}{2} 0.2^2\, 0.8^1 = \boxed{0.096}$

(e) Let X be the total number of calls received in a 5 minute period. Let Y be the number of complaints received in a 5 minute period. Construct the joint PMF of X and Y. If you choose to write the PMF as a table of values, complete the table only through $X = 2$ and $Y = 2$. (See below.)

X / Y	0	1	2	3…
0	0.1353	0.2165	0.1732	
1	0	0.0541	0.0866	
2	0	0	0.0217	
3…				

$$P_{X,Y}(x,y) = \underset{\text{poisson}(2)}{P_X(x)} \cdot \underset{\text{binomial}(n=x,\, p=0.2)}{P_{Y|X}(y|x)}$$

$$P_{X,Y}(0,0) = P_X(0)\cdot P_{Y|X}(0|0) = e^{-2}\cdot 1 = e^{-2} = 0.1353$$

$$P_{X,Y}(1,0) = P_X(1)\cdot P_{Y|X}(0|1) = 2e^{-2}\cdot 0.8 = 0.2165$$

$$P_{X,Y}(1,1) = P_X(1)\cdot P_{Y|X}(1|1) = 2e^{-2}\cdot 0.2 = 0.0541$$

$$P_{X,Y}(2,0) = P_X(2)\cdot P_{Y|X}(2|0) = \frac{4e^{-2}}{2}\cdot\binom{2}{0}(0.2)^0(0.8)^2 = 0.1732$$

$$P_{X,Y}(2,1) = P_X(2)\cdot P_{Y|X}(2|1) = \frac{4e^{-2}}{2}\cdot\binom{2}{1}(0.2)(0.8) = 0.0866$$

$$P_{X,Y}(2,2) = P_X(2)\cdot P_{Y|X}(2|2) = \frac{4e^{-2}}{2}\cdot\binom{2}{1}(0.2)^2(0.8)^0 = 0.0217$$

4. **[25 pts]** A drilling rig has two water pumps. The water pressure gauge attached to each pump is scaled from 0 to 1. A reading below 0.1 indicates a pump failure. If either gauge records a failure, the rig automatically shuts down. Let X be the pressure reading from pump 1 and Y be the pressure reading from pump 2. The two recordings vary randomly throughout the day. Their joint pdf is:

$$f_{X,Y}(x,y) = \begin{cases} 6x^2y & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the covariance of X and Y. Use the computational formula:

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

$$E[X] = \int_0^1\int_0^1 x \cdot 6x^2y \, dx\, dy = 6\left[\tfrac{1}{4}x^4\right]_0^1 \left[\tfrac{1}{2}y^2\right]_0^1 = \tfrac{3}{4}$$

$$E[Y] = \int_0^1\int_0^1 y \cdot 6x^2y \, dx\, dy = 6\left[\tfrac{1}{3}x^3\right]_0^1 \left[\tfrac{1}{3}y^3\right]_0^1 = \tfrac{2}{3}$$

$$E[XY] = \int_0^1\int_0^1 xy \cdot 6x^2y \, dx\, dy = 6\left[\tfrac{1}{4}x^4\right]_0^1 \left[\tfrac{1}{3}y^3\right]_0^1 = \tfrac{1}{2}$$

$$\mathrm{Cov}[X,Y] = \tfrac{1}{2} - \tfrac{3}{4}\cdot\tfrac{2}{3} = \boxed{0}$$

(b) Determine whether X and Y are independent. Justify your answer.

$$f_X(x) = \int_0^1 6x^2y\, dy = 6x^2\left[\tfrac{1}{2}y^2\right]_0^1 = 3x^2 \quad (0 \le x \le 1)$$

$$f_Y(y) = \int_0^1 6x^2y\, dx = 6y\left[\tfrac{1}{3}x^3\right]_0^1 = 2y \quad (0 \le y \le 1)$$

$$f_X(x)\cdot f_Y(y) = 3x^2 \cdot 2y = 6x^2y = f_{X,Y}(x,y)$$

X and Y <u>are</u> independent.

(c) Suppose we know the first pump has a reading of 0.8. Find the optimal linear estimate of the reading for pump 2.

$a = \frac{Cov[X,Y]}{Var[X]} = 0 \qquad b = E[Y] - aE[X] = \frac{2}{3}$

$\hat{y}_L(x) = 0x + \frac{2}{3} = \frac{2}{3}$

So $\hat{y}_L(0.8) = \frac{2}{3}$

(d) Suppose we know the first pump has a reading of 0.6. Find the optimal linear estimate of the reading for pump 2.

$\hat{y}_L(0.6) = \frac{2}{3}$

(e) How do your answers in parts (c) and (d) compare? What might explain the relationship between those 2 values?

They are the same. Because of independence, the first pump's reading has no bearing on the second pump's reading.

5. **[8 pts]** The weight of a component has a mean of 12.5g and a standard deviation of .05g. The manufacturing company offers a free sample of n components to potential customers. A particular customer weighs each component and computes the average. If this average deviates more than .01g from 12.5g, they will not buy the component from this company. Use Chebyshev's Inequality to find the smallest sample size n that **guarantees** the company will lose such a customer less than 10% of the time.

Want $P[|M_n - 12.5| > 0.01] < 0.10$.

$E[M_n] = 12.5, \quad Var[M_n] = \frac{0.05^2}{n}$

Chebyshev: $P[|M_n - 12.5| > 0.01] \leq \frac{\left(\frac{0.05^2}{n}\right)}{0.01^2} = \frac{25}{n^2}$

So set $\frac{25}{n^2} < 0.10$

$n^2 > \frac{25}{0.10} = 250$

$n > \sqrt{250} = 15.8$

$\boxed{n = 16}$

6. **[25 pts]** The fasting blood glucose level for a person with diabetes is approximately normally distributed with mean 105 milligrams per 100 milliliters and standard deviation 8 milligrams per 100 milliliters. For a healthy person (without diabetes), the fasting blood glucose level is approximately normally distributed with mean 95 milligrams per 100 milliliters and standard deviation 8 milligrams per 100 milliliters. Approximately 9% of people in the United States have diabetes.

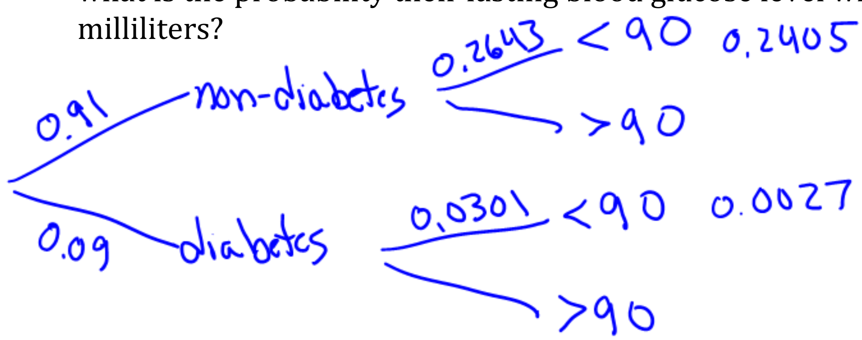
(a) What is the probability a healthy person will have a fasting blood glucose level below 90 milligrams per 100 milliliters?

$X = \text{normal}(95, 8)$

$P[X < 90] = P\left[Z < \frac{90-95}{8}\right] = \Phi(-0.625) = 1 - \Phi(0.63)$

$= 1 - 0.7357 = \boxed{0.2643}$

(b) If we select someone in the U.S.A. at random (not knowing whether they have diabetes), what is the probability their fasting blood glucose level will be below 90 milligrams per 100 milliliters?



for diabetes,

$Y = \text{normal}(105, 8)$

$P[Y<90] = P\left[Y < \frac{90-105}{8}\right]$

$= \Phi(-1.88)$

$= 1 - \Phi(1.88)$

$= 1 - 0.9699$

$= 0.0301$

$P[<90] = 0.2405 + 0.0027$

$= \boxed{0.2432}$

(c) If someone has a fasting blood glucose level below 90 milligrams per 100 milliliters, what is the probability that person has diabetes?

$P[\text{diabetes} \mid < 90] = \frac{P[\text{diabetes and } < 90]}{P[<90]} = \frac{0.0027}{0.2432}$

$= \boxed{0.0111}$

(d) Using a person's fasting blood glucose level, design a minimum cost test to determine whether a person has diabetes. Assume that mistakenly deciding a person with diabetes is healthy is twice as costly as concluding a healthy person has diabetes. Be sure to state your null and alternative hypotheses, H_0 and H_1. Find the decision threshold and state the decision rule.

H_0: person does not have diabetes
H_1: person has diabetes

$$P_{X|H_0}(x) \cdot P[H_0] \cdot C_{1|0} > P_{X|H_1}(x) \cdot P[H_1] \cdot C_{0|1}$$

$$\frac{1}{8\sqrt{2\pi}} e^{-\left(\frac{x-95}{8}\right)^2} \cdot 0.91 \cdot 1 > \frac{1}{8\sqrt{2\pi}} e^{-\left(\frac{x-105}{8}\right)^2} \cdot 0.09 \cdot 2$$

$$e^{-\left(\frac{x-95}{8}\right)^2} > e^{-\left(\frac{x-105}{8}\right)^2} \cdot \frac{0.09 \cdot 2}{0.91 \cdot 1}$$

$$-\left(\frac{x-95}{8}\right)^2 > -\left(\frac{x-105}{8}\right)^2 + \ln\left(\frac{0.09 \cdot 2}{0.91 \cdot 1}\right)$$

$$\left(\frac{x-95}{8}\right)^2 < \left(\frac{x-105}{8}\right)^2 - \ln\left(\frac{0.09 \cdot 2}{0.91 \cdot 1}\right)$$

$$(x-95)^2 < (x-105)^2 - 64 \ln\left(\frac{0.09 \cdot 2}{0.91 \cdot 1}\right)$$

$$x^2 - 190x + 95^2 < x^2 - 210x + 105^2 - 64 \ln\left(\frac{0.09 \cdot 2}{0.91 \cdot 1}\right)$$

$$20x < 105^2 - 95^2 - 64 \ln\left(\frac{0.09 \cdot 2}{0.91 \cdot 1}\right)$$

$$x < \frac{105^2 - 95^2 - 64 \ln\left(\frac{0.09 \cdot 2}{0.91 \cdot 1}\right)}{20} = 105.19$$

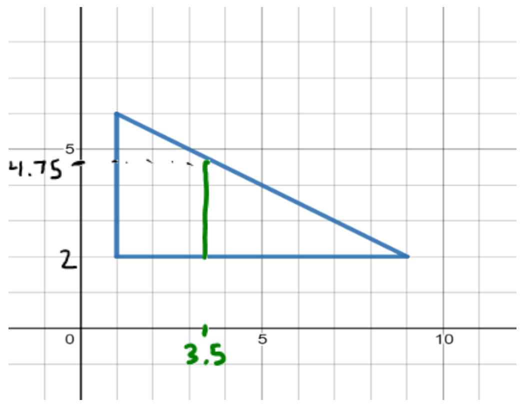
$x < 105.19 \rightarrow$ decide on H_0: no diabetes
$x > 105.19 \rightarrow$ decide on H_1: diabetes

7. **[20 pts]** Two companies are working on the same construction project. X is the time it takes the first company to complete its job, and Y is the time it takes the second company to complete its job. (Both X and Y are measured in days.) X and Y have the joint PDF given below.

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{16}, & 1 < x < 9, \quad 2 < y < 6, \quad x + 2y < 13 \\ 0 & \text{otherwise} \end{cases}$$

The marginal PDF of X is $f_X(x) = \frac{9-x}{32}$, $1 < x < 9$.

The border of the sample space is sketched below for your reference.



$x + 2y = 13$

$2y = 13 - x$

$y = \frac{13-x}{2}$

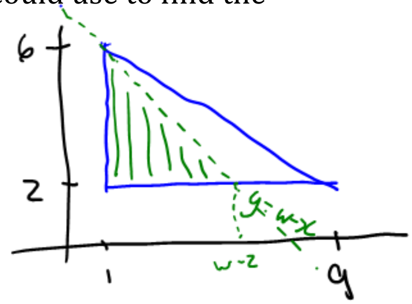
(a) Assuming it takes the first company 3.5 days to complete its job, find the minimum mean square error estimate of the time taken by the second company to complete its job.

$f_{Y|X}(y|3.5) = \frac{f_{X,Y}(3.5, y)}{f_X(3.5)} = \frac{\frac{1}{16}}{\frac{11}{64}} = \frac{4}{11} \quad (2 \le y \le 4.75)$

$\hat{y}_M(3.5) = E[Y|X=3.5] = \int_2^{4.75} y \cdot \frac{4}{11}\, dy = \frac{4}{11}\left[\frac{1}{2}y^2\right]_2^{4.75}$

$= \frac{2}{11}(4.75^2 - 2^2) = \boxed{3.375}$

(b) Let W be the total time spent by both companies on their jobs.
The sample space of W is $3 < W < 11$. **SET UP** (an) integral(s) that we could use to find the CDF of W. You do NOT need to compute the integrals or find the PDF.



$W = X + Y \quad 3 < w < 11$

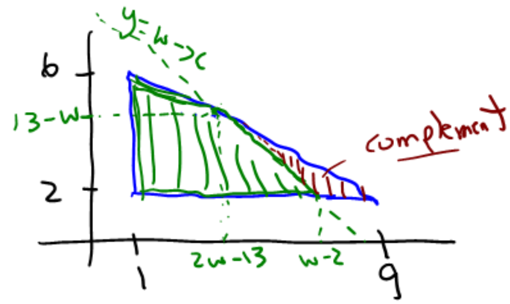
$F_W(w) = P[X + Y < w]$

If $3 < w < 7$, the region is as shown . . .

$$F_W(w) = \int_1^{w-2}\int_2^{w-x} \frac{1}{16}\,dy\,dx = \frac{1}{16}\cdot(\text{area of region})$$

$$= \frac{1}{16}\cdot\frac{1}{2}(w-3)(4) = \frac{1}{8}(w-3)$$

If $7 < w < 11$, the region is as shown



$$F_W(w) = 1 - P[X+Y > w]$$

$$= 1 - \frac{1}{16}(\text{area of region})$$

$$= 1 - \frac{1}{16}\cdot\frac{1}{2}\left(9-(w-2)\right)\left((13-w)-2\right)$$

$$= 1 - \frac{1}{32}(11-w)(11-w)$$

$$= 1 - \frac{1}{32}(11-w)^2$$

$w - x = \frac{13-x}{2}$

$2w - 2x = 13 - x$

$2w - 13 = x$

$y = w - x = w - (2w-13)$
$= 13 - w$

$$F_W(w) = \begin{cases} 0 & w < 3 \\ \frac{1}{8}(w-3) & 3 \le w \le 7 \\ 1 - \frac{1}{32}(11-w)^2 & 7 \le w \le 11 \\ 1 & w > 11 \end{cases}$$