M146 HW 1 Melody Chen #705120273

1) a)

#1	#2
4 red, 3 blue, 3 white	2 red, 4 blue, 4 white

p("picked two red")
= p("picked two red") ∩ p("urn 1") + p("picked two red") ∩ p("urn 2")
= p("picked two red" | "urn 1") p("urn 1") + p("picked two red" | "urn 2") p("urn 2")

p("picked two red" | "urn 1") = $\left(\frac{4}{10}\right)\left(\frac{3}{9}\right) = 0.13\overline{3}$

p("picked two red" | "urn 2") = $\left(\frac{2}{10}\right)\left(\frac{1}{9}\right) = 0.0\overline{2}$

p("urn 1") = 0.4 p("urn 2") = 0.6

⇒ p("picked two red") = $(0.13\overline{3})(0.4) + (0.0\overline{2})(0.6)$
$= 0.06\overline{6}$

b) Let 2B = event where 2nd ball is blue, 1B = 1st ball blue, 1R = 1st ball red

p(2B) = p(2B ∩ "urn 1") + p(2B ∩ "urn 2")

p(2B ∩ "urn 1") = p(2B ∩ "urn 1" ∩ 1B) + p(2B ∩ "urn 1" ∩ $1B^c$)
= p(2B ∩ 1B | "urn 1") p("urn 1") + p(2B ∩ $1B^c$ | "urn 1") p("urn 1")

p(2B ∩ 1B | "urn 1") = p((2B | "urn 1") ∩ (1B | "urn 1"))
= p((2B | "urn 1") | (1B | "urn 1")) p(1B | "urn 1")
= $\left(\frac{2}{9}\right)\left(\frac{3}{10}\right) = 0.06\overline{6}$

p(2B ∩ $1B^c$ | "urn 1") = p((2B | "urn 1") | ($1B^c$ | "urn 1")) p($1B^c$ | "urn 1")
= $\left(\frac{3}{9}\right)\left(\frac{7}{10}\right) = 0.2\overline{3}$

p(2B ∩ "urn 1") = $(0.6\overline{6})(0.4) + (0.2\overline{3})(0.4) = 0.12$

p(2B ∩ "urn 2") = use same logic as for urn 1 ...
= $\left(\frac{4}{10}\right)\left(\frac{3}{9}\right)(0.6) + \left(\frac{6}{10}\right)\left(\frac{4}{9}\right)(0.6) = 0.24$

p(2B) = 0.12 + 0.24 = 0.36

c) $p(2B \mid 1R) = \dfrac{p(2B \cap 1R)}{p(1R)}$

p(2B ∩ 1R) = p(2B ∩ 1R ∩ "urn 1") ∩ p(2B ∩ 1R ∩ "urn 2")
= p((2B ∩ 1R) | "urn 1") p("urn 1") + p(2B ∩ 1R | "urn 2") p("urn 2")
= $\left(\frac{4}{10}\right)\left(\frac{3}{9}\right)(0.4) + \left(\frac{2}{10}\right)\left(\frac{4}{9}\right)0.6 = 0.10\overline{6}$

p(1R) = p(1R ∩ "urn 1") ∩ p(1R ∩ "urn 2") = $0.4\left(\frac{4}{10}\right) + 0.6\left(\frac{2}{10}\right)$
= 0.28

$p(2B \mid 1R) = \dfrac{0.10\overline{6}}{0.28} = 0.38$

2) Total # of combinations: 6^6

Combinations that can result in 3 numbers each appear twice:

$$\underbrace{6C3}_{\text{ways to choose 3 numbers}} \cdot \underbrace{\frac{6!}{2!2!2!}}_{\text{ways to arrange the six dies}} = 1800$$

$$P(\text{"3 diff. number each appearing twice"}) = \frac{1800}{6^6} = 0.0386$$

3) $P(A) = 0.25 \quad P(B) = 0.35 \quad P(C) = 0.4$

Let D = event of defective

$P(D|A) = 0.05 \quad P(D|B) = 0.04 \quad P(D|C) = 0.02$

$$\begin{aligned} P(D) &= P(D\cap A) + P(D\cap B) + P(D\cap C) \\ &= P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C) \\ &= 0.05(0.25) + 0.04(0.35) + 0.02(0.4) = 0.0345 \end{aligned}$$

$$P\left(\text{"defective bolt manufactured by A"}\right) = P(A|D) = \frac{P(A\cap D)}{P(D)} = \frac{P(D|A)P(A)}{P(D)} = \frac{(0.05)(0.25)}{0.0345} = 0.3623$$

$$P\left(\text{"defective bolt manufactured by B"}\right) = P(B|D) = \frac{P(B\cap D)}{P(D)} = \frac{P(D|B)P(B)}{P(D)} = \frac{(0.04)(0.35)}{0.0345} = 0.4058$$

$$P\left(\text{"defective bolt manufactured by C"}\right) = P(C|D) = \frac{P(C\cap D)}{P(D)} = \frac{P(D|C)P(C)}{P(D)} = \frac{(0.02)(0.4)}{0.0345} = 0.2319$$

4) a) Show that $E(X+Y) = E(X) + E(Y)$

$$\begin{aligned} E(X+Y) &= \sum_X \sum_Y (x+y) P(Y=y, X=x) \\ &= \sum_x \sum_y x P(Y=y, X=x) + \sum_x \sum_y y P(Y=y, X=x) \\ &= \sum_x x \sum_y P(Y=y, X=x) + \sum_y y \sum_x P(Y=y, X=x) \\ &= \sum_x x P(X=x) + \sum_y y P(Y=y) = E(X) + E(Y) \end{aligned}$$

b) If X and Y are independent, show that $\text{var}[X+Y] = \text{var}(x) + \text{var}(y)$

$$\begin{aligned} \text{var}(X+y) &= E\left[\left((X+y) - E(X+y)\right)^2\right] = E\left[(X - \mu_x + y - \mu_y)^2\right] \\ &= \sum_x \sum_y \left((x-\mu_x)^2 + 2(x-\mu_x)(y-\mu_y) + (y-\mu_y)^2\right) p(x,y) \\ &= \sum_x \sum_y (x-\mu_x)^2 p(x,y) + 2\sum_x \sum_y (x-\mu_x)(y-\mu_y) p(x,y) + \sum_x \sum_y (y-\mu_y)^2 p(x,y) \\ &= \sum_x (x-\mu_x)^2 p(x) + 2\sum_x \sum_y xy\, p(x,y) - 2\sum_x \sum_y x \mu_y p(x,y) \\ &\quad - 2\sum_x \sum_y \mu_x y\, p(x,y) + 2\mu_x \mu_y \sum_x \sum_y p(x,y) + \sum_y (y-\mu_y)^2 p(y) \end{aligned}$$

$$= var(x) + 2E(xy) - 2\mu_y \sum_x x\,p(x) - 2\mu_x \sum_y y\,p(y) + 2\mu_x\mu_y + var(y)$$

$$= var(x) + 2E(x)E(y) - 2\mu_y\mu_x - 2\mu_x\mu_y + 2\mu_x\mu_y + var(y)$$

$$= var(x) + var(y)$$

5) $T \sim \text{Exponential}(\lambda)$

a) $P(T \geq d \mid T \geq r) = P(T \geq d)$ ($d \geq r$) because of memoryless property of exponentials

$$= 1 - P(T < d) = 1 - (1 - e^{-\lambda d}) = e^{-\lambda d}$$

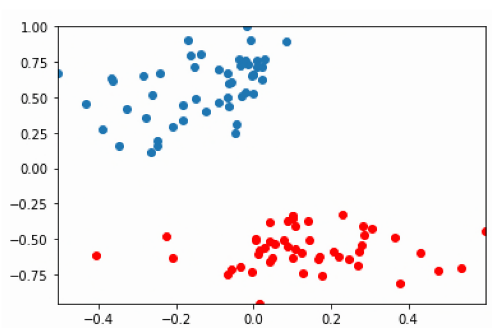
b) $E(T) = \int_0^\infty t(\lambda e^{-\lambda t})\,dt$ $\quad u = t \quad dv = e^{-\lambda t}dt$ $\quad du = dt \quad v = -\frac{1}{\lambda}e^{-\lambda t}$

$$= \lambda\left[-\frac{t}{\lambda}e^{-\lambda t} + \frac{1}{\lambda}\int_0^\infty e^{-\lambda t}\,dt\right]_0^\infty = \lambda\left[-\frac{t}{\lambda}e^{-\lambda t} - \frac{1}{\lambda^2}e^{-\lambda t}\right]_0^\infty$$

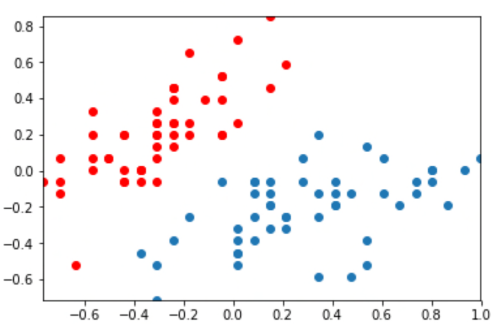
$$= \lambda\left(\frac{1}{\lambda^2}\right) = \frac{1}{\lambda}$$

6) a)

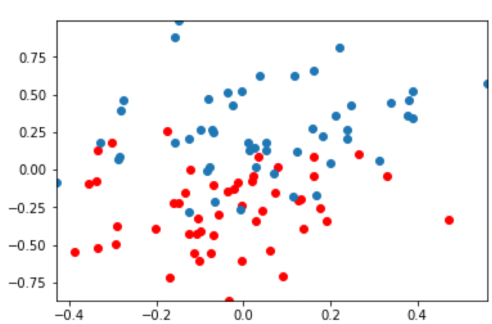
Data Set #1



Data Set #2



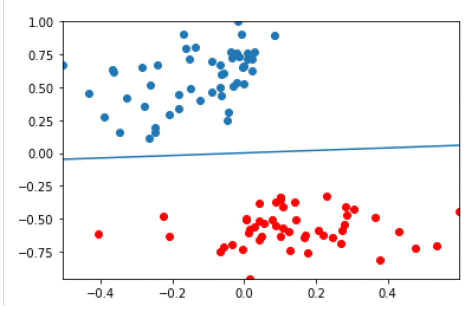
Data Set #3



Data set #1 and #2 are separable.

b) Data Set #1

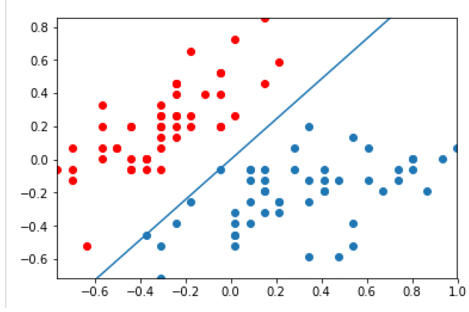
```
w vector: [0, 0.14217200000000002, -1.47323]
w1: 0.14217200000000002 w2: -1.47323
b bias: 0
# of updates: 2
```



Data Set #1 converges.

Data Set #2

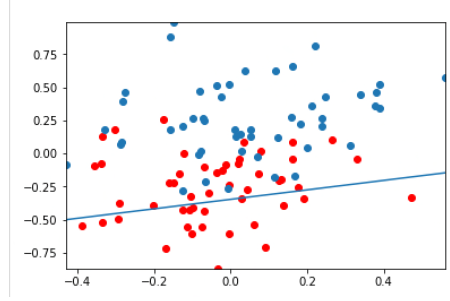
```
w vector: [0, -1.10918, 0.91344]
w1: -1.10918 w2: 0.91344
b bias: 0
# of updates: 4
```



Data Set #2 converges.

Data Set #3

```
w vector: [-1, 1.0291400000000066, -2.8796620000003377]
w1: 1.0291400000000066 w2: -2.8796620000003377
b bias: -1
# of updates: 4501
```



Data set #3 does not converge as data is not linearly separable, which is a creteria for using the perceptron algorithm. In addition, the # of updates is significantly large, unlike data set #1 and #2, showing that perceptron algorithm does not converge.

c) Data Set #1

$\gamma_{w,b} = 0.13687$

$\frac{1}{\gamma_{w,b}^2} = 53.3819$

Compare to the upperbound $\frac{1}{\gamma_{w,b}^2}$, our algorithm converged in 2 updates, which is significantly smaller.

Data Set #2

$\gamma_{w,b} = 0.000265$

$\frac{1}{\gamma_{w,b}^2} = 14251382.48$

Compare to the upper bound $\frac{1}{\gamma_{w,b}^2}$, our algorithm converged in 4 updates, which is significantly smaller.

Data Set #3 does not converge.