Sai Omkar Kandukuri

Problem Set 3

Question 1

①

A) Given $f(x) = \lambda e^{-\lambda x}$ for $0 \le x < \infty$

c.d.f $F(x) = \int_0^x f(x)\,dx$

$$F(x) = \int_0^x \lambda e^{-\lambda x}\,dx$$

$$= \left.\frac{\lambda e^{-\lambda x}}{-\lambda}\right]_0^x$$

$$= -e^{-\lambda x} - (-1)$$

$$\boxed{F(x) = 1 - e^{-\lambda x}}$$

B) Given $F(x) = u$.

$$1 - e^{-\lambda x} = u$$

$$1 - u = e^{-\lambda x}$$

$$\ln(1-u) = -\lambda x.$$

$$\boxed{x = \frac{\ln(1-u)}{-\lambda}}$$

Question 2

Given $f(x,y)$

	$y=1$	$y=2$	$y=3$	$y=4$	$f_x(x)$
$x=1$	0.15	0.05	0.05	0.05	0.30
$x=2$	0.05	0.05	0.05	0.25	0.40
$x=3$	0.05	0.05	0.10	0.10	0.30
$f_y(y)$	0.25	0.15	0.20	0.40	1

(2) Calculate $E(x)$, $Var(x)$, $Var(y)$ and $Cov(y,x)$

$$E(x) = 1\times 0.3 + 2\times 0.4 + 3\times 0.3$$

$$\boxed{E(x) = 2}$$

$$E(y) = 1\times 0.25 + 2\times 0.15 + 3\times 0.20 + 4\times 0.4$$

$$\boxed{E(y) = 2.75}$$

$$Var(x) = E(x^2) - \mu^2$$
$$= 1^2\times 0.3 + 2^2\times 0.4 + 3^2\times 0.3 - 2^2$$
$$= 0.3 + 1.6 + 2.7 - 4$$

$$\boxed{Var(x) = 0.6}$$

$$Var(y) = E(y^2) - \mu^2$$
$$= 1^2\times 0.25 + 2^2\times 0.15 + 3^2\times 0.2 + 4^2\times 0.4 - (2.75)^2$$
$$= 0.25 + 0.6 + 1.8 + 6.4 - 7.5625$$

$$\boxed{Var(y) = 1.4875}$$

xy	possibilities	$P(xy)$
1	(1,1)	0.15
2	(1,2) (2,1)	0.05+0.05
3	(1,3) (3,1)	0.05 +0.05
4	(1,4) (2,2)	0.05 +0.05
6	(2,3) (3,2)	0.05 +0.05
8	(2,4)	0.25
9	(3,3)	0.1
12	(3,4)	0.1

$$Cov(x,y) = Cov(y,x) = E(xy) - \mu_x \mu_y$$

$$= 1 \times 0.15 + 2(0.05+0.05) + 3(0.05+0.05) + 4(0.05+0.05) + 6(0.05+0.05) + 8(0.25) + 9(0.1) + 12(0.1) - (2 \times 2.75)$$

$$\boxed{Cov(y,x) = 5.75 - 5.5 = 0.25}$$

Question 3

③ Suppose $x=2$ Calculate $E(y|x=2)$ and $Var(y|x=2)$

	$y=1$	$y=2$	$y=3$	$y=4$	$f_x(x)$
$x=2$	0.05	0.05	0.05	0.25	0.4

$$E(y|x=2) = \frac{1\times0.05 + 2\times0.05 + 3\times0.05 + 4\times0.25}{0.4}$$

$$= \frac{1.3}{0.4}$$

$$\boxed{E(y|x=2) = 3.25}$$

$$Var(y|x=2) = E(y^2|x=2) - [E(y|x=2)]^2$$

$$= \frac{1^2\times0.05 + 2^2\times0.05 + 3^2\times0.05 + 4^2\times0.25}{0.4} - \left(\frac{13}{4}\right)^2$$

$$= \frac{0.05 + 0.2 + 0.45 + 4}{0.4} - \frac{169}{16}$$

$$\boxed{Var(y|x=2) = \frac{19}{6} = 1.1875}$$

Question 4

(4) Suppose $y<3$ Calculate $E(x|y<3)$, $Var(x|y<3)$

$$E(x|y<3) = \frac{f(x|y<3)}{f(y<3)}$$

$$= \frac{1\times(0.15+0.05)+2\times(0.05+0.05)+3(0.05+0.05)}{0.25+0.15}$$

$$= \frac{0.2+0.2+0.3}{0.4} = \frac{7}{4} = 1.75$$

$$\boxed{E(x|y<3) = 1.75}$$

$$Var(x|y<3) = E(x^2|y<3) - [E(x|y<3)]^2$$

$$= \frac{1^2(0.15+0.05)+2^2(0.05+0.05)+3^2(0.05+0.05)}{0.25+0.15} - \left(\frac{7}{4}\right)^2$$

$$= \frac{0.2+0.4+0.9}{0.4} - \left(\frac{7}{4}\right)^2$$

$$= \frac{15}{4} - \frac{49}{16} = \frac{11}{16}$$

$$\boxed{Var(x|y<3) = 0.6875}$$

Question 5

R

Sample size of 50, 500, 10000

```{r}
library(tidyverse)
set.seed(11182021)
grid <- expand.grid(x=c(1,2,3),y=c(1,2,3,4))
sample <- mutate(grid,p = c(0.15,0.05,0.05,0.05,0.05,0.05,0.05,0.05,0.1,0.05,0.25,0.1))

#sample_50n <- sample_n(sample,size = 50, replace = TRUE, weight = sample$p)
sample50n <- sample_n(sample,size = 50, replace = TRUE, weight = sample$p)


sample50n %>% summarise(mean(x), var(x), mean(y), var(y), cov(x,y))

```

Description: df [1 x 5]

mean(x) <dbl>	var(x) <dbl>	mean(y) <dbl>	var(y) <dbl>	cov(x, y) <dbl>
2.1	0.6632653	2.88	1.209796	0.2367347

1 row

```{r 500}

grid <- expand.grid(x=c(1,2,3),y=c(1,2,3,4))
sample <- mutate(grid,p = c(0.15,0.05,0.05,0.05,0.05,0.05,0.05,0.05,0.1,0.05,0.25,0.1))

sample500n <- sample_n(sample,size = 500, replace = TRUE, weight = sample$p)


sample500n %>% summarise(mean(x), var(x), mean(y), var(y), cov(x,y))
```

Description: df [1 x 5]

mean(x) <dbl>	var(x) <dbl>	mean(y) <dbl>	var(y) <dbl>	cov(x, y) <dbl>
2.022	0.6067295	2.726	1.533992	0.232493

1 row

```{r 10000}

grid <- expand.grid(x=c(1,2,3),y=c(1,2,3,4))
sample <- mutate(grid,p = c(0.15,0.05,0.05,0.05,0.05,0.05,0.05,0.05,0.1,0.05,0.25,0.1))

sample10000n <- sample_n(sample,size = 10000, replace = TRUE, weight = sample$p)


sample10000n %>% summarise(mean(x), var(x), mean(y), var(y), cov(x,y))

#Mean, variance, covariance gets fine tuned and closer to calculated values as we increase the sample size
```

Description: df [1 x 5]

mean(x) <dbl>	var(x) <dbl>	mean(y) <dbl>	var(y) <dbl>	cov(x, y) <dbl>
2.0064	0.6018192	2.7475	1.490093	0.2592419

1 row

Mean, Variance, covariance gets fine tuned and gets closer to calculated values as we increase the sample size.

Stata

Sample Size : 50,500, 10000

Mean, Variance, Covariance get closes to calculated values as we increase the sample size.

```
.
. sum x, detail

                              x
-------------------------------------------------------------
      Percentiles      Smallest
 1%            1              1
 5%            1              1
10%            1              1       Obs                  50
25%            1              1       Sum of Wgt.          50

50%            2                      Mean                1.9
                        Largest       Std. Dev.       .814411
75%            3              3
90%            3              3       Variance       .6632653
95%            3              3       Skewness       .1831898
99%            3              3       Kurtosis       1.560473

. sum y, detail

                              y
-------------------------------------------------------------
      Percentiles      Smallest
 1%            1              1
 5%            1              1
10%            1              1       Obs                  50
25%            1              1       Sum of Wgt.          50
```

```
50%            3                      Mean               2.62
                        Largest       Std. Dev.      1.193349
75%            4              4
90%            4              4       Variance       1.424082
95%            4              4       Skewness      -.1784097
99%            4              4       Kurtosis         1.5315
```

500

```
.
. sum x, detail

                              x
-------------------------------------------------------------
      Percentiles      Smallest
 1%            1              1
 5%            1              1
10%            1              1       Obs                 500
25%            1              1       Sum of Wgt.         500

50%            2                      Mean              2.028
                        Largest       Std. Dev.      .7748657
75%            3              3
90%            3              3       Variance       .6004168
95%            3              3       Skewness      -.0481971
99%            3              3       Kurtosis       1.670152

. sum y, detail

                              y
-------------------------------------------------------------
      Percentiles      Smallest
 1%            1              1
 5%            1              1
10%            1              1       Obs                 500
25%          1.5              1       Sum of Wgt.         500

50%            3                      Mean              2.744
                        Largest       Std. Dev.      1.212189
75%            4              4
90%            4              4       Variance       1.469403
95%            4              4       Skewness       -.345174
99%            4              4       Kurtosis       1.545386

.
```

10000

```
.
. sum x, detail

                              x
-------------------------------------------------------------
      Percentiles      Smallest
 1%            1              1
 5%            1              1
10%            1              1       Obs              10,000
25%            1              1       Sum of Wgt.      10,000

50%            2                      Mean             1.9962
                        Largest       Std. Dev.      .7733339
75%            3              3
90%            3              3       Variance       .5980454
95%            3              3       Skewness       .0065246
99%            3              3       Kurtosis       1.672305

. sum y, detail

                              y
-------------------------------------------------------------
      Percentiles      Smallest
 1%            1              1
 5%            1              1
10%            1              1       Obs              10,000
25%            1              1       Sum of Wgt.      10,000

50%            3                      Mean             2.7465
                        Largest       Std. Dev.      1.216053
75%            4              4
90%            4              4       Variance       1.478786
95%            4              4       Skewness      -.3424367
99%            4              4       Kurtosis       1.532367
```

Question 6

(6) $x = \{1, 2, 3\}$ $y = \{1, 2, 3\}$

for x, y to be statistically independent, knowledge of one variable dictates nothing about the other variable

	$y=1$	$y=2$	$y=3$	$f_x(x)$
$x=1$	1/18	2/18	3/18	1/3
$x=2$	1/18	2/18	3/18	1/3
$x=3$	1/18	2/18	3/18	1/2
$f_y(y)$	3/18	6/18	9/18	1

In above table, we can see same distribution of y for every value of x (and) same distribution of x for every value of y.

Hence y, x are independent

$$Y \perp\!\!\!\perp X$$

This also implies $f_y(y|x) = f_y(y)$

Question 7

(7) Given

notation

$t_{+ve} \Rightarrow$ test +ve

$t_{-ve} \Rightarrow$ test neg.

Covid $\Rightarrow$ Has Covid

$-$Covid $\Rightarrow$ No Covid

$$P(t_{+ve} \mid Covid) = \frac{80}{100} = \frac{4}{5}$$

$$P(t_{-ve} \mid Covid) = \frac{20}{100} = 1/5$$

$$P(t_{+ve} \mid -Covid) = \frac{2}{100} = \frac{1}{50}$$

$$P(t_{-ve} \mid -Covid) = \frac{98}{100} = \frac{49}{50}$$

$P(Covid) = \frac{3}{100}$ (3% of population has disease)

$$P(-Covid) = \frac{100-3}{100} = \frac{97}{100}$$

A) Patient that tests -ve, probabibility of having covid

$$P(Covid \mid t_{-ve}) = \frac{P(t_{-ve} \mid Covid) \times P(Covid)}{P(t_{-ve} \mid Covid) \times P(Covid) + P(t_{-ve} \mid -Covid) \times P(-Covid)}$$

$$= \frac{\frac{1}{5} \times \frac{3}{100}}{\frac{1}{5} \times \frac{3}{100} + \frac{49}{50} \times \frac{97}{100}} = \frac{\frac{3}{500}}{\frac{4783}{5000}}$$

$$\boxed{P(Covid \mid t_{-ve}) = 0.00627 \text{ i.e } 0.62\%}$$

Question 7B

7B) for a patient that tests +ve, what is probability of having the disease

$$P(\text{Covid} \mid t_{+ve}) = \frac{P(t_{+ve} \mid \text{Covid}) \times P(\text{Covid})}{P(t_{+ve} \mid \text{Covid}) \times P(\text{Covid}) + P(t_{+ve} \mid -\text{Covid}) \times P(-\text{Covid})}$$

$$= \frac{\frac{4}{5} \times \frac{3}{100}}{\frac{4}{5} \times \frac{3}{100} + \frac{1}{50} \times \frac{97}{100}}$$

$$= \frac{\frac{12}{500}}{\frac{217}{5000}} = 0.553$$

$$\boxed{P(\text{Covid} \mid t_{+ve}) = 0.553 \text{ i.e } 55.3\%}$$