Robin Baldeo St 540
Home Work 2

Question # 2

(a) For valid pdf we need to prove
1. $f(x) \geq 0$
2. $\int_a^b f(x) = 1$

1. We know for
$f(x) \geq 0$ where $S = [a,b]$ $x \in S$

2. $\int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} \left[x \right] \Big|_a^b = \frac{1}{b-a}(b-a) = 1.$

There because $f(x) \geq 0$ and $\int_a^b \frac{1}{b-a} dx = 1$. The pdf for the uniform distribution is a valid pdf.

(b)
$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_a^b x \frac{1}{b-a} dx$$
$$= \frac{1}{b-a} \int_a^b x\, dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right] \Big|_a^b$$
$$= \frac{1}{b-a} \left[\frac{b^2}{2} - \frac{a^2}{2} \right] = \frac{(b+a)(b-a)}{2(b-a)}$$
$$= \frac{b+a}{2}$$

Find Variance

$$E(x^2)=\int_{-\infty}^{\infty}x^2 f(x)\,dx = \int_a^b x^2\cdot\frac{1}{b-a}dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]\Big|_a^b = \frac{1}{b-a}\cdot\frac{(b^3-a^3)}{3}$$

$$=\frac{(b^3-a^3)}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

Using Variance computting formula

$$Var(x)=E(x)^2-E(x^2)$$

$$=\frac{(b+a)^2}{4}-\frac{(a^2+ab+b^2)}{3} = \frac{3b^2+6ab+3a^2-4a^2-4ab-4b^2}{12}$$

$$=\frac{-a^2+2ab-b^2}{12} = \frac{(b-a)^2}{12}$$

Question #3

A distrubtion that meets the criteria is a gama distrubtion

$\frac{a}{b}=5 \qquad \frac{a}{b^2}=3$

$a=5b \qquad 5b=3b^2$

$b=\frac{5}{3},\ a=\frac{25}{3}$

Therefore Gama$\left(\frac{25}{3},\frac{5}{3}\right)$ satisfy the requirement.

Question #4

		X_2		
		0	1	
X_1	0	.15	.15	.3
	1	.15	.2	.35
	2	.15	.2	.35
		.45	.55	

(a) $P_{X_1}(x_1) = \begin{cases} .3 & x_1=0 \\ .35 & x_2=1 \\ .35 & x_3=2 \end{cases}$

$P_{X_1}(X_1=0) = \sum_{0}^{1} P(x_{1i}, x_{2i}) = P(X_1=0, X_2=0) + P(X_1=0, X_2=1)$
$= .15 + .15 = .3$

$P_{X_1}(x_1=1) = \sum_{0}^{1} P(x_{1i}, x_{2i}) = P(X_1=1, X_2=0) + P(X_1=1, X_2=1)$
$= .15 + .20 = .35$

$P_{X_1}(X_1=2) = \sum_{i=0}^{1} P(x_i, x_{2i}) = P(X_1=2, X_2=0) + P(X_1=2, X_2=1)$
$= .15 + .2 = .35$

(b) $P_{X_2}(X_2) = \begin{cases} .45 & x_2=0 \\ .55 & x_2=1 \end{cases}$

$P_{X_2}(X_2=0) = \sum_{i=0}^{2} P(x_{1i}, x_i) = P(X_1=0, X_2=0) + P(X_1=1, X_2=0) + P(X_1=2, X_2=0)$
$= .15 + .15 + .15 = .45$

$P_{X_2}(X_2=1) = \sum_{i=0}^{2} P(x_{1i}, x_{2i}) = P(X_1=0, X_2=1) + P(X_1=1, X_2=1) + P(X_1=2, X_2=1)$
$= .15 + .2 + .2 = .55$

(c) $P(X_1|X_2)$

X_1 \ X_2	0	1
0	$\frac{1}{3}$	$\frac{3}{11}$
1	$\frac{1}{3}$	$\frac{4}{11}$
2	$\frac{1}{3}$	$\frac{4}{11}$
	1	1

$P(x=0|x_2=0) = \frac{.15}{.45}$ $\quad P(x=0|x_2=1) = \frac{.15}{.55}$

$P(x=1|x_2=0 = \frac{.15}{.45}$ $\quad P(x=1|x_2=1) = \frac{.2}{.55}$

$P(x=2|x_2=0) = \frac{.15}{.45}$ $\quad P(x=2|x_2=1) = \frac{.2}{.55}$

$P(X_2|X_1)$ (d)

X_1 \ X_2	0	1	
0	$\frac{1}{2}$	$\frac{1}{2}$	1
1	$\frac{3}{7}$	$\frac{4}{7}$	1
2	$\frac{3}{7}$	$\frac{4}{7}$	1

$P(X_2=0|X_1=0) = \frac{.15}{.3}$ $\quad P(X_2=0|X_1=2) = \frac{.15}{.35}$

$P(X_2=1|X_1=0) = \frac{.15}{.3}$ $\quad P(X_2=1|X=2) = \frac{.2}{.35}$

$P(X_2=0|X_1=1) = \frac{.15}{.35}$

$P(X_2=1|X_1=1) = \frac{.2}{.35}$

(c) X_1 and X_2 are dependent because

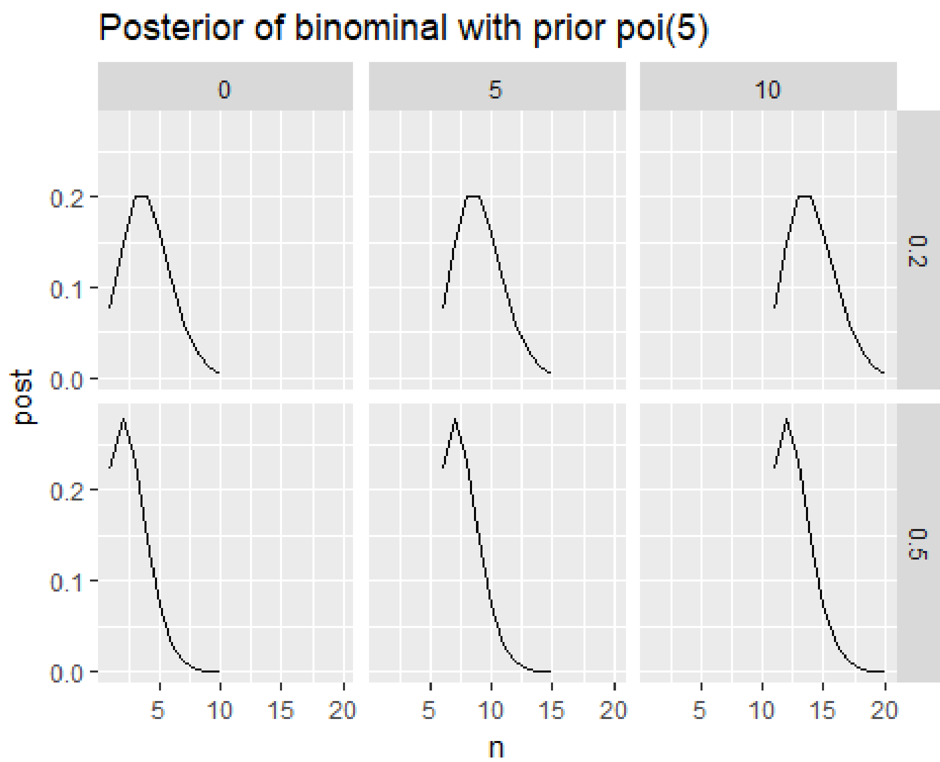
$P(X_1=2 \mid X_2=0) = \frac{.15}{.45} \neq .35$

Question7

```
library(purrr)
library(tidyr)
library(ggplot2)
```

Question 7

```
#building the posterior function with parameters
#theta given in the question
#y given in the question

postE<- function(theta, y){
  n= seq(1:10) + y
  like =dbinom(y,n,theta)
  prior= dpois(n,5)
  fy = sum(like*prior)
  post= like*prior/fy
  return(cbind(theta= theta, y = y, n = n, post = post))
}

#loop throught theta and y to create list of various combinations
df<-
map2(c(.2,.5)%>%rep(3)%>%map(~..1),c(0,5.,10.)%>%rep(2)%>%sort()%>%map(~..1),
~postE(..1, ..2))

#flaten out dataframes in list as one dataframe
data<- df[[1]]%>%rbind(df[[2]])

for(i in 3:6){
  data= rbind(data, df[[i]])
}

#plot the dataframes
ggplot(as.data.frame(data), aes(y = post,x = n))+ geom_line()+
facet_grid(theta~y)+labs(title = "Posterior of binominal with prior poi(5)")
```



We see with the different theta the post exterior curve peaks with theta .2 at .2, but with .5 the curve peaks at about 2.75. Also, the posterior curve stays the same for each combination of y with the same theta. The only difference between the 3 values of y is that the curve shifts further away from zero as y increases.