Problem 1: X_1 and X_2 have joint PMF

x_1	x_2	$Prob(X_1 = x_1; X_2 = x_2)$
0	0	0.15
1	0	0.15
2	0	0.15
0	1	0.15
1	1	0.20
2	1	0.20

a) Compute the marginal distribution of X_1.

$$f_{X_1}(0) = P(X_1 = 0) = \sum_{x_2} f(0, x_2) = 0.15 + 0.15 = 0.3$$
$$f_{X_1}(1) = P(X_1 = 1) = \sum_{x_2} f(1, x_2) = 0.15 + 0.20 = 0.35$$
$$f_{X_1}(2) = P(X_1 = 2) = \sum_{x_2} f(2, x_2) = 0.15 + 0.20 = 0.35$$

b) Compute the marginal distribution of X_2.

$$f_{X_2}(0) = P(X_2 = 0) = \sum_{x_1} f(x_1, 0) = 0.15 + 0.15 + 0.15 = 0.45$$

$$f_{X_2}(1) = P(X_2 = 1) = \sum_{x_2} f(x_1, 1) = 0.15 + 0.20 + 0.20 = 0.55$$

c) Compute the conditional distribution of $X_1|X_2$.

$$f(X_1 = 0|X_2 = 0) = \frac{f(X_2 = 0, X_1 = 0)}{f_{X_2}(0)} = \frac{0.15}{0.45} = \frac{1}{3}$$
$$f(X_1 = 1|X_2 = 0) = \frac{f(X_2 = 1, X_1 = 0)}{f_{X_2}(0)} = \frac{0.15}{0.45} = \frac{1}{3}$$
$$f(X_1 = 2|X_2 = 0) = \frac{f(X_2 = 2, X_1 = 0)}{f_{X_2}(0)} = \frac{0.15}{0.45} = \frac{1}{3}$$
$$f(X_1 = 0|X_2 = 1) = \frac{f(X_2 = 0, X_1 = 1)}{f_{X_2}(1)} = \frac{0.15}{0.55} = \frac{3}{11}$$
$$f(X_1 = 1|X_2 = 1) = \frac{f(X_2 = 1, X_1 = 1)}{f_{X_2}(1)} = \frac{0.20}{0.55} = \frac{4}{11}$$
$$f(X_1 = 2|X_2 = 1) = \frac{f(X_2 = 2, X_1 = 1)}{f_{X_2}(1)} = \frac{0.20}{0.55} = \frac{4}{11}$$

d) Compute the conditional distribution of $X_2|X_1$.

$$
\begin{aligned}
f(X_2=0|X_1=0) &= \frac{f(X_1=0,X_2=0)}{f_{X_1}(0)} = \frac{0.15}{0.30} = \frac{1}{2} \\
f(X_2=1|X_1=0) &= \frac{f(X_1=0,X_2=1)}{f_{X_1}(0)} = \frac{0.15}{0.30} = \frac{1}{2} \\
f(X_2=0|X_1=1) &= \frac{f(X_1=1,X_2=0)}{f_{X_1}(1)} = \frac{0.15}{0.35} = \frac{3}{7} \\
f(X_2=1|X_1=1) &= \frac{f(X_1=1,X_2=1)}{f_{X_1}(1)} = \frac{0.20}{0.35} = \frac{4}{7} \\
f(X_2=0|X_1=2) &= \frac{f(X_1=2,X_2=0)}{f_{X_1}(2)} = \frac{0.15}{0.35} = \frac{3}{7} \\
f(X_2=1|X_1=2) &= \frac{f(X_1=2,X_2=1)}{f_{X_1}(2)} = \frac{0.20}{0.35} = \frac{4}{7}
\end{aligned}
$$

e) Are X_1 and X_2 independent? Justify your answer.

Since

$$P(X_1=1,X_2=1) = 0.2 \neq 0.1925 = .35 \times .55 = P(X_1=1) \times P(X_2=1),$$

X_1 and X_2 are dependent.

Problem 2: Assume (X_1, X_2) have bivariate PDF

$$f(x_1,x_2) = \frac{1}{2\pi}(1+x_1^2+x_2^2)^{-\frac{3}{2}}$$

a) Plot the conditional distribution of $X_1|X_2=x_2$ for $x_2 \in \{-3,-2,-1,0,1,2,3\}$ (preferably on the same plot).

Conditional probability for continuous variable is equal to

$$f(X_1|X_2) = \frac{f(X_1=x_1,X_2=x_2)}{f_{X_2}(x_2)} = \frac{\frac{1}{2\pi}(1+x_1^2+x_2^2)^{-\frac{3}{2}}}{\frac{1}{(x_2^2+1)\pi}} = \frac{1}{2}(1+x_2^2)(1+x_1^2+x_2^2)^{-\frac{3}{2}}$$

where

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} \frac{1}{2\pi}(1+x_1^2+x_2^2)^{-\frac{3}{2}}\,dx_1 = \frac{1}{(x_2^2+1)\pi}$$

```r
library(tibble)
library(ggplot2)
library(tidyr)
library(dplyr)
```

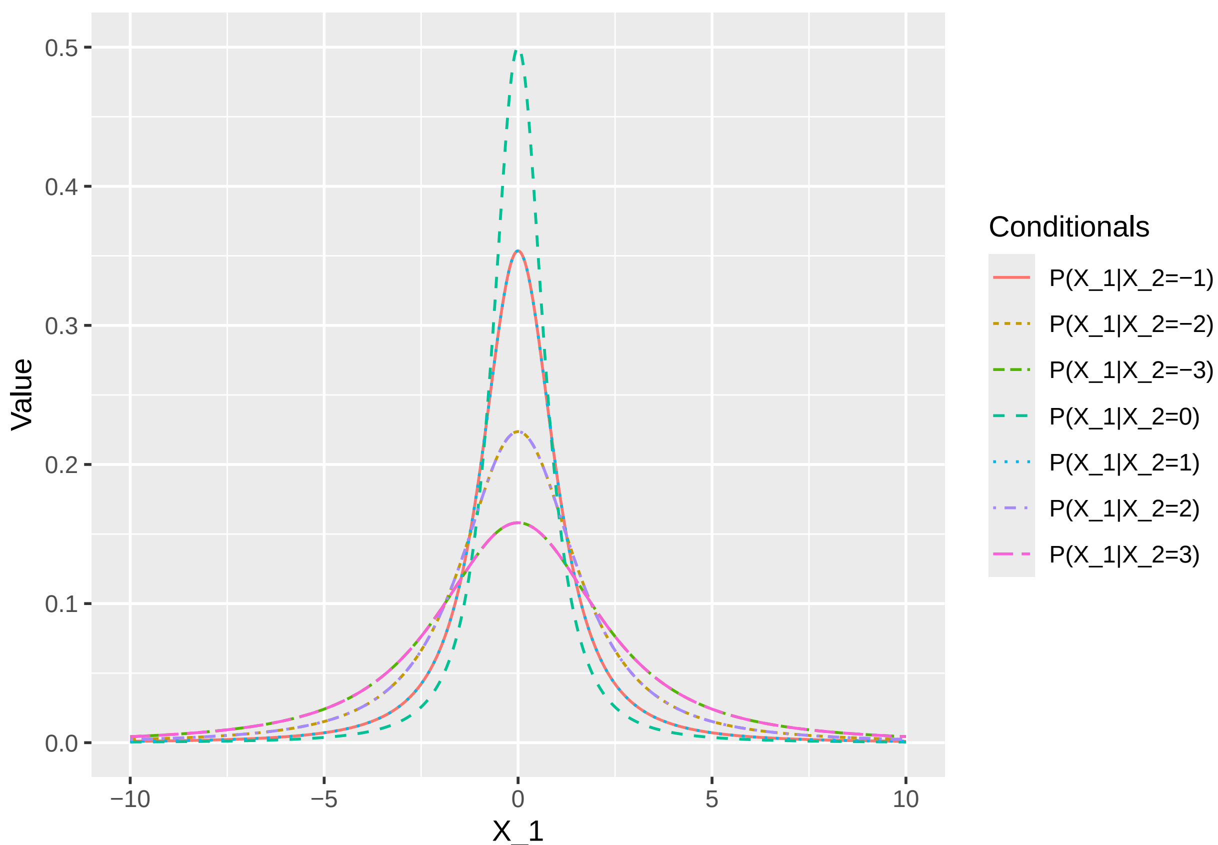
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
cd_func=function(x2){
  #conditional probability
  cond_prob=(((1+x1^2+x2^2)^(-1.5))*(1+x2^2))/2
}
x1=seq(-10,10,.01)
x2=c(-3,-2,-1,0,1,2,3)
con_dist=sapply(x2,cd_func)
df=data.frame(x1, con_dist)
colnames(df)=c("X_1","P(X_1|X_2=-3)","P(X_1|X_2=-2)","P(X_1|X_2=-1)",
                "P(X_1|X_2=0)","P(X_1|X_2=1)","P(X_1|X_2=2)","P(X_1|X_2=3)")
df=df%>%
pivot_longer(cols = "P(X_1|X_2=-3)":"P(X_1|X_2=3)", names_to = "Conditionals",
             values_to = "Value")
#I used this to create graphs in one figure

ggplot(df, aes(x=X_1, y=Value, color=Conditionals, linetype = Conditionals))+
         geom_line(aes(color = Conditionals))
```



b) Do X_1 and X_2 appear to be correlated? Justify your answer.

We know that correlation coefficient is equal to

$$\rho = \frac{cov(X_1, X_2)}{\sqrt{var(X_1)var(X_2)}}$$

where

$$cov(X_1, X_2) = E(X_1, X_2) - E(X_1)E(X_2).$$

We calculate marginal distribution with respect to both X_1 and X_2.

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} \frac{1}{2\pi}(1 + x_1^2 + x_2^2)^{-\frac{3}{2}} dx_1 = \frac{1}{(x_2^2 + 1)\pi}$$

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi}(1 + x_1^2 + x_2^2)^{-\frac{3}{2}} dx_2 = \frac{1}{(x_1^2 + 1)\pi}$$

$$E(X_1) = \int_{-\infty}^{+\infty} \frac{1}{(x_1^2 + 1)\pi} x_1 dx_1 = 0$$

$$E(X_2) = \int_{-\infty}^{+\infty} \frac{1}{(x_2^2 + 1)\pi} x_2 dx_2 = 0$$

$$E(X_1, X_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{x_1 x_2}{2\pi(x_1^2 + x_2^2 + 1)} dx_1 dx_2 = 0$$

Since covariance is equal to zero, correlation coefficient will be zero, which concludes that X_1 and X_2 are uncorrelated.

c) Do X_1 and X_2 appear to be independent? Justify your answer.

They are not independent it can be seen from the plot. As the values of X_2 changes, distribution of X_1 changes. We can additionally show that $\frac{1}{2\pi}(1 + x_1^2 + x_2^2)^{-\frac{3}{2}} = f(X_1 = x_1, X_2 = x_2) \neq f_{X_1}(x_1) f_{X_2}(x_2) = \frac{1}{(1+x_1^2)\pi} \cdot \frac{1}{(1+x_2^2)\pi}$ which indicates they are dependent.

Problem 3: For this problem pretend we are dealing with a language with a six-word dictionary

{fun, sun, sit, sat, fan, for}.

An extensive study of literature written in this language reveals that all words are equally likely except that "for" is α times as likely as the other words. Further study reveals that:

i. Each keystroke is an error with probability θ.

ii. All letters are equally likely to produce errors.

iii. Given that a letter is typed incorrectly it is equally likely to be any other letter.

iv. Errors are independent across letters.

For example, the probability of correctly typing "fun" (or any other word) is $(1-\theta)^3$, the probability of typing "pun" or "fon" when intending to type is "fun" is $\theta(1-\theta)^2$, and the probability of typing "foo" or "nnn" when intending to type "fun" is $\theta^2(1-\theta)$. Use Bayes' rule to develop a simple spell checker for this language. For each of the typed words "sun", "the", "foo", give the probability that each word in the dictionary was the intended word. Perform this for the parameters below:

a) $\alpha = 2$ and $\theta = 0.1$
b) $\alpha = 50$ and $\theta = 0.1$
c) $\alpha = 2$ and $\theta = 0.95$

Comment on the changes you observe in these three cases.

a)

```r
likelihood= function (theta,p){
  f=theta^(3-p)*(1-theta)^p
}
alpha=2
theta=0.1
typed=c("sun", "the","foo")
intended=c("fun","sun","sit","sat","fan","for")
p=matrix(0,3,6)
for (i in 1:length(typed)){
  for (j in 1:length(intended)){
     a=unlist(strsplit(typed[i],""))
     b=unlist(strsplit(intended[j],""))
     p[i,j]=sum(!is.na(pmatch(a,b)))
     }
}
data=data.frame(prior=round(ifelse(c("fun","sun","sit","sat","fan","for")=="for",
                             alpha/(5+alpha),1/(5+alpha)),3))
row.names(data) = c("fun","sun","sit","sat","fan","for")

data=data%>%
    mutate(sun_likelihood = round(likelihood(theta,p[1,]),3))%>%
    mutate(sun_posterior=round(sun_likelihood*prior/sum(sun_likelihood*prior),3))%>%
    mutate(the_likelihood = round(likelihood(theta,p[2,]),3))%>%
    mutate(the_posterior =round(the_likelihood*prior/sum(the_likelihood*prior),3))%>%
    mutate(foo_likelihood = round(likelihood(theta,p[3,]),3))%>%
    mutate(foo_posterior= round(foo_likelihood*prior/sum(foo_likelihood*prior),3))

data[,c(1,3,5,7)]
```

```
##     prior sun_posterior the_posterior foo_posterior
## fun 0.143         0.097         0.043         0.049
## sun 0.143         0.869         0.043         0.005
## sit 0.143         0.011         0.391         0.005
## sat 0.143         0.011         0.391         0.005
## fan 0.143         0.011         0.043         0.049
## for 0.286         0.002         0.087         0.885
```

(b)

```r
alpha=50
theta=0.1
data=data.frame(prior=round(ifelse(c("fun","sun","sit","sat","fan","for")=="for",
                             alpha/(5+alpha),1/(5+alpha)),3))
row.names(data) = c("fun","sun","sit","sat","fan","for")
data=data%>%
    mutate(sun_likelihood = round(likelihood(theta,p[1,]),3))%>%
    mutate(sun_posterior=round(sun_likelihood*prior/sum(sun_likelihood*prior),3))%>%
    mutate(the_likelihood = round(likelihood(theta,p[2,]),3))%>%
    mutate(the_posterior =round(the_likelihood*prior/sum(the_likelihood*prior),3))%>%
    mutate(foo_likelihood = round(likelihood(theta,p[3,]),3))%>%
    mutate(foo_posterior= round(foo_likelihood*prior/sum(foo_likelihood*prior),3))
data[,c(1,3,5,7)]
```

```
##     prior sun_posterior the_posterior foo_posterior
## fun 0.018         0.091         0.014         0.002
## sun 0.018         0.821         0.014         0.000
## sit 0.018         0.010         0.126         0.000
## sat 0.018         0.010         0.126         0.000
## fan 0.018         0.010         0.014         0.002
## for 0.909         0.057         0.706         0.995
```

Comparing the result obtained in part (a) and (b), as we increase the value of α, prior values has been decreased except the word "for". Posterior values for the typed word "sun" and intended words are almost the same for both values of α except for the intended word "for" because for part (a) its value much smaller than in part(b). Regarding the posterior values for the typed word "the" , "foo" and intended words, posterior values has been decreased in part (b) except for the intended word "for" because for part (a) its value smaller than in part(b).

(c)

```r
alpha=2
theta=0.95
data=data.frame(prior=round(ifelse(c("fun","sun","sit","sat","fan","for")=="for",
                             alpha/(5+alpha),1/(5+alpha)),3))
row.names(data) = c("fun","sun","sit","sat","fan","for")
data=data%>%
    mutate(sun_likelihood = round(likelihood(theta,p[1,]),3))%>%
    mutate(sun_posterior=round(sun_likelihood*prior/sum(sun_likelihood*prior),3))%>%
    mutate(the_likelihood = round(likelihood(theta,p[2,]),3))%>%
    mutate(the_posterior =round(the_likelihood*prior/sum(the_likelihood*prior),3))%>%
    mutate(foo_likelihood = round(likelihood(theta,p[3,]),3))%>%
    mutate(foo_posterior= round(foo_likelihood*prior/sum(foo_likelihood*prior),3))
data[,c(1,3,5,7)]
```

```
##     prior sun_posterior the_posterior foo_posterior
## fun 0.143         0.001         0.196         0.017
## sun 0.143         0.000         0.196         0.322
## sit 0.143         0.024         0.010         0.322
## sat 0.143         0.024         0.010         0.322
## fan 0.143         0.024         0.196         0.017
## for 0.286         0.926         0.392         0.002
```

When we increase θ, having all letters are different case gives the bigger posterior probability compared to part (a) and (b), whereas having partially match or all letters correct cases give decrease in the posterior probability.

Problem 4: If 70% of a population is vaccinated, and the hospitalization rate is 5 times higher for an unvaccinated person than a vaccinated person, what is the probability that a person is vaccinated given they are hospitalized?

Let V be vaccinated population and H be hospitalization. We are given that $P(V) = 0.7$, $P(H|notV) = 5P(H|V)$. Using Bayes theorem, we get

$$P(V|H) = \frac{P(H|V) \times P(V)}{P(H|V) \times P(V) + P(H|notV) \times P(notV)} = \frac{0.2P(H|notV) \times 0.7}{0.2P(H|notV)0.7 + 0.3P(H|notV)} \approx 0.32$$

or 32%.