

# Homework 2: Bayesian Networks

*Soumya Smruti Mishra*

*March 20, 2016*

```
source("BayesianNetworks-template.r")
library(ggplot2)
riskFactor = read.csv("RiskFactors.csv")
varNames = colnames(riskFactor)
riskCPT = createCPT.fromData(riskFactor, colnames(riskFactor))

i = createCPT.fromData(riskFactor, c("income"))
s.i = createCPT.fromData(riskFactor, c("smoke", "income"))
e.i = createCPT.fromData(riskFactor, c("exercise", "income"))
bmi.e.i = createCPT.fromData(riskFactor, c("bmi", "exercise", "income"))
bp.s.i.e = createCPT.fromData(riskFactor, c("bp", "smoke", "income", "exercise"))
c.s.i.e = createCPT.fromData(riskFactor, c("cholesterol", "smoke", "income", "exercise"))
d.bmi = createCPT.fromData(riskFactor, c("diabetes", "bmi"))
s.bmi.bp.c = createCPT.fromData(riskFactor, c("stroke", "bmi", "bp", "cholesterol"))
attack.bmi.bp.c = createCPT.fromData(riskFactor, c("attack", "bmi", "bp", "cholesterol"))
angina.bmi.bp.c = createCPT.fromData(riskFactor, c("angina", "bmi", "bp", "cholesterol"))

riskNet = list(i, s.i, e.i, bmi.e.i, bp.s.i.e, c.s.i.e, d.bmi, s.bmi.bp.c, attack.bmi.bp.c, angina.bmi.bp.c)
```

1. what is the total number of probabilities needed to store the full joint distribution?

```
nrow(riskCPT)
```

```
## [1] 32768
```

2. For each of the four health outcomes (diabetes, stroke, heart attack, angina), answer the following by querying your network (using your infer function):
  - (a) What is the probability of the outcome if I have bad habits (smoke and don't exercise)? How about if I have good habits (don't smoke and do exercise)?

Diabetes with Bad Habits:-

```
dbh = infer(riskNet, setdiff(varNames, c("diabetes", "smoke", "exercise")), c("smoke", "exercise"), c(1, 0))
print(knitr::kable(dbh, caption = "Table 2adbh - Probabilities of having daibetes when having bad habits"))
```

Table 2adbh - Probabilities of having daibetes when having bad habits.

exercise	smoke	diabetes	probs
2	1	1	0.1557026
2	1	2	0.0087836
2	1	3	0.8156906
2	1	4	0.0198232

Diabetes with Good Habits:-

```
dgh = infer(riskNet, setdiff(varNames, c("diabetes", "smoke", "exercise")), c("smoke", "exercise"), c(2,1))
print(knitr::kable(dgh, caption = "Table 2adgh- Probabilities of having daibetes when having good habits."))
```

Table 2adgh- Probabilities of having daibetes when having good habits.

exercise	smoke	diabetes	probs
1	2	1	0.1317945
1	2	2	0.0086636
1	2	3	0.8417994
1	2	4	0.0177425

Stroke with Bad Habits:-

```
sbh = infer(riskNet, setdiff(varNames, c("stroke", "smoke", "exercise")), c("smoke", "exercise"), c(1,2))
print(knitr::kable(sbh, caption = "Table 2asbh- Probabilities of having stroke when having bad habits."))
```

Table 2asbh- Probabilities of having stroke when having bad habits.

exercise	smoke	stroke	probs
2	1	1	0.050901
2	1	2	0.949099

Stroke with Good Habits:-

```
sgb = infer(riskNet, setdiff(varNames, c("stroke", "smoke", "exercise")), c("smoke", "exercise"), c(2,1))
print(knitr::kable(sgb, caption = "Table 2asgb- Probabilities of having stroke when having good habits."))
```

Table 2asgb- Probabilities of having stroke when having good habits.

exercise	smoke	stroke	probs
1	2	1	0.0371955
1	2	2	0.9628045

Heart Attack with Bad Habits:-

```
hbh = infer(riskNet, setdiff(varNames, c("attack", "smoke", "exercise")), c("smoke", "exercise"), c(1,2))
print(knitr::kable(hbh, caption = "Table 2ahbh- Probabilities of having heart attack when having bad habits."))
```

Table 2ahbh- Probabilities of having heart attack when having bad habits.

exercise	smoke	attack	probs
2	1	1	0.0755818
2	1	2	0.9244182

Heart Attack with Good Habits:-

```
hgh = infer(riskNet, setdiff(varNames, c("attack", "smoke", "exercise")), c("smoke", "exercise"), c(2,1))
print(knitr::kable(hgh, caption = "Table 2ahgh- Probabilities of having heart attack when having good habits."))
```

Table 2ahgh- Probabilities of having heart attack when having good habits.

exercise	smoke	attack	probs
1	2	1	0.0533643
1	2	2	0.9466357

Angina with Bad Habits:-

```
abh = infer(riskNet, setdiff(varNames, c("angina", "smoke", "exercise")), c("smoke", "exercise"), c(1,2))
print(knitr::kable(abh, caption = "Table 2aabh- Probabilities of having Angina when having bad habits."))
```

Table 2aabh- Probabilities of having Angina when having bad habits.

exercise	smoke	angina	probs
2	1	1	0.080087
2	1	2	0.919913

Angina with Good Habits:-

```
agh = infer(riskNet, setdiff(varNames, c("angina", "smoke", "exercise")), c("smoke", "exercise"), c(2,1))
print(knitr::kable(agh, caption = "Table 2agh- Probabilities of having angina when having good habits."))
```

Table 2agh- Probabilities of having angina when having good habits.

exercise	smoke	angina	probs
1	2	1	0.0538606
1	2	2	0.9461394

- (b) What is the probability of the outcome if I have poor health (high blood pressure, high cholesterol, and overweight)? What if I have good health (low blood pressure, low cholesterol, and normal weight)?

Diabetes with poor health:-

```
dph = infer(riskNet, setdiff(varNames, c("diabetes", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"))
print(knitr::kable(dph, caption = "Table 2bdph- Probabilities of having daibetes when having poor health."))
```

Table 2bdph- Probabilities of having daibetes when having poor health.

bmi	bp	cholesterol	diabetes	probs
3	1	1	1	0.1198953
3	1	1	2	0.0074724

bmi	bp	cholesterol	diabetes	probs
3	1	1	3	0.8549105
3	1	1	4	0.0177218

Diabetes with Good Health:-

```
ddgh = infer(riskNet, setdiff(varNames, c("diabetes", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"))
print(knitr::kable(ddgh, caption = "Table 2bddgh- Probabilities of having daibetes when having good health."))
```

Table 2bddgh- Probabilities of having daibetes when having good health.

bmi	bp	cholesterol	diabetes	probs
2	3	2	1	0.0608828
2	3	2	2	0.0093110
2	3	2	3	0.9191494
2	3	2	4	0.0106567

Stroke with poor health:-

```
sph = infer(riskNet, setdiff(varNames, c("stroke", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"))
print(knitr::kable(sph, caption = "Table 2bsph- Probabilities of having stroke when having poor health."))
```

Table 2bsph- Probabilities of having stroke when having poor health.

bmi	bp	cholesterol	stroke	probs
3	1	1	1	0.0837514
3	1	1	2	0.9162486

Stroke with Good Health:-

```
ssgh = infer(riskNet, setdiff(varNames, c("stroke", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"))
print(knitr::kable(ssgh, caption = "Table 2bssgh- Probabilities of having stroke when having good health."))
```

Table 2bssgh- Probabilities of having stroke when having good health.

bmi	bp	cholesterol	stroke	probs
2	3	2	1	0.0137527
2	3	2	2	0.9862473

Heart Attack with Poor health:-

```
hph = infer(riskNet, setdiff(varNames, c("attack", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"))
print(knitr::kable(hph, caption = "Table 2bhph- Probabilities of having heart attack when having poor health."))
```

Table 2bhph- Probabilities of having heart attack when having poor health.

bmi	bp	cholesterol	attack	probs
3	1	1	1	0.1398783
3	1	1	2	0.8601217

Heart Attack with Good Health:-

```
hagh = infer(riskNet, setdiff(varNames, c("attack", "bp", "cholesterol", "bmi")), c("bp", "cholesterol")
print(knitr::kable(hagh, caption = "Table 2hagh- Probabilities of having heart attack when having good health."))
```

Table 2hagh- Probabilities of having heart attack when having good health.

bmi	bp	cholesterol	attack	probs
2	3	2	1	0.0153899
2	3	2	2	0.9846101

Angina with Poor health:-

```
aph = infer(riskNet, setdiff(varNames, c("angina", "bp", "cholesterol", "bmi")), c("bp", "cholesterol")
print(knitr::kable(aph, caption = "Table 2baph- Probabilities of having angina when having poor health."))
```

Table 2baph- Probabilities of having angina when having poor health.

bmi	bp	cholesterol	angina	probs
3	1	1	1	0.1550232
3	1	1	2	0.8449768

Angina with Good Health:-

```
aagh = infer(riskNet, setdiff(varNames, c("angina", "bp", "cholesterol", "bmi")), c("bp", "cholesterol")
print(knitr::kable(aagh, caption = "Table 2baagh- Probabilities of having angina when having good health."))
```

Table 2baagh- Probabilities of having angina when having good health.

bmi	bp	cholesterol	angina	probs
2	3	2	1	0.0128474
2	3	2	2	0.9871526

- Evaluate the effect a person's income has on their probability of having one of the four health outcomes (diabetes, stroke, heart attack, angina). For each of these four outcomes, plot their probability given income status (your horizontal axis should be  $i = 1, 2, \dots, 8$ , and your vertical axis should be  $P(y = 1 \mid \text{income} = i)$ , where  $y$  is the outcome). What can you conclude?

```

diabetes <- integer()
stroke <- integer()
attack <- integer()
angina <- integer()
for(i in 1:8){
  db <- infer(riskNet, setdiff(varNames, c("diabetes", "income")), c("income"), c(i))
  diabetes <- c(diabetes, db[db$diabetes == 1, ]$probs)

  str <- infer(riskNet, setdiff(varNames, c("stroke", "income")), c("income"), c(i))
  stroke <- c(stroke, str[str$stroke == 1, ]$probs)

  att <- infer(riskNet, setdiff(varNames, c("attack", "income")), c("income"), c(i))
  attack <- c(attack, att[att$attack == 1, ]$probs)

  ang <- infer(riskNet, setdiff(varNames, c("angina", "income")), c("income"), c(i))
  angina <- c(angina, ang[ang$angina == 1, ]$probs)
}

```

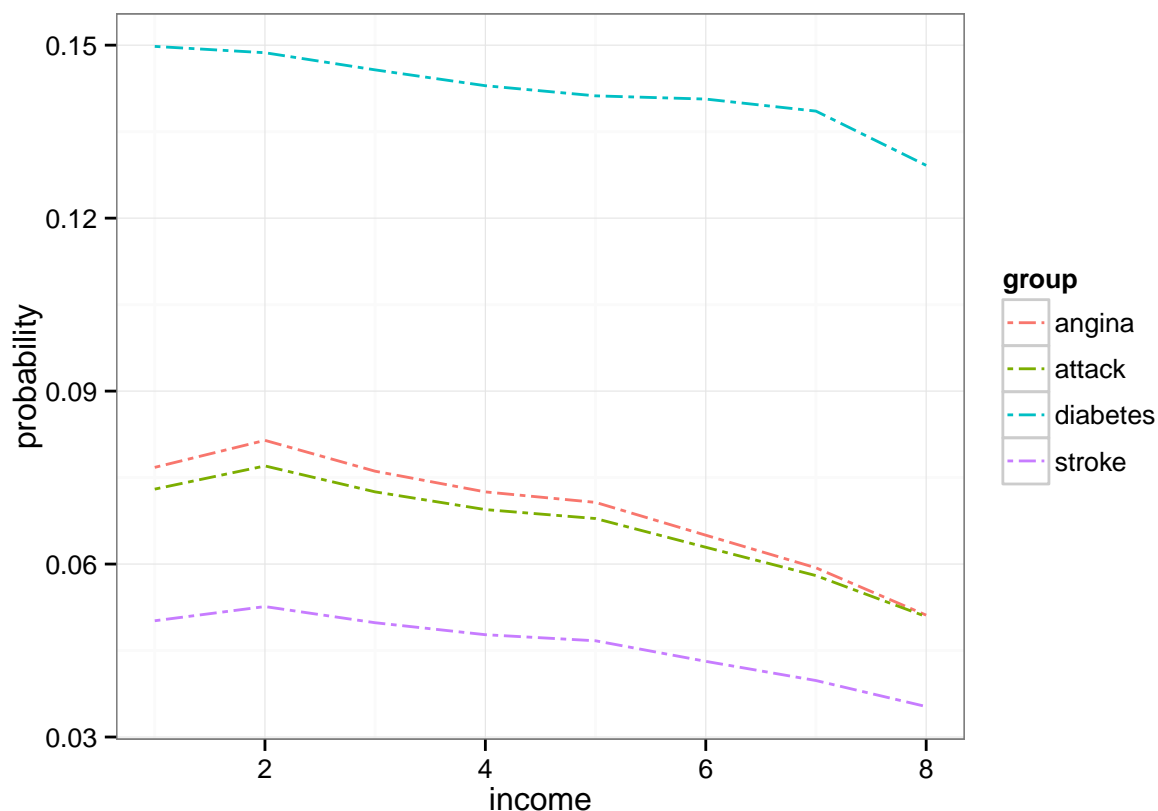
```

dat <- list(diabetes=diabetes, stroke=stroke, attack=attack, angina=angina)
dat <- lapply(dat, function(income) cbind(income = seq_along(income), probability = income))

list.names <- names(dat)
lns <- sapply(dat, nrow)
dat <- as.data.frame(do.call("rbind", dat))
dat$group <- rep(list.names, lns)

ggplot(dat, aes(x = income, y = probability, colour = group)) +
  theme_bw() +
  geom_line(linetype = "twodash")

```



Inference from the Graph: From the above graph we can conclude that with increase in income the probability of having any of the four health outcomes (diabetes, stroke, heart attack, angina) decreases.

For diabetes with increase in income from 1 to 8 the probability of having diabetes decreases from 15% to 13%, the absolute decrease is around 2% approximately.

For stroke with increase in income from 1 to 8 the probability of having stroke decreases from 5.02% to 3.53%, the absolute decrease is 1.49%.

For heart attack with increase in income from 1 to 8 the probability of having heart attack decreases from 7.3% to 5.1%, the absolute decrease is 2.2%.

For Angina with increase in income from 1 to 8 the probability of having angina decreases from 7.68% to 5.11%, the absolute decrease is 2.57%.

4. Notice there are no links in the graph between the habits (smoking and exercise) and the outcomes. What assumption is this making about the effects of smoking and exercise on health problems? Let's test the validity of these assumptions. Create a second Bayesian network as above, but add edges from smoking to each of the four outcomes and edges from exercise to each of the four outcomes. Now redo the queries in Question 2. What was the effect, and do you think the assumptions of the first graph were valid or not?

```
i <- createCPT.fromData(riskFactor,"income")
e.i <- createCPT.fromData(riskFactor,c("exercise","income"))
b.i.e <- createCPT.fromData(riskFactor,c("bmi","exercise","income"))
b.e.i.s <- createCPT.fromData(riskFactor,c("bp","exercise","income","smoke"))
s.i <- createCPT.fromData(riskFactor,c("smoke","income"))
c.e.i.s <- createCPT.fromData(riskFactor,c("cholesterol","exercise","income","smoke"))
```

```
d.b.s.e <- createCPT.fromData(riskFactor,c("diabetes","bmi","smoke","exercise"))
s.b.b.c.s.e <- createCPT.fromData(riskFactor,c("stroke","bmi","bp","cholesterol","smoke","exercise"))
at.b.b.c.s.e <- createCPT.fromData(riskFactor,c("attack","bmi","bp","cholesterol","smoke","exercise"))
an.b.b.c.s.e <- createCPT.fromData(riskFactor,c("angina","bmi","bp","cholesterol","smoke","exercise"))
riskNetNew = list(i, e.i, b.i.e, b.e.i.s, s.i, c.e.i.s, d.b.s.e, s.b.b.c.s.e, at.b.b.c.s.e, an.b.b.c.s.e)
```

- (a) What is the probability of the outcome if I have bad habits (smoke and don't exercise)? How about if I have good habits (don't smoke and do exercise)?

Diabetes with Bad Habits:-

```
debh = infer(riskNetNew, setdiff(varNames, c("diabetes", "smoke", "exercise")), c("smoke", "exercise"),
print(knitr::kable(debh, caption = "Table 4adebh- Probabilities of having diabetes when having bad habits"))
```

Table 4adebh- Probabilities of having diabetes when having bad habits after adding extra edges from smoking and exercise to the outcomes.

smoke	exercise	diabetes	probs
1	2	1	0.2191865
1	2	2	0.0066634
1	2	3	0.7511638
1	2	4	0.0229863

Diabetes with Good Habits:-

```
degh = infer(riskNetNew, setdiff(varNames, c("diabetes", "smoke", "exercise")), c("smoke", "exercise"),
print(knitr::kable(degh, caption = "Table 4degh- Probabilities of having diabetes when having good habits"))
```

Table 4degh- Probabilities of having diabetes when having good habits after adding extra edges from smoking and exercise to the outcomes.

smoke	exercise	diabetes	probs
2	1	1	0.1004388
2	1	2	0.0093148
2	1	3	0.8750927
2	1	4	0.0151536

Comparing these two above table and tables 2adbh and 2adgh we can conclude that, on adding edges from smoking and exercise to diabetes, the probability of having diabetes with bad habits increases, and probability of having diabetes with good habits decreases. Hence our assumption is not valid.

Stroke with Bad Habits:-

```
sebh = infer(riskNetNew, setdiff(varNames, c("stroke", "smoke", "exercise")), c("smoke", "exercise"), c
print(knitr::kable(sebh, caption = "Table 4asebh - Probabilities of having stroke when having bad habits"))
```



Table 4asebh - Probabilities of having stroke when having bad habits after adding extra edges from smoking and exercise to the outcomes.

smoke	exercise	stroke	probs
1	2	1	0.0801954
1	2	2	0.9198046

Stroke with Good Habits:-

```
segh = infer(riskNetNew, setdiff(varNames, c("stroke", "smoke", "exercise")), c("smoke", "exercise"), c
print(knitr::kable(segh, caption = "Table 4asegh - Probabilities of having stroke when having good habi
```

Table 4asegh - Probabilities of having stroke when having good habits after adding extra edges from smoking and exercise to the outcomes.

smoke	exercise	stroke	probs
2	1	1	0.0255604
2	1	2	0.9744396

Comparing these two above table and tables 2asbh and 2asgh we can conclude that, on adding edges from smoke and exercise to stroke, the probability of having stroke increases given bad habits and decreases given good habits. Hence our assumption was not valid.

Heart Attack with Bad Habits:-

```
hebh = infer(riskNetNew, setdiff(varNames, c("attack", "smoke", "exercise")), c("smoke", "exercise"), c
print(knitr::kable(hebh, caption = "Table 4ahebh - Probabilities of having heart attack when having bad
```

Table 4ahebh - Probabilities of having heart attack when having bad habits after adding extra edges from smoking and exercise to the outcomes.

smoke	exercise	attack	probs
1	2	1	0.1205514
1	2	2	0.8794486

Heart Attack with Good Habits:-

```
hegh = infer(riskNetNew, setdiff(varNames, c("attack", "smoke", "exercise")), c("smoke", "exercise"), c
print(knitr::kable(hegh, caption = "Table 4ahegh - Probabilities of having heart attack when having goo
```

Table 4ahegh - Probabilities of having heart attack when having good habits after adding extra edges from smoking and exercise to the outcomes.

smoke	exercise	attack	probs
2	1	1	0.0318827
2	1	2	0.9681173

Comparing these two above table and tables 2ahbh and 2ahgh we can conclude that, on adding edges from smoke and exercise to attack, the probability of having attack increases given bad habits and decreases given good habits. Hence our assumption was not valid.

Angina with Bad Habits:-

```
aebh = infer(riskNetNew, setdiff(varNames, c("angina", "smoke", "exercise")), c("smoke", "exercise"), c
print(knitr::kable(aebh, caption = "Table 4aaebh- Probabilities of having angina when having bad habits"))
```

Table 4aaebh- Probabilities of having angina when having bad habits after adding extra edges from smoking and exercise to the outcomes.

smoke	exercise	angina	probs
1	2	1	0.1170839
1	2	2	0.8829161

Angina with Good Habits:-

```
aegh = infer(riskNetNew, setdiff(varNames, c("angina", "smoke", "exercise")), c("smoke", "exercise"), c
print(knitr::kable(aegh, caption = "Table 4aaegh- Probabilities of having angina when having good habits"))
```

Table 4aaegh- Probabilities of having angina when having good habits after adding extra edges from smoking and exercise to the outcomes.

smoke	exercise	angina	probs
2	1	1	0.0357139
2	1	2	0.9642861

Comparing these two above table and tables 2aabh and 2aagh we can conclude that, on adding edges from smoke and exercise to angina, the probability of having angina increases given bad habits and decreases given good habits. Hence our assumption was not valid.

- (b) What is the probability of the outcome if I have poor health (high blood pressure, high cholesterol, and overweight)? What if I have good health (low blood pressure, low cholesterol, and normal weight)?

Diabetes with poor health:-

```
ddeph = infer(riskNetNew, setdiff(varNames, c("diabetes", "bp", "cholesterol", "bmi")), c("bp", "choles
print(knitr::kable(ddeph, caption = "Table 4bddeph - Probabilities of having diabetes when having poor health"))
```

Table 4bddeph - Probabilities of having diabetes when having poor health after adding extra edges from smoking and exercise to the outcomes.

bmi	bp	cholesterol	diabetes	probs
3	1	1	1	0.1288173
3	1	1	2	0.0074469
3	1	1	3	0.8452901
3	1	1	4	0.0184457

Diabetes with Good Health:-

```
ddegh = infer(riskNetNew, setdiff(varNames, c("diabetes", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"))
print(knitr::kable(ddegh, caption = "Table 4bddegh- Probabilities of having diabetes when having good health after adding extra edges from smoking and exercise to the outcomes.")
```

Table 4bddegh- Probabilities of having diabetes when having good health after adding extra edges from smoking and exercise to the outcomes.

bmi	bp	cholesterol	diabetes	probs
2	3	2	1	0.0570344
2	3	2	2	0.0094863
2	3	2	3	0.9233090
2	3	2	4	0.0101703

Comparing these two above table and tables 2bdph and 2bddgh we can conclude that, on adding edges from smoking and exercise to diabetes, the probability of having diabetes with poor health increases, and probability of having diabetes with good health decreases. Hence our assumption is not valid.

Stroke with poor health:-

```
steph = infer(riskNetNew, setdiff(varNames, c("stroke", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"))
print(knitr::kable(steph, caption = "Table 4bsteph- Probabilities of having stroke when having poor health after adding extra edges from smoking and exercise to the outcomes.")
```

Table 4bsteph- Probabilities of having stroke when having poor health after adding extra edges from smoking and exercise to the outcomes.

bmi	bp	cholesterol	stroke	probs
3	1	1	1	0.0854038
3	1	1	2	0.9145962

Stroke with Good Health:-

```
stegh = infer(riskNetNew, setdiff(varNames, c("stroke", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"))
print(knitr::kable(stegh, caption = "Table 4bstegh- Probabilities of having stroke when having good health after adding extra edges from smoking and exercise to the outcomes.")
```

Table 4bstegh- Probabilities of having stroke when having good health after adding extra edges from smoking and exercise to the outcomes.

bmi	bp	cholesterol	stroke	probs
2	3	2	1	0.0132257
2	3	2	2	0.9867743

Comparing these two above table and tables 2bsph and 2bssgh we can conclude that, on adding edges from smoking and exercise to stroke, the resultant probabilities are almost same. So our assumptions are valid in this case.

Heart Attack with Poor health:-

```
haeph = infer(riskNetNew, setdiff(varNames, c("attack", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"),
print(knitr::kable(haeph, caption = "Table 4bhaeph - Probabilities of having heart attack when having poor health after adding extra edges from smoking and exercise to the outcomes."))
```

Table 4bhaeph - Probabilities of having heart attack when having poor health after adding extra edges from smoking and exercise to the outcomes.

bmi	bp	cholesterol	attack	probs
3	1	1	1	0.1415756
3	1	1	2	0.8584244

Heart Attack with Good Health:-

```
haegh = infer(riskNetNew, setdiff(varNames, c("attack", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"),
print(knitr::kable(haegh, caption = "Table 4bhaegh - Probabilities of having heart attack when having good health after adding extra edges from smoking and exercise to the outcomes."))
```

Table 4bhaegh - Probabilities of having heart attack when having good health after adding extra edges from smoking and exercise to the outcomes.

bmi	bp	cholesterol	attack	probs
2	3	2	1	0.0148278
2	3	2	2	0.9851722

Comparing these two above table and tables 2bhph and 2bhagh we can conclude that, on adding edges from smoking and exercise to heart attack, the resultant probabilities are almost same. Hence our assumption is almost valid.

Angina with Poor health:-

```
aneph = infer(riskNetNew, setdiff(varNames, c("angina", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"),
print(knitr::kable(aneph, caption = "Table 4baneph- Probabilities of having angina when having poor health after adding extra edges from smoking and exercise to the outcomes."))
```

Table 4baneph- Probabilities of having angina when having poor health after adding extra edges from smoking and exercise to the outcomes.

bmi	bp	cholesterol	angina	probs
3	1	1	1	0.1567437
3	1	1	2	0.8432563

Angina with Good Health:-

```
anegh = infer(riskNetNew, setdiff(varNames, c("angina", "bp", "cholesterol", "bmi")), c("bp", "cholesterol", "bmi"),
print(knitr::kable(anegh, caption = "Table 4banegh- Probabilities of having angina when having good health after adding extra edges from smoking and exercise to the outcomes."))
```

Table 4baneph- Probabilities of having angina when having good health after adding extra edges from smoking and exercise to the outcomes.

bmi	bp	cholesterol	angina	probs
2	3	2	1	0.0125751
2	3	2	2	0.9874249

Comparing these two above table and tables 2baph and 2baagh we can conclude that, on adding edges from smoking and exercise to angina, the resultant probabilities are same. Hence our assumption is valid.

- Also notice there are no edges between the four outcomes. What assumption is this making about the interactions between health problems? Make a third network, starting from the network in Question 4, but adding an edge from diabetes to stroke. For both networks, evaluate the following probabilities:  $P(\text{stroke} = 1 \mid \text{diabetes} = 1)$  and  $P(\text{stroke} = 1 \mid \text{diabetes} = 3)$  Again, what was the effect, and was the assumption about the interaction between diabetes and stroke valid?

```
s.b.b.c.s.e.d <- createCPT.fromData(riskFactor,c("stroke","bmi","bp","cholesterol","smoke","exercise"),
riskNet5 = list(i, e.i, b.i.e, b.e.i.s, s.i, c.e.i.s, d.b.s.e, s.b.b.c.s.e.d, at.b.b.c.s.e, an.b.b.c.s.e)
ds1o <-infer(riskNetNew, setdiff(varNames, c("stroke", "diabetes")), c("diabetes"), c(3))
ds3o <- infer(riskNetNew, setdiff(varNames, c("stroke", "diabetes")), c("diabetes"), c(1))

ds3 <- infer(riskNet5, setdiff(varNames, c("stroke", "diabetes")), c("diabetes"), c(3))
ds1 <- infer(riskNet5, setdiff(varNames, c("stroke", "diabetes")), c("diabetes"), c(1))
```

```
print(knitr::kable(ds3o, caption = "Table - Probabilities before adding edge between diabetes and stroke"))
```

Table - Probabilities before adding edge between diabetes and stroke. When diabetes is 1, i.e, no diabetes

diabetes	stroke	probs
1	1	0.0455432
1	2	0.9544568

```
print(knitr::kable(ds1o, caption = "Table - Probabilities before adding edge between diabetes and stroke"))
```

Table - Probabilities before adding edge between diabetes and stroke. When diabetes is 3, i.e, has diabetes

diabetes	stroke	probs
3	1	0.0418514
3	2	0.9581486

```
print(knitr::kable(ds1, caption = "Table - Probabilities after adding edge between diabetes and stroke"))
```

Table - Probabilities after adding edge between diabetes and stroke. When diabetes is 1, i.e, has diabetes

diabetes	stroke	probs
1	1	0.0769877

diabetes	stroke	probs
1	2	0.9230123

```
print(knitr::kable(ds3, caption = "Table - Probabilities after adding edge between diabetes and stroke"))
```

Table - Probabilities after adding edge between diabetes and stroke. When diabetes is 3, i.e, no diabetes

diabetes	stroke	probs
3	1	0.0361551
3	2	0.9638449

From the above tables we can conclude that, on considering diabetes as one of the causes of stroke in Bayesian network the probability of having stroke given diabetes increases. There is 40% increase in relative probability. The absolute difference in probabilities is small, but relatively they are high. Hence here our assumption was invalid.