init

```
library (faraway)
library(reshape)
```

data load

```
data(femsmoke)

femdata = femsmoke
```

joint dist:

```
v <- c("smoke.yes", "smoke.no")
u <- c("dead.yes", "dead.no")
w <- c("age.1", "age.2", "age.3", "age.4", "age.5", "age.6", "age.7")

mat.femdata <- cast(femdata, smoker+dead ~ age, sum, value = 'y')
mat.dead.y <- mat.femdata[c(1,3), c(1, 3:9)]
mat.dead.n <- mat.femdata[c(2,4), c(1, 3:9)]
```

array

```
rownames(mat.dead.y)<-rownames(mat.dead.y)<-v
mat.dead.y <- mat.dead.y[,c(2:8)]
rownames(mat.dead.n)<-rownames(mat.dead.n)<-v
mat.dead.n <- mat.dead.n[,c(2:8)]
```

reshape

```
femdata_array<-array(rep(NA,28),
                     dim=c(2,7,2),
                     dimnames=list(paste("smoke",c("yes", "no"),
                                         sep="."),
                                   paste("age",1:7,sep="."),
```

```
                                    paste("dead",c("yes", "no"),sep=".")))

femdata_array[,,1] <- as.vector(unlist(mat.dead.y))
femdata_array[,,2] <- as.vector(unlist(mat.dead.n))
femdata_array
```

```
## , , dead.yes
##
##           age.1 age.2 age.3 age.4 age.5 age.6 age.7
## smoke.yes     2     3    14    27    51    29    13
## smoke.no      1     5     7    12    40   101    64
##
## , , dead.no
##
##           age.1 age.2 age.3 age.4 age.5 age.6 age.7
## smoke.yes    53   121    95   103    64     7     0
## smoke.no     61   152   114    66    81    28     0
```

```
N<-sum(femdata_array)
femdata.joint.p<-femdata_array/N

head(femdata.joint.p)
```

```
## [1] 0.001522070 0.000761035 0.002283105 0.003805175 0.010654490 0.005327245
```

```
# check on sum
sum(femdata.joint.p)
```

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

marginal distribution

marginal distribution for u_dead

```
(uMarginal.dead <- apply(femdata.joint.p, 3, sum))
```

```
##  dead.yes   dead.no
## 0.2808219 0.7191781
```

```
# check
sum(uMarginal.dead)
```

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

marginal distribution for v_smoke

```
(vMarginal.smoke <- apply(femdata.joint.p, 1, sum))
```

```
## smoke.yes  smoke.no
## 0.4429224 0.5570776
```

```
#check
sum(vMarginal.smoke)
```

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

marginal distribution for w_age.

```
(wMarginal.age <- apply(femdata.joint.p, 2, sum))
```

```
##     age.1     age.2     age.3     age.4     age.5     age.6     age.7
## 0.0890411 0.2138508 0.1750381 0.1582953 0.1796043 0.1255708 0.0585997
```

```
sum(wMarginal.age)
```

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

Create conditional distribution

p(w,v|u="alive")=p(smoke,age|alive).

```
(cond.v.w.given.uAlive <- femdata.joint.p[,,"dead.no"]/uMarginal.dead["dead.no"])
```

```
##               age.1     age.2     age.3      age.4      age.5       age.6
## smoke.yes 0.05608466 0.1280423 0.1005291 0.10899471 0.06772487 0.007407407
## smoke.no  0.06455026 0.1608466 0.1206349 0.06984127 0.08571429 0.029629630
##           age.7
## smoke.yes     0
## smoke.no      0
```

```
sum(cond.v.w.given.uAlive)
```

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

p(v|u="alive")=p(smoke|alive)

```
(cond.v.given.uAlive <- apply(femdata.joint.p[,,"dead.no"], 1, sum)/uMarginal.dead["dead.no"])
```

```
## smoke.yes  smoke.no
## 0.4687831 0.5312169
```

```
sum(cond.v.given.uAlive)
```

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

p(w|u="alive",v="smoker")=p(age|alive,smoke)

```
(cond.w.given.uAlive.vSmoke <- femdata.joint.p[,,"dead.no"]["smoke.yes",]/sum(femdata.joint.p[,,"dead.no"]["smoke.yes",]))
```

```
##      age.1      age.2      age.3      age.4      age.5      age.6
## 0.11963883 0.27313770 0.21444695 0.23250564 0.14446953 0.01580135
##      age.7
## 0.00000000
```

```
sum(cond.w.given.uAlive.vSmoke)
```

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

Compare the vectors p(w|v2,u1)p(v2|u1)p(u1) and p(w,v,u)[,v2,u1]

```
rbind(uMarginal.dead["dead.no"]*cond.v.given.uAlive["smoke.yes"]*cond.w.given.uAlive.vSmoke,
      femdata.joint.p["smoke.yes",,"dead.no"])
```

```
##           age.1      age.2      age.3      age.4      age.5       age.6
## [1,] 0.04033486 0.09208524 0.07229833 0.07838661 0.04870624 0.005327245
## [2,] 0.04033486 0.09208524 0.07229833 0.07838661 0.04870624 0.005327245
##      age.7
## [1,]     0
## [2,]     0
```

Let the marginal distribution for age group be p(w) estimated marginal distribution from the sample:

Given simulated age group, simulate variable v using conditional distribution p(v|w),i.e. using probabilities p(smoke.yes|age), p(smoke.no|age).

given simulated variables for age and for smoke, simulate mortality variable using distribution p(dead|v,w), p(alive|v,w).

Using the described procedure simulate outcomes for 100 participants. Use seed set.seed(284) for comparison.

Simulate age group

```
set.seed(284)
sim.age <- c(rep(NA,100))
sim.age <- sapply(sim.age, function(z) sample(w,1,replace=T, prob = wMarginal.age))
```

Simulate variable v using conditional distribution p(v|w): p(smoke.yes|age), p(smoke.no|age).

```
set.seed(284)
sim.v <- sapply(sim.age, function(z) sample(c("smoke.yes", "smoke.no"),1, replace=T,
                                            prob = vMarginal.smoke) )
```

Given simulated variables for age and smoke, simulate mortality

```
set.seed(284)
sim.w <- apply(cbind(sim.age, sim.v),1,
               function(z) sample(c("dead.yes", "dead.no"), 1, replace=T,
                                                      prob=femdata.joint.p[z[2], z[1], ]))
```

```
simulatedData <- cbind(sim.age, sim.v, sim.w)
rownames(simulatedData) <- 1:100
head(simulatedData, 25)
```

```
##    sim.age sim.v       sim.w
## 1  "age.3" "smoke.no"  "dead.no"
## 2  "age.6" "smoke.yes" "dead.yes"
## 3  "age.3" "smoke.no"  "dead.no"
## 4  "age.3" "smoke.no"  "dead.no"
## 5  "age.4" "smoke.yes" "dead.no"
## 6  "age.2" "smoke.no"  "dead.no"
## 7  "age.7" "smoke.yes" "dead.yes"
## 8  "age.5" "smoke.no"  "dead.no"
## 9  "age.5" "smoke.no"  "dead.no"
## 10 "age.5" "smoke.no"  "dead.no"
## 11 "age.5" "smoke.no"  "dead.no"
```

```
## 12 "age.3" "smoke.no"  "dead.no"
## 13 "age.2" "smoke.no"  "dead.no"
## 14 "age.7" "smoke.yes" "dead.yes"
## 15 "age.3" "smoke.no"  "dead.no"
## 16 "age.1" "smoke.yes" "dead.no"
## 17 "age.6" "smoke.yes" "dead.no"
## 18 "age.2" "smoke.no"  "dead.no"
## 19 "age.5" "smoke.no"  "dead.no"
## 20 "age.2" "smoke.no"  "dead.no"
## 21 "age.4" "smoke.yes" "dead.no"
## 22 "age.7" "smoke.yes" "dead.yes"
## 23 "age.3" "smoke.no"  "dead.no"
## 24 "age.3" "smoke.no"  "dead.no"
## 25 "age.1" "smoke.yes" "dead.no"
```