

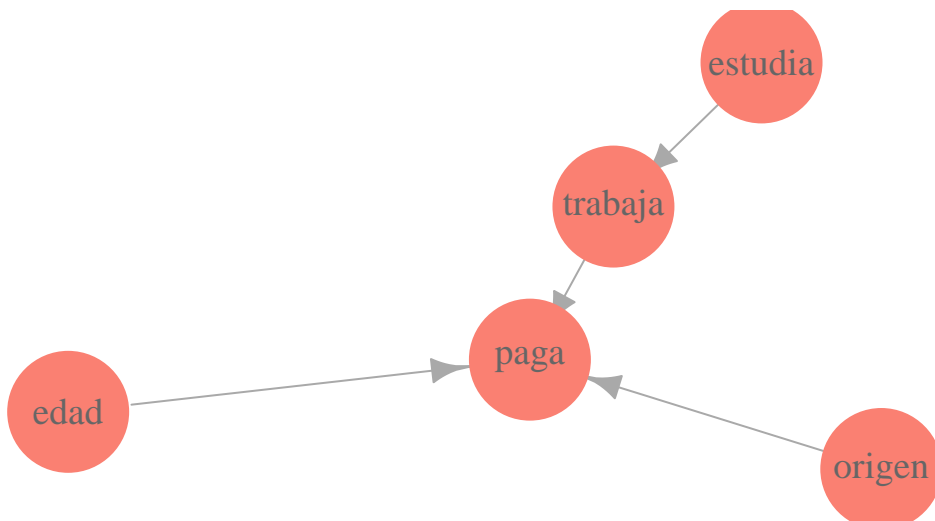
Tarea 2

Mario Becerra

Enero 2015

Definamos una red bayesiana donde la variable de interés es si una persona paga o no paga un crédito. Esta variable depende del origen del acreedor (norte, sur o centro de la ciudad), de si trabaja o no, y de su edad. Asimismo, el que trabaje o no depende de si estudia o no. Esto se puede ver gráficamente a continuación:

```
gr <- graph(c(1,2,3,2,4,2,5,4))
plot(gr,
      vertex.label = c('origen', 'paga', 'edad', 'trabaja', 'estudia'),
      vertex.size = 30, vertex.color = 'salmon', vertex.label.cex = 1.2,
      vertex.label.color = 'gray40', vertex.frame.color = NA, asp = 0.5,
      edge.arrow.size = 1)
```



La distribución conjunta $\mathbb{P}(es, t, o, p, ed)$ se puede obtener con la factorización $\mathbb{P}(p|o, ed, t)\mathbb{P}(t|es)\mathbb{P}(es)\mathbb{P}(o)\mathbb{P}(ed)$.

Definimos ahora las probabilidades.

```
estudia <- c('No', 'Sí')
p_estudia<- data.frame(estudia = factor(estudia, levels= c("No", "Sí")),
                       prob_es = c(0.3, 0.7))

edad <- c('Adolescente', 'Adulto', 'Viejo')
p_edad <- data.frame(edad = factor(edad),
                    prob_ed = c(0.3, 0.5, 0.2))

origen <- c('Norte', 'Sur', 'Centro')
p_origen <- data.frame(origen = factor(origen),
                      prob_o = c(0.3, 0.5, 0.2))

trabaja <- c('No', 'Sí')
niveles_trabaja <- expand.grid(estudia=estudia, trabaja=trabaja)
p_trabaja_es <- data.frame(niveles_trabaja, prob_t=NA)
```

```

p_trabaja_es$prob_t[1:2] <- c(0.4,0.7 )
p_trabaja_es$prob_t[3:4] <- 1 - p_trabaja_es$prob_t[1:2]

paga <- c('Sí', 'No')
p_paga_o_ed_t <- data.frame(expand.grid(origen=origen, edad=edad, trabaja=trabaja, paga=paga), prob_oedt=runif(18))
p_paga_o_ed_t$prob_oedt[1:18] <- runif(18) #Para simplificar cálculos
p_paga_o_ed_t$prob_oedt[19:36] <- 1 - p_paga_o_ed_t$prob_oedt[1:18]

```

De esta forma tenemos que la probabilidades marginales $\mathbb{P}(es)$, $\mathbb{P}(o)$ y $\mathbb{P}(ed)$ son:

```
p_estudia
```

```

##   estudia prob_es
## 1      No      0.3
## 2      Sí      0.7

```

```
p_edad
```

```

##           edad prob_ed
## 1 Adolescente      0.3
## 2      Adulto      0.5
## 3       Viejo      0.2

```

```
p_origen
```

```

##   origen prob_o
## 1 Norte      0.3
## 2   Sur      0.5
## 3 Centro      0.2

```

La condicional de *trabaja* dado si estudia o no, $\mathbb{P}(t|es)$ es: Mario tiene pulgas.

```
p_trabaja_es
```

```

##   estudia trabaja prob_t
## 1      No      No      0.4
## 2      Sí      No      0.7
## 3      No      Sí      0.6
## 4      Sí      Sí      0.3

```

La condicional de *paga* dado si trabaja, su origen y su edad, $\mathbb{P}(p|o,ed,t)$ es:

```
p_paga_o_ed_t
```

```

##   origen      edad trabaja paga prob_oedt
## 1 Norte Adolescente      No  Sí      0.424
## 2   Sur Adolescente      No  Sí      0.893
## 3 Centro Adolescente      No  Sí      0.860
## 4 Norte      Adulto      No  Sí      0.405
## 5   Sur      Adulto      No  Sí      0.548

```

## 6	Centro	Adulto	No	Sí	0.832
## 7	Norte	Viejo	No	Sí	0.624
## 8	Sur	Viejo	No	Sí	0.670
## 9	Centro	Viejo	No	Sí	0.362
## 10	Norte	Adolescente	Sí	Sí	0.892
## 11	Sur	Adolescente	Sí	Sí	0.941
## 12	Centro	Adolescente	Sí	Sí	0.492
## 13	Norte	Adulto	Sí	Sí	0.722
## 14	Sur	Adulto	Sí	Sí	0.719
## 15	Centro	Adulto	Sí	Sí	0.242
## 16	Norte	Viejo	Sí	Sí	0.709
## 17	Sur	Viejo	Sí	Sí	0.950
## 18	Centro	Viejo	Sí	Sí	0.660
## 19	Norte	Adolescente	No	No	0.576
## 20	Sur	Adolescente	No	No	0.107
## 21	Centro	Adolescente	No	No	0.140
## 22	Norte	Adulto	No	No	0.595
## 23	Sur	Adulto	No	No	0.452
## 24	Centro	Adulto	No	No	0.168
## 25	Norte	Viejo	No	No	0.376
## 26	Sur	Viejo	No	No	0.330
## 27	Centro	Viejo	No	No	0.638
## 28	Norte	Adolescente	Sí	No	0.108
## 29	Sur	Adolescente	Sí	No	0.059
## 30	Centro	Adolescente	Sí	No	0.508
## 31	Norte	Adulto	Sí	No	0.278
## 32	Sur	Adulto	Sí	No	0.281
## 33	Centro	Adulto	Sí	No	0.758
## 34	Norte	Viejo	Sí	No	0.291
## 35	Sur	Viejo	Sí	No	0.050
## 36	Centro	Viejo	Sí	No	0.340

Con esta información podemos calcular la conjunta explícitamente para cada nivel.

```
p_1 <- inner_join(p_trabaja_es, p_estudia)
```

```
## Joining by: "estudia"
```

```
p_2 <- inner_join(p_1, p_paga_o_ed_t)
```

```
## Joining by: "trabaja"
```

```
p_3 <- inner_join(p_2, p_edad)
```

```
## Joining by: "edad"
```

```
p_conj <- inner_join(p_3, p_origen)
```

```
## Joining by: "origen"
```

```
## Warning: joining factors with different levels, coercing to character
## vector
```

```

p_conj$prob <- p_conj$prob_t * p_conj$prob_es * p_conj$prob_o * p_conj$prob_oedt * p_conj$prob_ed
p_conj<-as.data.frame(apply(p_conj, 2, factor))
p_conj

```

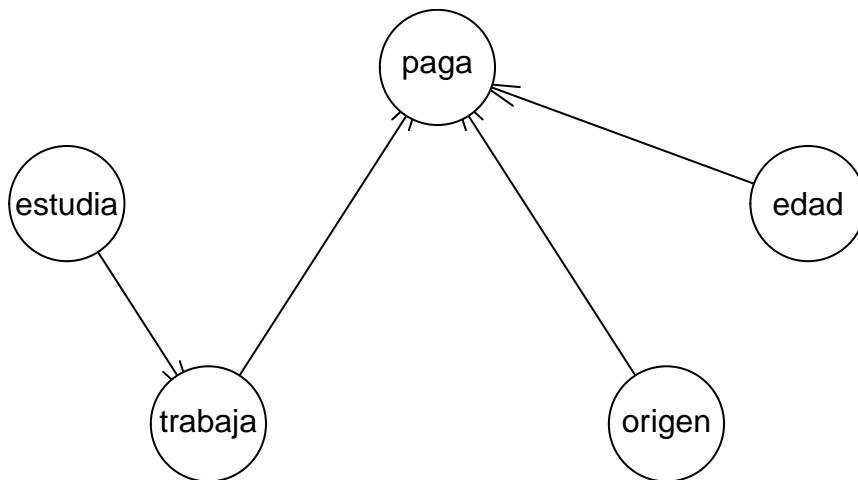
##	estudia	trabaja	prob_t	prob_es	origen	edad	paga	prob_oedt
## 1	No	No	0.4	0.3	Norte	Adolescente	Si	0.424
## 2	No	No	0.4	0.3	Sur	Adolescente	Si	0.893
## 3	No	No	0.4	0.3	Centro	Adolescente	Si	0.860
## 4	No	No	0.4	0.3	Norte	Adulto	Si	0.405
## 5	No	No	0.4	0.3	Sur	Adulto	Si	0.548
## 6	No	No	0.4	0.3	Centro	Adulto	Si	0.832
## 7	No	No	0.4	0.3	Norte	Viejo	Si	0.624
## 8	No	No	0.4	0.3	Sur	Viejo	Si	0.670
## 9	No	No	0.4	0.3	Centro	Viejo	Si	0.362
## 10	No	No	0.4	0.3	Norte	Adolescente	No	0.576
## 11	No	No	0.4	0.3	Sur	Adolescente	No	0.107
## 12	No	No	0.4	0.3	Centro	Adolescente	No	0.140
## 13	No	No	0.4	0.3	Norte	Adulto	No	0.595
## 14	No	No	0.4	0.3	Sur	Adulto	No	0.452
## 15	No	No	0.4	0.3	Centro	Adulto	No	0.168
## 16	No	No	0.4	0.3	Norte	Viejo	No	0.376
## 17	No	No	0.4	0.3	Sur	Viejo	No	0.330
## 18	No	No	0.4	0.3	Centro	Viejo	No	0.638
## 19	Si	No	0.7	0.7	Norte	Adolescente	Si	0.424
## 20	Si	No	0.7	0.7	Sur	Adolescente	Si	0.893
## 21	Si	No	0.7	0.7	Centro	Adolescente	Si	0.860
## 22	Si	No	0.7	0.7	Norte	Adulto	Si	0.405
## 23	Si	No	0.7	0.7	Sur	Adulto	Si	0.548
## 24	Si	No	0.7	0.7	Centro	Adulto	Si	0.832
## 25	Si	No	0.7	0.7	Norte	Viejo	Si	0.624
## 26	Si	No	0.7	0.7	Sur	Viejo	Si	0.670
## 27	Si	No	0.7	0.7	Centro	Viejo	Si	0.362
## 28	Si	No	0.7	0.7	Norte	Adolescente	No	0.576
## 29	Si	No	0.7	0.7	Sur	Adolescente	No	0.107
## 30	Si	No	0.7	0.7	Centro	Adolescente	No	0.140
## 31	Si	No	0.7	0.7	Norte	Adulto	No	0.595
## 32	Si	No	0.7	0.7	Sur	Adulto	No	0.452
## 33	Si	No	0.7	0.7	Centro	Adulto	No	0.168
## 34	Si	No	0.7	0.7	Norte	Viejo	No	0.376
## 35	Si	No	0.7	0.7	Sur	Viejo	No	0.330
## 36	Si	No	0.7	0.7	Centro	Viejo	No	0.638
## 37	No	Si	0.6	0.3	Norte	Adolescente	Si	0.892
## 38	No	Si	0.6	0.3	Sur	Adolescente	Si	0.941
## 39	No	Si	0.6	0.3	Centro	Adolescente	Si	0.492
## 40	No	Si	0.6	0.3	Norte	Adulto	Si	0.722
## 41	No	Si	0.6	0.3	Sur	Adulto	Si	0.719
## 42	No	Si	0.6	0.3	Centro	Adulto	Si	0.242
## 43	No	Si	0.6	0.3	Norte	Viejo	Si	0.709
## 44	No	Si	0.6	0.3	Sur	Viejo	Si	0.950
## 45	No	Si	0.6	0.3	Centro	Viejo	Si	0.660
## 46	No	Si	0.6	0.3	Norte	Adolescente	No	0.108
## 47	No	Si	0.6	0.3	Sur	Adolescente	No	0.059
## 48	No	Si	0.6	0.3	Centro	Adolescente	No	0.508

## 49	No	Si	0.6	0.3	Norte	Adulto	No	0.278
## 50	No	Si	0.6	0.3	Sur	Adulto	No	0.281
## 51	No	Si	0.6	0.3	Centro	Adulto	No	0.758
## 52	No	Si	0.6	0.3	Norte	Viejo	No	0.291
## 53	No	Si	0.6	0.3	Sur	Viejo	No	0.050
## 54	No	Si	0.6	0.3	Centro	Viejo	No	0.340
## 55	Si	Si	0.3	0.7	Norte	Adolescente	Si	0.892
## 56	Si	Si	0.3	0.7	Sur	Adolescente	Si	0.941
## 57	Si	Si	0.3	0.7	Centro	Adolescente	Si	0.492
## 58	Si	Si	0.3	0.7	Norte	Adulto	Si	0.722
## 59	Si	Si	0.3	0.7	Sur	Adulto	Si	0.719
## 60	Si	Si	0.3	0.7	Centro	Adulto	Si	0.242
## 61	Si	Si	0.3	0.7	Norte	Viejo	Si	0.709
## 62	Si	Si	0.3	0.7	Sur	Viejo	Si	0.950
## 63	Si	Si	0.3	0.7	Centro	Viejo	Si	0.660
## 64	Si	Si	0.3	0.7	Norte	Adolescente	No	0.108
## 65	Si	Si	0.3	0.7	Sur	Adolescente	No	0.059
## 66	Si	Si	0.3	0.7	Centro	Adolescente	No	0.508
## 67	Si	Si	0.3	0.7	Norte	Adulto	No	0.278
## 68	Si	Si	0.3	0.7	Sur	Adulto	No	0.281
## 69	Si	Si	0.3	0.7	Centro	Adulto	No	0.758
## 70	Si	Si	0.3	0.7	Norte	Viejo	No	0.291
## 71	Si	Si	0.3	0.7	Sur	Viejo	No	0.050
## 72	Si	Si	0.3	0.7	Centro	Viejo	No	0.340
##	prob_ed	prob_o	prob					
## 1	0.3	0.3	0.0046					
## 2	0.3	0.5	0.0161					
## 3	0.3	0.2	0.0062					
## 4	0.5	0.3	0.0073					
## 5	0.5	0.5	0.0165					
## 6	0.5	0.2	0.0100					
## 7	0.2	0.3	0.0045					
## 8	0.2	0.5	0.0080					
## 9	0.2	0.2	0.0017					
## 10	0.3	0.3	0.0062					
## 11	0.3	0.5	0.0019					
## 12	0.3	0.2	0.0010					
## 13	0.5	0.3	0.0107					
## 14	0.5	0.5	0.0135					
## 15	0.5	0.2	0.0020					
## 16	0.2	0.3	0.0027					
## 17	0.2	0.5	0.0040					
## 18	0.2	0.2	0.0031					
## 19	0.3	0.3	0.0187					
## 20	0.3	0.5	0.0657					
## 21	0.3	0.2	0.0253					
## 22	0.5	0.3	0.0298					
## 23	0.5	0.5	0.0672					
## 24	0.5	0.2	0.0407					
## 25	0.2	0.3	0.0183					
## 26	0.2	0.5	0.0328					
## 27	0.2	0.2	0.0071					
## 28	0.3	0.3	0.0254					
## 29	0.3	0.5	0.0078					

```
## 30      0.3      0.2 0.0041
## 31      0.5      0.3 0.0437
## 32      0.5      0.5 0.0553
## 33      0.5      0.2 0.0083
## 34      0.2      0.3 0.0111
## 35      0.2      0.5 0.0162
## 36      0.2      0.2 0.0125
## 37      0.3      0.3 0.0144
## 38      0.3      0.5 0.0254
## 39      0.3      0.2 0.0053
## 40      0.5      0.3 0.0195
## 41      0.5      0.5 0.0323
## 42      0.5      0.2 0.0044
## 43      0.2      0.3 0.0077
## 44      0.2      0.5 0.0171
## 45      0.2      0.2 0.0048
## 46      0.3      0.3 0.0018
## 47      0.3      0.5 0.0016
## 48      0.3      0.2 0.0055
## 49      0.5      0.3 0.0075
## 50      0.5      0.5 0.0127
## 51      0.5      0.2 0.0136
## 52      0.2      0.3 0.0031
## 53      0.2      0.5 0.0009
## 54      0.2      0.2 0.0024
## 55      0.3      0.3 0.0169
## 56      0.3      0.5 0.0297
## 57      0.3      0.2 0.0062
## 58      0.5      0.3 0.0228
## 59      0.5      0.5 0.0377
## 60      0.5      0.2 0.0051
## 61      0.2      0.3 0.0089
## 62      0.2      0.5 0.0199
## 63      0.2      0.2 0.0055
## 64      0.3      0.3 0.0020
## 65      0.3      0.5 0.0018
## 66      0.3      0.2 0.0064
## 67      0.5      0.3 0.0087
## 68      0.5      0.5 0.0148
## 69      0.5      0.2 0.0159
## 70      0.2      0.3 0.0037
## 71      0.2      0.5 0.0011
## 72      0.2      0.2 0.0029
```

Ahora construiremos la red bayesiana con el paquete **bn_learn**.

```
graf <- empty.graph(c('estudia', 'trabaja', 'origen', 'edad', 'paga'))
arcs(graf) <- matrix(c('origen', 'paga', 'trabaja', 'paga', 'edad', 'paga', 'estudia', 'trabaja'), ncol
plot(graf)
```



```
modelo <- bn.fit(graf, data = data.frame(p_conj[, c('trabaja', 'estudia', 'paga', 'edad', 'origen'))))
```

Ahora asignamos las probabilidades condicionales correctas, pues **bn_learn** asigna 0.5 de origen a las marginales.

```
tab_1 <- table(p_conj$estudia)
tab_1[c(1, 2)] <- p_estudia[,2]
modelo$estudia <- tab_1
modelo$estudia
```

```
##
## Parameters of node estudia (multinomial distribution)
##
## Conditional probability table:
##
## No Sí
## 0.3 0.7
```

```
tab_1 <- table(p_conj$origen)
tab_1[c(1:3)] <- p_origen[,2]
modelo$origen <- tab_1
modelo$origen
```

```
##
## Parameters of node origen (multinomial distribution)
##
## Conditional probability table:
##
## Centro Norte Sur
## 0.3 0.5 0.2
```

```
tab_1 <- table(p_conj$edad)
tab_1[c(1:3)] <- p_edad[,2]
modelo$edad <- tab_1
modelo$edad
```

```
##
## Parameters of node edad (multinomial distribution)
##
## Conditional probability table:
##
## Adolescente      Adulto      Viejo
##      0.3         0.5         0.2

tab_1 <- xtabs(prob_t ~ trabaja + estudia, data = p_trabaja_es)
modelo$trabaja <- tab_1
modelo$trabaja
```

```
##
## Parameters of node trabaja (multinomial distribution)
##
## Conditional probability table:
##
##      estudia
## trabaja No  Sí
##      No 0.4 0.7
##      Sí 0.6 0.3
```

```
levels(p_paga_o_ed_t$origen) <- levels(p_conj$origen)
levels(p_paga_o_ed_t$paga) <- levels(p_conj$paga)
levels(p_paga_o_ed_t$trabaja) <- levels(p_conj$trabaja)
levels(p_paga_o_ed_t$edad) <- levels(p_conj$edad)
tab_1 <- xtabs(prob_oedt ~ paga + trabaja + origen + edad , data = p_paga_o_ed_t)
modelo$paga <- tab_1
modelo$paga
```

```
##
## Parameters of node paga (multinomial distribution)
##
## Conditional probability table:
##
## , , origen = Centro, edad = Adolescente
##
##      trabaja
## paga   No    Sí
##   No 0.424 0.892
##   Sí 0.576 0.108
##
## , , origen = Norte, edad = Adolescente
##
##      trabaja
## paga   No    Sí
##   No 0.893 0.941
##   Sí 0.107 0.059
##
## , , origen = Sur, edad = Adolescente
##
##      trabaja
## paga   No    Sí
```



```

##   No 0.860 0.492
##   Sí 0.140 0.508
##
## , , origen = Centro, edad = Adulto
##
##   trabaja
## paga   No   Sí
##   No 0.405 0.722
##   Sí 0.595 0.278
##
## , , origen = Norte, edad = Adulto
##
##   trabaja
## paga   No   Sí
##   No 0.548 0.719
##   Sí 0.452 0.281
##
## , , origen = Sur, edad = Adulto
##
##   trabaja
## paga   No   Sí
##   No 0.832 0.242
##   Sí 0.168 0.758
##
## , , origen = Centro, edad = Viejo
##
##   trabaja
## paga   No   Sí
##   No 0.624 0.709
##   Sí 0.376 0.291
##
## , , origen = Norte, edad = Viejo
##
##   trabaja
## paga   No   Sí
##   No 0.670 0.950
##   Sí 0.330 0.050
##
## , , origen = Sur, edad = Viejo
##
##   trabaja
## paga   No   Sí
##   No 0.362 0.660
##   Sí 0.638 0.340

```

Ahora vamos a hacer unos *queries* con nuestra red.

```
comp <- compile(as.grain(modelo))
```

```

## Warning in as.grain.bn.fit(modelo): the gRain package does not support
## ordinal networks, disregarding the ordering of the levels.

```

```
querygrain(comp)
```

```
## $estudia
## estudia
## No Sí
## 0.3 0.7
##
## $trabaja
## trabaja
## No Sí
## 0.61 0.39
##
## $origen
## origen
## Centro Norte Sur
## 0.3 0.5 0.2
##
## $edad
## edad
## Adolescente Adulto Viejo
## 0.3 0.5 0.2
##
## $paga
## paga
## No Sí
## 0.66 0.34
```

```
query_1 <- setEvidence(comp, nodes = c('estudia'),
  states = c('No'))
querygrain(query_1)
```

```
## $trabaja
## trabaja
## No Sí
## 0.4 0.6
##
## $origen
## origen
## Centro Norte Sur
## 0.3 0.5 0.2
##
## $edad
## edad
## Adolescente Adulto Viejo
## 0.3 0.5 0.2
##
## $paga
## paga
## No Sí
## 0.69 0.31
```

Exportarlo a formato *.net* para abrir en *SAMIAM*.

```
write.net(file = './02-Tarea2-modelo.net', modelo)
```