Hard evidence and virtual/likelihood evidence with gRain

Søren Højsgaard
Aalborg University, Denmark

`gRain` version 1.2-1 as of 2013-08-08

Contents

1 Introduction

The `gRain` package implements propagation in [gra]phical [i]ndependence [n]etworks (hereafter abbreviated `grain`). Such networks are also known as probabilistic networks and Bayesian networks.

To cite `gRain` in publications, please use:

> Søren Højsgaard (2012). Graphical Independence Networks with the gRain Package for R. Journal of Statistical Software, 46(10), 1-26. `http://www.jstatsoft.org/v46/i10/`.

More information about the package, other graphical modelling packages and development versions is available from

> `http://people.math.aau.dk/~sorenh/software/gR`

This document describes how to work with virtual evidence (also known as likelihood evidence) in `gRain`. This is done via the function `setEvidence()`. The `setEvidence()`

function is an extension of the function `setFinding()` (but with a slightly different syntax). Users of `gRain` are recommended to use `setEvidence()` instead of `setFinding()` in the future.

2 An excerpt of the chest clinic network

Consider the following excerpt of the chest clinic network which is described in the paper mentioned above.[1] (We admit that a better example for illustrating the various type of evidence would be desirable.)

```
> yn <- c("yes","no")
> a    <- cptable(~asia, values=c(1,99),levels=yn)
> t.a  <- cptable(~tub|asia, values=c(5,95,1,99),levels=yn)
> ( plist1 <- compileCPT( list( a, t.a ) ) )

CPTspec with probabilities:
 P( asia )
 P( tub | asia )

> plist1[[1]]

asia
 yes   no
0.01 0.99

> plist1[[2]]

     asia
tub    yes   no
  yes 0.05 0.01
  no  0.95 0.99

> ( chest1 <- grain(plist1) )

Independence network: Compiled: FALSE Propagated: FALSE
  Nodes: chr [1:2] "asia" "tub"

> querygrain( chest1 )

$asia
asia
 yes   no
0.01 0.99


$tub
tub
   yes      no
0.0104 0.9896
```

---

[1]Think of a better example.

2.1 Specifying hard evidence

Suppose we want to make a diagnosis about tuberculosis given the evidence that a person has recently been to Asia. The functions `setFinding()` (which has been in `gRain` for years) and `setEvidence()` (which is a recent addition to `gRain`) can both be used for this purpose. The following forms are equivalent.

```
> setFinding(  chest1, nodes="asia", states="yes")
```

```
Independence network: Compiled: TRUE Propagated: TRUE
  Nodes: chr [1:2] "asia" "tub"
  Findings: chr "asia"
```

```
> setEvidence( chest1, nodes="asia", states="yes")
```

```
Independence network: Compiled: TRUE Propagated: TRUE
  Nodes: chr [1:2] "asia" "tub"
  Findings: chr "asia"
```

```
> setEvidence( chest1, nslist=list(asia="yes"))
```

```
Independence network: Compiled: TRUE Propagated: TRUE
  Nodes: chr [1:2] "asia" "tub"
  Findings: chr "asia"
```

```
> querygrain( setEvidence( chest1, nslist=list(asia="yes")) )
```

```
$tub
tub
 yes   no
0.05 0.95
```

2.2 What is virtual evidence (also called likelihood evidence) ?

Suppose we do not know with certainty whether a patient has recently been to Asia (perhaps the patient is too ill to tell). However the patient (if he/she is Caucasian) may be unusually tanned and this lends support to the hypothesis of a recent visit to Asia.

To accommodate we create an extended network with an extra node for which we enter evidence. However, it is NOT necessary to do so in practice, because we may equivalently enter the virtual evidence in the original network.

We can then introduce a new variable `guess.asia` with `asia` as its only parent.

```
> g.a <- parray(c("guess.asia", "asia"), levels=list(yn, yn),
+               values=c(.8,.2, .1,.9))
```

```
          asia
guess.asia yes  no
       yes 0.8 0.1
       no  0.2 0.9
```

This reflects the assumption that for patients who have recently been to Asia we would guess so in 80% of the cases, whereas for patients who have not recently been to A we would (erroneously) guess that they have recently been to Asia in 10% of the cases.

```
> ( plist2 <- compileCPT( list( a, t.a, g.a ) ) )
```

```
CPTspec with probabilities:
 P( asia )
 P( tub | asia )
 P( guess.asia | asia )
```

```
> ( chest2 <- grain(plist2) )
```

```
Independence network: Compiled: FALSE Propagated: FALSE
  Nodes: chr [1:3] "asia" "tub" "guess.asia"
```

```
> querygrain( chest2 )
```

```
$asia
asia
 yes   no
0.01 0.99
```

```
$tub
tub
    yes      no
0.0104 0.9896
```

```
$guess.asia
guess.asia
  yes    no
0.107 0.893
```

Now specify the guess or judgment, that the person has recently been to Asia:

```
> querygrain( setEvidence( chest2, nslist=list(guess.asia="yes")) )
```

```
$asia
asia
       yes          no
0.07476636 0.92523364
```

```
$tub
tub
       yes          no
0.01299065 0.98700935
```

2.3 Specifying virtual evidence

The same guess or judgment can be specified as virtual evidence (also called likelihood evidence) for the original network:

```
> querygrain( setEvidence( chest1, nslist=list(asia=c(.8, .1))) )
```

```
$tub
tub
       yes         no
0.01299065 0.98700935
```

This also means that specifying that specifying asia='yes' can be done as

```
> querygrain( setEvidence( chest1, nslist=list(asia=c(1, 0))) )
```

```
$tub
tub
 yes    no
0.05 0.95
```