Sensitivity and robustness analysis in Bayesian networks with the bnmonitor R package

Manuele Leonelli

School of Human Sciences and Technology, IE University

09/03/2022

- ▶ Describe the package and an illustrative dataset
- ▶ Bayesian networks (BNs)
- ▶ Model validation of BNs and its implementation
- ▶ Discussion

bnmonitor

- ▶ The package can be downloaded from CRAN

```r
# install.packages("bnmonitor")
library(bnmonitor)
```

- ▶ It has its own GitHub repository with a user guide at https://github.com/manueleleonelli/bnmonitor
- ▶ A preprint available at ArXiv illustrates its use at https://arxiv.org/pdf/2107.11785.pdf
- ▶ A more comprehensive webpage is under development
- ▶ Package jointly created with Ramsiya Ramanathan and Rachel L. Wilkerson

SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

An illustrative dataset

We consider the `diabetes` dataset comprising information about 392 Pima-Indian women of at least 21 years of age.

- ▶ `PREG`: number of times pregnant (low/high);
- ▶ `GLUC`: plasma glucose concentration (low/high);
- ▶ `PRES`: diastolic blood pressure (low/high);
- ▶ `TRIC`: triceps skin fold thickness (low/high);
- ▶ `INS`: 2-hour serum insulin (low/high);
- ▶ `MASS`: body mass index (low/high);
- ▶ `PED`: diabetes pedigree function (low/high);
- ▶ `AGE`: age (low/high);
- ▶ `DIAB`: test for diabetes (neg/pos)

SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

Bayesian Networks

```
data(diabetes)
dag <- bnlearn::hc(diabetes)
qgraph::qgraph(dag)
```
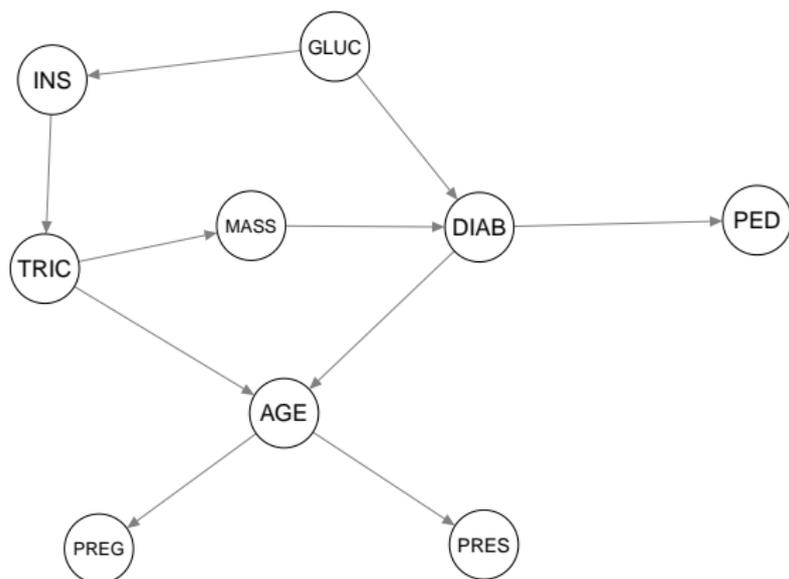
- ▶ Broad array of methods to check the compatibility between data and an assumed model.
- ▶ This is routinely done in many areas of statistics!!
- ▶ For BNs we implement methods based on the *prequential approach*: the fit of a model is based on the quality of the predictions it sequentially makes.

SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

Node and Global Monitors

```
global_monitor(dag = dag, df = diabetes)
##   Vertex    Score
## 1   PREG 236.2658
## 2   GLUC 274.3482
## 3   PRES 250.0871
## 4   TRIC 267.1841
## 5    INS 219.8782
## 6   MASS 231.8470
## 7    PED 272.6041
## 8    AGE 246.5046
## 9   DIAB 214.0108
```

SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

Sequential Node Monitors

```r
p1 <- plot(seq_marg_monitor(dag, diabetes, "DIAB"))
p2 <- plot(seq_marg_monitor(dag, diabetes, "PED"))
p3 <- plot(seq_cond_monitor(dag, diabetes, "DIAB"))
p4 <- plot(seq_cond_monitor(dag, diabetes, "PED"))
gridExtra::grid.arrange(p1,p2,p3,p4,ncol=2)
```
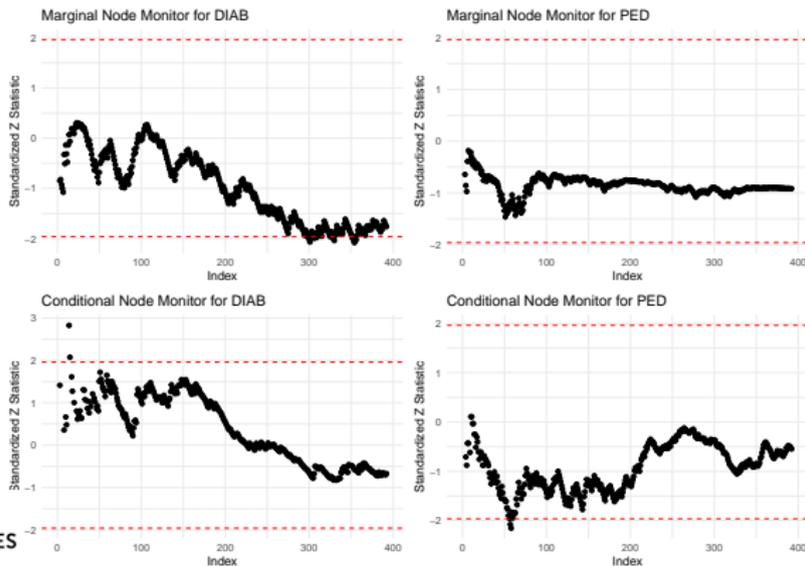
Sequential Parent-Child Monitors

```
p1 <- plot(seq_pa_ch_monitor(dag, diabetes, "PED",
       pa.names = c("DIAB") , pa.val = c("neg")))
p2 <- plot(seq_pa_ch_monitor(dag, diabetes, "PED",
       pa.names = c("DIAB") , pa.val = c("pos")))
gridExtra::grid.arrange(p1,p2,ncol=2)
```

Model Validation - Sensitivity Analysis
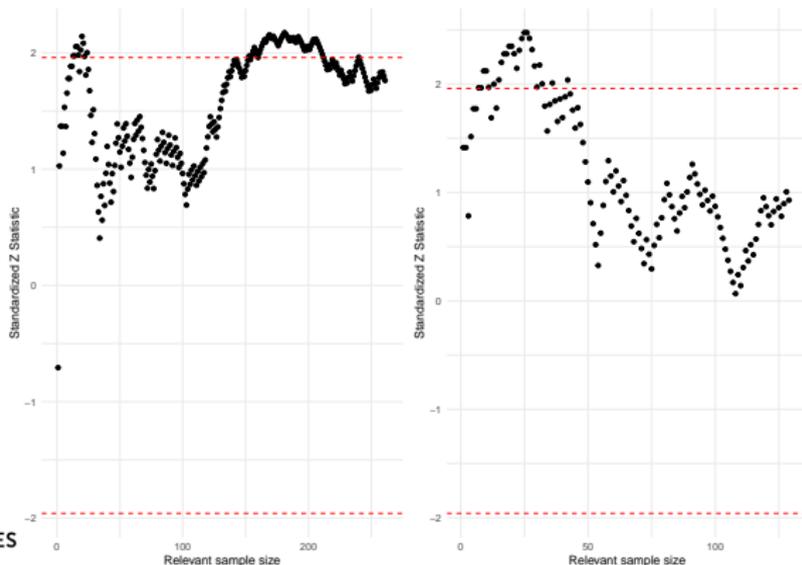
- ▶ Study of the effects of the parameters of a model to its outputs.
- ▶ All methods implemented are *local*: one parameter is changes while others are kept fixed.
- ▶ This theory has been mostly developed by computer scientists (there is no data involved)
- ▶ Changing more than one parameter at the time becomes challenging. . .

Sensitivity Functions

```
bn <- bnlearn::bn.fit(dag, diabetes)
sens <- sensitivity(bn, interest_node = "DIAB",
        interest_node_value = "pos",
        evidence_nodes = "INS",
        evidence_states = "low",
        node = "GLUC", value_node = "high",
        value_parents = NULL, new_value = "all")
plot(sens)
```
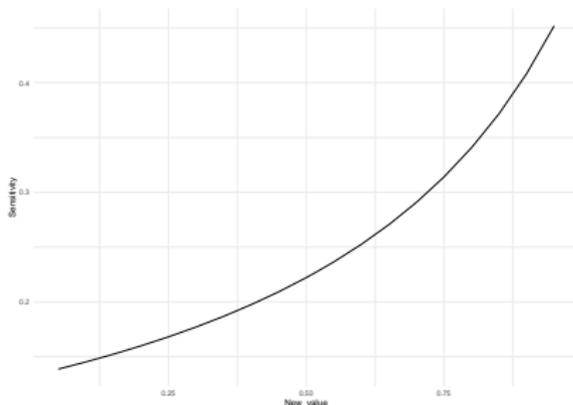
CD Distance

```
cd <- CD(bn, node = "GLUC", value_node = "high",
        value_parents = NULL, new_value = "all")
plot(cd)
```


CD distance

SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY
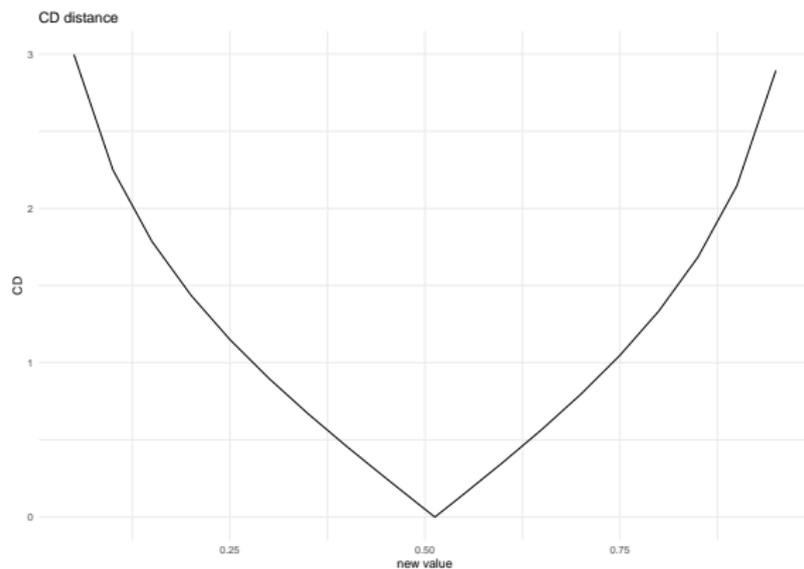
Changing Parameters

```r
bnlearn::cpquery(bn, DIAB == "pos", PRES == "high")
## [1] 0.3738819
```

```r
out <- sensquery(bn, interest_node = "DIAB",
        interest_node_value = "pos",
        new_value = 0.4, evidence_nodes = "PRES",
        evidence_states = "high")
knitr::kable(out)
```

Node	Val_node	Val_par	Orig_value	New_value	CD
GLUC	low		0.49	0.44	0.19
PRES	low	low	0.61	0.70	0.40
DIAB	pos	low,low	0.06	0.13	0.87
MASS	high	low	0.26	0.48	0.97
AGE	low	low,neg	0.70	0.93	1.67

Discussion

- `bnmonitor` supports the validation of a BN model using a variety of techniques
- In this talk I only considered binary discrete BNs
- There are implemented methods also for Gaussian BNs
- For discrete variables that take more than two values, `bnmonitor` implements methods for *covariation*
- Newer model validation methods for BNs are currently being investigated (e.g. Sobol indexes), which are planned to be soon implemented in the package

SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY