Introduction to Decision Trees in R

3 vessel coronary artery disease (CAD) example

Petros Pechlivanoglou

Wednesday, September 13, 2023

R software

The R Project for Statistical Computing: https://www.r-project.org/

Installing R

Install R

Download R version 4.3.1 from https://cran.r-project.org/bin/windows/base/old/4.3.1/

Download the file `R-4.3.1-win.exe` and follow the installation procedure.

Install RStudio

Download and install the free version of RStudio Desktop from: https://posit.co/download/rstudio-desktop/#download

Building a Tree

This R Markdown document provides a foundation for constructing a decision tree model for CAD:

Scenario - 70 yr old man with 3 vessel coronary artery disease (CAD).

The two strategies being compared are:

1. Medical therapy
2. Surgery

```r
# Unsuccessful medicine, die from CAD (probability) = 0.3
# Life expectancy after unsuccessful medicine on CAD (life years) = 5
# Life expectancy after successful medicine on CAD (life years) = 10
Medicine <- 0.3 * 5 + (1 - 0.3) * 10
Medicine
```

```
## [1] 8.5
```

```r
# Perioperative death (probability) = 0.05
# Life expectancy after perioperative death (life years) = 0
# Unsuccessful surgery, die from CAD (probability) = 0.2
# Life expectancy after unsuccessful surgery on CAD (life years) = 5
# Life expectancy after successful surgery on CAD (life years) = 10
Surgery  <- 0.05 * 0 + (1 - 0.05) * (0.2 * 5 + (1 - 0.2) * 10)
Surgery
```

```
## [1] 8.55
```

This is an example of a tree built using NUMERIC values for probabilities and "payoffs". The tree can be directly calculated to yield an expected value - surgery is slightly preferred with an expected value of 8.55 life years, compared with 8.50 life years for medicine.

Limited options, though, for sensitivity analysis.

Building a Tree with Variables

Much greater flexibility is possible if numeric quantities in the tree are defined as variables or expressions.

Variables can be globally defined and live in the Global Environment in R and applied to the whole tree.

Steps for using variables:

1. Declare name — Based on its intended function in your model, decide on a clear name for the variable.
2. Use — Anywhere the corresponding value is used in the tree (e.g., payoffs or probabilities).

Simplest way to create a variable is to type it in a R chunk like below:

```
# Life expectancy (life years)
LE_death_CAD        <- 5    # Life expectancy after unsuccessful surgery on CAD
LE_death_ORSurgery  <- 0    # Life expectancy after perioperative death
LE_death_otherCauses <- 10  # Life expectancy after successful surgery on CAD

# Probabilities
p_death_CADMedicine <- 0.3  # Probability of death from CAD due to unsuccessful medicine
p_death_CADSurgery  <- 0.2  # Probability of death from CAD due to unsuccessful surgery
p_death_ORSurgery   <- 0.05 # Probability of perioperative death

# Utility (of CAD)
u_CADMedicine       <- 0.5  # Utility after CAD under Medicine
u_CADSurgery        <- 0.8  # Utility after CAD under Surgery
```

Our previous tree, now with variables:

```
Medicine <-     p_death_CADMedicine  * LE_death_CAD +
          (1 - p_death_CADMedicine) * LE_death_otherCauses
Medicine
```

```
## [1] 8.5
```

```
Surgery  <-     p_death_ORSurgery * LE_death_ORSurgery +
          (1 - p_death_ORSurgery) * (     p_death_CADSurgery  * LE_death_CAD +
                                    (1 - p_death_CADSurgery) * LE_death_otherCauses
                                    )
Surgery
```

```
## [1] 8.55
```

Now, among other things, sensitivity analysis is possible.

```
v_p_death_ORSurgery <- c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)
# v_p_death_ORSurgery <- seq(0, 1, 0.1) # simpler
Medicine_sens <-     p_death_CADMedicine  * LE_death_CAD +
                (1 - p_death_CADMedicine) * LE_death_otherCauses

Surgery_sens  <-     v_p_death_ORSurgery  * LE_death_ORSurgery +
                (1 - v_p_death_ORSurgery) * (     p_death_CADSurgery  * LE_death_CAD +
                                            (1 - p_death_CADSurgery) * LE_death_otherCauses
                                            )
plot(v_p_death_ORSurgery, Surgery_sens, type="l", main="Sensitivity Analysis",
     xlab="probability of perioperative death", ylab="Expected value", col="red")
points(v_p_death_ORSurgery, Surgery_sens, col="red")
lines(v_p_death_ORSurgery, rep(Medicine_sens, 11), col="blue")
points(v_p_death_ORSurgery, rep(Medicine_sens, 11), col="blue")
legend("right", c("Medicine", "Surgery"), col=c("blue", "red"), lty=c(1,1), bty="n")
```

Sensitivity Analysis
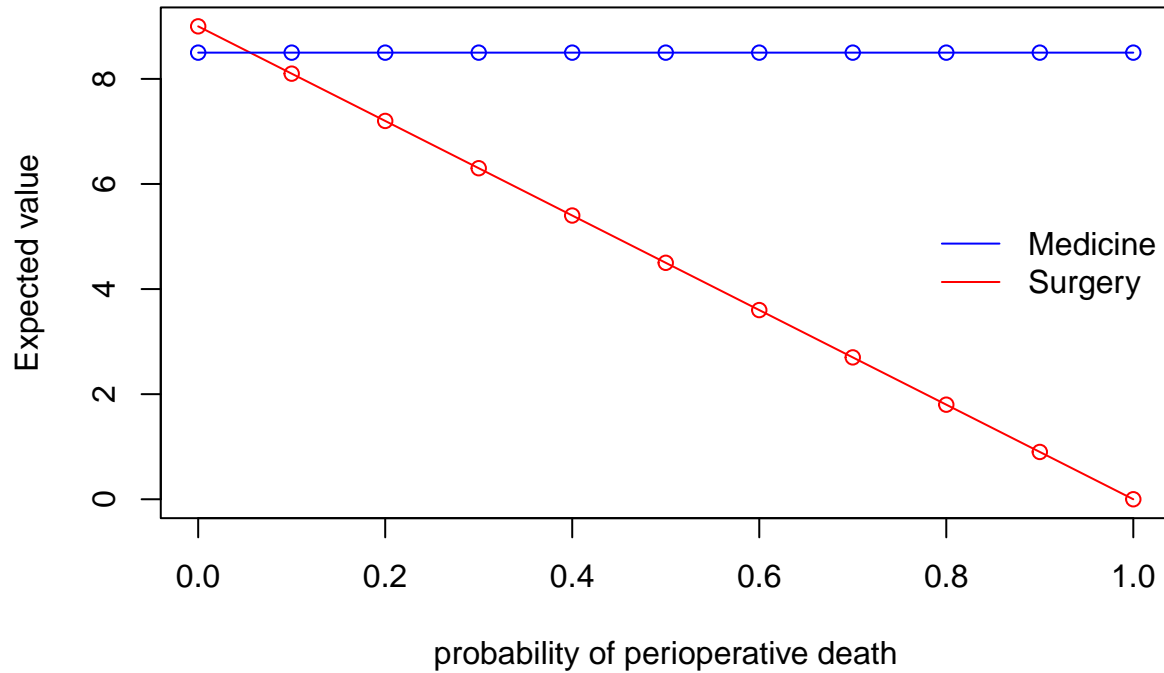
Using Subtrees

In R you can create subtrees and assign them to functions with desired names. For example, the subtree describing outcomes for those alive post OR can be assigned into a function and we can call it `Effectiveness_Subtree`. The arguments of this function include probability of death if treatment is unsuccessful, life expectancy if treatment is unsuccessful and successful, and utility after CAD. We assume utilities for those with successful treatment to have a utility of 1.

```r
# p_death_Tx: probability of death after unsuccessful treatment
# LE_death_Tx: life expectancy after unsuccessful treatment
# LE_death_noTx: life expectancy after successful treatment
# u_CAD: utility after CAD
Effectiveness_Subtree <- function(p_death_Tx, LE_death_Tx, LE_death_noTx, u_CAD) {

                    p_death_Tx  * LE_death_Tx   * u_CAD +
              (1 - p_death_Tx) * LE_death_noTx * 1
}
```

To use this subtree function, you just need to pass values to the arguments. Below we calculate QALY outcomes for both medicine and surgery while using subtree `Effectiveness_Subtree` for both strategies.

```r
Medicine <- Effectiveness_Subtree(p_death_Tx    = p_death_CADMedicine,
                                  LE_death_Tx   = LE_death_CAD,
                                  LE_death_noTx = LE_death_otherCauses,
                                  u_CAD         = u_CADMedicine)
Medicine
```

```
## [1] 7.75
```

```
Surgery  <-       p_death_ORSurgery  * LE_death_ORSurgery +
          (1 - p_death_ORSurgery) * Effectiveness_Subtree(p_death_Tx    = p_death_CADSurgery,
                                                          LE_death_Tx   = LE_death_CAD,
                                                          LE_death_noTx = LE_death_otherCauses,
                                                          u_CAD         = u_CADSurgery)
Surgery
```

```
## [1] 8.36
```