Aalto University
School of Science

Decision trees and Bayesian networks

Jussi Leppinen

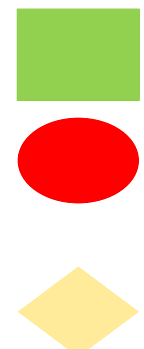
MS-E2191 Graduate seminar on operations research

18.9.2020

Decision trees

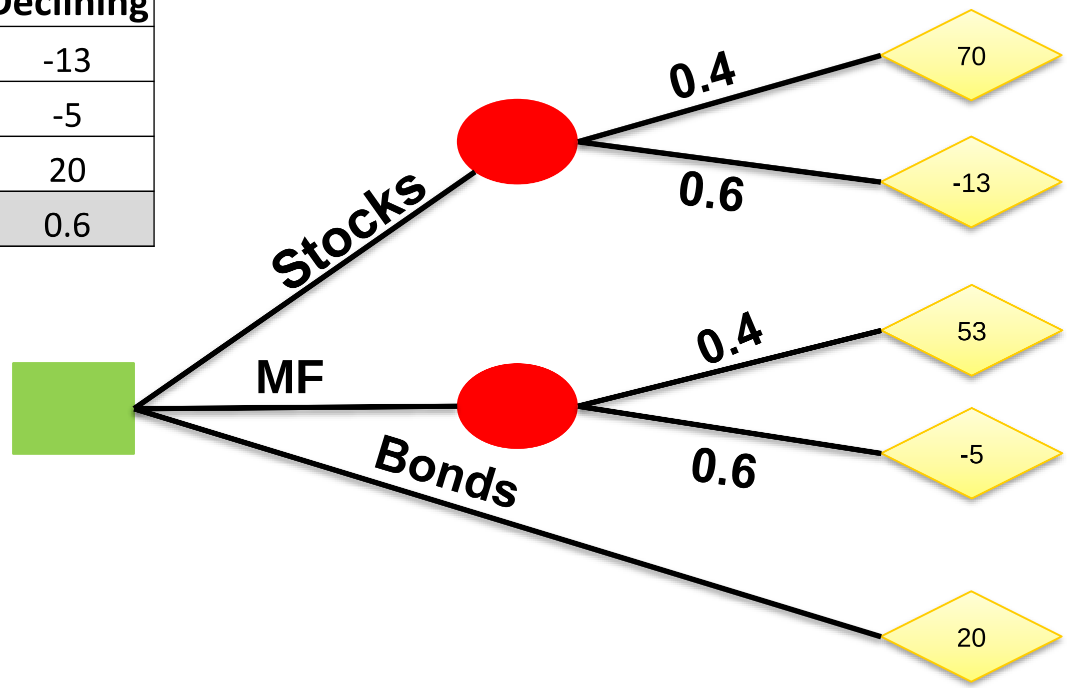
A classical way of representing decision scenarios with several decisions

- Nonleaf nodes
  - Decision nodes, rectangular boxes
  - Chance nodes, circles or ellipses
- Leaves
  - Utility/outcome nodes
- Links
  - Labeled

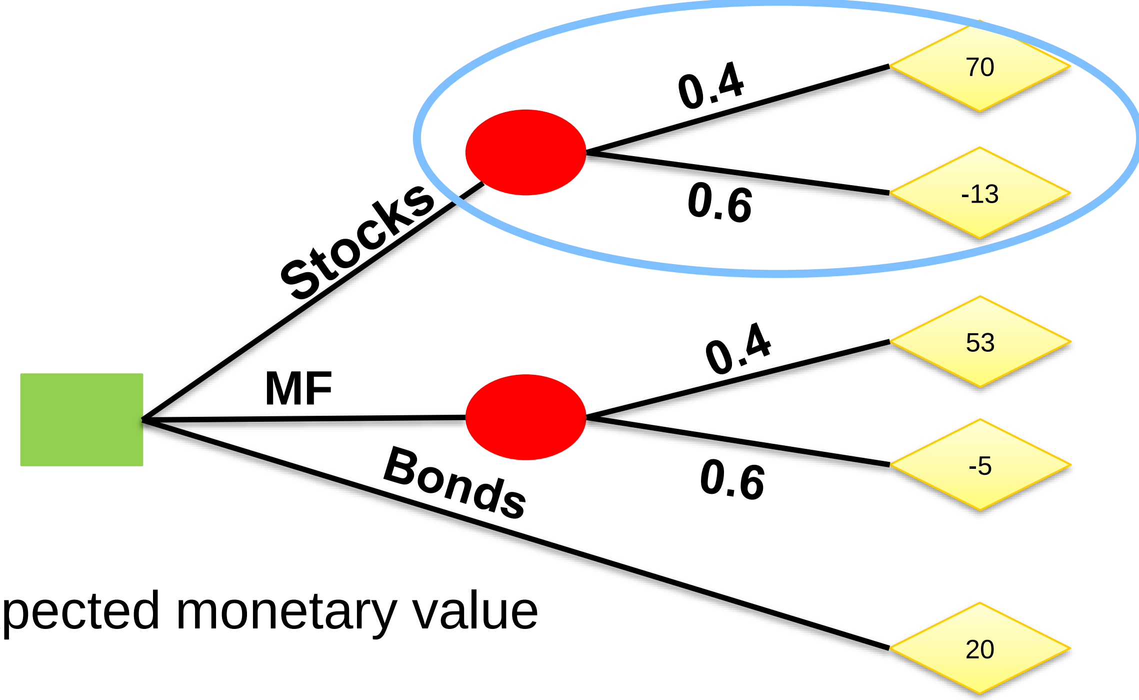


Example

Alternatives	Growing	Declining
Stocks	70	-13
Mutual Funds	53	-5
Bonds	20	20
Probability	0.4	0.6



Example



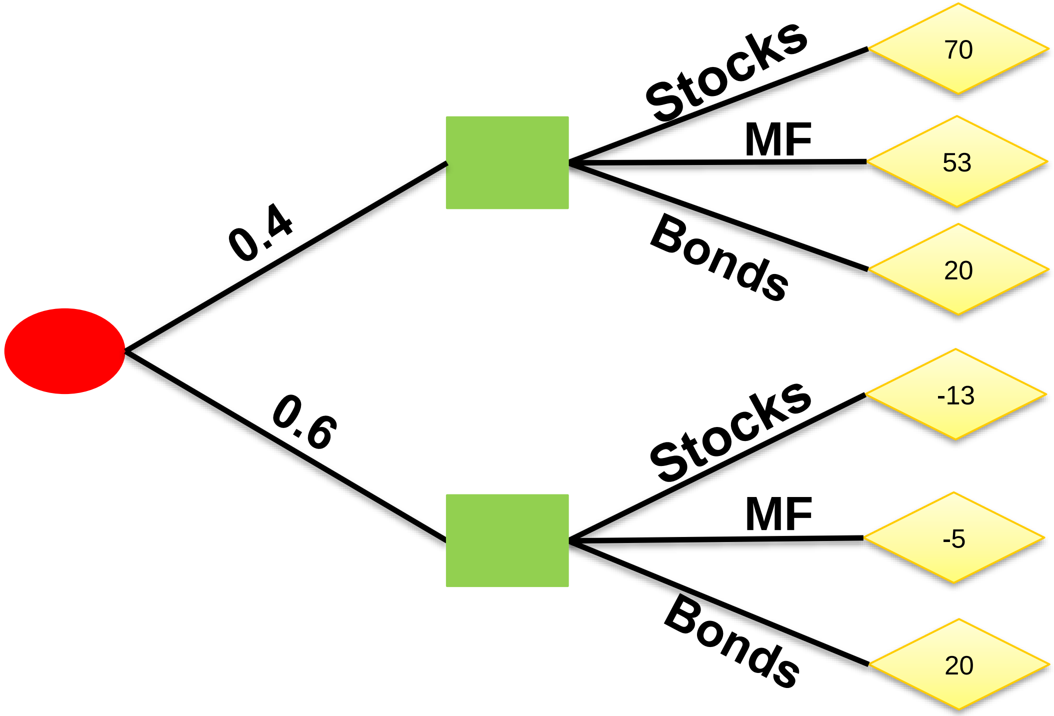
Each decision has its own expected monetary value
- Stocks: 20.2
- MF: 18.2
- Bonds: 20

Choose the decision with highest expected value

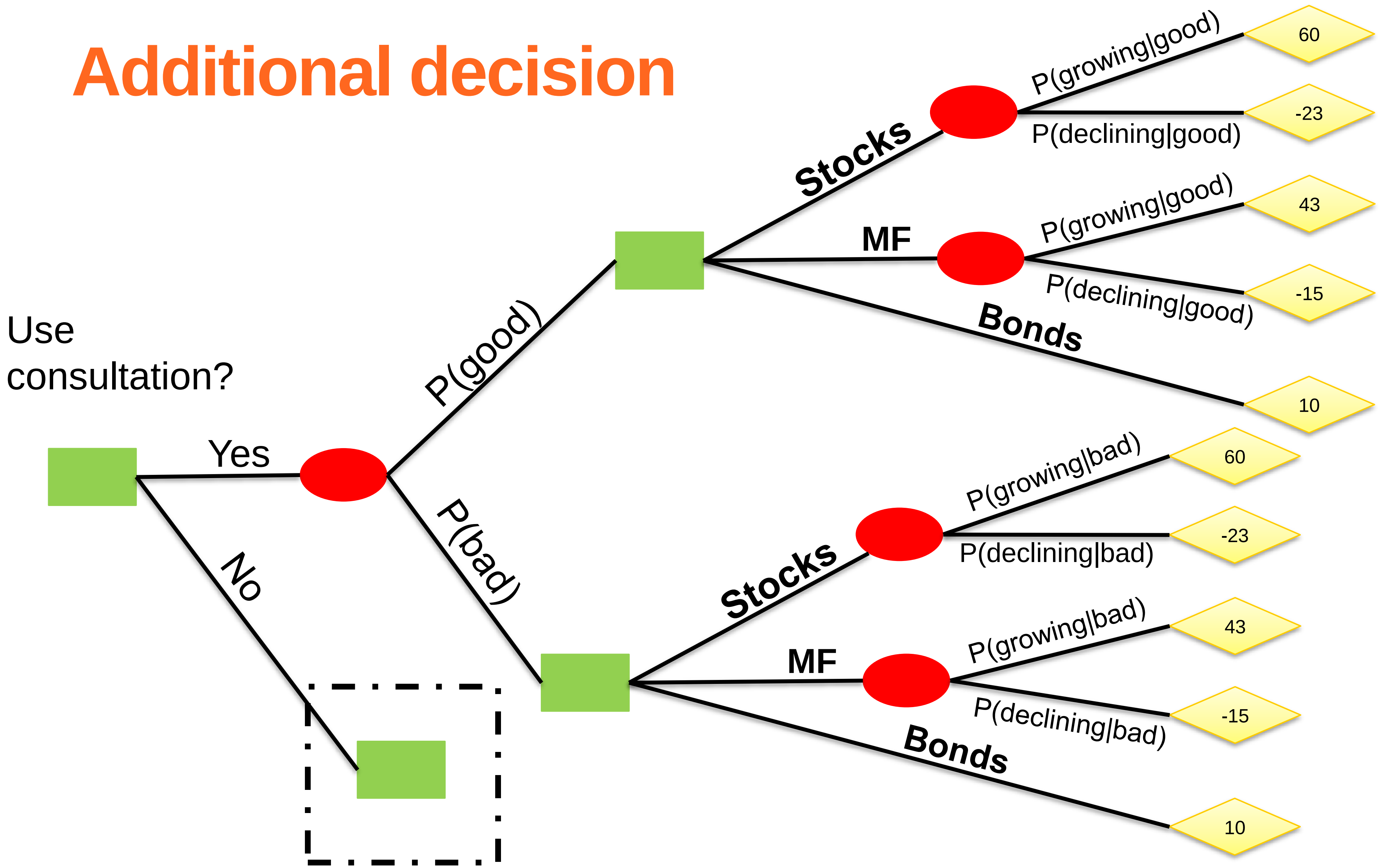
Expected value of perfect information

$\mathrm{EVPI} = 0.4 \cdot 70 + 0.6 \cdot 20 = 40$

Difference is $40 - 20.2 = 19.8$



Additional decision



Probability calculus

Based on the estimations of $P(good|growing)$ and $P(bad|declining)$

Law of total probability

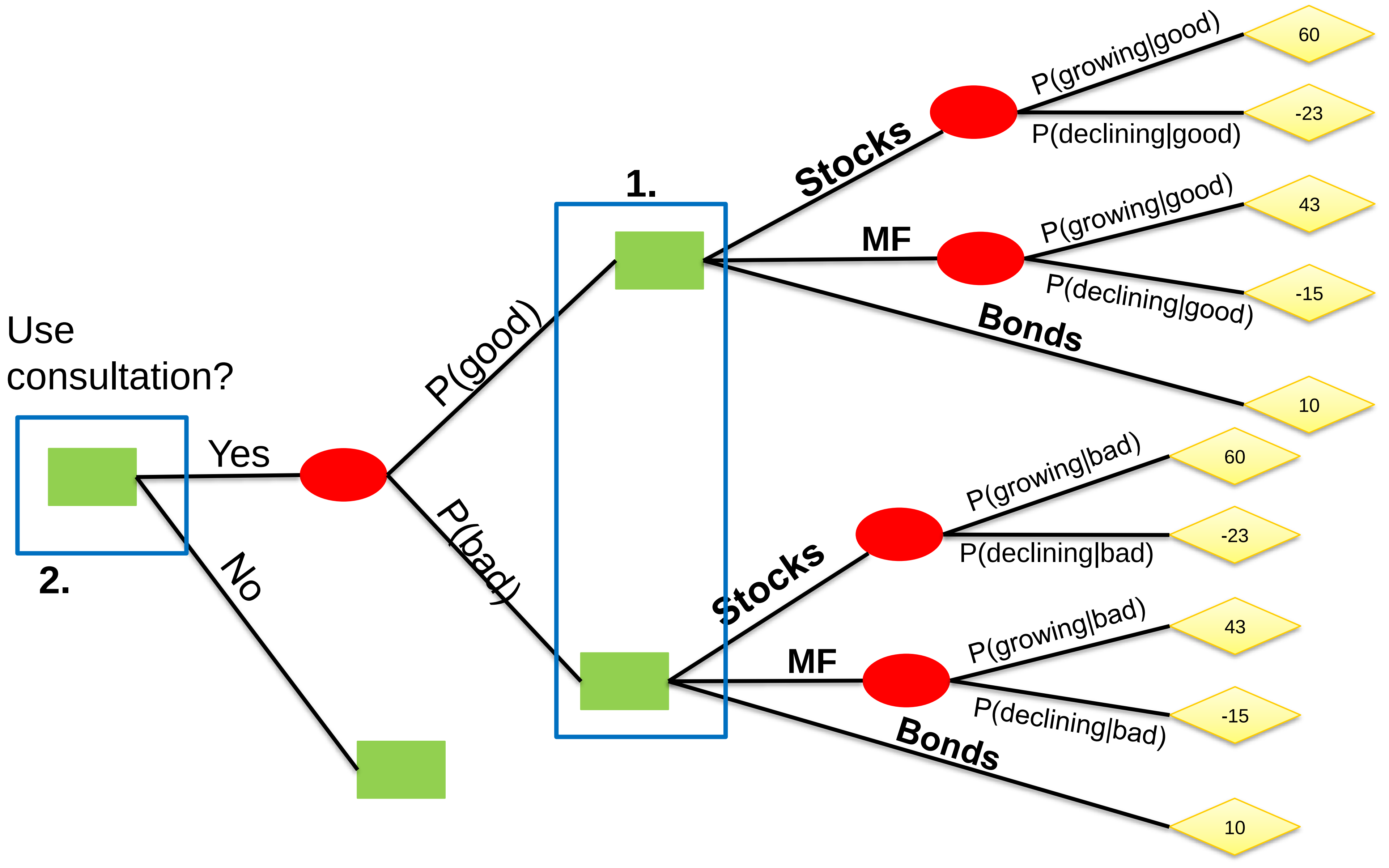
$P(good) = P(good|growing)P(growing) + P(good|declining)P(declining)$

$P(bad) = 1 - P(good)$

Bayesian update

$$P(growing|good) = \frac{P(good|growing)P(growing)}{P(good)}$$

After the probability calculations the expected values of different decisions are obtained, and we can choose the most optimal decision.

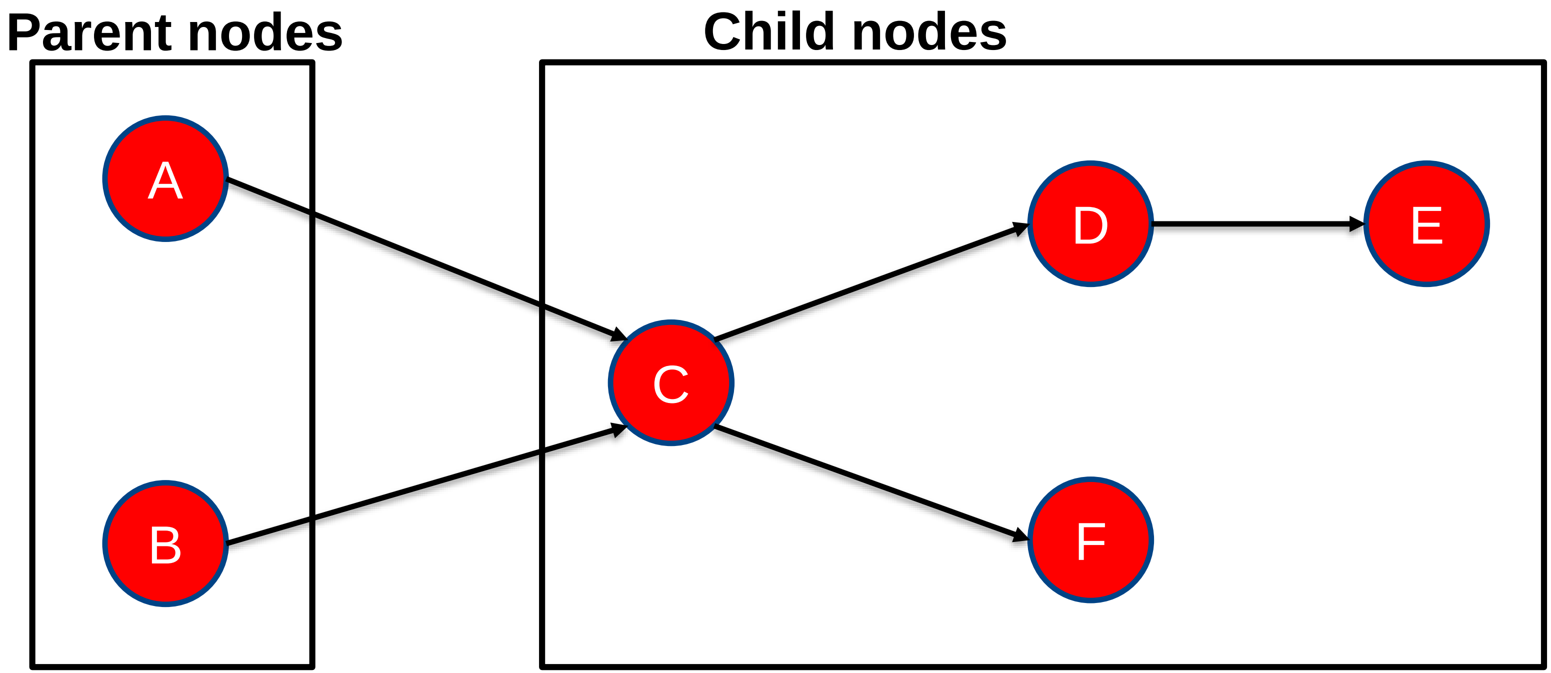


Use consultation?
2.
Yes
No
P(good)
P(bad)
1.
Stocks
MF
Bonds
P(growing|good)
P(declining|good)
P(growing|good)
P(declining|good)
60
-23
43
-15
10
Stocks
MF
Bonds
P(growing|bad)
P(declining|bad)
P(growing|bad)
P(declining|bad)
60
-23
43
-15
10

Decision trees – pros and cons

- **Pros**
  - Informative way of modeling: the outcomes of decision are visible and easy to track if the problem is small
- **Cons**
  - Tree gets big and messy if there are many consecutive decisions with many possible outcomes

Bayesian networks (BN)

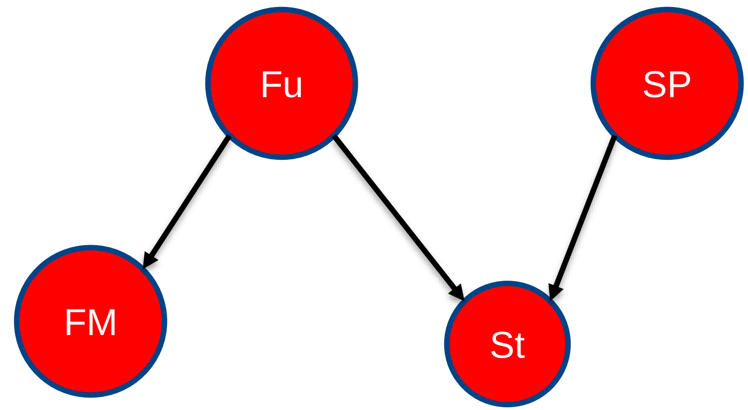


BN definition

- A set of variables and a set of directed edges between variables
- Each variable has a finite set of mutually exclusive states
- The variables and the directed edges for a *directed acyclic graph* (DAG)
  - A graph without cycles
- To each variable A with parents $B_1, \ldots, B_n$ there is attached the potential table $P(A|\, B_1, \ldots, B_n)$

BN example: Car not starting

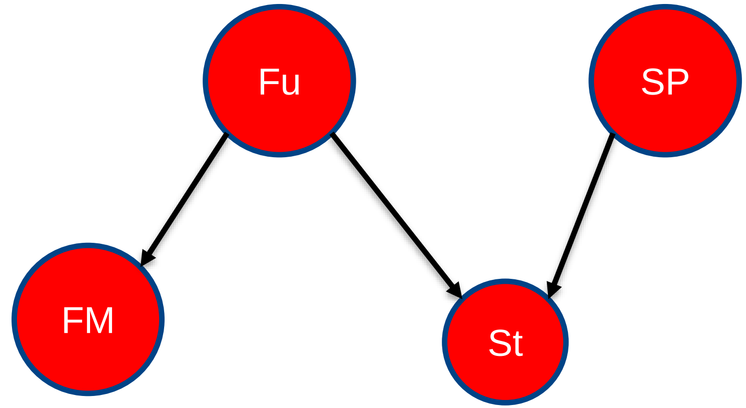
- **Nodes and states**
  - Fuel: Fu {yes, no}
  - Fuel meter standing: FM {full, ½, empty}
  - Clean spark plugs: SP {yes, no}
  - Start: St {yes, no}



BN example

P(Fu)	
yes	0.98
no	0.02

P(SP)	
yes	0.96
no	0.04



P(FM|Fu)		
	Fu =yes	Fu=no
FM=full	0.39	0.001
FM=1/2	0.6	0.001
FM=empty	0.01	0.998

P(St|Fu,SP)				
	Fu=yes		Fu=No	
	St=yes	St=no	St=yes	St=no
Sp=yes	0.99	0.01	0	1
Sp=no	0.01	0.99	0	1

Chain rule for BNs

Let BN be a Bayesian network over $U = \{A_1, \ldots, A_n\}$. Then the joint probability distribution $P(U)$ is the product of all the potentials specified in BN

$$P(U) = \prod_i P(A_i | pa(A_i)),$$

where $pa(A_i)$ is the parent set of A_i.

Chain rule in practise

P(Fu)	
yes	0.98
no	0.02

P(SP)	
yes	0.96
no	0.04

P(FM|Fu)		
	Fu =yes	Fu=no
FM=full	0.39	0.001
FM=1/2	0.6	0.001
FM=empty	0.01	0.998

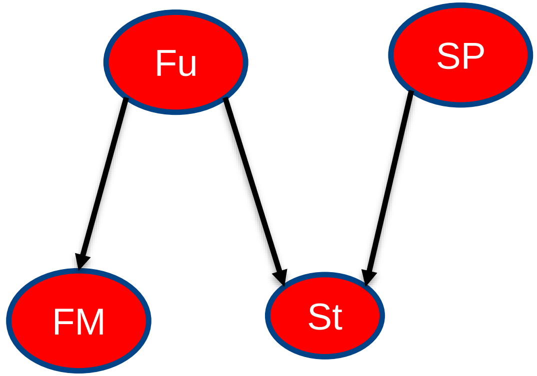
P(St|Fu,SP)				
	Fu=yes		Fu=No	
	St=yes	St=no	St=yes	St=no
Sp=yes	0.99	0.01	0	1
Sp=no	0.01	0.99	0	1

$$P(Fu = yes, FM = full, Sp = yes, St = yes)$$
$$= P(Fu = yes)P(SP = yes)\ P(FM = full|Fu = yes)$$
$$P(St = yes|Fu = yes, SP = yes)$$
$$= 0.98 \cdot 0.96 \cdot 0.39 \cdot 0.99$$
$$= 0.363$$

The results

P(Fu,FM,SP,St=yes)						
	FM=full		FM=1/2		FM=empty	
	Fu=yes	Fu=no	Fu=yes	Fu=no	Fu=yes	Fu=no
Sp=yes	0.36324288	0	0.5588352	0	0.00931392	0
Sp=no	0.00015288	0	0.0002352	0	0.00000392	0

P(Fu,FM,SP,St=no)						
	FM=full		FM=1/2		FM=empty	
	Fu=yes	Fu=no	Fu=yes	Fu=no	Fu=yes	Fu=no
Sp=yes	0.00366912	0.0000192	0.0056448	0.0000192	0.00009408	0.0191616
Sp=no	0.01513512	0.0000008	0.0232848	0.0000008	0.00038808	0.0007984



Evidence

Let A be a variable with n states. A *finding* on A is an n-dimensional table of zeros and ones, denoted by $\underline{e}$.

Theorem Let BN be a Bayesian network over the universe U and let $\underline{e}_1, \ldots, \underline{e}_n$ be findings; then

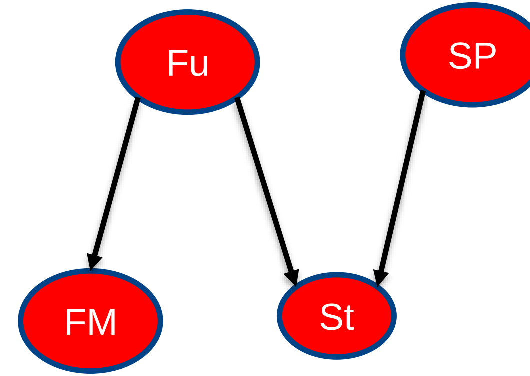
$$P(U, e) = \prod_{A \in U} P(A|pa(A)) \cdot \prod_i \underline{e}_i,$$

and for $A \in U$ we have

$$P(A|e) = \frac{\sum_{U \setminus \{A\}} P(U,e)}{P(e)}.$$

Evidence that St=no

P(Fu,FM,SP,St=no)						
	FM=full		FM=1/2		FM=empty	
	Fu=yes	Fu=no	Fu=yes	Fu=no	Fu=yes	Fu=no
Sp=yes	0.00366912	0.0000192	0.0056448	0.0000192	0.00009408	0.0191616
Sp=no	0.01513512	0.0000008	0.0232848	0.0000008	0.00038808	0.0007984



	P(SP,St=no)	P(SP|St=no)
Sp=yes	0.029	0.419
Sp=no	0.040	0.581

	P(Fu,St=no)	P(Fu|St=no)
Fu=yes	0.048	0.707
Fu=no	0.020	0.293

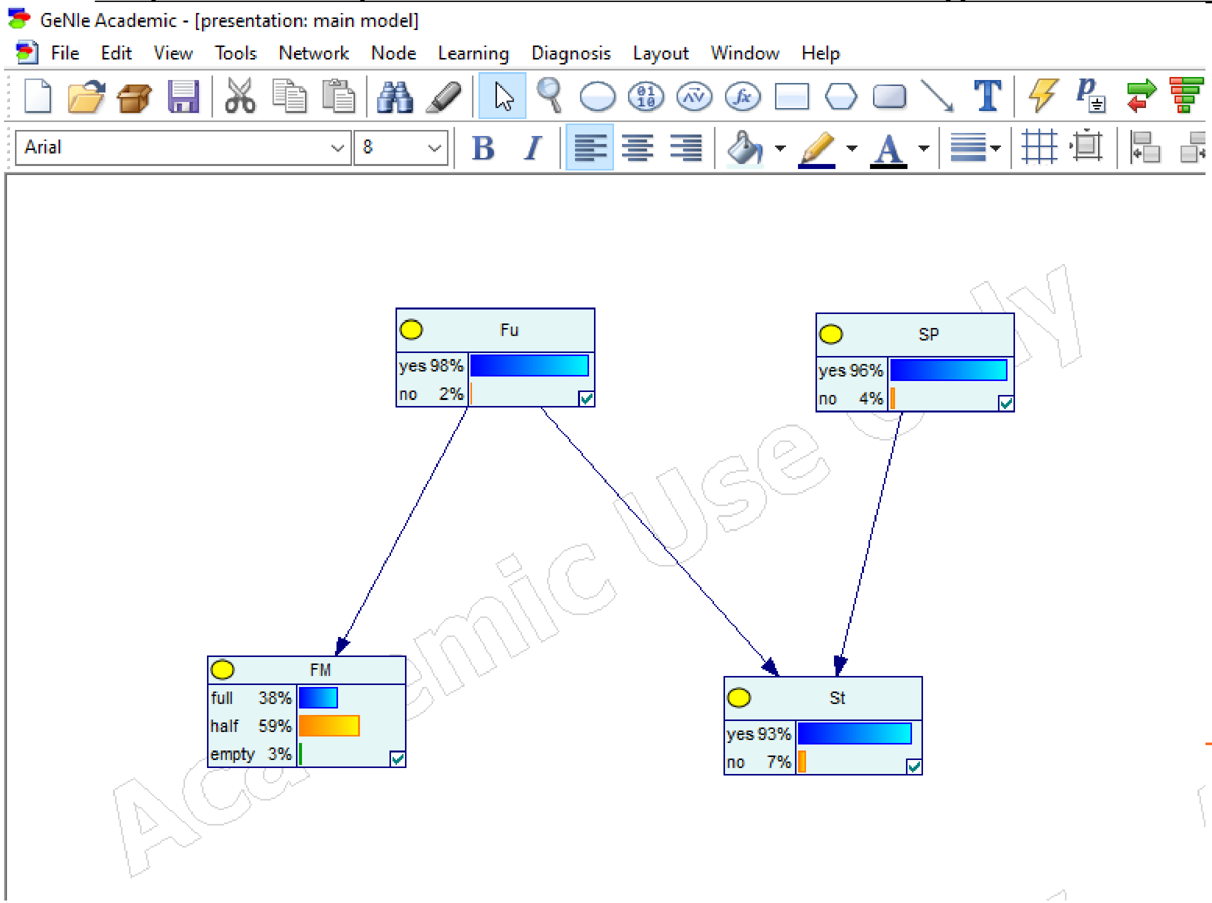
GeNIe

- Sofware from BayesFusion
- Graphical interface for Bayesian networks
- Free academic download
  - Univesity level information needed during the installation

- https://www.bayesfusion.com/

Demonstrate with GeNIe

Some basic information:
https://www.youtube.com/watch?v=aW3gxE6XB9E&ab_channel=mdb



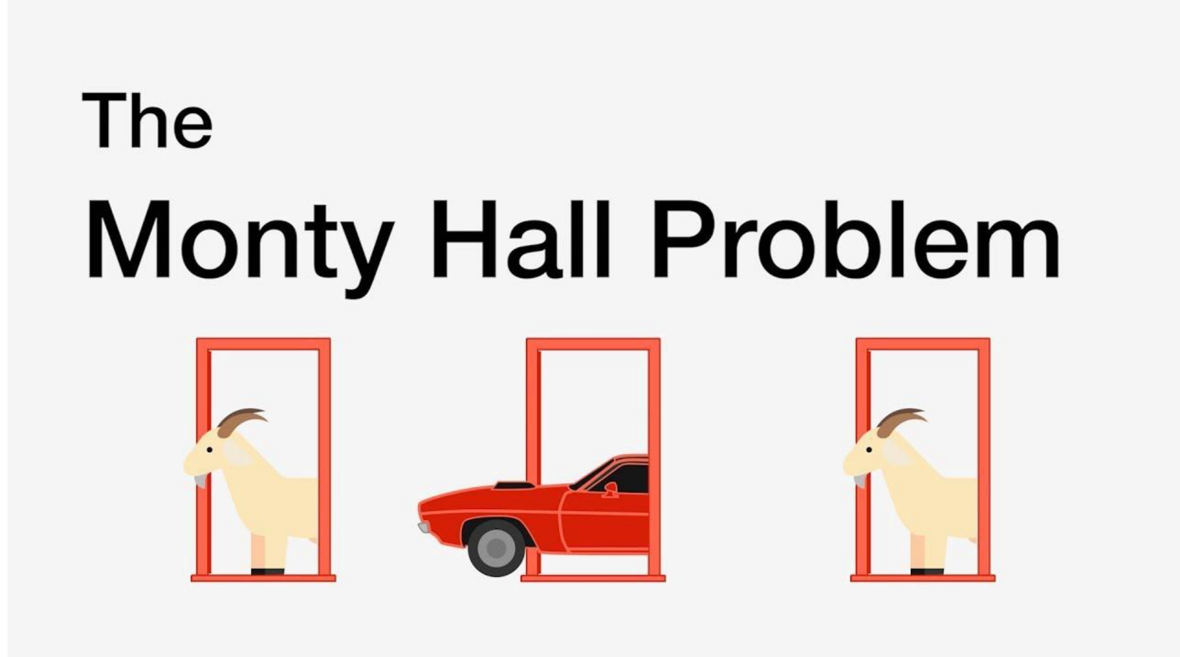
References

Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs.* Springer, New York

https://medium.com/swlh/explaining-the-monty-hall-problem-ce8152b7cbc0

https://en.wikipedia.org/wiki/Monty_Hall_problem

Homework



Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice when you are trying to choose the door that has car behind it?

Homework: questions

1. Present the Bayesian network graph for the problem
2. Fill in the missing information P(Monty opens door x|competitor chooses door y, prize behind door z) in the Excel Template
3. Calculate joint probabilities using the chain rule
4. Given the information that the competitor chose door 1 and Monty opened the door 2 what are the winning odds of doors 1-3?
5. Given the previous information how much should the prior winning distribution of door 1 increase (when doors 2 and 3 still win equally likely) so that there is not difference between doors 1 and 3 winning?
6. Return solutions to jussi.leppinen@aalto.fi