Modeling with Bayesian Networks

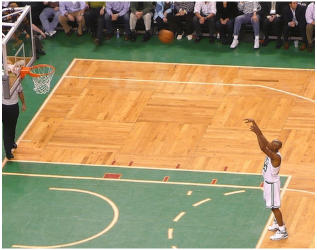
Javier Larrosa

UPC Barcelona Tech

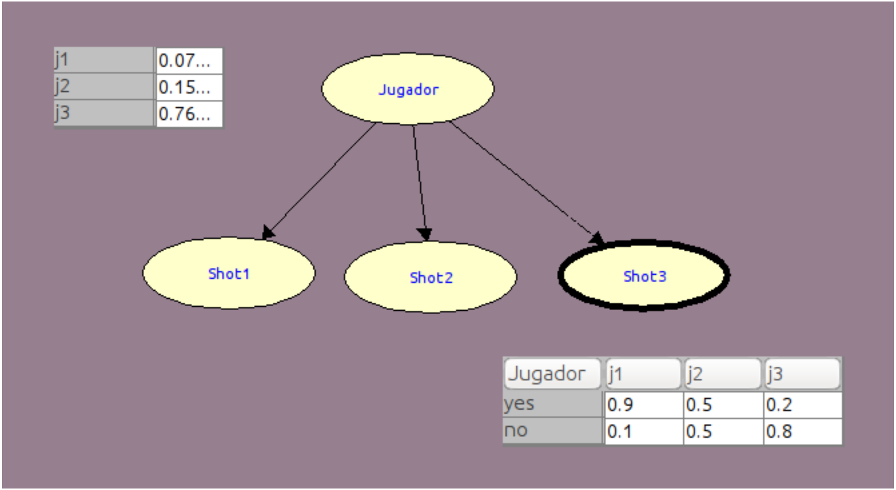
.

Basketball

We turn on the radio and they are broadcasting a basketball game. Somebody in our favorite team is shooting 3 free shots, but we could not hear who it was. What is our best guess and how do we update it depending on how the shots go? (for simplicity, let suppose that the team is made of three players and they are all on court)
Player one has scored 90 out of 100 shots, player two 100 out of 200 and player three 200 out of 1000

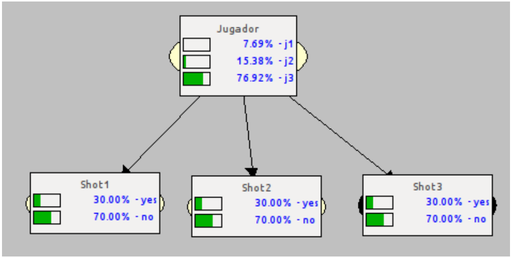


Basquetball

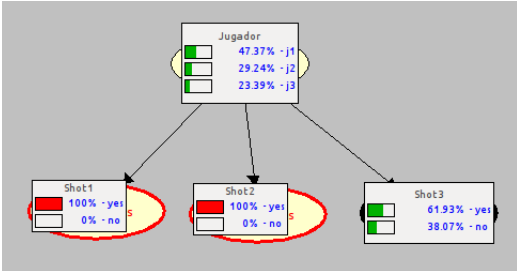


$$P(J, S1, S2, S3) = P(S1|J)P(S2|J)P(S3|J)P(J)$$

Basquetball

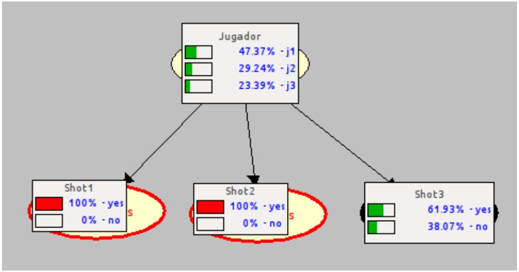


$P(J), P(S1), P(S2), P(S3)$



$$P(J|S1 = yes, S2 = yes), P(S3|S1 = yes, S2 = yes)$$

Basquetball



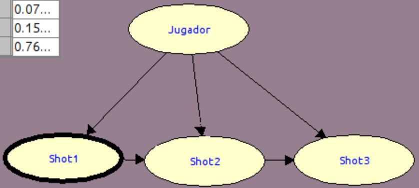
Given evidence $S1 = yes, S2 = yes$,

- MAP over $S3$ is *yes* because (see monitor)
- MAP over J is $j1$ because (see monitor)
- MAP over $\{S3, J\}$ is $(yes, j1)$ because it is the entry with the highes value in

$S1$	$S2$	$S3$	$S4$	$P(S3, J \mid S1 = yes, S2 = yes)$
yes	yes	yes	j1	0.426
yes	yes	yes	j2	...
yes	yes	yes	j3	...
yes	yes	no	j1	...
yes	yes	no	j2	...
yes	yes	no	j3	...

Basquetball

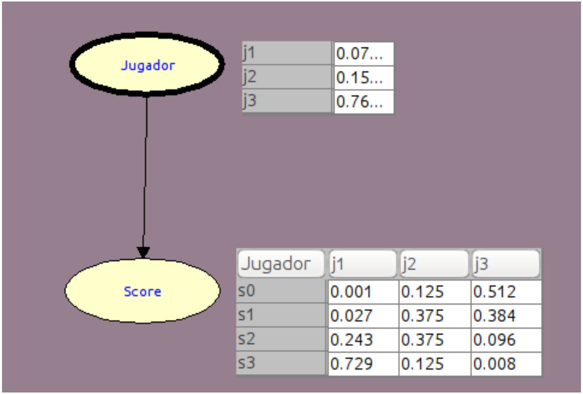
j1	0.07...
j2	0.15...
j3	0.76...



Jugador	j1	j2	j3
yes	0.9	0.5	0.2
no	0.1	0.5	0.8

Jugador	j1		j2		j3	
Shot2	yes	no	yes	no	yes	no
yes	0.95	0.85	0.55	0.45	0.25	0.15
no	0.05	0.15	0.45	0.55	0.75	0.85

$$P(J, S1, S2, S3) = P(S3|J, S2)P(S2|J, S1)P(S1|J)P(J)$$



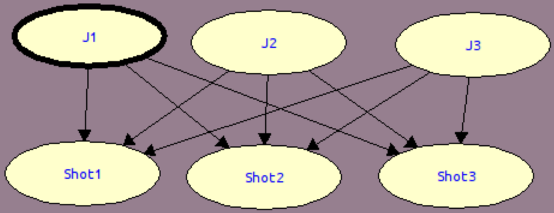
j1	0.07...
j2	0.15...
j3	0.76...

Jugador	j1	j2	j3
s0	0.001	0.125	0.512
s1	0.027	0.375	0.384
s2	0.243	0.375	0.096
s3	0.729	0.125	0.008

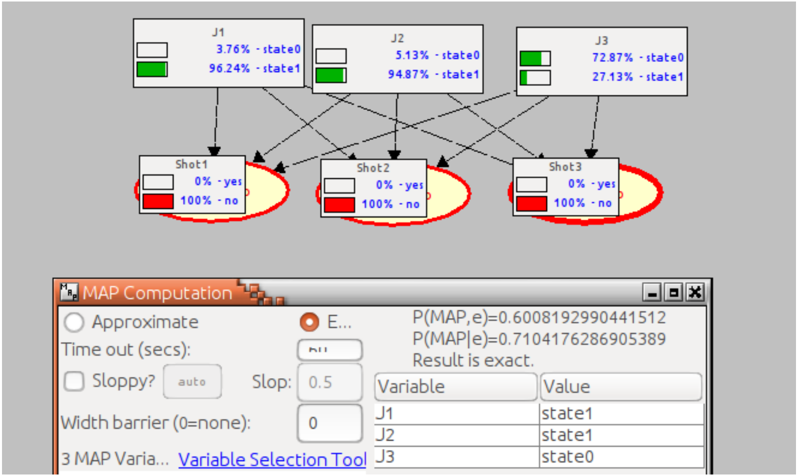
Score: Binomial distribution $P(S = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

Basquetball

state0	0.07...
state1	0.92...

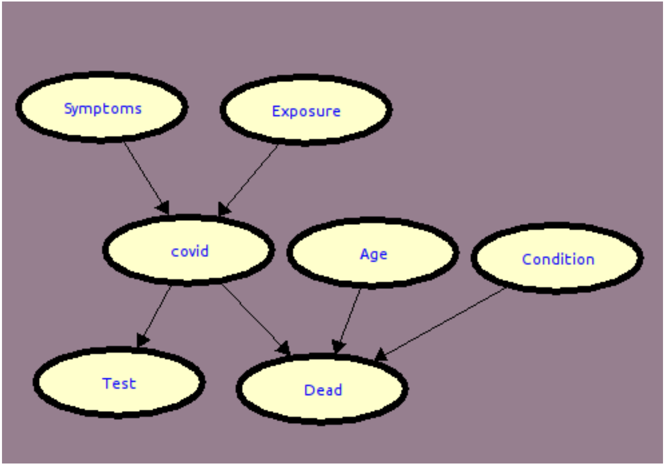


J1	state0				state1			
J2	state0		state1		state0		state1	
J3	state0	state1	state0	state1	state0	state1	state0	state1
yes	0.0	0.2	0.5	0.0	0.9	0.0	0.0	0.0
no	1.0	0.8	0.5	1.0	0.1	1.0	1.0	1.0



MAP Sobre $J1, J2, J3$ de evidencia $S1 = S2 = S3 = no$

Covid before vaccination



Late For Dinner

My wife usually takes the 6:30pm bust to get home after work and she arrives around 7pm. On her way from the bus stop she sometimes gets bread from the bakery. Some times she is late because she gets some unexpected thing to do in the last moment and she misses the bus. Some other times the bus is caught in a traffic jam. Traffic may get bad due to weather conditions or car accidents. Some times there is a special event in the city that also makes the traffic to collapse. Sometimes she is late because she stays longer than expected at the bakery or because she comes across some acquaintance at the street.
Build a Bayesian Network to model this problem. You may add new variables not mentioned in the previous description that you may find useful to make the model more accurate.

Late for Dinner

- Late $\leftarrow$ Miss Bus, Traffic Jam, Slow after bus
- Miss Buss $\leftarrow$Late Work, Watch Problem
- Watch Problem$\leftarrow$Broken Watch, Confusion, Yearly Day light saving
- Traffic Jam$\leftarrow$Weather, Accident, City Special Event
- City Event$\leftarrow$Barça Match, World Mobile, Special Summit, Demonstration
- Barça$\leftarrow$Champions

Observations: Weather can be dry, rain, snow. Weather influences accident. Barça Match influences World mobile. World Mobile influences Special Summit. If Barça is in the Champions chance of match increases.

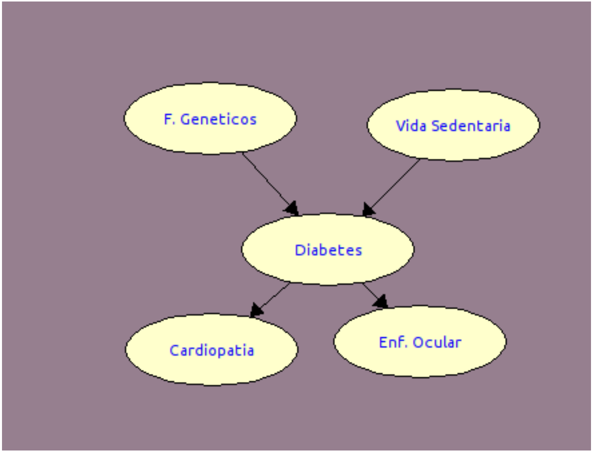
Diabetes

Diabetes

1 de cada 10 adultos sufren diabetes de tipo 2 (D). Se han identificado factores genéticos (G) que triplican la probabilidad de padecer D. Estos G se dan en el 10 por cien de la población. Se sabe que una vida sedentaria (S) duplica la probabilidad de padecer la enfermedad y que el 30 por ciento de la población hace S

La diabetes tiene muchas consecuencias. las dos más importantes son cardiopatías (C) y enfermedades oculares (O). Si bien una persona sin D tiene una probabilidad de 0.05 de tener C, en una persona con D esta probabilidad es 0.2. una persona sin D tiene una probabilidad de 0.01 de padecer O, pero si tiene D la probabilidad se multiplica por 10.

Diabetes


F. Geneticos
Vida Sedentaria
Diabetes
Cardiopatia
Enf. Ocular

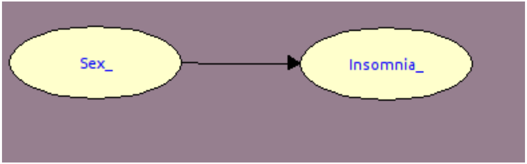
Diabetes

- $P(d|\neg g, \neg s) = \alpha$, $P(d|g, \neg s) = 3\alpha$
- $P(d|\neg g, s) = 2\alpha$, $P(d|g, s) = 6\alpha$
- $P(d) = \sum_{g,s} P(d, g, s) = \sum_{g,s} P(d|g, s)P(g)P(s) = 0.1$
- $P(g) = 0.1$, $P(s) = 0.3$
- $0.1 = \alpha(0.9)(0.7) + 3\alpha(0.1)(0.7) + 2\alpha(0.9)(0.3) + 6\alpha(0.1)(0.3)$

Sex and Insomnia

Mental Health research shows that one out of four males and one out of three females have insomnia.

The following network represents this knowledge, but it is useless in the sense that we cannot make an intervention in the Sex variable.

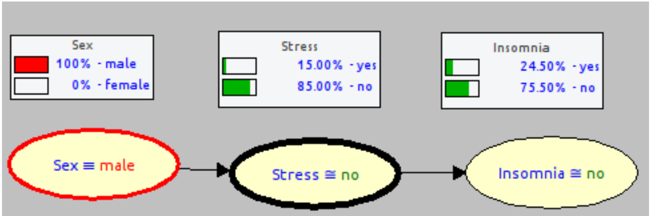


Bayesian Networks as Decision Making tools

Sex, Stress and Insomnia

Mental Health research shows that stress is highly related insomnia.
Women are typically more stressed than man.

The following network is compatible with the previous one, but represents more detailed and useful knowledge. We can make an intervention in the stress variable (go to yoga class!)



Sex	male	female
yes	0.15	0.45
no	0.85	0.55

Stress	yes	no
yes	0.5	0.2
no	0.5	0.8

Books and school success

A study shows that children do better at school if there are books in their home

The following network represents this knowledge. It may seem useful

yes	0.53
no	0.47

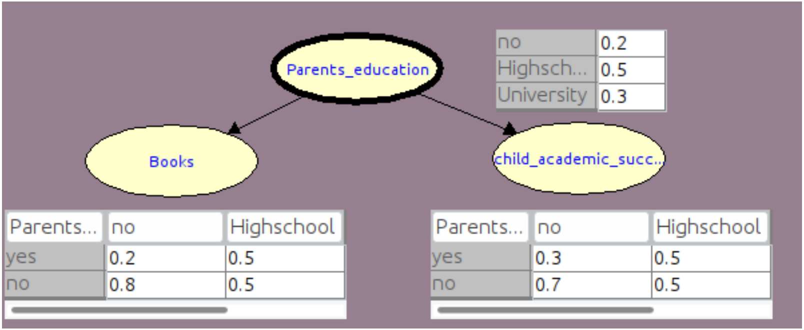
books_	yes	no
yes	0.62	0.47
no	0.38	0.53

Bayesian Networks as Decision Making tools

Books and school success

The education level of parents influences both having books at home and children success

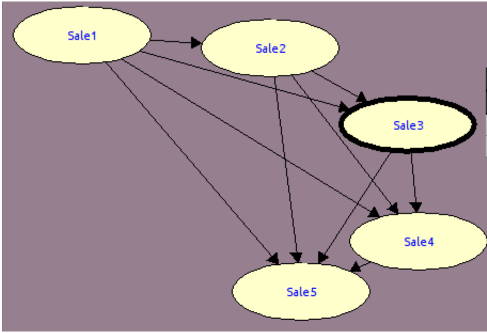
The following network is compatible with the previous one, but represents more detailed knowledge.



Cards

We have a shuffled deck of cards. We uncover them one by one. What is the probability of the Ace of Hearts appearing in the $i-$th position?

Cards (Naive Solution)

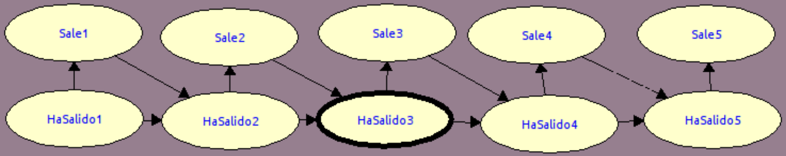


Sale2	state0		state1	
Sale1	state0	state1	state0	state1
state0	0.6666...	1.0	1.0	0.5
state1	0.3333...	0.0	0.0	0.5

Unfeasible for a real deck

Cards (solution with auxiliary variable)

HaSalid...	state0	state1
state0	0.666...	1.0
state1	0.333...	0.0



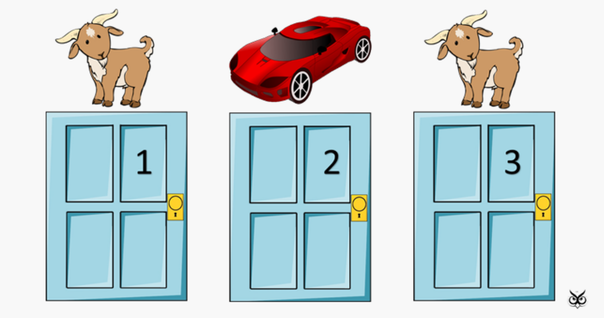
HaSalido2	state0		state1	
Sale2	state0	state1	state0	state1
state0	1.0	0.0	0.0	0.0
state1	0.0	1.0	1.0	1.0

Cards (variation)

We have a shuffled deck of cards. We uncover them one by one. What is the probability of a red card appearing in the $i-$th position? (note that half of the cards are red)

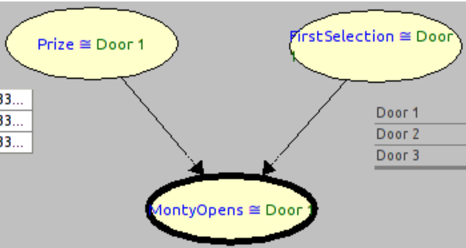
Monty Hall

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?



Monty Hall

Door 1	0.3333...
Door 2	0.3333...
Door 3	0.3333...



Door 1	0.3333...
Door 2	0.3333...
Door 3	0.3333...

FirstSelection	Door 1			Door 2			Door 3		
Prize	Door 1	Door 2	Door 3	Door 1	Door 2	Door 3	Door 1	Door 2	Door 3
Door 1	0.0	0.0	0.0	0.0	0.5	1.0	0.0	1.0	0.5
Door 2	0.5	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.5
Door 3	0.5	1.0	0.0	1.0	0.5	0.0	0.0	0.0	0.0

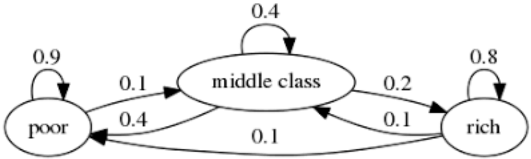
Deterministic Monty

Suppose that you have followed the show every week and noticed that the host, when having two options, always opens the door with the lowest number.
Adapt the previous network to the current version and use it to give you a winning strategy

Markov Model

Markov Model

A markov model is a set of states $S = \{s_1, s_2, \dots, s_n\}$ and a CPT $Prob(s_i|s_j)$ which says the probability of going to state s_i given that we are in state s_j

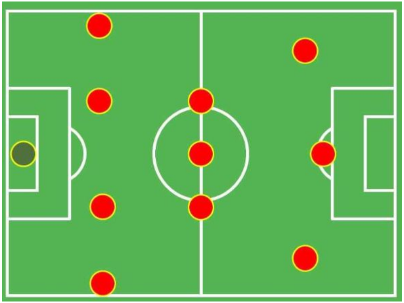


Property

A markov model can be represented as a Bayesian Network with variables $X_1, X_2, \dots$. All variables have the same domain S and the same CPT $P(X_{i+1}|X_i)$

Football (Markov Model)

Markov Models are used in **sport analytics**



Serengeti

Examen 2022

Un equipo de biólogos están estudiando la fauna del parque de Serengeti. En particular, les interesa saber el tamaño de un grupo de un tipo raro de rinoceronte que viven en una zona muy concreta del parque. Saben que todos los rinocerontes del grupo pasan al amanecer por un desfiladero para llegar al rio y poder beber. Para estimar el número de ejemplares (R) pondrán a dos oteadores(O1, O2) en los dos extremos del desfiladero para que cada uno de ellos diga cuántos ve. Obviamente, los oteadores pueden no ver alguno de los ejemplares. En el improbable caso de que haya niebla (N), el margen de error de los oteadores será mucho mayor. Plantea una Red Bayesiana que tenga como objetivo estimar el número de ejemplares (R). Incluye valores para las CPTs que sean razonables y consistentes con el enunciado. Aunque en la vida real el tamaño de los grupos de rinocerontes puede ser grande, por simplicidad, en este ejercicio puedes asumir que R nunca será mayor de 3.