# Uncertainty

Not Knowing Knowledge for sure

Rarely we are gonna know the information for sure

So we want it to make the best possible decission.

## Probability theory

- Possible worlds $\omega$

  ↳ Every possible situation, rolling a dice gives 6 possible worlds

$P(\omega) \rightarrow$ Probability of a world.

$$\boxed{0 \leq P(\omega) \leq 1 \qquad \sum_{\omega \in \Omega} P(\omega) = 1}$$

$$P(\boxed{\cdot \;\, \cdot}) = \frac{1}{6}$$

2 Dice

there are $n^2$ possible worlds equally equal the sums (results are not equally likely)

$P(\text{sum of } 12) = 1/36$

$P(\text{sum of } 7) = 6/36 = 1/6$

$$\frac{1}{6} > \frac{1}{36}$$

HNB203-57

**Bayes' Rule** -> Used commonly to ~~calculate~~ compute conditional probability

$$\boxed{P(b|a) = \frac{P(b)\,P(a|b)}{P(a)}}$$

80% of rainy afternoons start with cloudy mornings
40% of days have cloudy mornings
10% of days have rainy afternoons

$$P(\text{rain}|\text{clouds}) = \frac{P(\text{clouds}|\text{rain})\,P(\text{rain})}{P(\text{clouds})}$$

$$P(\text{rain}|\text{clouds}) = \frac{0.8 \cdot 0.1}{0.4} = 0.2 = 20\%$$

Knowing:

$P(\text{visible effect} \mid \text{unknown cause})$

We can calculate

$P(\text{unknown cause} \mid \text{visible effect})$

Knowing -> P(medical ~~test~~ result | disease)
We can -> P(disease | medical test result)

HNB203-57

Random variables → a variable in probability theory with a domain of possible values it can take on.

Roll = {1, 2, 3, 4, 5, 6}

$\text{Weather} = \underbrace{\{\text{sun, cloud, rain, wind, snow}\}}_{\text{Different probabilities}}$.

Probability distribution → the probability of each ces in a random variable.

$P(\text{Flight}) = \langle 0.6, 0.3, 0.1 \rangle$

Independence → The knowledge that ~~one event~~ the occurance of one event does not affect the probability of another event.

Two ~~var~~iables are independent if:

$P(a \wedge b) = P(a) P(b)$

because

$P(a \wedge b) = P(a) P(b \mid a)$

a does not change the probability of b



HNB203-57

Unconditional probability -> degree of belief in a proposition in the absence of any other evidence

Conditional Probability -> degree of belief in a proposition given some evidence that has already been revealed

$P(a \mid b)$ -> Probability of a knowing b

P(rain today / rain yesterday)

P(disease | test)

$$\boxed{P(a \mid b) = \frac{P(a \wedge b)}{P(b)}}$$

$$P(\text{sum } 12 \mid ⚅) = \frac{P(\text{sum } 12 \wedge ⚅)}{P(⚅)} =$$

$$= \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{36} \cdot \frac{6}{1} = \boxed{\frac{1}{6}}$$

Another common way:

$$\boxed{\begin{aligned} P(a \wedge b) &= P(b)\,P(a \mid b) \\ P(a \wedge b) &= P(a)\,P(b \mid a) \end{aligned}}$$

# Joint Probability

| AM | | PM | |
|---|---|---|---|
| C = cloud | C = ¬cloud | R = rain | R = ¬rain |
| 0.4 | 0.6 | 0.1 | 0.9 |

| | R = rain | R = ¬rain |
|---|---|---|
| C = clouds | 0.08 | 0.32 |
| C = ¬clouds | 0.02 | 0.58 |

(constant normalization constant)

$$P(\text{clouds} \mid \text{rain}) = \frac{P(\text{clouds} \wedge \text{rain})}{P(\text{rain})} = \alpha P(\text{clouds}, \text{rain})$$

This probability is proportional to P(clouds, rain)

$$P(\text{clouds} \mid \text{rain}) = \alpha P(\text{clouds}, \text{rain}) = \alpha\, 0.08$$

We know it needs to sum to one

$$P(\neg\text{clouds} \mid \text{rain}) = \alpha P(\neg\text{clouds}, \text{rain}) = \alpha 0.02$$

$$\alpha 0.02 + \alpha 0.08 = 1 \qquad \boxed{\alpha = 10}$$

## Rules

Negation rule

$P(\neg a) = 1 - P(a)$

Inclusion Exclusion

$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

Marginalization

$P(a) = P(a \wedge b) + P(a \wedge \neg b)$

$$P(X = x_i) = \sum_j P(X = x_i \wedge Y = y_j)$$

P(C=clouds) = P(cloud ∧ rain) + P(cloud ∧ ¬ Rain)

Conditioning (Equivalent to marginalization with condition)

$P(a) = P(a \mid b)\, P(b) + P(a \mid \neg b)\, P(\neg b)$

$$P(X = x_i) = \sum_j P(X = x_i \mid Y = y_j)\, P(Y = y_j)$$

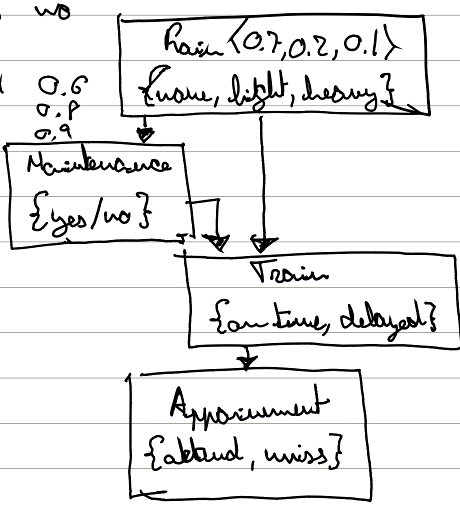
# Bayesian Networks

data structure that represents the dependencies among random variables

- directed graph
- each node represents a random variable
- arrow from X to Y means X is a parent of Y
- Each node has a probability distribution of:

$$P(X \mid Parents(X))$$

| R \ M | yes | no |
|---|---|---|
| none | 0.4 | 0.6 |
| light | 0.2 | 0.8 |
| heavy | 0.1 | 0.9 |




| R | M | on time | delayed |
|---|---|---|---|
| none | yes | 0.8 | 0.2 |
| none | no | 0.9 | 0.1 |
| light | yes | 0.6 | 0.4 |
| light | no | 0.7 | 0.3 |
| heavy | yes | 0.4 | 0.6 |
| heavy | no | 0.5 | 0.5 |

| T | attend | miss |
|---|---|---|
| on time | 0.9 | 0.1 |
| delayed | 0.6 | 0.4 |

# Computing Joint Probabilities

$P(\text{light}) = 0.2$

$P(R = \text{light}, M = \text{no}) = P(R = \text{light}) \cdot P(M = \text{no} \mid R = \text{light})$

$P(R = \text{light}, M = \text{no}, T = \text{delayed}) =$

$\ldots$

## Inference

Query $X$ = variable for which to compute distribution
Evidence variables $\mathbf{E}$: observed variables of even $e$
Hidden variables $\mathbf{Y}$: non-evidence, non-query
Goal: Calculate $P(X \mid e)$

$P(\text{Appointment} \mid \text{light}, \text{no}) \Rightarrow$ Query = Appointment

$\alpha P(\text{Appointment}, \text{light}, \text{no}) =$

$= \alpha [P(\text{Appointment}, \text{light}, \text{no}, \text{on time})$
$P(\text{Appointment}, \text{light}, \text{no}, \text{delayed})]$

Hidden variable = Train

## Inference by enumeration

$$P(X|e) = \alpha P(X,e) = \underbrace{\alpha \sum_y P(X,e,y)}$$

↳ Marginalization.

$X$ is query
$e$ is evidence
$y$ ranges over values of hidden variables
$\alpha$ normalizes the result

## Approximate inference

Sampling → sample a value of every variable

samples → take randomly

Likelihood Weighting (a way of sampling)

- Start by fixing the value of evidence values
- Sample non-evidence variables using conditional probabilities in the Bayesian network
- Weight each sample by it's likelihood = the probability of all the evidence

↳ if the evidence is not likely, the weight of this node will be lower.

Each sample has a weight equal to the probability of evidence ~~variables~~ values.

## Markov Models

$X_t$ = weather at time $t$

**Markov assumption**

the assumption that the current state depends on only a finite fixed number of previous states

**Markov Chain**

A Markov chain is a sequence of random variables where the distribution of each variable follows Markov's assumption

Tomorrow $X_{t+1}$

Today $X_t$

| | | |
|---|---|---|
| | 0.8 | 0.2 |
| | 0.3 | 0.7 |

Transition model to build Markov Chain

HNB203-57

## Sensor models

| Hidden States | Observation |
|---|---|
| robot's position | robot's sensor data |
| words spoken | audio waveforms |
| User engaging | website data |
| weather | umbrella |

## Hidden Markov Model

model for a system with hidden states that generate some observed event

for what I'm seeing all this probability charts eventually will come from training

| | Observation ($E_t$) umbrella | no umbrella |
|---|---|---|
| State ($X_t$)  | 0.2 | 0.8 |
|  | 0.9 | 0.1 |

Sensor Markov assumption -- the assumption that the evidence variable depends only in the corresponding state