1. Reasoning Under Uncertainty using Bayes' Theorem

problem: the patient shows the symptoms $S_1, ..., S_m$ and no other symptoms — what is the probability that the patient has the disease D_i?

available data:

$P(D_i)$:= the a priori probability that the patient sufferes from disease D_i; the probability that the patient has the disease *before* any symptoms have been observed.

$\lambda_{ij} = \frac{P(S_j|D_i)}{P(S_j)}$:= Estimates the relationship between the occurrence of the symptom S_j and the occurrence of the disease D_i. For example: $\lambda_{ij} = 8$ expresses that the symptom occurs 8 times more frequent if the patient has the disease (in this case the symptom provides some positive evidence for the disease). On the other hand, $\lambda_{ij} = 0.125$ expresses that the symptom occurs 8 times less frequent together with the disease(in this case the symptom provides some negative evidence for the occurrence of the disease). Finally, $\lambda_{ij} \approx 1$ expresses that the symptom S_j does not provide any evidence at all for the occurrence of the disease D_i.

solution: Under the conditional independence assumption (concerning the symptoms involved in the reasoning process and concerning the symptoms assuming the disease is present) the probability of having the disease D_i when showing the symptoms $S_1, ..., S_m$ can be calculated as follows:

$$P(D_i|S_1, ..., S_m) = P(D_i) \times \prod_{k=1}^{m} \lambda_{ik}$$

Another formulation that facilitates calculations is the following:

$$log_2(P(D_i|S_1, ..., S_m)) = log_2(P(D_i)) + \sum_{k=1}^{m} log_2(\lambda_{ik})$$

remark: in most systems λ_{ij} is called *the "new evidence multiplier"* — warning: do not mix up with $\frac{P(S_j|D_i)}{P(S_j|\neg D_i)}$ the *odds-multiplier* for rules in PROSPECTOR.

2. Bayesian Reasoning in PROSPECTOR

2.1 PROSPECTOR Rules

Syntax: `(r) if E then H with S=s1, N=n1`

Semantics

- PROSPECTOR rules perform Bayesian reasoning relying on Bayes' Theorem and the Conditional Independence Assumption.
- However, in contrast to classical statistical approaches, PROSPECTOR rules perform computations on odds rather than on probabilities.
- Odds and probabilities are related as follows:
  - $O(H) = \frac{P(H)}{1-P(H))}$
  - $P(H) = \frac{O(H)}{O(H)+1}$
- The processing of a PROSPECTOR rule computes an odds-multiplier based on the probability of P(E'). The odds-multipler indicates how the currents odds of H (O(H)) change based on observing P(E').
- If $P(E')$=0 the multiplier N is used; that is, $O(H|E') = N * O(H)$.
- If $P(E')$=1 the multiplier S is used; that is, $O(H|E') = S * O(H)$.
- If $0 < P(E') < 1$ holds the multiplier of the rule is computed by interpolating between S and N.

2.2 More about S and N

syntax: **if E then (to degree S,N) H**

S:= estimates the belief in the **sufficiency** of H for E. S is estimated by:

$$S = \frac{P(E|H)}{P(E|\neg H)}$$

S measures how strongly the presence E is related to the presence of H; this is, *how stongly the presense of H increases the probability of E.*

N:= estimates the belief in the **necessarity** of H for E. N is estimated by:

$$N = \frac{P(\neg E|H)}{P(\neg E|\neg H)}$$

N measures, *how strongly the presense of H increases the probability of the absense of E*(P(¬E)).

The following relationship must hold between the values given for S and N:

$$\mathbf{S} = \frac{1 - P(\neg E|H)}{1 - P(\neg E|\neg H)}$$

$$= \frac{1 - \mathbf{N} \times P(\neg E|\neg H)}{1 - P(\neg E|\neg H)}$$

special cases:

1) $0 \leq \frac{S-1}{S-N} \leq 1 \Rightarrow$ S and N are legal combinations.
2) $N \geq 1 \Leftrightarrow S \leq 1$

2.3 Thoughts on Interpolation

Interpolation can be done by

- interpolating between multipliers (the log-function and linear interpolation approach)
- interpolating between probabilities (PROSPECTOR's approach)
  - case 1: The prior odds of a rule's left-hand side are not known: use 2-point interpolation function.
  - case 2: The prior odds of a rules left hand side are known: 3-point interpolation function. Main idea: If P(E')=P(E) a rule `if E then H S=s1,N=n1` should have a multiplier of 1!
- **but not** by interpolating between odds.

2.4 Steps of PROSPECTOR's Rule Processing

```
(r) if E then H with S=s1,N=n1
```

Goal of PROSPECTOR's rule processing: compute a rule's odds multiplier λ_r.

Steps:

1. Compute $P(H|E)$ and $P(H|\neg E)$.
2. Compute $P(H|E')$ based on $P(E')$ using PROSPECTOR's 2-point or 3-point interpolation function.
3. Compute $O(H|E')$ by converting $P(H|E')$ to odds.
4. Compute $\lambda_r := \frac{O(H|E')}{O(H)}$.

Remark: PROSPECTOR's interpolation approach is more complicated than the log-interpolation approach (to be discussed later) which directly computes a rule's multiplier based on $P(E')$ by interpolating between S and N.

2.5 Multiplier Computation in PROSPECTOR

We assume that the a priori probabilities $P(E), P(H)$ are given. Furthermore, additional(a posteriori) observations E' have been made. Using the inference techniques of PROSPECTOR we compute $P(H|E \wedge E')$ for a rule

```
(r) if E then H with S=s1,N=n1
```

as follows:

We calculate from the a priori data:

$O(H) = \frac{P(H)}{1-P(H)}$

PROSPECTOR works with odds and not probabilities.

$O(H|E) = S \times O(H)$

$O(H|\neg E) = N \times O(H)$

$P(H|E) = \frac{O(H|E)}{O(H|E)+1}$

$P(H|\neg E) = \frac{O(H|\neg E)}{O(H|\neg E)+1}$

A 2-Point Interpolation Function to Compute $P(H|E')$

$P(H|E') = P(H|E) \times P(E|E') + P(H|\neg E) \times P(\neg E|E') =$
$P(H|E) \times P(E') + P(H|\neg E) \times (1 - P(E'))$

In the above formula we estimate $P(E')$ by the evidence we have got so far for E for the current case; that is, P(E') denotes the a posteriori probability for E — in contrast to P(E) which denotes the a priori probability of E.

A 3-Point Interpolation Function to compute $P(H|E')$

We can compute $P(H|E')$ depending on P(E') using the following 3-point interpolation function **f**:

$$f : [0,1] \rightarrow [0,1]$$

$$f = \begin{cases} P(H|\neg E) + \frac{x}{P(E)} \times (P(H) - P(H|\neg E)) & x \leq P(E) \\ P(H|E) - \frac{1-x}{1-P(E)} \times (P(H|E) - P(H)) & P(E) \leq x \end{cases}$$

f guarantees the consistency of the inferred probability with the a priori probabilities; this is:
$f(P(E)) = P(H)$
$f(0) = P(H|\neg E)$
$f(1) = P(H|E)$

2.6 Evidence Combination in PROSPECTOR

Consider, the following rules provide evidence for a hypothesis H.
$if\ E_1\ then\ H \quad S = s_1; N = n_1$
...
$if\ E_m\ then\ H \quad S = s_m; N = n_m$

$O(H|...)$ *can be computed as follows:*

1) Compute

$$\lambda_i = \frac{O(H|E_i')}{O(H)} \quad for\ i = 1, m.$$

2) Compute

$$O(H|E_1' \wedge ... \wedge E_m') = O(H) \times \prod_{i=1}^{m} \lambda_i$$

remark: λ_i is similar to the "new evidence multipliers" in Bayes's theorem.

2.7 Example: Rule Processing in PROSPECTOR

Consider the following rule set is given:

(R1) If A then B S=10,N=0
(R2) If B then C S=5,N=0.01
(R3) If D then C S=4,N=0.5

The a priori probabilities are: P(B)=0.3; P(C)=0.01.
Moreover, the following a posteriori probabilities have been observed: P(A')=0.9; P(D')=0.2

O(A')= 9
O(B)= 0.43
O(C)= 0.01
O(D')= 0.25
$O(B|A) = 10 * O(B) = 10 * 0.43 = 4.3$
O(B|¬A)= $0 * O(B) = 0$
$O(B|A') = O(B') = 2.7$
$\lambda_{R1}(B) = \frac{2.7}{0.43} = 6.27$
*see (*1)*

$O(C|B) = 5 * O(C) = 0.05$
$O(C|\neg B) = 0.01 * O(C) = 0.0001$
$O(C|B') = 0.03284$
$\lambda_{R2}(C) = 3.3$
*see (*2)*

$O(C|D) = 4 * O(C) = 0.04$
$O(C|\neg D) = 0.5 * O(C) = 0.005$
$O(C|D') = 0.0118$
$\lambda_{R3}(C) = 1.18$
*see (*3)*

$O(C|B' \wedge D') = O(C') = O(C) * \lambda_{R2}(C) * \lambda_{R3}(C) \approx$
$(0.01 * 3.3 * 1.18) = 0.039$

(*1)

$P(A') = 0.9$
$P(B|A) = \frac{4.3}{5.3} = 0.81$
$P(B|\neg A) = 0$
$P(B|A') = 0.9 * 0.81 = 0.73$
$O(B|A') = O(B') = 2.7$

(*2)

$P(B') = 0.73$
$P(C|B) = \frac{0.05}{1.05} = 0.048$
$P(C|\neg B) = 0.0001$
$P(C) = 0.01$
$P(B) = 0.3$
$0.73 > 0.3$, therefore:

$$P(C|B') = 0.048 - \frac{1 - 0.73}{1 - 0.3} \times (0.048 - 0.01) = 0.0318$$

$O(C|B') = 0.033$

(*3)

$P(D') = 0.2$
$P(C|D) = \frac{0.04}{1.04} = 0.0384$
$P(C|\neg D) = 0.005$
$P(C|D') = 0.2 * 0.038 + 0.8 * 0.005 = 0.0116$
$O(C|D') = 0.0118$

2.8 Other Interpolation Functions

The linear multiplier and the log-interpolation perform interpolation based on odds-multipliers rather than based on probabilities. If compared with PROSPECTOR's approach to interpolation, multiplier interpolation is much simpler in terms of computational cost.

The Log-Interpolation Function

Assume we have the following PROSPECTOR-style rule r1 is given:

```
(r1)  If A  then X  S=s1,N=n1
```

The *log-interpolation function* computes r1's multiplier λ_{r1}, based on $P(A')$ as follows:

$$\lambda_{r1} = 10^{log_{10}(s1)*P(A')+log_{10}(n1)*(1-P(A'))}$$

The log-interpolation function should be used if no prior odds for the left-hand hand side condiditions are available, which is frequently the case in pragmatic applications of the Bayesian approach in which S and N approximate weights, rather than real probabilities.

The Linear Interpolation Function

The *linear odds multiplier interpolation function* computes r1's multiplier λ_{r1}, based on $P(A')$ as follows:

$$\lambda_{r1} = s1 * P(A') + n1 * (1 - P(A'))$$

2.9 PROSPECTOR and Bayes' Theorem

$$O(D|S_1 \wedge ... \wedge S_n) =$$

$$\frac{P(D|S_1 \wedge ... \wedge S_n)}{P(\sim D|S_1 \wedge ... \wedge S_n)} =$$

Using Bayes's Theorem we receive:

$$\frac{P(D) * \frac{P(S_1|D)}{P(S_1)} * ... * \frac{P(S_n|D)}{P(S_n)}}{P(\sim D) * \frac{P(S_1|\sim D)}{P(S_1)} * ... * \frac{P(S_n|\sim D)}{P(S_n)}} =$$

Using $\frac{P(S_j|D)}{P(S_j)} = \frac{P(D|S_j)}{P(D)}$ we receive:

$$O(D) * \frac{\frac{P(D|S_1)}{P(D)}}{\frac{P(\sim D|S_1)}{P(\sim D)}} * ... =$$

$$O(D) * \frac{O(D|S_1)}{O(D)} * ... * \frac{O(D|S_n)}{O(D)}$$

The last formula is identical with PROSPECTOR's formula for combination; that is, PROSPECTOR's combination formula is based on Bayes' Theorem: More precisely, it is odds-based reformulation of the probability based definition of Bayes' Theorem.

3. The Conditional Independence Assumption

Applying Bayes' theorem the a posterori probability of a disease D_j can be computed as follows:

$$P(D_j|S_1 \wedge ...S_n) = \frac{P(D_j) * P(S_1 \wedge ... \wedge S_n|D_j)}{P(S_1 \wedge ... \wedge S_n)}$$

Assuming that we have a diagnostic problem that involves 50 diseases and 500 symptoms, more than

$50 * 2^{500} \approx 10^{150}$

conditional probabilities are needed for a dignostic expert system of this size. However, frequently, the following simplified formula is used instead, which relies on the *conditional independence assumption*:

$$P(D_j|S_1 \wedge ...S_n) \approx \frac{P(D_j) * P(S_1|D_j) * ... * P(S_n|D_j)}{P(S_1) * ... * P(S_n)}$$

With the second formula only appoximately 25000 conditional probabilities are needed; it sacrifices precision in order to reduce the enourmous knowledge acquisition costs in diagnostic expert systems. In general, if the conditional independence assumption is not valid, computation errors will occur, especially if small probabilities are involved.

Final Remarks:

- *In summary, we frequently have to tolerate imprecision by making assumptions (such as the conditional independence assumption) in the design of a system in order to make it feasible to get such a system running, only spending a limited amount of time and money.*
- Bayesian approaches that rely on the conditional independence assumption are nowadays called *naive Bayesian approaches* in the literature. More complicated approaches, such as *Bayesian networks* (that were introduced by Judea Pearl) and *influence networks*, that do not rely on the condition independence assumption have been introduced in the last decade in the literature.

4. Automating Decision Making using Bayesian Rules

This can be done

- *in the true spirit of Bayesian approaches.* This requires that prior odds, and conditional probabilities are known for the task to be solved. In this case, S and N and prior odds can be directly computed from statistical data.
- *pragmatically* — in the spirit of soft computing. Frequently, prior odds and conditional probabilities are not known, or too expensive to obtain. In the pragmatic approach, odds multipliers approximate weights a particular piece of evidence carries for a particular decision candidate, but not precise probabilities.

Remark: Most real world expert systems that employ Bayesian techniques relying on the pragmatic approach.

5. Bayesian versus Fuzzy Approaches

- Bayesian approaches rely on *probabilities* in their theories. Probabilities cope with randomness in our world and enable us to make predictions with which frequency events occur in the future.
- Fuzzy set theory relies on *possibilities.* Possibilities cope with the degree an object belongs to a particular class / set — possibilities are used to describe the semantics of terms in a terminology, measuring to which extend an object matches the description of a particular term. Possibilities approximate the meaning of terms, and they have nothing to do with the frequency a particular object or event occurs in reality.
- Some approaches try to combine Bayesian reasoning with fuzzy sets by supporting fuzzy sets on the left-hand side of Bayesian rules, whose possibilities are be interpreted probabilistically by the underlying Bayesian rules; that is, possibilities are directly converted into probabilities and decision making by evidence combination is provided under the umbrella of Bayes' Theorem and the Conditional Independence Assumption. The scholarship committee rule-set is a good example of a rule-set that uses fuzzy sets on the left-hand side of Bayesian, PROSPECTOR-style rules.

6. *Application to the A-B-C Example in PROSPECTOR*

A PROSPECTOR-Ruleset

`decision candidates:  D1, D2, D3`

`(r1) if A then D1 with` $S = s_1, N = n_1$
`(r2) if C then D1` $S = s_2, N = n_2$
`(r3) if B then D2` $S = s_3, N = n_3$

solutions parameters: $n_1, n_2, n_3, s_1, s_2, s_3, O(D1), O(D2), O(D3)$.

Choose the 9 parameters so that the following 5 equations are satisfied:

Equations:

$(I) s_1 \oplus s_2 \oplus O(D1) > s_3 \oplus O(D2)$
$(II) s_3 \oplus O(D2) > s_1 \oplus n_2 \oplus O(D1) > O(D3)$
$(III) s_3 \oplus O(D2) > s_2 \oplus n_1 \oplus O(D1) > O(D3)$
$(IV) O(D3) > n_3 \oplus O(D2)$
$(V) O(D3) > n_1 \oplus n_2 \oplus O(D1)$

In the case of PROSPECTOR $\oplus := *$.