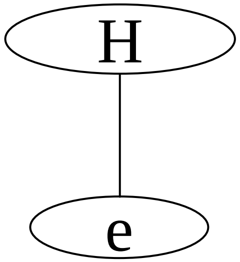
ECE 566

Bayesian Inference I

<u>Combining Predictive and Diagnostic Supports:</u>



$$\underbrace{\frac{P(H\,|\,e)}{P(\neg H\,|\,e)}}_{\text{posterior odds}} = \underbrace{\frac{P(e\,|\,H)}{P(e\,|\,\neg H)}}_{\text{likelihood ratio}} \underbrace{\frac{P(H)}{P(\neg H)}}_{\text{prior odds}}$$

$$O(H|e) = L(e|H)\; O(H)$$

This formula allows us to update our belief about *H* once we have observed evidence *e*.

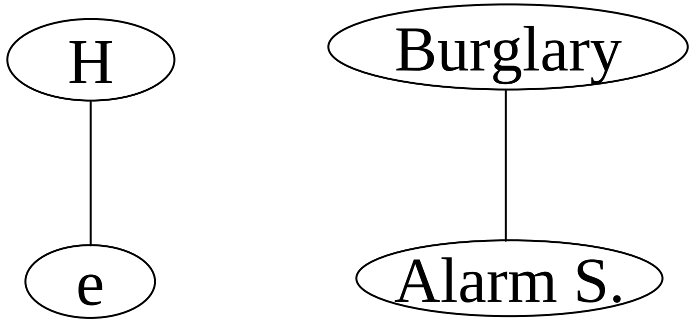
Ex:

You are awakened one night by the sound of your house alarm. Every night one in ten thousand homes gets burglarized. There is a 95% chance that a burglary attempt triggers the alarm, there is a 1% chance that the alarm triggers by other reasons such as malfunction. What is the probability that your house is being burglarized?

$$P(\text{Alarm} \mid \text{Burglary}) = 0.95$$

$$P(\text{Alarm} \mid \neg\text{Burglary}) = 0.01$$

$$P(\text{Burglary}) = 10^{-4}$$



$$O(\text{Burglary} \mid \text{Alarm}) = L(\text{Alarm} \mid \text{Burglary}).O(\text{Burglary})$$

$$L(\text{Alarm} \mid \text{Burglary}) = \frac{P(\text{Alarm} \mid \text{Burgalry})}{P(\text{Alarm} \mid \neg\text{Burglary})}$$

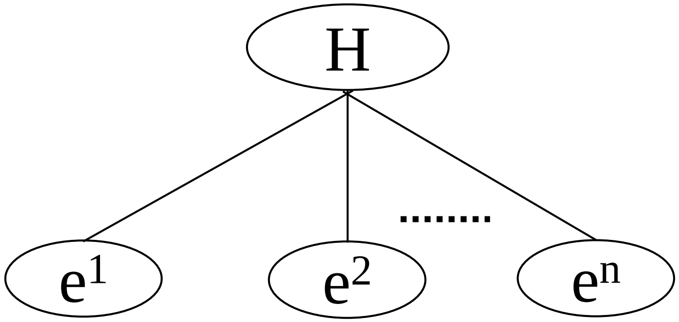
$$\mathrm{L(Alarm \mid Burglary)} = \frac{0.95}{0.01}$$

$$\mathrm{O(Burglary)} = \frac{\mathrm{P(Burglary)}}{\mathrm{P(\neg Burglary)}} = \frac{10^{-4}}{1 - 10^{-4}}$$

$$\mathrm{O(Burglary \mid Alarm)} = 0.0095$$

$$\mathrm{P(Burglary \mid Alarm)} = \frac{0.0095}{1 + 0.0095} = 0.00941$$

Pooling of Evidences



Assume that the alarm systems consists of *n* devices, and each produces a different sign.

Let e^k stand for evidence k (k[th] detector):

e_1^k evidence k confirms the hypothesis

e_0^k evidence k disconfirms

$$L(e_1^k \mid H) = \frac{P(e_1^k \mid H)}{P(e_1^k \mid \neg H)}$$

The combined belief is obtained from:

$$\begin{aligned} O(H \mid e^1, e^2, \ldots, e^n) &= L(e^1, e^2, \ldots, e^n \mid H)\, O(H) \\ &= L(e^1 \mid H).L(e^2 \mid H) \ldots L(e^n \mid H)\, O(H) \\ &= O(H) \prod_{k=1}^{n} L(e^k \mid H) \end{aligned}$$

assuming that the *n* devices operate independent of each other.

Recursive Bayesian Updating

Suppose we have observed n evidences $\vec{e}^{\,n} = e^1, e^2, \ldots, e^n$ regarding a hypothesis H.

Now, a new evidence e’ becomes available. It needs to be incorporated into the previous results.

Since evidences are assumed to be independent:

$$P(e' \mid \vec{e}^{\,n}, H) = P(e' \mid H)$$

$$P(e' \mid \vec{e}^{\,n}, \neg H) = P(e' \mid \neg H)$$

Thus:

$$O(H \mid \vec{e}^{\,n}, e') = O(H \mid \vec{e}^{\,n})\, L(e' \mid H)$$

So to update the belief, multiply the current posterior odds by the likelihood ration of e’.

If we take the log of the above formula, we get an incremental updating process.

$$\log O(H \mid \vec{e}^{\,n}, e') = \log O(H \mid \vec{e}^{\,n}) + \log L(e' \mid H)$$

This is the weight carried by evidence e'.

Evidence supporting the hypothesis carries a positive weight and that opposing it carries a negative weight.

If we later find that one of the evidences was erroneous, we can rectify the error using:

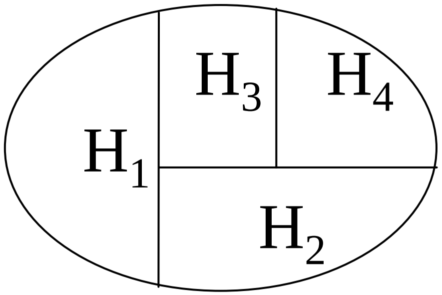
$$\Delta = \log L(e^{c} \mid H) - \log L(e^{w} \mid H)$$

$$\text{where } e^{c} = e^{\text{correct}}$$

$$e^{w} = e^{\text{wrong}}$$

Multi-Valued Hypotheses

The outcome of a hypothesis could be one of several states.



For example, burglary could be break-in through the door, or break-in through the window. Similarly evidence may have several modes.

- Refine the hypothesis space, and group the hypotheses into multi-valued variables. Represent conditional probabilities relating the hypothesis outcomes and evidences with a matrix.

Ex:

Using burglary, assign H_1, H_2, H_3 and H_4 as follows:

H_1 = No burglary, animal entry.

H_2 = Attempted burglary, window break-in.

H_3 = Attempted burglary, door break-in.

H_4 = No burglary, no entry.

Each evidence, e^k has the following possible values:

$$e_1^k = \text{no sound}$$

$$e_2^k = \text{low sound}$$

$$e_3^k = \text{high sound}$$

Represent the conditional probabilities by a matrix:

$P(e_j^k \mid H_i) =$ element *i*, *j* in the matrix represents the conditional probability between the j^{th} value of evidence *k* and hypothesis H_i.

$$P(e_j^k \mid H_i) = \begin{array}{c} \\ H_1 \\ H_2 \\ H_3 \\ H_4 \end{array} \begin{array}{c} \begin{matrix} e_1^k & e_2^k & e_3^k \end{matrix} \\ \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.06 & 0.5 & 0.44 \\ 0.5 & 0.1 & 0.4 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

To compute total belief from a set of n evidences, do the following:

Let

$$\vec{\lambda}_i^k = [P(e_i^k \mid H_1)\ P(e_i^k \mid H_2) \ldots P(e_i^k \mid H_m)]$$

In this case 4 outcomes for the hypothesis

$$\Lambda_i = \prod \vec{\lambda}_i^k$$

This is not traditional vector product, it is the product of vectors term by term.

then:

$$P(H_i \mid e^1, e^2, \ldots, e^n) = \alpha\, P(H_i)\, \Lambda_i$$

α is a normalizing factor which will be set to ensure the posterior probabilities for H_i sum up to 1.

Ex: In our last burglary example, assume we have two alarms each with properties given by the previous matrix. Let's assume the prior probabilities are :

$$\vec{P}(H_i) = \begin{bmatrix} 0.099 \\ 0.009 \\ 0.001 \\ 0.891 \end{bmatrix}$$

We hear our first detector issuing a high sound. The second detector in our system is silent.

e^1 = high sound
e^2 = silent

$$\vec{\lambda}_3^1 = \begin{bmatrix} P(e_3^1 \mid H_1) \\ P(e_3^1 \mid H_2) \\ P(e_3^1 \mid H_3) \\ P(e_3^1 \mid H_4) \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.44 \\ 0.4 \\ 0 \end{bmatrix}$$

$$\vec{\lambda}_1^2 = \begin{bmatrix} P(e_1^2 \mid H_1) \\ P(e_1^2 \mid H_2) \\ P(e_1^2 \mid H_3) \\ P(e_1^2 \mid H_4) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.06 \\ 0.5 \\ 1 \end{bmatrix}$$

$$\Lambda = \vec{\lambda}^{\,1} \vec{\lambda}^{\,2} = \begin{bmatrix} 0.1 \\ 0.44 \\ 0.4 \\ 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.06 \\ 0.5 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.05 \\ 0.0264 \\ 0.2 \\ 0 \end{bmatrix}$$

$$P(H_i \mid e^1, e^2) = \alpha\, \vec{P}(H_i)\, \Lambda$$

$$= \alpha \begin{bmatrix} 0.099 \\ 0.009 \\ 0.001 \\ 0.891 \end{bmatrix} \begin{bmatrix} 0.05 \\ 0.0264 \\ 0.2 \\ 0 \end{bmatrix}$$

$$= \alpha\, .10^{-3} \begin{bmatrix} 4.95 \\ 0.238 \\ 0.2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.919 \\ 0.0439 \\ 0.0375 \\ 0 \end{bmatrix}$$

Arrival of information at different times

We can update belief incrementally by using earlier posterior probabilities as priors for later arriving information. Let's say that we first observe a high sound from our 1st device.

$$P(H_i \mid e^1) = \alpha \, \vec{\lambda}^1 \, \vec{P}(H_i) = \alpha \begin{bmatrix} 0.0099 \\ 0.00396 \\ 0.0004 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.694 \\ 0.277 \\ 0.028 \\ 0 \end{bmatrix}$$

Later we obtain information from our 2nd device:

$$P(H_i \mid e^1, e^2) = \alpha' \vec{\lambda}^2 \vec{P}(H_i \mid e^1)$$

$$= \alpha' \begin{bmatrix} 0.347 \\ 0.0166 \\ 0.014 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.919 \\ 0.0439 \\ 0.0375 \\ 0 \end{bmatrix}$$