BAYESIAN STATISTICS, INFORMATION, AND SIGNAL DETECTION

M. M. Taylor Feb 5, 1966

SOME DEFINITIONS AND DERIVATIONS

$P_A(B) \equiv P(B|A)$ = probability of B given A is true

Bayes' Theorem

$$P_D(H) = \frac{P(H)\,P_H(D)}{P(D)} \quad \text{derived from} \quad P(D)\,P_D(H) = P(H,D) = P(H)\,P_H(D)$$

whence

$$\frac{P_D(H)}{P_D(J)} = \frac{P(H)}{P(J)}\;\frac{P_H(D)}{P_J(D)}$$

In words

posterior odds = prior odds x likelihood ratio (2 hypotheses)

or

posterior assessment = prior assessment x likelihood gain (any number of hypotheses)

ASSESSMENT FUNCTIONS — definitions

$$A_D(H_1, H_2, \text{----}\, H_n) = \{cP_D(H_1), cP_D(H_2), \text{----}\, cP_D(H_n)\}$$

where c is an arbitrary constant.

If A is normalized, the notation is unchanged, but $1/c = \sum_i P_D(H_i)$

$$A_D(\mathbb{H}) = \{cP_D(H)\}$$

where $\mathbb{H}$ is a hypothesis space with members H

$$A_D(\mathbb{H}|E) = \{cP_{D,E}(H)\}$$

where E is some observation taken before observation D

$$\mathcal{A}_D(\mathbb{H}) = \{b + \log P_D(H)\}$$

if A is normalized, $\mathcal{A}$ is normalized by taking $b = \log c$

GAIN FUNCTIONS – definitions

$$G_D(\mathbb{H}) = \{k\,P_{\underset{H}{\cancel{\mathbb{H}}}}(D)\}$$

$$G_D(\mathbb{H}|E) = \{k\,P_{H,E}(D)\}$$

note the essential difference between $A_D(\mathbb{H}|E)$ and $G_D(\mathbb{H}|E)$

$$\mathscr{g}_D(\mathbb{H}) = \{m + \log P_H(D)\}$$

REVISION OF ASSESSMENT PROBABILITIES

$$A_D(\mathbb{H}) = A(\mathbb{H})\,G_D(\mathbb{H})$$

or

$$\mathscr{A}_D(\mathbb{H}) = \mathscr{A}(\mathbb{H}) + \mathscr{g}_D(\mathbb{H})$$

for multiple observations

$$\mathscr{A}_{D_{1\cdot n}}(\mathbb{H}) = \mathscr{A}(\mathbb{H}) + \mathscr{g}_{D_1}(\mathbb{H}) + \mathscr{g}_{D_2}(\mathbb{H}|D_1) + \text{-----} + \mathscr{g}_{D_n}(\mathbb{H}|D_1, D_2, \cdots D_{n-1})$$

INFORMATION or UNCERTAINTY

The uncertainty about the world of hypotheses $\mathbb{H}$ is given by

$$U(\mathbb{H}) = -\sum_{\mathbb{H}} A(\mathbb{H})\,\mathscr{A}(\mathbb{H}) \qquad \text{when } A(\mathbb{H}) \text{ is normalized}$$

$$= -\sum_{\mathbb{H}} P(H)\log P(H)$$

after the 'observation' with result D, the uncertainty is

$$U(\mathbb{H}|D) = -\sum_{\mathbb{H}} A_D(\mathbb{H})\,\mathscr{A}_D(\mathbb{H})$$

The observation of D has resulted in a reduction of uncertainty

$$U_D(\mathbb{H}) = U(\mathbb{H}) - U(\mathbb{H}|D)$$

which I shall call the "informativeness" of D about $\mathbb{H}$

MMT 5Feb 1966

INFORMATIVENESS IN A FINITE EXPERIMENT

Hypotheses in the experiment are $\mathbb{H} = \{H_i\}$

Possible results of an observation are $\mathbb{D} = \{D_j\}$

The experiment may be characterized, before the observation, by the matrix whose elements contain the joint probabilities that the i^{th} hypothesis is true and that the j^{th} datum will result from the observation

$$
\begin{array}{c|c|cccc|c}
 & & & & \mathbb{D} & & \\
 & j = & 1 & 2 & \cdots & m & \Sigma \\
\hline
 & i = 1 & p_{11} & p_{12} & & p_{1m} & P(H_1) \\
 & 2 & p_{21} & p_{22} & & p_{2m} & P(H_2) \\
\mathbb{H} & \vdots & & & & & \\
 & n & p_{n1} & p_{n2} & & p_{nm} & P(H_n) \\
\hline
 & \Sigma & P(D_1) & P(D_2) & & P(D_m) & 1
\end{array}
$$

Setting $c_k = \dfrac{1}{P(D_k)}$

$$U(\mathbb{H}|D_k) = -\sum_i c_k p_{ik} \log c_k p_{ik}$$

$$= -c_k \sum_i p_{ik} \log p_{ik} \quad - \log c_k$$

Expected value of $U(\mathbb{H}|D_k)$ is

$$E[U(\mathbb{H}|D_k)] = \sum_k P(D_k)\, U(\mathbb{H}|D_k)$$

$$= -\sum_k \sum_i p_{ik} \log p_{ik} + \cancel{\sum_k p_k \log p} \sum_k P(D_k) \log P(D_k)$$

$$= U(\mathbb{H}, \mathbb{D}) - U(\mathbb{D})$$ in Garner's terms

But

$$U(\mathbb{H},\mathbb{D}) = U(\mathbb{H}) + U(\mathbb{D}) - U(\mathbb{H}:\mathbb{D})$$

and the informativeness of D_k about $\mathbb{H}$ is

$$U_{D_k}(\mathbb{H}) = U(\mathbb{H}) - U(\mathbb{H}|D_k)$$

$$\underset{k}{E}[U_{D_k}(\mathbb{H})] = E[U(\mathbb{H})] - \underset{k}{E}[U(\mathbb{H}|D_k)]$$

$$= U(\mathbb{H}) + U(\mathbb{D}) - U(\mathbb{H},\mathbb{D})$$

$$= U(\mathbb{H}:\mathbb{D})$$

The expected reduction in the uncertainty of the assessment function over the hypothesis space $\mathbb{H}$ is just the contingent uncertainty between $\mathbb{H}$ and $\mathbb{D}$.

An immediate consequence of this equality is

$$\underset{k}{E}[U_{D_k}(\mathbb{H})] \leq U(\mathbb{D}) \qquad \text{or}$$

An observation cannot be expected to be more informative than it is uncertain.