Chapter 5

Bayesian Inference

Bayesian inference is a probability-based reasoning discipline grounded in Bayes' rule. When used to support data fusion, Bayesian inference belongs to the class of data fusion algorithms that use *a priori* knowledge about events or objects in an observation space to make inferences about the identity of events or objects in that space. Bayesian inference provides a method for calculating the conditional *a posteriori* probability of a hypothesis being true given supporting evidence. Thus, Bayes' rule offers a technique for updating beliefs in response to information or evidence that would cause the belief to change.

5.1 Bayes' rule

Bayes' rule may be derived by evaluating the probability of occurrence of an arbitrary event E assuming that another event H has occurred. The probability is given by[1]

$$P(E \mid H) = \frac{P(EH)}{P(H)}, \tag{5-1}$$

where H is an event with positive probability. The quantity $P(E|H)$ is the probability of E conditioned on the occurrence of H. The conditional probability is not defined when H has zero probability. The factor $P(EH)$ represents the probability of the intersection of events E and H.

To illustrate the meaning of Eq. (5-1), consider a population of N people that includes N_E left-handed people and N_H females as shown in the Venn diagram of Figure 5.1. Let E and H represent the events that a person chosen at random is left-handed or female, respectively. Then

$$P(E) = N_E/N \tag{5-2}$$

and

$$P(H) = N_H/N. \tag{5-3}$$

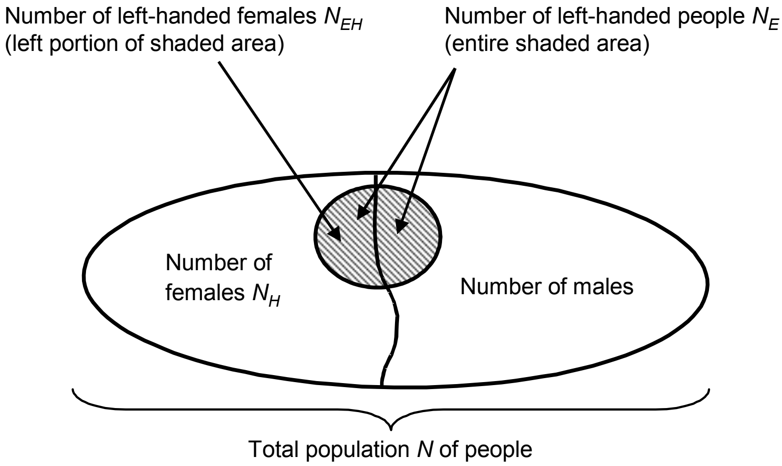

Figure 5.1 Venn diagram illustrating intersection of events *E* (person chosen at random is left-handed) and *H* (person chosen at random is female).

The probability that a female chosen at random is left-handed is N_{EH}/N_H, where N_{EH} is the number of left-handed females. In this example, $P(E|H)$ denotes the probability of selecting a left-handed person at random assuming the person is female. In terms of population parameters, $P(E|H)$ is

$$P(E \mid H) = \frac{N_{EH}}{N_H} = \frac{P(EH)}{P(H)}. \tag{5-4}$$

Returning to the derivation of Bayes' rule, Eq. (5-1) may be rewritten as

$$P(EH) = P(E \mid H)\, P(H), \tag{5-5}$$

which is referred to as the theorem on compound probabilities.

When H consists of a set of mutually exclusive and exhaustive hypotheses $H_1, \ldots,$ H_n, conditional probabilities, which may be easier to evaluate than unconditional probabilities, can be substituted for $P(EH)$ as follows. The mutually exhaustive property implies that one hypothesis necessarily is true, i.e., the union of $H_1, \ldots,$ H_n is the entire sample space. Under these conditions, any event E can occur only in conjunction with some H_j such that

$$E = EH_1 \cup EH_2 \cup \ldots \cup EH_n. \tag{5-6}$$

Since the EH_j are mutually exclusive, their probabilities add as

$$P(E) = \sum_{i=1}^{n} P(E H_i) . \tag{5-7}$$

Upon substituting H_j for H and summing over i, Eq. (5-5) becomes

$$P(E) = \sum_i [P(E \mid H_i)\ P(H_i)] , \tag{5-8}$$

when the identity in Eq. (5-7) is applied.

Equation (5-8) states that the belief in any event E is a weighted sum over all the distinct ways that E can be realized.

In Bayesian inference, we are interested in the probability that hypothesis H_i is true given the existence of evidence E. This statement is expressed as

$$P(H_i \mid E) = \frac{P(EH_i)}{P(E)} . \tag{5-9}$$

If Eqs. (5-5) and (5-8) are introduced into Eq. (5-9), Eq. (5-9) takes the form of Bayes' rule as

$$P(H_i \mid E) = \frac{P(E \mid H_i)\ P(H_i)}{P(E)} = \frac{P(E \mid H_i)\ P(H_i)}{\sum_i [P(E \mid H_i)\ P(H_i)]} , \tag{5-10}$$

where

$P(H_i \mid E)$ = *a posteriori* or posterior probability that hypothesis H_i is true given evidence E,

$P(E \mid H_i)$ = probability of observing evidence E given that H_i is true (sometimes referred to as the likelihood function),

$P(H_i)$ = *a priori* or prior probability that hypothesis H_i is true,

$$\sum_i P(H_i) = 1 , \tag{5-11}$$

and

$\sum_i P(E \mid H_i) P(H_i) =$ preposterior or probability of observing evidence E given that hypothesis H_i is true, summed over all hypotheses i.

The likelihood functions represent the extent to which the posterior probability is subject to change. These functions are evaluated either through offline experiments or by analyzing the information available to support the problem at hand. The preposterior is simply the sum of the products of the likelihood functions and the *a priori* probabilities and serves as a normalizing constant.[2]

5.2 Bayes' rule in terms of odds probability and likelihood ratio

Further insight into the interpretation of Bayes' rule is gained when Eq. (5-10) is divided by $P(\overline{H}_i \mid E)$, where $\overline{H}_i$ represents the negation of H_i. Thus,

$$\frac{P(H_i \mid E)}{P(\overline{H}_i \mid E)} = \frac{P(E \mid H_i)\, P(H_i)}{P(E)\, P(\overline{H}_i \mid E)]} = \frac{P(E \mid H_i)\, P(H_i)}{P(E)\dfrac{P(E\overline{H}_i)}{P(E)}} = \frac{P(E \mid H_i)\ P(H_i)}{P(E \mid \overline{H}_i)\ P(\overline{H}_i)}, \tag{5-12}$$

where Eq. (5-5) has been applied to convert $P(E\overline{H}_i)$ into the form shown in the last iteration of the equation.

If the prior odds are defined as

$$O(H_i) = P(H_i)/[1 - P(H_i)] = P(H_i)/\,P(\overline{H}_i)\,, \tag{5-13}$$

the likelihood ratio as

$$L(E \mid H_i) = P(E \mid H_i)/P(E \mid \overline{H}_i)\,, \tag{5-14}$$

and the posterior odds as

$$O(H_i \mid E) = P(H_i \mid E)/P(\overline{H}_i \mid E)\,, \tag{5-15}$$

then the posterior odds can also be written in product form as

$$O(H_i \mid E) = L(E \mid H_i)\ O(H_i). \tag{5-16}$$

Thus, Bayes' rule implies that the overall strength of belief in hypothesis H_i, based on previous knowledge and the observed evidence E, is based on two factors: the prior odds $O(H_i)$ and the likelihood ratio $L(E|H_i)$. The prior odds factor is a measure of the predictive support given to H_i by the background knowledge alone, while the likelihood ratio represents the diagnostic or retrospective support given to H_i by the evidence actually observed.[2]

Although the likelihood ratio may depend on the content of the knowledge base, the relationship that controls $P(E|H_i)$ is dependent on somewhat local factors when causal reasoning is used. Thus, when H_i is true, the probability of event E can be estimated in a natural way that is not dependent on many other propositions in the knowledge base. Accordingly, the conditional probabilities $P(E|H_i)$ (i.e., the likelihood function), as opposed to the posterior probabilities $P(H_i|E)$, are the fundamental relationships in Bayesian analysis. The conditional probabilities $P(E|H_i)$ possess features that are similar to logical production rules. They convey a degree of confidence stated in rules such as "If H then E," a confidence that persists regardless of what other rules or facts reside in the knowledge base.[2]

As an example of how to compute the posterior probability using the prior odds and likelihood ratio, consider a patient who visits a physician who administers a low cost screening test for cancer. Assume: (1) there is a 95 percent chance that the test administered to detect cancer is correct when the patient has cancer, i.e., P(test positive|cancer) = 95 percent; (2) based on previous false alarm history, there is a slight chance (4 percent) that the positive test result will occur when the patient does not have cancer, i.e., P(test positive|no cancer) = 4 percent; (3) historical data indicate that cancer occurs in 5 out of every 1,000 people in the general population, i.e., P(cancer) = 0.005. What is the probability that the patient has cancer given a positive test result?

Applying Eq. (5-16) gives

$$\begin{aligned} O(\text{cancer}|\text{test positive}) &= L(\text{test positive}|\text{cancer})\, O(\text{cancer}) \\ &= \frac{0.95}{0.04}\,\frac{0.005}{1-0.005} = 0.119. \end{aligned} \tag{5-17}$$

The general relation between $P(A)$ and $O(A)$ is obtained by rearranging the factors in Eq. (5-13) as

$$P(A) = O(A)/[1 + O(A)]. \tag{5-18}$$

Therefore,

$$P(\text{cancer}|\text{test positive}) = 0.119/[1 + 0.119] = 10.7 \text{ percent}. \tag{5-19}$$

Thus, the retrospective support given to the cancer hypothesis by the test evidence (through the likelihood ratio) has increased its degree of belief by approximately a factor of 20, from five in a thousand to 107 in a thousand.

5.3 Direct application of Bayes' rule to cancer screening test example

In Section 5.2, the prior odds and likelihood ratio were used to compute the probability of a patient having cancer given a positive test result. The same type of calculation may be made by directly applying Bayes' rule.[3] In this formulation, the problem statement is as follows. Suppose a patient visits his physician who proceeds to administer a low cost screening test for cancer. The test has an accuracy of 95 percent (i.e., the test will indicate positive 95 percent of the time if the patient has the disease) with a 4 percent false alarm probability. Furthermore, suppose that cancer occurs in 5 out of every 1,000 people in the general population. If the patient is informed that he has tested positively for cancer, what is the probability he actually has cancer?

The Bayesian formulation of Eq. (5-10) predicts the required probability as

$$P(\text{patient has cancer} \mid \text{test positive}) = \frac{P(\text{test positive} \mid \text{cancer}) \;\; P(\text{cancer})}{P(\text{test positive})}, \tag{5-20}$$

where

$$\begin{aligned} P(\text{test positive}) = \; & P(\text{test positive} \mid \text{cancer})\, P(\text{cancer}) \\ & + P(\text{test positive} \mid \text{no cancer})\, P(\text{no cancer}). \end{aligned} \tag{5-21}$$

The probability $P(\text{test positive} \mid \text{no cancer})$ is the false alarm probability or Type 1 error. The Type 2 error is the probability of missing the detection of cancer in a patient with the disease.

Upon substituting the statistics for this example into Eq. (5-20), we find

$$P(\text{patient has cancer} \mid \text{test positive}) = \frac{(0.95)(0.005)}{(0.95)(0.005) + (0.04)(0.995)} = 0.107 \tag{5-22}$$

or 10.7 percent, the same value as found using the prior odds and likelihood ratio formulation of the problem.

Intuitively, this result may appear smaller than expected. It asserts that in only 10.7 percent of the cases in which the test gives a positive result and declares cancer to be present is it actually true that cancer is present. Further testing is thus required when this type of initial test is administered. The screening test may be said to be reliable because it will detect cancer in 95 percent of the cases in which cancer is present. However, the critical Type 2 error is 0.05, implying that the test will not diagnose 1 in 20 cancers.

To increase the probability of the patient actually having cancer, given a positive test, and concurrently reduce the Type 2 error requires a test with a greater accuracy. A more effective method of increasing the *a posteriori* probability is to reduce the false alarm probability. If, for example, the test accuracy is increased to 99.9 percent and the false alarm probability reduced to 1 percent, the probability of the patient actually having cancer, given a positive test, is increased to 33.4 percent. The Type 2 error now implies a missed diagnosis in only 1 out of 1,000 patients. Increasing the test accuracy to 99.99 percent has a minor effect on the *a posteriori* probability, but reduces the Type 2 error by another order of magnitude.

In other situations, the Type 1 error may be the more serious error. Such a case occurs if an innocent man is tried for a crime and his freedom relied on the outcome of a certain experiment. If a hypothesis corresponding to his innocence was constructed and was rejected by the experiment, then an innocent man would be convicted and a Type 1 error would result. On the other hand, if the man was guilty and the experiment accepted the hypothesis corresponding to innocence, the guilty man would be freed and a Type 2 error would result.[4]

5.4 Comparison of Bayesian inference with classical inference

Bayes' formulation of conditional probability is satisfying for several reasons. First, it provides a determination of the probability of a hypothesis being true, given the evidence. By contrast, classical inference gives the probability that an observation can be attributed to an object or event, given an assumed hypothesis. Second, Bayes' formulation allows incorporation of *a priori* knowledge about the likelihood of a hypothesis being true at all. Third, Bayes permits the use of subjective probabilities for the *a priori* probabilities of hypotheses and for the probability of evidence given a hypothesis when empirical data are not available. This attribute permits a Bayesian inference process to be applied to multisensor fusion since probability density functions are not required. However, the output of such a process is only as good as the input *a priori* probability data.

Bayesian inference therefore resolves some of the difficulties that occur with classical inference methods as shown in Table 5.1. However, Bayesian methods require the *a priori* probabilities and likelihood functions be defined, introduce complexities when multiple hypotheses and multiple conditional dependent events are present, require that competing hypotheses be mutually exclusive, and cannot support an uncertainty class as does Dempster-Shafer.[5,6] The types of information needed to apply classical inference, Bayesian inference, Dempster-Shafer evidential theory, and other classification and identification data fusion algorithms to a target identification application are compared and summarized in Chapter 11.

Table 5.1 Comparison of classical and Bayesian inference.

Classical	Bayesian
Features of the model:	
Probability model links observed data and a population.	Probability of a hypothesis being true is determined from known evidence.
Probability model is usually empirically based on parameters calculated from a large number of samples.	Likelihood of a hypothesis is updated using a previous likelihood estimate and additional evidence.
A number of decision rules may be used to decide between the null hypothesis H_0 and an opposing hypothesis H_1, including maximum likelihood, Neyman-Pearson, and minimax. Other cost functions available for use with Bayesian inference are maximum *a posteriori* and Bayes.[4,7,8] *	Either classical probabilities or subjective probability estimates may be used (i.e., probability density functions are not necessarily required). Subjective probabilities are inferred from experience and vary from person to person. Supports more than two hypotheses at a time.
Disadvantages:	
When generalized to include multi-dimensional data from multiple sensors, *a priori* knowledge and multidimensional probability density functions are required.	*A priori* probabilities and likelihoods must be defined.
Generally, only two hypotheses can be assessed at a time, namely H_0 and H_1.	Complexities are introduced when multiple hypotheses and multiple conditional dependent events are present.
Multivariate data produce evaluation complexities.	Competing hypotheses must be mutually exclusive.
A priori assessments are not utilized.	Cannot support an uncertainty class.

**Maximum likelihood:* Accepts hypothesis H_0 as true if the probability $P(H_0)$ of H_0 multiplied by $P(y\,|\,H_0)$ is greater than $P(H_1) \times P(y\,|\,H_1)$.

Neyman-Pearson: Accepts the hypothesis H_0 if the ratio of the likelihood function for H_0 to the likelihood function for H_1 is less than or equal to a constant c. The constant is selected to give the desired significance level.

Minimax: A cost function is constructed to quantify the risk or loss associated with choosing a hypothesis or its alternative. The minimax approach selects H_0 such that the maximum possible value of the cost function is minimized.

Maximum a posteriori: Accepts hypothesis H_0 as true if the probability $P(H_0\,|\,y)$ of H_0 given observation y is greater than the probability $P(H_1\,|\,y)$ of H_1 given observation y.

Bayes: A cost function is constructed that provides a measure of the consequences of choosing hypothesis H_0 versus H_1. This decision rule selects the hypothesis that minimizes the cost function based on detection and false alarm probabilities.

5.5 Application of Bayesian inference to fusing information from multiple sources

Figure 5.2 illustrates the Bayesian inference process as applied to the fusion of multisensor identity information. In this example, multiple sensors observe parametric data [e.g., infrared signatures, radar cross section, pulse repetition interval, rise and fall times of pulses, frequency spectrum signal parameters, and identification-friend-or-foe (IFF)] about an entity whose identity is unknown.

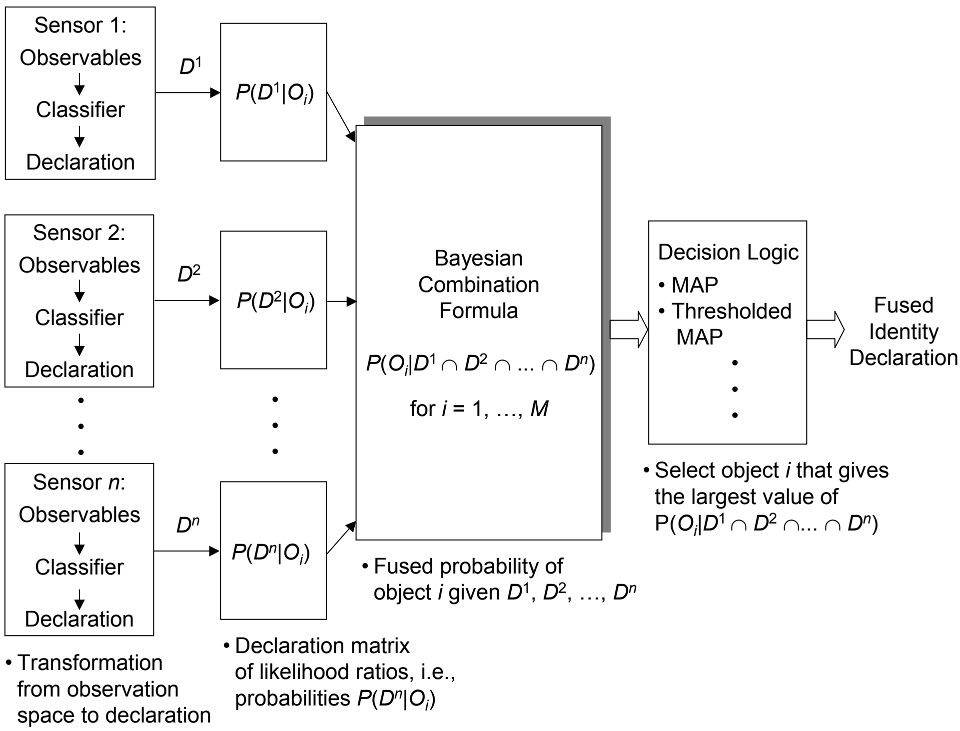

Figure 5.2 Bayesian fusion process. (Adapted from E. Waltz and J. Llinas, *Multisensor Data Fusion*, Artech House, Norwood, MA [1990].)

Each of the sensors provides an identity declaration D or hypothesis about the object's identity based on the observations and a sensor-specific algorithm. The previously established performance characteristics of each sensor's classification algorithm (developed either theoretically or experimentally) provide estimates of the likelihood function, that is, the probability $P(D|O_i)$ that the sensor will declare the object to be a certain type, given that the object is in fact type i. These declarations are then combined using a generalization of Eq. (5-10) to produce an updated, joint probability for each entity O_i based on the multisensor declarations.

Thus, the probability of having observed object i from the set of M objects given declaration (evidence) D^1 from Sensor 1, declaration D^2 from Sensor 2, etc., is

$$P(O_i \,|\, D^1 \cap D^2 \cap D^3 \cap \ldots \cap D^n),\ i = 1, \ldots, M. \tag{5-23}$$

By applying a decision logic, a joint declaration of identity can be selected by choosing the object whose joint probability given by Eq. (5-23) is greatest. The choice of the maximum value of Eq. (5-23) is referred to as the maximum *a posteriori* probability (MAP) decision rule. Other decision rules exist as indicated in Figure 5.2 and Table 5.1. The Bayes formulation, therefore, provides a method to combine identity declarations from multiple sensors to obtain a new and hopefully improved joint identity declaration. Required inputs for the Bayes method are the ability to compute or model $P(E|H_i)$, i.e., $P(D|O_i)$, for each sensor and entity and the *a priori* probabilities that the hypotheses $P(H_i)$, i.e., $P(O_i)$, are true. When *a priori* information is lacking concerning the relative likelihood of H_i, the principle of indifference may be invoked in which $P(H_i)$ for all i are set equal to one another.

The application of Bayes' rule is often contrasted in modern probability theory with the application of confidence intervals.[9] While Bayes' rule provides an inference approach suitable for some data fusion applications, the theory of confidence intervals is better suited when it is desired to assert, with some specified probability, that the true value of a certain parameter (e.g., mean and variance) that characterizes a known distribution is situated between two limits.

5.6 Combining multiple sensor information using the odds probability form of Bayes' rule

The odds probability formulation of Bayes' rule leads to a convenient method for combining information from a number of sensors. Assume that the sensors respond to different signature-generating phenomenologies and that the output of each sensor is unambiguous (e.g., activated or deactivated) and independent of the outputs of the other sensors.

Let H represent some hypothesis and E^k represent the evidence obtained from the k^{th} sensor, where E_1^k denotes that Sensor k is activated (i.e., produces an output in support of hypothesis H) and E_0^k denotes that Sensor k is deactivated (i.e., does not produce an output in support of hypothesis H). The reliability and sensitivity of each sensor to H are characterized by the probabilities $P(E_1^k \,|\, H)$ and $P(E_1^k \,|\, \overline{H})$, or by their ratio as

$$L(E_1^k \,|\, H) = \frac{P(E_1^k \,|\, H)}{P(E_1^k \,|\, \overline{H})}. \tag{5-24}$$

If some of the sensors are activated and others deactivated, there is conflicting evidence concerning hypothesis H. The combined belief in H is computed from Eq. (5-16) as

$$O(H|E^1, E^2, \ldots, E^n) = L(E^1, E^2, \ldots, E^n|H)\, O(H). \tag{5-25}$$

When the state of each sensor depends only on whether it has detected and responded to the hypothesized event, independently of the response of the other sensors, the probability of sensor activation or deactivation given hypothesis H is expressed as

$$P(E^1, E^2, \ldots, E^n|H) = \prod_{k=1}^{n} P(E^k \mid H)\,. \tag{5-26}$$

Similarly, the probability of a sensor being activated or deactivated given the negation of H is

$$P(E^1, E^2, \ldots, E^n|\,\overline{H}\,) = \prod_{k=1}^{n} P(E^k \mid \overline{H})\,. \tag{5-27}$$

From Eq. (5-25), the posterior odds or belief in hypothesis H becomes

$$O(H|E^1, E^2, \ldots, E^n) = O(H) \prod_{k=1}^{n} L(E^k \mid H)\,. \tag{5-28}$$

Thus, the individual characteristics of each sensor are sufficient for determining the combined impact of any group of sensors.[2]

5.7 Recursive Bayesian updating

The Bayesian approach to recursive computation updates the posterior probability by using the previous posteriors as the new values for the prior probabilities. In Eq. (5-29), H_i denotes a hypothesis as before. The vector $\mathbf{E^N} = E^1, E^2, \ldots, E^N$ represents a sequence of data observed from N sources in the past, while E represents a new fact (or new datum). Once $P(H_i|\mathbf{E^N})$ is calculated and past data discarded, the impact of the new datum E is expressed as[2,5,6]

$$P(H_i \mid \mathbf{E^N}, E) = \frac{P(E \mid \mathbf{E^N}, H_i)\quad P(H_i \mid \mathbf{E^N})}{P(E \mid \mathbf{E^N})} = \frac{P(E \mid \mathbf{E^N}, H_i)\quad P(H_i \mid \mathbf{E^N})}{\sum_i [P(E \mid \mathbf{E^N}, H_i)\quad P(H_i)]}, \tag{5-29}$$

where

$P(H_i\,|\mathbf{E^N}, E) =$ *a posteriori* or posterior probability of H_i for the current period, given the evidence or data $\mathbf{E^N}, E$ available at the current period,

$P(E\,|\mathbf{E^N}, H_i) =$ probability of observing evidence E given H_i and the evidence $\mathbf{E^N}$ from past observations (i.e., the likelihood function),

$P(H_i\,|\mathbf{E^N}) =$ *a priori* or prior probability of H_i, set equal to the posterior probability calculated using the evidence $\mathbf{E^N}$ from past observations,

and

$\sum_i P(E|\,\mathbf{E}^{\mathrm{N}}, H_i)\ \ P(H_i) =$ preposterior or probability of the evidence E occurring given the evidence $\mathbf{E^N}$ from past observations, conditioned on all possible outcomes H_i.

The old belief $P(H_i|\mathbf{E^N})$ assumes the role of the prior probability when computing the new posterior. It completely summarizes past experience. Thus, updating of the posterior is accomplished by multiplying the old belief by the likelihood function $P(E\,|\mathbf{E^N}, H_i)$, which is equal to the probability of the new datum E given the hypothesis and the past observations.

A further simplification of Eq. (5-29) is possible when the conditional independence described by Eqs. (5-26) and (5-27) holds and the likelihood function is independent of the past data and involves only E and H_i. In this case,

$$P(E\,|\mathbf{E^N}, H_i) = P(E\,|\,H_i). \tag{5-30}$$

Similarly,

$$P(E\,|\mathbf{E^N},\, \overline{H}_i\,) = P(E\,|\,\overline{H}_i\,). \tag{5-31}$$

Upon dividing Eq. (5-29) by the complementary equation for $P(\,\overline{H}_i\,|\mathbf{E}^N, E)$, the equation for the posterior odds in recursive form is obtained as

$$O(H_i|\mathbf{E^{N+1}}) = O(H_i|\mathbf{E^N})\; L(E|H_i). \tag{5-32}$$

The recursive procedure expressed by Eq. (5-32) for computing the posterior odds is to multiply the current posterior odds $O(H_i|\mathbf{E^N})$ by the likelihood ratio of E upon arrival of each new datum E. The posterior odds can be viewed as the prior odds relative to the next observation, while the prior odds are the posterior odds that have evolved from previous observations not included in $\mathbf{E^N}$.[2]

5.8 Posterior calculation using multivalued hypotheses and recursive updating†

Suppose several hypotheses $\mathbf{H} = \{H_1, H_2, H_3, H_4\}$ exist where each represents one of four possible conditions, such as

H_1 = enemy fighter aircraft, H_2 = enemy bomber aircraft,

H_3 = enemy missile, H_4 = no threat.

Assume that the evidence variable $\mathbf{E}^k$ produced by a sensor can have one of several output states in response to an event. For example, when a multispectral sensor is used, three types of outputs may be available as represented by

E_1^k = evidence from detected emission in radiance spectral band 1,

E_2^k = evidence from detected emission in radiance spectral band 2, and

E_3^k = evidence from detected emission in radiance spectral band 3.

The causal relations between $\mathbf{H}$ and $\mathbf{E}^k$ are quantified by a $q \times r$ matrix $\mathbf{M}^k$, where q is the number of hypotheses under consideration and r is the number of output states or output values of the sensor. The $(i, j)^{\text{th}}$ matrix element of $\mathbf{M}^k$ represents

$$M_{ij}^k = P(E_j^k \mid H_i). \tag{5-33}$$

For example, the sensitivity of the k^{th} sensor having $r = 3$ output states to $\mathbf{H}$ containing $q = 4$ hypotheses is represented by the 4×3 evidence matrix in Table 5.2.

Table 5.2 $P(\mathbf{E}^k|\mathbf{H}_i)$: likelihood functions corresponding to evidence produced by k^{th} sensor with 3 output states in support of 4 hypotheses.

	E_1^k : **detection of emission in spectral band 1**	E_2^k : **detection of emission in spectral band 2**	E_3^k : **detection of emission in spectral band 3**
$\boldsymbol{H_1}$	0.35	0.40	0.10
$\boldsymbol{H_2}$	0.26	0.50	0.44
$\boldsymbol{H_3}$	0.35	0.10	0.40
$\boldsymbol{H_4}$	0.70	0	0

† This discussion is based in large part on material from Pearl.[2]

Based on the given evidence, the overall belief in the i^{th} hypothesis H_i is [from Eq. (5-10)]

$$P(H_i|E_1, \ldots, E_r) = \alpha\, P(E_1, \ldots, E_r|H_i)\, P(H_i), \tag{5-34}$$

where $\alpha = [P(E_1, \ldots, E_r)]^{-1}$ is a normalizing constant computed by requiring Eq. (5-34) to sum to unity over i. When a sensor's response is conditionally independent, i.e., each sensor's response is independent of that of the other sensors, Eq. (5-26) can be applied to give

$$P(H_i \mid E_1, \ldots, E_r) = \alpha\ P(H_i)[\prod_{k=1}^{N} P(E^k \mid H_i)]. \tag{5-35}$$

Therefore, the matrices $P(E^k|H_i)$ are analogous to the likelihood ratios in Eq. (5-28).

A likelihood vector λ^k can be defined for the evidence produced by each sensor E^k as

$$\boldsymbol{\lambda}^k = (\lambda_1^k,\ \lambda_2^k,\ \ldots\ ,\ \lambda_q^k), \tag{5-36}$$

where

$$\lambda_i^k = P(E^k \mid H_i). \tag{5-37}$$

Now Eq. (5-35) can be evaluated using a vector-product process as follows:

1. The individual likelihood vectors from each sensor are multiplied together, term by term, to obtain an overall likelihood vector $\boldsymbol{\Lambda} = \lambda_1, \ldots, \lambda_n$ given by

$$\Lambda_i = \prod_{k=1}^{n} P(E^k \mid H_i). \tag{5-38}$$

2. The overall belief vector $P(H_i|E^1, \ldots, E^N)$ is computed from the product

$$P(H_i \mid E^1, \ldots, E^N) = \alpha\, P(H_i)\ \Lambda_i, \tag{5-39}$$

which is analogous to Eq. (5-28).

Only estimates for the relative magnitudes of the conditional probabilities in Eq. (5-37) are required. Absolute magnitudes do not affect the outcome because α can be found later from the requirement

$$\sum_i P(H_i \mid E^1, \ldots, E^N) = 1. \tag{5-40}$$

To model the behavior of a multisensor system, let us assume that two sensors are deployed, each having the identical evidence matrix shown in Table 5-2. Furthermore, the prior probabilities for the hypotheses $\mathbf{H} = \{H_1, H_2, H_3, H_4\}$ are assigned as

$$P(H_i) = (0.42, 0.25, 0.28, 0.05), \tag{5-41}$$

where Eq. (5-11) is satisfied by this distribution of prior probabilities.

If Sensor 1 detects emission in spectral band 3 and Sensor 2 detects emission in spectral band 1, the elements of the likelihood vector are

$$\lambda^1 = (0.10, 0.44, 0.40, 0) \tag{5-42}$$

and

$$\lambda^2 = (0.35, 0.26, 0.35, 0.70). \tag{5-43}$$

Therefore, the overall likelihood vector is

$$\Lambda = \lambda^1 \lambda^2 = (0.035, 0.1144, 0.140, 0) \tag{5-44}$$

and from Eq. (5-39),

$$\begin{aligned} P(H_i|E^1, E^2) &= \alpha\,(0.42, 0.25, 0.28, 0.05) \cdot (0.035, 0.1144, 0.140, 0) \\ &= \alpha\,(0.0147, 0.0286, 0.0392, 0) = (0.178, 0.347, 0.475, 0), \end{aligned} \tag{5-45}$$

where α is found from Eq. (5-40) as 12.1212.

From Eq. (5-45), it can be concluded that the probability of an enemy aircraft attack, H_1 or H_2, is 0.178 + 0.347 = 0.525 or 52.5 percent and the probability of an enemy missile attack is 47.5 percent. The combined probability for some form of enemy attack is 100 percent.

The updating of the posterior belief does not have to be delayed until all the evidence is collected, but can be implemented incrementally. For example, if it is first observed that Sensor 1 detects emission in spectral band 3, the belief in $\mathbf{H}$ becomes

$$P(H_i|E^1) = \alpha\,(0.042, 0.110, 0.112, 0) = (0.1591, 0.4167, 0.4242, 0) \tag{5-46}$$

with $\alpha = 3.7879$.

These values of the posterior are now used as the new values of the prior probabilities when the next datum arrives, namely evidence from Sensor 2, which detects emission in spectral band 1. Upon incorporating this evidence, the posterior updates to

$$P(H_i|E^1, E^2) = \alpha' \lambda_i^2 \cdot P(H_i \mid E^1)$$

$$= \alpha' \,(0.35, 0.26, 0.35, 0.70) \cdot (0.1591, 0.4167, 0.4242, 0)$$

$$= \alpha' \,(0.0557, 0.1083, 0.1485, 0) = (0.178, 0.347, 0.475, 0), \quad (5\text{-}47)$$

where $\alpha' = 3.2003$. This is the same result given by Eq. (5-45) for $P(H_i|E^1, E^2)$.

Thus, the evidence from Sensor 2 lowers the probability of an enemy aircraft attack slightly from 57.6 percent to 52.5 percent, but increases the probability of an enemy missile attack by the same amount from 42.4 percent to 47.5 percent. The result specified by Eq. (5-45) or (5-47) is unaffected by which sensor's evidence arrives first and is subsequently used to update the priors for incorporation of the evidence from the next datum.

5.9 Enhancing underground mine detection using two sensors that produce data generated by uncorrelated phenomena

The detection of buried mines may be enhanced by fusing data from multiple sensors that respond to signatures generated by independent phenomena. Two sensors that meet this criterion are metal detectors and ground penetrating radars. The metal detector (MD) indicates the presence of metal fragments larger than 1 cm with weight exceeding a few grams. The ground penetrating radar (GPR) detects objects larger than approximately 10 cm that differ in electromagnetic properties from the soil or background material. While the metal detector simply distinguishes between objects that contain or do not contain metal, the GPR supports object classification since it is responsive to multiple characteristics of the object such as size, shape, material type, and internal design.

In an experiment reported by Brusmark et al., a low metal content mine, metal fragments, plastic, beeswax (an explosive simulant), and stone were buried in sand at a 5-cm depth.[10] The metal detector provided a signal whose amplitude was proportional to the metal content of the object. The GPR transmitted a broadband waveform covering 300 to 3000 MHz. The antenna footprint consisted of four separate lobes, with a common envelope of about 30 cm. An artificial neural network was trained to classify the buried objects that were detected by the GPR. The inputs to the neural network were features produced by Fourier-transform analysis, bispectrum-transform analysis, wavelet-transform analysis, and local frequency analysis of the GPR signals.

Bayesian inference was used to compute and update the *a posteriori* probabilities that the detected object belonged to one of the object classes represented by mine (MINE), not mine ($\overline{\text{MINE}}$), or background (BACK). Figure 5.3 contains an influence diagram that models the Bayesian decision process.

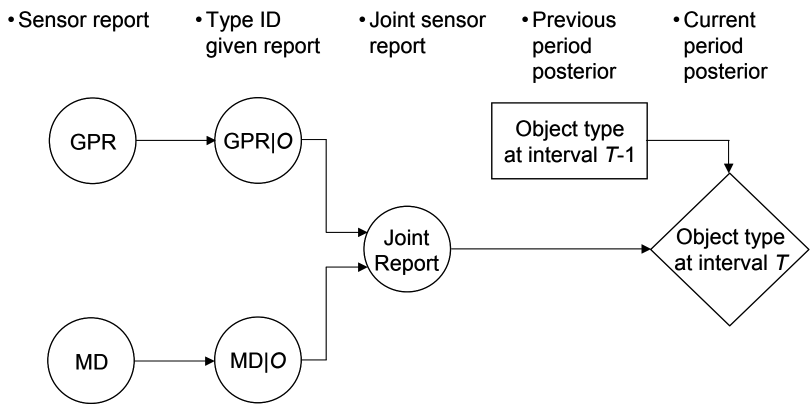

Figure 5.3 Influence diagram for two-sensor mine detection.

Influence diagrams are generally used to capture causal, action sequence, and normative knowledge in one graphical representation. Each type of knowledge is based on different principles, namely:

1. Causal knowledge deals with how events influence each other in the domain of interest.

2. Action sequence knowledge describes the feasibility of actions and their sequence in any given set of circumstances.

3. Normative knowledge encompasses how desirable the consequences are.

Influence diagrams are drawn as directed acyclic graphs with three types of nodes: decision, chance, and value.[2] Decision nodes, depicted as squares, represent choices available to the decision maker. Chance nodes, depicted as circles, represent random variables or uncertain quantities. The value node, shown as a diamond, represents the objective to be maximized.

The probability of the sensors observing data conditioned on object type is given by

$$\begin{aligned} P_{MD}(\text{data} \,|\, O_i) = \; & P_{MD}(\text{data} \,|\, \text{MINE})\; P(\text{MINE} \,|\, O_i) \\ & + P_{MD}(\text{data} \,|\, \overline{\text{MINE}})\; P(\overline{\text{MINE}} \,|\, O_i) \\ & + P_{MD}(\text{data} \,|\, \text{BACK})\; P(\text{BACK} \,|\, O_i) \end{aligned} \tag{5-48}$$

and

$$\begin{aligned} P_{GPR}(\text{data}\,|\,O_i) &= P_{GPR}(\text{data}\,|\,\text{MINE})\;\; P(\text{MINE}\,|\,O_i) \\ &\quad + P_{GPR}(\text{data}\,|\,\overline{\text{MINE}})P(\overline{\text{MINE}}\,|O_i) \\ &\quad + P_{GPR}(\text{data}\,|\,\text{BACK})\;\; P(\text{BACK}\,|\,O_i), \end{aligned} \tag{5-49}$$

where MD denotes the mine sensor, GPR the ground penetrating radar, and O_i an object of type i. The set of arrows from "sensor report" to "type identification given report" in Figure 5.3 represents the probability calculations defined by Eqs. (5-48) and (5-49).

The values of the likelihood functions for the metal detector, namely $P_{MD}(\text{data}\,|\,\text{MINE})$, $P_{MD}(\text{data}\,|\,\overline{\text{MINE}})$, and $P_{MD}(\text{data}\,|\,\text{BACK})$, and for the ground penetrating radar, namely, $P_{GPR}(\text{data}\,|\,\text{MINE})$, $P_{GPR}(\text{data}\,|\,\overline{\text{MINE}})$, and $P_{GPR}(\text{data}\,|\,\text{BACK})$, are found through *a priori* measurements. The mine detector "data" are equal to the preprocessed signal amplitude, and $P_{MD}(\text{data}\,|\,O_i)$ is equal to the probability of receiving a signal of some amplitude given the object is of type O_i. These probabilities are found from extensive experiments with buried mine-like objects consisting of different materials and sizes. The ground penetrating radar signal profile data in the scanned area are input to an artificial neural network trained to identify antipersonnel mines. The output of the neural network over many experiments gives $P_{GPR}(\text{data}\,|\,O_i)$. Quantitative values for $P(\text{MINE}\,|\,O_i)$, $P(\overline{\text{MINE}}\,|O_i)$, and $P(\text{BACK}\,|\,O_i)$ are dependent on the types and numbers of objects in the mine-infected area.

Next, the joint sensor report shown in Figure 5.3 is computed for a given time interval as the product of Eqs. (5-48) and (5-49) since the sensors respond to signatures generated by independent phenomena, i.e., they are uncorrelated. Thus, the joint probability of detection is [analogous to Eq. (5-38)]

$$P(\text{data}\,|\,O_i) = \prod_k P^k(\text{data}\,|\,O_i), \tag{5-50}$$

where k is the sensor index, here equal to 1 and 2.

Finally, Bayes' rule is applied to calculate the current period *a posteriori* probability $P(O_i\,|\,\text{data})$ that the detected object is of type i based on the value of $P(\text{data}|O_i)$ and the posterior probabilities evaluated in the previous period. Accordingly, from Eq. (5-29),

$$P(O_i\,|\,\text{data}) = \frac{P(\text{data}\,|\,O_i)\,P(O_i)}{P(\text{data})}, \tag{5-51}$$

where

$$P(\text{data}\,|\,O_i) = \prod_k P^k(\text{data}\,|\,O_i), \text{ i.e., value from Eq. (5-50),} \tag{5-52}$$

$$P(O_i) = \text{value of } P(O_i\,|\,\text{data}) \text{ from the previous period, and} \tag{5-53}$$

$$P(\text{data}) = \sum_i P(\text{data}\,|\,O_i)\,P(O_i) \tag{5-54}$$

is the preposterior or probability of observing the data collected during the previous period given that objects O_i are present. Larger values of P(data) imply that the previous period values are more predictive of the situation as it evolves. When the sensors do not report an object type for the current time interval, updating is not performed and the values of $P(O_i\,|\,\text{data})$ for the current interval are set equal to those from the previous period.

Since the primary task in this example is to locate mines, the second and third terms in Eqs. (5-48) and (5-49) are combined into a single declaration $\overline{\text{MINE}}$ that represents the absence of a mine. The problem is further simplified by choosing O_1 = MINE (in this experiment, the mine was an antipersonnel mine) and $O_2 = \overline{O_1}$. Therefore, the required probabilities are only dependent on $P(\text{data}\,|\,O_1)$ since

$$P(\text{data}\,|\,O_2) = 1 - P(\text{data}\,|\,O_1) \tag{5-55}$$

and

$$P(O_2) = 1 - P(O_1). \tag{5-56}$$

An initial value for $P(O_1)$ and lower and upper bounds inside the interval (0, 1) for admissible values of $P(O_1\,|\,\text{data})$ are needed to evaluate Eq. (5-51). Since 5 different types of objects were buried, $P(O_1)$ was initially set equal to 1/5. The boundaries for $P(O_1\,|\,\text{data})$ were limited to (0.01, 0.99) to prevent the process that computes the *a posteriori* probability from terminating prematurely at the limiting endpoint values of 0 and 1. The updated joint probability of detection by the sensors is found by applying Eq. (5-50) to the joint MD and GPR reports as represented by a matrix formed by the scanned data. This process is enhanced by passing the GPR signatures through a matched filter to remove the distortion caused by the antenna pattern.[11]

The posterior probabilities for object classes mine, not mine, and background are computed from the posterior probabilities for object type and the scenario defined values for $P(\text{MINE}\,|\,O_i)$, etc., as

$$P(\text{MINE}\,|\,\text{data}) = \sum_i [P(O_i\,|\,\text{data})\;P(\text{MINE}\,|\,O_i)], \tag{5-57}$$

$$P(\overline{\text{MINE}}\,|\,\text{data}) = \sum_i [P(O_i\,|\,\text{data})\; P(\overline{\text{MINE}}\,|\,O_i)], \tag{5-58}$$

and

$$P(\text{BACK}\,|\,\text{data}) = \sum_i [P(O_i\,|\,\text{data})\; P(\text{BACK}\,|\,O_i)]. \tag{5-59}$$

Thus, the probability of locating a mine is the sum of individual probabilities that are dependent on the identification of various features. The term $P(\text{MINE}\,|\,O_i)$ expresses the *a priori* probability of finding a mine conditioned on object type O_i being present. In this particular application where metal detector and ground penetrating radar data were fused, it was assumed that very low-metal-content mines could be detected by the metal detector alone. Two cautions were mentioned by the authors, however. The first was that the data fusion algorithms should be robust in their ability to identify objects other than those expected to be found. Second, since the metal detector may often not detect metal, the multisensor system must be designed to rely on ground penetrating radar detections alone to identify objects.

An application of Bayesian inference to freeway incident detection is described in Appendix B.

5.10 Summary

Bayes' rule has been derived from the classical expression for the conditional probability of the occurrence of an event given supporting evidence. Bayes' formulation of conditional probability provides a method to compute the probability of a hypothesis being true, given supporting evidence. It allows incorporation of *a priori* knowledge about the likelihood of a hypothesis being true at all. Bayes also permits the use of subjective probabilities for the *a priori* probabilities of hypotheses and for the probability of evidence given a hypothesis. These attributes let Bayesian inference be applied to multisensor fusion since probability density functions are not required. However, the output of such a process is only as good as the input *a priori* probability data. Bayesian inference can be used in an iterative manner to update *a posteriori* probabilities for the current time period by using the posterior probabilities calculated in the previous period as the new values for the prior probabilities. This method is applicable when past data can be discarded after calculating the posterior and information from only the new datum used to update the posterior for the current time period. A procedure for updating posterior probabilities in the presence of multivalued hypotheses and supporting evidence from sequentially obtained sensor data was described. A result of this analysis is that the updating of the posterior belief does not have to be delayed until all the evidence is collected, but can be implemented incrementally. An application of Bayesian inference to data gathered by a metal detector and ground-penetrating radar was presented to demonstrate recursive updating of the posterior probability to enhance the detection of buried mines.

References

1. W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd Ed., John Wiley and Sons, New York, NY (1962).
2. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., San Mateo, CA (1988).
3. E. Parzen, *Modern Probability Theory and Its Applications*, John Wiley and Sons, New York, NY (1964).
4. P.G. Hoel, *Introduction to Mathematical Statistics*, John Wiley and Sons, New York, NY (1947).
5. E. Waltz and J. Llinas, *Multisensor Data Fusion*, Artech House, Norwood, MA (1990).
6. D.L. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Norwood, MA (1992).
7. G.S. Robinson and A.O. Aboutalib, "Trade-off analysis of multisensor fusion levels," *Proc. of the 2nd National Symposium on Sensors and Sensor Fusion*, Vol. II, GACIAC PR 89-01, 21-34, IIT Research Institute, Chicago, IL, (1990).
8. D.C. Lai and R.D. McCoy, "Optimal classification fusion in multi-sensor target recognition systems," *Proc. 2nd National Symposium on Sensors and Sensor Fusion*, Vol. II, GACIAC PR89-01, 259-266, IIT Research Institute, Chicago, IL (1990).
9. H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ (1961).
10. B. Brusmark, A.-L. Christiansen, P. Jägerbro, and A. Lauberts, "Combination of ground penetrating radar and metal detector data for mine detection," *Proc. of 7th International Conference on Ground Penetrating Radar* (*GPR98*), Radar Systems and Remote Sensing Laboratory, Univ. of Kansas, Lawrence, KS, 1, 331-336 (June 1998).
11. P. Jägerbro, B. Brusmark, A.-L. Christiansen, and A. Lauberts, "Combination of GPR and metal detector for mine detection," *IEE 2nd International Conference on the Detection of Abandoned Land Mines*, Edinburgh, 177-181 (1998).