LA-UR-00-4849

Title: Information Integration Technology

Author(s): Mark S. McNulty, TSA-11
Alyson G. Wilson, TSA-1
Sallie Keller-McNulty, TSA-1

Submitted to: Army Conference on Applied Statistics

Los Alamos

NATIONAL LABORATORY

Form **836 (10/96)**

Information Integration Technology

Mark McNulty, Sallie Keller-McNulty, and Alyson Wilson
Los Alamos National Laboratory, Statistical Sciences

1. Introduction

Information Integration Technology (IIT) is a combination of processes, methods, and tools for collecting, organizing, and analyzing diverse information and for utilizing that information to guide optimal decision making. The diversity of the information that can be focused on a given problem is its defining characteristic of IIT. Effective use of diverse information is a prerequisite for optimal decision making. Optimal decisions require information about many different things: information about the structure and behavior of the system the decision-maker is controlling, information about the constraints and options that the decision-maker faces, information about the decision-maker's preferences. Optimal decisions also require information from many different sources: information from theoretical models, information from test data, information from the performance of similar systems, information from computer simulations, and information from all the individuals (scientists, engineers, field personnel, managers, and the decision-makers themselves) who are stakeholders in the decision making process.

The desire to support decision making at any point in the life of a project or process was a driving force behind the development of IIT. Fulfilling this desire resulted in several important features in the technology. First, the technology can function with limited information. This is important because crucial decisions must often be made precisely when hard data are most scarce. IIT utilizes all available information, even if the amount is small, to guide decisions whenever they must be made. Because information is often scarce, IIT strives to employ information that may be ignored by other decision-support analyses. For example, the technology utilizes information from similar systems. Expert opinion is another important source of information that can be exploited when hard data are hard to come by. Expert opinion is used to some extent during all decision-making processes. However, it is typically not formally elicited and recorded as data. IIT formalizes elicitation where it has been used in the past, introduces elicitation where it has not previously been employed, and explicitly makes use of the resulting information. Finally, uncertainty quantification is never more important than when decisions are to be based upon limited information. IIT treats uncertainty quantification as an integral, critical component of the decision-making process.

Another important feature of IIT is the dynamic nature of the technology, which is required to support continuous and comprehensive evaluation and decision making The technology is structured so that new information can be quickly incorporated to produce complete model updates. The dynamics allow for changes in process design and decision objectives, as well as the more typical data arising from additional tests and experiments.

IIT have been successfully applied in industrial environments and with the nuclear weapons program at Los Alamos National Laboratory. Most of these applications have focused on reliability and performance estimation for problems where traditional statistical approaches that are grounded in the estimation of metrics based principally on test data were not sufficient to

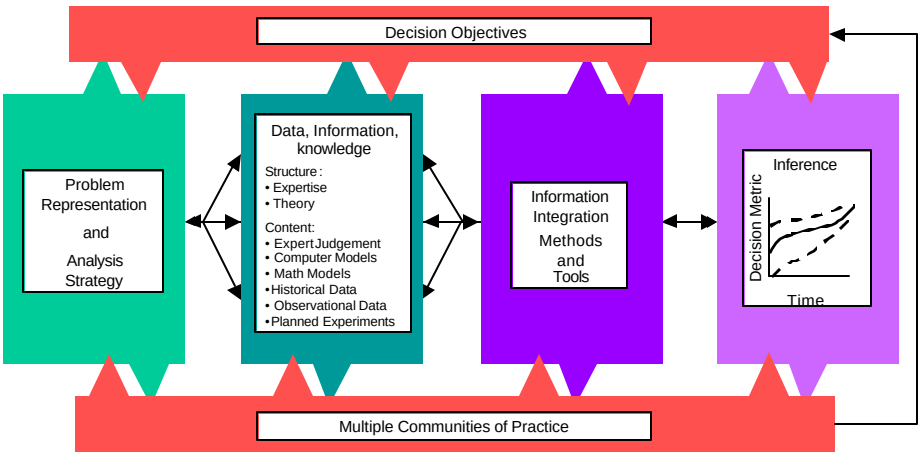
guide the decision making, e.g., nuclear weapons certification of aging missiles without full system tests. These tradition methods ignore the vast amount of knowledge and intuition in the communities that have developed (or are developing) the systems under study. Also, the traditional methods are not amenable to the integration of information from computer models or other data on similar systems, nor do the flexible adapt to system changes as occurs with aging or re-design.

The driving force behind the development of IIT has been to overcome the problems stated above. Several applications of IIT have led to a number of practical considerations that are critical for the successful implementation of the technology. One is that the generic processes, methods, and tools that make up the technology must be couched in the nomenclature of the specific organization and communities being served. Interactions with engineers must be in the terms of an engineer, interactions with managers must be in the terms of a manager. Attempts to interact using statistical or decision-theoretic jargon are destined to fail. Another consideration is that state-of-the art computationally intensive statistical methods must be brought to bear on the problem, e.g., Monte Carlo methods. Finally, since these problems are about decision making and not simply modeling, highly collaborative multidisciplinary teams of individuals must come together in the development of the problem solutions.

2. IIT Framework

The major elements of the IIT framework are illustrated in Figure 1. The connections between the figure elements emphasize the recursive nature of the technology. New information about any particular element generates updates of all the other elements. This continuous updating makes the technology highly dynamic, allowing decisions based upon the best and most complete information to be made at any point in time.

Figure 1. IIT Framework



The bottom element of Figure 1 represents an understanding that all stakeholders in the decision process have important perspectives on the problem that must be integrated into the IIT process. Groups of stakeholders who think in similar ways are often called "communities of practice". Within nuclear weapons the communities of practice include physicists, who approach problems by trying to understand the physical processes involved; weapons designers, who are more concerned with harnessing the physical processes; engineers, who think about problems in terms of interacting components; statisticians, who consider issues of uncertainty; computer scientists, who want to understand how complex computer codes work; managers, who work to coordinate activities; and politicians, who care about the policy implications of the science. Each of these communities approaches the problem from a different viewpoint, and each represents its information in a different way. Understanding the various communities of practice allows them to be approached on their own turf, rather than forcing them onto the unfamiliar playing fields of the statistician or decision scientist. Experience has demonstrated that the former approach makes the technology far more likely to succeed. IIT strives to understand all of the communities of practice within the organization being served. The bottom element of Figure 1 also emphasizes the need for a broad understanding of all the disciplines that may be called upon to help solve the organization's problems.

The remaining five elements in Figure 1 represent the more technical aspects of the technology. The development of these elements is not a sequential process, they all progress simultaneously. However, the development is most efficient if the top element, the definition of Decision Objectives, is at the front of the pack. This step is frequently overlooked by analysts in the rush to collect, model, and analyze data, a natural inclination because analysts are well-trained in modeling and data analysis, and not so well-trained in the complexities of understanding decision contexts. However, the definition of the decision objectives forms the foundation upon which the remainder of the IIT process is built. The decision objectives guide the kind of models that should be built, the kind of information that should be gathered, and the kinds of analyses that should be performed.

The Analysis Strategy and Problem Representation is extremely important. Before any information is collected, it must be determined how this information will be analyzed and integrated, and how the results will bring better resolution to the decision objectives. These determinations should drive the requirements regarding what is collected, what computer models to build, what experiments to run, etc.

The collection of data, information, and knowledge is represented in the Information element. Notice that this information includes more that just "data" in its traditional, narrow sense. All decisions incorporate more than just data: they also include the information and knowledge required to understand the problem, structure the representations, find data sources, and select appropriate models. Even "data" in its narrower sense is quite broad, including opinions elicited from experts, outputs from computer codes, etc.

The Information Integration Methods and Tools element is used to tie all of the decision objectives, community representations, and information together. If these methods are effective, they lead to the Inference element in which the quantitative results required to make decisions are actually produced.

3. Organization of Information

The way in which information is organized has a major influence on how that information is used. Because IIT employs such a wide array of information, the proper organization of information is critical for the successful implementation of the technology. This section provides details on how information is organized within IIT.

Within IIT information is grouped into two broad classes, information that is used to define structure and information that is used to provide content. Structure is a complete qualitative description of the problem and problem solution. Because the emphasis of IIT is on decision making, the decision maker is the proper starting point for the collection of structural information. The decision objectives are defined by the answers to questions such as What are the objectives of the decision maker? What metrics are meaningful? What options are available? What constraints do they face? What analysis capabilities would they like? Given a description of decision objectives, information about the processes that must be modeled is gathered next. This information emphasizes complete descriptions of process components and their relationships. Defining structure requires considerable information about information. What data are available? Who knows what? Are there relevant data on related processes? What is the nomenclature of the different communities of practice? Finally, all this information is used to identify the best system representation and the kinds of analyses that are going to be performed. Information about structure can be qualitative in nature, driven by experience and expertise, or it can be very concrete, driven by theory and the physical laws of nature.

Content, the second broad information classification, provides the quantification of the structure. For example, content includes the mathematical representation of decision maker's objective function, the actual process representation, parameter estimates, probability distributions, etc. There are three broad types of information used to create content. The first is sample data, which are observations taken on the relevant processes. These observations may arise from experimentation, tests, or be purely observational. The second type of information used to create content are data taken from similar systems. These are observations that are taken on processes that are related to, yet distinctly different from, the processes defined in the structure. Examples of similar systems are prototypes, modified processes, and processes developed by the same design team. Also included as similar systems are mathematical and computer models. The third type of information used to produce content is expert opinion, produced through elicitation. Elicited information pertaining to content will be called judgement.

Because the use of expert opinion as information may be novel to some, a brief introduction to this important source of information is provided. Individuals with backgrounds in relevant subject matter and qualified to answer questions are experts and their informed opinions are obtained through elicitation. Quantitative expert opinion can be considered to be data, and like data from other sources, expert opinion must be handled with certain considerations. Expert judgment is affected by how it is gathered. Elicitation methods must take advantage of the body of knowledge on human cognition and motivation and include procedures for aiding memory and countering effects arising from the phrasing of the questions, response modes, the influence of

the elicitor, and the expert's personal agenda. Expert judgment has uncertainty, which can be characterized and subsequently analyzed. Analysts should be aware of experts' natural tendency to underestimate uncertainty. Expert judgment can be conditioned on various factors including the phrasing of the question, the information the experts considered, the experts' methods of solving the problem, and the experts' assumptions. A formal structured approach to elicitation provides analysts a better handle on conditioning effects. A complete description of elicitation details and techniques is available in Meyer and Booker (1991) and Meyer et al. (2001).

An important part of the organization of the information includes tracking all the diverse sources being brought to bear on the problem and the dynamic changes in the IIT processes. An effective way to do this is to create a knowledge system. A knowledge system is a browser-accessed electronic repository that has been customized to the cognition and culture of the technical communities involved in the decision process. The knowledge system brings together their data and knowledge in structured quantitative ways. In addition, these systems integrate all of the data, knowledge, and information with the relevant statistical methods and tools. The purpose is to provide distributed communities with electronic access to information, methods, and tools they seek to perform their problem solving/decision making. As a result, knowledge system serves to rapidly evolve the knowledge in science and technology environments.

4. System Representation

An important component of IIT is the representation of the system under study. This representation must be flexible to accommodate change in the underlying processes and understandable by all of the communities of practice. The representation must include all the factors that are of interest to the decision maker, including the ways in which the decision maker can control and influence the system and metrics for the decision maker's objectives. Finally, the representation must be accurate. It must correctly portray the system and the ways in which the system components interact. This section presents an introduction to some of many different ways of representing a given system. For a more complete discussion see Wilson and Keller-McNulty(2000).

The general forms of the representations used by the information integration technologies have three parts: icons/pictures/diagrams, rules/statements, and abstract mathematics (Paton et al. 1994). Rules and statements are used when we have observable phenomena that have been characterized by physical laws or statistical relationships; abstract mathematics capture physical laws about unobservable phenomena. There are four primary purposes of these representations: capture all of the factors that affect the measures of performance; outline the components and subsystems of the larger system and how they interact; identify the information that feeds into the estimation of the metrics; and specify the methods for combining and aggregating the information and quantifying uncertainty about the metrics.

There are many types of diagrams that can be used to accomplish the four primary purposes of diagramming and representing the "system" under study. However, they share common features. The basic components of the diagrams are "boxes," which represent where data, information, and knowledge can be collected, and arcs/arrows/lines, which represent the information flow between boxes. In attempting to represent a system, many different diagrams are often drawn.

This may be because the members of a multi-disciplinary team draw different diagrams to represent the different ways they think about the problem, or it may be because different parts of the system are better or more easily represented in different ways. In the integration process, it is constructive to combine these different representations into a single diagram, not forcing it into the structure of a specific diagram type, but integrating the different representations into a complex diagram that captures the unique features, relationships, and information flows in the system under study.

5. Statistical Methods

There are many types of data that can be used to populate the boxes/nodes in the system representation: expert judgment, historical test data, data from similar or relevant systems, design specifications, computer simulation model outputs, physical test data. Each type of data and information has natural ways that it can be represented (Keller-McNulty and McNulty 2000). These include graphs (e.g. histograms, boxplots, and scatterplots), tables, mathematical models, probability distribution functions, and fuzzy distributions and membership functions. For mathematically rigorous information integration, all of the system and data representations must be collected and transformed into probability distributions, probabilistic dependency relationships, and statistical models. These models and distributions may not be parametric, but they must be in the language of distributions, dependencies, correlations, conditional independencies, and statistical models of various forms. The statistical methods described in this section constitute a primer on how to integrate the diverse information sources relevant to a box or a node in the system representation.

Suppose that the operation of one component (one box) in a system is represented by a variable Y. Y might denote whether the component is functioning (y = 1) or failed (y = 0), the lifetime of the component, or the output of the component (in units of horsepower, volts, floating-point operations per second, etc.). While it is the performance of the entire system that is of interest, this can only be derived by understanding the behavior of each component. This section examines the use of diverse information to describe the behavior of Y. Three specific sources of information are considered: sample data, expert opinion, and data arising from similar components. While only the simplest cases are considered and many simplifying assumptions are made for clarity of exposition, the methods described below form the basis of the IIT statistical processes.

5.1 Classical Methods

It will be assumed that Y is random variable so that knowledge of the probability density function (pdf) of Y completely describes its behavior. Using standard notation the pdf of Y is written $f(y;\theta)$, where θ is assumed to be a scalar parameter for simplicity.

In order to proceed here a functional form of $f(y;\theta)$ will be specified[1]. In some situations the nature of the random variable suggests a natural choice for $f(y;\theta)$. There are a wide range of probability distributions that have been developed to describe random variables in specific

[1] In practice, nonparametric and empirical representations of $f(y;\theta)$ are allowed.

situations. If Y can be interpreted as the number of successes in a given number of trials, a binomial pdf ($\theta \equiv p$) may be appropriate. If Y measures time until failure of some component, an exponential pdf ($\theta \equiv \lambda$) would merit consideration. When there is no natural candidate distribution, the examination of sample data may suggest a form, as may discussions with experts familiar with the behavior of the random variable. While the choice of function form is an interesting and important topic, extended discussion will be deferred so that the paper may focus on the issue of primary interest, which is using diverse information to learn about θ.

The classic source of information is a sample of observations on the random variable of interest, Y. A sample of n observations on Y is represented by $\{y_1,\ldots,y_n\}$. The observations may occur in a controlled, experimental setting (i.e. testing) or under more uncontrolled, subjective conditions (i.e. in the field). Examples of sample data are fracture counts from a component being tested or consumer satisfaction rankings. The problem of analyzing sample data has historically dominated statistical thinking, and a plethora of methods for estimating θ given a sample have been developed. Two of the more common estimation strategies are maximum likelihood estimation and method of moments (Mood, Graybill, and Boes 1974). Whatever strategy is pursued, the end result will be an estimate of θ based upon the sample, denoted $\hat{\theta}_S$, and an (estimated) variance of the estimate, $\mathrm{Var}(\hat{\theta}_S)$. It will be assumed that $\hat{\theta}_S$ is an unbiased estimator so that $E(\hat{\theta}_S) = \theta$.

Experts are another source of information about θ, and this information can be acquired through elicitation. One goal of the elicitation is to obtain an estimate of θ based upon the expert's knowledge. This estimate will be denoted $\hat{\theta}_{EJ}$ (EJ standing for "expert judgement"). Expert elicitation will generally focus on features that define the pdf of Y rather than on the random variable itself[2]. Because it is an estimate of θ that we seek, it is natural to first consider extracting the expert's knowledge in terms of that parameter. This is most likely to be fruitful when the parameter θ has a real meaning for the expert. For example, $1/\lambda$ in the exponential distribution may be interpreted as mean lifetime, and p in the binomial distribution may be interpreted as the probability of success. In these cases, the expert may be comfortable providing an estimate of λ or p such as "I believe the mean lifetime is 10 hours" ($\hat{\theta}_{EJ} \equiv \hat{\lambda}_{EJ} = 0.10$) or "I believe the component will work 70% of the time ($\hat{\theta}_{EJ} \equiv \hat{p}_{EJ} = 0.70$). The elicitations are rarely this simple, and they may require experts to study documents, conduct experiments, etc. before an answer is provided.

In other cases the expert may not relate to the parameter θ. For example, if Y has a t-distribution then θ is the degrees-of-freedom parameter k. Degrees-of-freedom probably means little to the non-statistician, and attempting to elicit information about it will likely result in frustration for all concerned. When the elicitation cannot proceed in terms of θ directly, other parameters, from which θ may be inferred, may be investigated. The elicitation of percentiles is an important strategy, and may be approached from several angles. The expert may be willing to provide a range of likely Y values ("I am quite sure Y will be in the range [-2, 2]"). Further

[2] An important exception is when y depends upon some other variable x so that the pdf of y is $f(y;x,\theta)$. For example measured crack length in a seal, y, depends upon temperature, x. In this case one may elicit information about the value of y that the expert expects for different values of x.)

discussions will focus on what "quite sure" means to the expert. Are [-2, 2] the 10th and 90th percentiles or are they the 1st and 99th percentiles? Once the percentiles have been elicited they can be transformed into an estimate of θ: if [-2, 2] the 10th and 90th percentiles of a t-distribution, then $\hat{\theta}_{EJ} = \hat{k}_{EJ} = 1.76$.

Information about percentiles can be derived in other ways (e.g. consideration of log-odds ratios), and if the expert's knowledge or disposition precludes providing quantitative estimates, tools derived from fuzzy control theory (Smith et al., 1998) can be exploited. For example, if Y is the crack length in a seal, which depends upon operating temperature, the expert may share knowledge in the form of rules such as "If the temperature is too hot, the cracks will be so long that the seal will fail." These rules can be quantified using membership functions that capture the experts beliefs about the relationship between crack length and temperature..

Discussion of all the possible methods for eliciting an estimate of θ is beyond the scope of this paper. Elicitation is a formal process, and one of the keys to success is to couch the elicitation in terms the expert is comfortable with and then derive what is statistically needed. Forcing the expert to respond to questions heavy with statistical jargon is a recipe for disaster. However, whatever form the elicitation takes, the final result is an estimate, $\hat{\theta}_{EJ}$, that is based upon the expert's knowledge. It will be assumed that $\hat{\theta}_{EJ}$ is also an unbiased estimator of θ.

In addition to the expert's best estimate of θ, a properly conducted elicitation will also produce information about the expert's confidence in that estimate. The expert may have a high level of confidence in their estimate or they may be very uncertain. Just as the expert's beliefs about θ were obtained through elicitation, so can the expert's beliefs about their confidence. In reality, the two elicitations will generally proceed simultaneously. Obtaining information from the expert about their confidence can become quite complicated, and emphasizes even more the need for careful construction of the elicitation tools. In addition, the data analyst may have other information concerning the quality of the expert's estimate that can be factored into the confidence measure. The information concerning confidence in $\hat{\theta}_{EJ}$ may be quantified in a number of ways (e.g. tail percentiles of a pdf of $\hat{\theta}_{EJ}$). It will be assumed here that confidence is measured by a variance of $\hat{\theta}_{EJ}$, $\mathrm{Var}(\hat{\theta}_{EJ})$, constructed from the expert's and analyst's beliefs. The interpretation of this variance is the usual one: the probability that the $\hat{\theta}_{EJ} \pm 2\sqrt{\mathrm{Var}(\hat{\theta}_{EJ})}$ contains the true θ is at least 0.75 (using Chebyshev's inequality, see Higgins and Keller-McNulty 1995).

Similar components are the third source of information that will be addressed here. A similar component is a process that exhibits commonality with the process that generates the Y random variable, yet is distinctly different from the Y-generating process. One example is a computer model of the Y-process. Another is a prototype created during the development of the Y-process. Another would be a different product created by the same design team responsible for the product underlying Y.

In order to make use of information arising from a similar component to learn about θ, the commonality between the two processes must be quantified. In order to accomplish this, let the similar component generate a random variable x that has a pdf $f(x;\delta)$. It will be assumed here

that the source of commonality between the two components is a relationship between the Y-parameter θ and the x-parameter δ. In particular, it is assumed that θ and δ are independent, identically distributed random variables[3]. The upshot of this assumption is that the parameter difference $\varepsilon = \theta - \delta$ is a random variable with mean 0 and variance σ^2_ε. The parameter σ^2_ε measures the degree of commonality between the components. If the components are very similar we will expect θ and δ to be very close, on average, which is equivalent to small value of σ^2_ε. If the commonality is weak, we will not be surprised if θ and δ are quite different, which is reflected in a large σ^2_ε.

Given knowledge of σ^2_ε, a sample $\{x_1,\ldots,x_n\}$ from the similar component can be used to learn about θ. It is generally possible to write the probability density function of the similar component data as $f(x_1,\ldots,x_n|\ \theta\ ,\ \sigma^2_\varepsilon)$ and then apply one of the usual sample estimation techniques (e.g. maximum likelihood) to estimate θ. However, a simpler method is the following. Use the sample $\{x_1,\ldots,x_n\}$ to estimate the similar component parameter δ in one of the usual ways, giving an estimate $\hat{\delta}_{SC}$ and (estimated) variance $\mathrm{Var}(\hat{\delta}_{SC})$. Assume that $\hat{\delta}_{SC}$ is an unbiased estimator of δ. Now, viewing θ and δ as random variables, $\hat{\delta}_{SC}$ is an unbiased predictor of θ because $E(\theta) = E(\delta)$ (because θ and δ are identically distributed). Consequently, the similar component parameter estimate is used as the component estimate, $\hat{\theta}_{SC} = \hat{\delta}_{SC}$. This makes sense, because we have assumed that the components are alike, but we have made no assumptions about whether we expect $\theta > \delta$ or $\theta < \delta$, although such an assumption could be incorporated into the analysis. The fact that we are using data from a different component is reflected in the variance of the θ estimate[4], which is $\mathrm{Var}(\hat{\theta}_{SC}) = \mathrm{Var}(\hat{\delta}_{SC}) + \sigma^2_\varepsilon$. The greater the disparity between the components, the greater is σ^2_ε, and the less is the confidence placed in $\hat{\theta}_{SC}$ as an estimate of θ.

The key variable that admits the use of information from a similar component is σ^2_ε, which is inversely related to the degree of similarity. In some cases this quantity may be estimated. For example, there may exist a set of separate, similar components and parameter variations within this set may be used to estimate σ^2_ε. For mathematical or computer models of the Y-process, varying input parameters may provide the required information. In other cases, there will be no direct data concerning σ^2_ε, in which case estimates can be based upon expert opinion.

The use of information arising from samples, expert elicitation, and similar components to learn about θ have been considered. The ways in which information from these sources may be integrated is now explored. Suppose we have an estimate from each of the three sources, $\hat{\theta}_S$, $\hat{\theta}_{EJ}$, $\hat{\theta}_{SC}$. The problem is to combine these estimates into a single estimate of θ. A natural solution would be to take the simple average of the three estimates. However, suppose it is believed that one of the estimates is a more precise than the other two. The more precise estimate should receive a greater weight in the average. This chain of logic is precisely what leads to the definition of Best Linear Unbiased Estimator (BLUE) as a criterion for constructing an estimator. There are three requirements for an estimator to be BLUE. The estimator must be

[3] More generally it is assumed there is some relationship between $f(y;\theta)$ and $f(x;\delta)$ that must be modeled.

[4] This variance calculation requires that the true (not estimated) variance of δ_{SC} does not depend upon the δ.

Unbiased. The estimator must be a Linear function of the data, i.e. it is constructed as a (weighted) average. The requirement that the estimator be Best is, essentially, that it have the minimum possible variance (in the class of unbiased linear estimators). The way that minimum variance is achieved is by using weights in the average that are inversely related to precision, and where precision is measured by variance. For example, assuming the estimates $\hat{\theta}_S$, $\hat{\theta}_{EJ}$, $\hat{\theta}_{SC}$ are statistically independent, the BLUE estimator of θ is given by $\hat{\theta}_{BLUE} = w_S\hat{\theta}_S + w_{EJ}\hat{\theta}_{EJ} + w_{SC}\hat{\theta}_{SC}$ where the weights w_S, w_{EJ}, and w_{SC} are well defined functions of $\mathrm{Var}(\hat{\theta}_S)$, $\mathrm{Var}(\hat{\theta}_{EJ})$, and $\mathrm{Var}(\hat{\theta}_{SC})$, where for example $w_S = \dfrac{\mathrm{Var}(\hat{\theta}_{EJ})\mathrm{Var}(\hat{\theta}_{SC})}{\mathrm{Var}(\hat{\theta}_S)\mathrm{Var}(\hat{\theta}_{EJ}) + \mathrm{Var}(\hat{\theta}_S)\mathrm{Var}(\hat{\theta}_{SC}) + \mathrm{Var}(\hat{\theta}_{EJ})\mathrm{Var}(\hat{\theta}_{SC})}$.

The expression for the BLUE estimator of θ given above is based upon the assumption of independence the different estimates, which is equivalent to independence of the information used to construct those estimates. This is a critical assumption that must be carefully evaluated in a given situation. It is perhaps easiest to describe what is meant by information independence by considering information dependence and the problems dependence creates. It is hard to imagine that the expert's beliefs would not be influenced by this information. Or, suppose two estimates of θ are elicited from two different experts. Speaking loosely, these two estimates will be counted as two "pieces of information" when they are integrated under the assumption of independence. Now, how did the experts acquire their knowledge? Did they receive similar training? Where they exposed to the same reports, experiments, presentations, etc? Did they see the sample data that is also to be integrated? If so, their estimates contain redundant information, and this redundancy must be accounted for, otherwise the analysis results will be biased toward their particular knowledge. It is certainly not correct to give these dependent estimates the same worth as two independent estimates based upon completely separate knowledge sets. Dependence of information is very likely to occur, can have a major impact on analysis results, and can and must be dealt with.

A common statistical method of quantifying dependency is through covariance, which is a measure of the linear relationship between two random variables. For a vector of random variables all pairwise covariances can be arranged in a variance-covariance matrix, typically denoted Σ. Let the estimates from the three information sources be arranged in a vector $\underline{\hat{\theta}} = \left[\hat{\theta}_s, \hat{\theta}_{EJ}, \hat{\theta}_{SC}\right]^T$ with variance-covariance matrix Σ, and let $\boldsymbol{1} = [1,1,1]^T$. Then the BLUE estimator[5] of θ is $\hat{\theta} = \left[1^T\Sigma^{-1}1\right]^{-1}1^T\Sigma^{-1}\underline{\hat{\theta}}$.

5.2 Bayesian Methods

The methodology outlined in Section 5.1 is from the standpoint of a classical statistician. The Bayesian perspective (Berger 1985, and Berry 1996) is another point of view that lends itself very naturally to the kinds of problems being considered here. In the classical framework the parameter θ is a fixed constant, data and beliefs are random, and the goal of the analysis is to estimate θ. Bayesians make the eminently reasonable point that beliefs and data are what are known and that it is the unknown θ that is the source of uncertainty. Consequently, Bayesians

[5] This expression is also appropriate if the estimates are uncorrelated.

view θ as a random variable, information as fixed, and the goal is to estimate the pdf of θ. The result of a Bayesian analysis is a pdf of the form $\pi(\theta \mid y)$, which is the probability distribution of θ conditional upon information y. Rather than estimating θ, as is in the classical approach, Bayesians think about predicting θ. Given the pdf $\pi(\theta \mid y)$ there are a number of natural predictions that can be computed such as the mean, median, or mode of $\pi(\theta \mid y)$. Both classical and Bayesian approaches have advantages and disadvantages, discussed in Section 5.3, as well as die-hard supporters and opponents. Our recommendation is to use that method that best suits the particular problem at hand.

The fundamental tool of Bayesian analysis is Baye's rule. The starting point for using Baye's rule is a pdf $\pi(\theta)$ that captures initial beliefs about θ and is known as the prior distribution. Let information be the observed random variable Y with pdf $f(y \mid \theta)$ where the notation reflects the fact that the behavior of Y depends upon a specific, although unknown, realization of θ. Because the behavior of Y depends upon θ, the observed value y must contain information about θ. The problem is how to combine that information with the original beliefs expressed in $\pi(\theta)$. The solution is given by Baye's rule, which states that the probability distribution of θ given y is $\pi(\theta \mid y) = f(y \mid \theta)\, \pi(\theta) / f(y)$. The distribution $\pi(\theta \mid y)$ is known as the posterior distribution, and provides a distribution of θ that incorporates the information contained in y. Given the observed information, the quantity f(y) is a fixed constant, so the result is often written as $\pi(\theta \mid y) = kf(y \mid \theta)\, \pi(\theta)$ where k is the constant required to make $\pi(\theta \mid y)$ a valid pdf (in particular, so that $\pi(\theta|y)$ integrates to 1). Baye's' rule is the basis of almost all Bayesian analysis, and is an incredibly simple rule for incorporating new information. It simply says the prior belief, $\pi(\theta)$, is multiplied by the pdf of the data $f(y \mid \theta)$ to obtain the updated, posterior distribution of θ.

Two general types of information were discussed in the previous section. The first was sample data whose probability distribution depended upon θ. This may be viewed as indirect information because the nature of θ is inferred by studying the information. The two cases of indirect data discussed were sample data from the component under study, with pdf $f(y|\theta)$, and sample data from a different but related component, with pdf $f(x|\theta, \sigma^2_\varepsilon)$[6]. The information structures $f(y|\theta)$ and $f(x|\theta, \sigma^2_\varepsilon)$ associated with indirect information are unchanged in Bayesian analysis. The second general type of information is direct information about θ. With direct information, no inference is required to learn about θ. Direct information was provided by expert elicitation, which revealed the expert's beliefs about θ. In classical analysis, those beliefs are interpreted as estimates of θ. In Bayesian analysis those beliefs are interpreted as estimates of a probability distribution for θ.

Because Bayesian analysis is built upon the idea of updating beliefs according to new information, it is particularly well suited to combining information, which is the main topic of this paper. In order to begin an initial prior distribution $\pi(\theta)$ is required. In the absence of any information, there exist probability distributions, known as diffuse priors, that reflect nearly complete uncertainty about the likely values of θ. However, in the current setting there is a natural choice for the initial prior, which is the pdf derived from the expert elicitation. Let the

[6] Recall that σ^2_ε is a parameter that measures the commonality between the components.

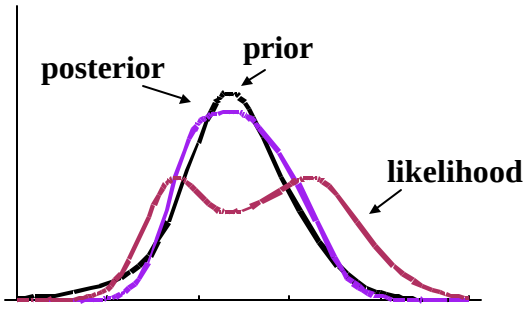
elicited pdf be $\pi(\theta)$. Now, the (potentially very complicated) problem is how to combine the expert information, contained in $\pi(\theta)$, with the sample information, contained in $f(y|\theta)$. The straightforward Bayesian solution is a direct application of Baye's rule, giving $\pi(\theta|y) = kf(y|\theta)\pi(\theta)$. The information from the similar component needs to be incorporated next. The simple trick to achieve this is to use the posterior distribution that results from incorporating the sample data as the prior distribution for incorporating the similar component data. That is, information is updated sequentially, with the posterior for a given update serving as the prior for the next. The only modification that must be made is that a prior distribution for the similarity measure σ^2_ε, given by $\pi(\sigma^2_\varepsilon)$, must also be included. This illustrates an important part of Bayesian analysis: prior distributions for all unknown parameters must be specified. In this setting, this is a probably a good thing. There is likely a great deal of uncertainty about the degree of similarity between the components, and the prior distribution $\pi(\sigma^2_\varepsilon)$ allows this uncertainty to be explicitly expressed. Combining the data from the three sources using Baye's rule, and assuming independent information, the final posterior distribution of θ is:

$$\pi(\theta, \sigma^2_\varepsilon|x_1,\ldots,x_n,y_1,\ldots,y_n) = k\, f(x_1,\ldots,x_n|\theta, \sigma^2_\varepsilon)\, f(y_1,\ldots,y_n|\theta)\, \pi(\theta)\, \pi(\sigma^2_\varepsilon).$$

This process generalizes immediately, so that it is very easy to incorporate new information as it becomes available. Figure 2 depicts the process.

Figure 2. Bayesian Estimation

$$\pi(\theta|y) = k\, f(y|\theta) * \pi(\theta)$$



Independence of information is of the same concern here as under classical analysis. Information that is dependent is redundant, and will be given improper weights if the expression given above is used to combine the information. If the dependence arises among sample data, the problem can be handled by including the variance-covariance matrix (as described above) in the probability distributions of the data. If the dependence arises from multiple elicitations, the nature of the solution is that a weighted aggregation of the various beliefs is required (Berger 1985). The best way to construct and utilize such weights is an area of active research.

5.3 Classical or Bayesian?

The previous two sections described classical and Bayesian methods for combining information. While the two methods may appear quite distinct, there are more similarities between them than differences. Both are reasonable methods of combining information, and should not give wildly differing results. Applying each to a given problem is a good way of investigating the robustness of the final conclusions. The reasons for any large differences should be explored, which may reveal particular model sensitivities or errors in formulation or computation.

Numerous criticisms of the Bayesian methodology can be levied from a philosophical level. However, an equal number can be charged against the classical methodology by those who are so inclined. The philosophical debate will be left to others, and this comparison of the two methodologies will proceed on more technical ground.

The major advantage of the Bayesian approach is that it optimally[7] combines the given information, even when information is coming from diverse sources. The Bayesian approach is also appealing because it directly accommodates information that comes in the form of beliefs about the likely values of θ. When sample data on the actual component are scarce, such beliefs are likely to be a major source of information. The major criticism of the Bayesian approach is that it can be very difficult to specify prior distributions (Berger 1985), which require functional forms as well as parameter values. Because of this difficulty, it is recommended that sensitivity analyses be performed to determine the dependence of the results on the prior distributions. Also, Bayesian calculations can become extremely complicated, even in light of toady's computing resources.

The particular classical estimation method discussed above was based upon the BLUE criterion. One advantage of this approach is that it is relatively distribution free. That is, an estimator developed under the BLUE criterion remains the same regardless of the underlying probability distributions. The price paid for this generality is that all the information in the observed data is not utilized, which is a considerable liability. This tradeoff reoccurs throughout statistical theory. Assuming the form of underlying probability distributions allows one to fully utilize information, but creates the risk of substantial error if the assumption is incorrect. Another advantage of classical estimators methods is that they are relatively easy to compute. Robustness and ease of computation are likely the reasons the majority of applied statistical analyses are currently performed under the classical umbrella.

6. System Analysis/Decision Making

Using diverse information to derive the probability distribution of a metric for a single system component was discussed in Section 5. The same can be performed for each component in the system. The selected system representation, as described in Section 4, can then be used to derive the probability distributions of system-wide metrics. Given the decision maker's objective function, the vast body of decision theory can then be utilized to guide decisions.

[7] Optimal in the sense that it minimizes certain loss functions (???ref).

As an extremely simple example, consider a system of two components, A and B, arranged in series. The decision to be made centers on the reliability of the system. Is the system reliability great enough? If not, what should be done? Upgrade A? Upgrade B? Upgrade them both? Do more testing? In order to answer these questions the behavior of the system must be described. The functioning of the components and the system as a whole is represented by the simple Bernoulli random variables Y_A, Y_B, and Y_S that take on the value 1 for functioning and 0 for failed. The probability distributions of the components are $f(y_A; p_A)$ and $f(y_B;p_B)$, where p_A and p_B are the probabilities that the components function, and represent the integration of all available information about each of the components. The definition of the system gives the relationship $Y_S = \min[Y_A, Y_B]$, which can be used to derive the probability distribution $f(y_S; p_S)$. In this simple example the result can be derived analytically.

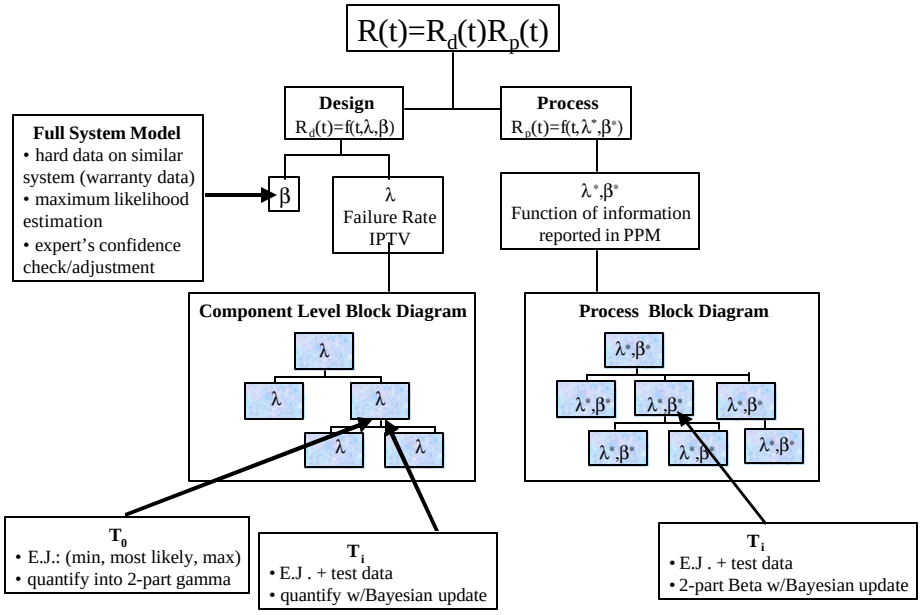
Suppose the decision maker needs to be 90% confident that reliability is greater than 0.99. If the pdf of Y_S provides this confidence then no further work needs to be done. However, what if the pdf of Y_S does not support the requirement? Sensitivity analysis can be used to guide the next step. Is the problem lack of information? This can be investigated by reducing the uncertainties in the pdf's associated with A and B. Is it A or B that needs to be upgraded? This can be investigated by increasing the reliability of each component and examining the impact on results.

More complicated, and realistic, problems require the distribution of system-wide metrics be derived through Monte Carlo methods. For example, consider the systems represented in Figure 3. In these systems, Monte Carlo simulation is used to roll the component and subsystem level up to the system level, or to propagate new information from a higher level down into the system to study the drivers of uncertainty. For example Figure 3 shows the integration of two very different processes, manufacturing and design (Booker et al. 2001) where such propagation has been done.

7. Conclusions

This paper has provided a short introduction into Information Integration Technology. While the name refers to the utilization of diverse system information to solve complex decision problems, it should be clear that constructing such a solution requires a corresponding level of integration of the information possessed by a diverse group of statisticians, social scientists, natural scientists, etc. There is no shortage of interesting problems that remain to be solved for a more complete implementation of IIT.

Figure 3. Automotive System Representation and Analysis



REFERENCES

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

Berry, D. (1996). *Statistics A Bayesian Perspective.* Duxbury, New York.

Higgins, J. and S. Keller-McNulty (1995). *Concepts in Probability and Stochastic Modeling*. Duxbury Press, New York.

Keller-McNulty, S. and M. McNulty (2000). Show Me The Data: Statistical Representations. To appear in *Theoria et Historia Scientarium*.

Meyer, M. and J. Booker (1991). *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Academic Press, San Diego.

Meyer, M., K. Butterfield, W. Murray, R. Smith, and J. Booker (2001). Guidelines for Eliciting Expert Judgement As Probabilities or Fuzzy Logic. *Fuzzy Logic and Probability Applications,* SIAM Series.

Mood, A., F. Graybill, and D. Boes (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

Paton, R.C., D. M. Jones, and M. J. R. Shave (1994). The Role of Verbal and Visual Metaphors in Scientific Theories, Conference on Information-Oriented Approaches to Language, Logic, and Computation, St. Mary's College, Moraga, California, June 13-15.

Smith, R.E., J.M. Booker, T. R. Bement, M.A. Meyer, W.J. Parkinson, M. Jamshidi (1998). The Use of Fuzzy Control System Methods for Characterizing Expert Judgment Uncertainty Distributions. *Proceedings PSAM 4 International Conference*.

Wilson, A. and S. Keller-McNulty (2000). Statistical Representations For Information Integration. Los Alamos National Laboratory Technical Report, LA-UR xxxx.