

Assessing the Value of Information for Multiple, Correlated Design Alternatives

Shawn P. Capser and Efstratios Nikolaidis
University of Toledo

ABSTRACT

Design optimization occurs through a series of decisions that are a standard part of the product development process. Decisions are made anywhere from concept selection to the design of the assembly and manufacturing processes. The effectiveness of these decisions is based on the information available to the decision maker. Decision analysis provides a structured approach for quantifying the value of information that may be provided to the decision maker. This paper presents a process for determining the value of information that can be gained by evaluating linearly correlated design alternatives. A unique approach to the application of Bayesian Inference is used to provide simulated estimates in the expected utility with increasing observations sizes. The results provide insight into the optimum observation size that maximizes the expected utility when assessing correlated decision alternatives.

CITATION: Capser, S. and Nikolaidis, E., "Assessing the Value of Information for Multiple, Correlated Design Alternatives," *SAE Int. J. Commer. Veh.* 10(1):2017, doi:10.4271/2017-01-0208.

INTRODUCTION

When making design decisions, there becomes an intrinsic value to the information related to the decision making criterion. In the case of a reliability related criterion, the value of information may come from the understanding of a failure probability or failure rate. This value of information may also be referred to as the value of research; which represents the level of investment by the "decision maker" to learn more about the design alternatives. For example, if the decision maker intends to compare the failure probabilities of two circuit boards he or she may consider investing in an Environmental Stress Test or performing some level of simulations to evaluate the thermal stresses on the components that make up the circuit board. These activities will increase the information going into the decision making process.

It is important to be able to properly quantify the value of information that may come from the data collection efforts across the decision alternatives. A critical part of this analysis is properly accounting for the level of correlation that may exist between linearly correlated design alternatives. The value of information that can be obtained by evaluating correlated design alternatives will be inversely proportional to the strength of the correlation that may exist between them. That is to say, for two strongly correlated design alternatives the information gained by evaluating one alternative will likely provide insight into the performance of the other alternative(s).

There are standard means for determining the sample size for a test as described above; however, these methods require the assumption that the competing alternatives are independent of one another with regards to the decision making criterion, e.g. failure probability. One such alternative method would be to use Two-Sample Log Rank Test for determining differences in survival rates of just two **independent** reliability distributions ([Mathews \(2010\)](#)), e.g. two design alternatives.

The method just described is restricted to just two alternatives. There very well could be a need of the decision maker to have an approach for determining the level of analysis or testing that would be required to properly evaluate and compare 2 or more **correlated** alternatives. In most real-world applications, the competing design alternatives are **not** independent. It could be hypothesized that the competing designs share many of the same failure modes and mechanisms and, as a result, have some level of correlation among them.

This paper is intended to address the above issues through the application of Decision Analysis. The approach will address the decision maker's risk tolerance, as well as the cost of information, when determining the sample size for comparing correlated design alternatives. The approach will use Bayesian Inference to incorporate *a priori* knowledge of the design alternatives that support the elicited belief of the decision maker before obtaining any information on the design alternatives. The approach then extends the *a priori* belief across a random field of events in order to evaluate the expected utility and value of information that maximizes the decision.

The rest of the paper compares two methods for applying Bayesian inference in estimation of the value of information; the first uses numerical integration and the second the Metropolis-Hastings algorithm.

LITERATURE REVIEW

Bayesian Approach

Raiffa, Howard, and Schlaifer (1961), and [Raiffa \(1968\)](#) formulated and solved a basic problem of estimation of the value of information. Raiffa demonstrated the calculation of the expected increase in value

of a decision. He also presented a formulation that accounts for the risk attitude of a decision maker by maximizing the expected utility of a decision.

Lindley (1997) studied the fundamental problem of finding the sample size required in order to make an inference or decision about a population. Lindley suggested that Raiffa and Schlaifer (1961) could be the first to consider this sample size determination problem. In order to find the optimum sample size, Raiffa and Schlaifer formulated and solved the following optimization problem, which consists of two loops:

1. Outer loop: Find the optimum sample size n to maximize the expected utility of decision d . $Utility(n, \mathbf{x}, d, \boldsymbol{\theta})$ depends on the number observations, (n), evidence ($\mathbf{x}$), decisions (d) and parameter ($\boldsymbol{\theta}$).
2. Inner loop: Find the optimum design, which has maximum expected utility, given that n tests were performed and $\mathbf{x}$ failures were observed. The optimization problem in the inner loop is solved repeatedly for many observed failure counts $\mathbf{x}$.

In principle, solving the above set of nested optimization problems is expensive. The cost increases exponentially with the number of alternative designs and tests. For example, consider five alternative designs ($d = 5$). The decision maker tests 100 realizations of each design. Then decision maker could observe from zero to 100 failures in each test. The total number of possible outcomes of the series of the 100 tests is $5 \cdot 101^5 = 5.255 \cdot 10^{10}$.

Lindley (2000) wrote a comprehensive review of the disciplines of probabilities and statistics. He stated three principles for finding the best course of action in a decision under uncertainty:

1. Model uncertainties by using probabilities.
2. Quantify the value of each possible consequence of a decision by the utility of that consequence.
3. Select the option that has the maximum expected utility.

The strength of this approach is its axiomatic foundation. One can prove that a rational decision maker must maximize the expected utility of a decision (French (1986)).

Clemen and Winkler (1995) presented a method for the estimation of the value of information using data from depended experts. Each expert provides a point estimate of a parameter of interest. The method requires the prior joint PDF of the errors. The method assumes that the errors follow a Gaussian probability density function (PDF) with zero mean value and some known correlation matrix. The method is very efficient because it exploits the conjugacy of the prior and the error in the estimates of the parameter of interest, which are assumed normal.

Clemen and Winkler (1995) provided an example where estimates of the parameter of interest were elicited from correlated experts. An example of such a parameter is a probability. It was observed that positive dependence reduces dramatically the value of information.

These authors developed the method for a single parameter. They quantified the value of a decision with dependent experts by the equivalent number of independent experts whose estimate provided the same variance as that of the independent experts. Clemens and Winklers method is not applicable to decisions where the parameter of interest must be positive (for example, a probability of success or failure), which is the case in this paper.

The computational cost of Clemen's and Winkler's method will increase if the users want to account for their risk attitude by using utility functions because the estimation of the expected utility requires numerical integration.

Galvan et al. (2015) presented a study for calculation of the expected utility of a decision in automotive design. They only calculated the expected value of perfect information (EVPI) because of the large computational cost of their method.

Chávez and Henrion (1994) studied the problem of the high computational cost of the estimation of the value of information to support decisions involving noncombatant evacuations.

Brennan and Kharroubi (2007) recommended that Markov Chain Monte Carlo (MCMC) be the standard for the estimation of the value of information. Their study suggested that the estimation of the value of information with dependent design options can be addressed by using the Metropolis-Hastings algorithm (Metropolis et. al., 1953). This observation is consistent with the findings of the authors of this paper.

Nikolaidis, et al. (2012) estimated the value of information using standard Monte Carlo simulation. They considered alternative designs with statistically independent failure probabilities.

Nikolaidis et al. (2013) extended their method to studies in which the failure probability is calculated efficiently using Importance Sampling (2013). They demonstrated that Importance Sampling enables a decision maker to estimate probabilities as low as 10^{-4} using a modest number of states based on the performance function (e.g. 10,000).

Frazier and Powell (2010) studied the problem of optimizing a budget allocation for collection of information in order to maximize the expected utility of a decision. In their example the expected value of information was a convex function of the number of observations, for fewer than approximately 10 observations. This means each additional observation increases the value of a decision more than the previous observation. For more than about 10 observations the expected utility function became concave. In the same study, Frazier and Powell (2010) considered decisions with as many as 100 alternatives.

Often, models of uncertain variables are based on expert judgments because of scarcity of data. Hammitt and Shlyakhter (1999) studied the effect of overconfidence in expert opinions. They observed that experts often underestimate variability in the problem parameters and in the probability of obtaining surprising data.

Frequentist Approach

This approach seeks to determine the sample size needed to estimate a statistic such as an expectation or a relative frequency using only observations.

Adcock (1997) reviewed both frequentist and Bayesian methods for determining the required sample size for estimating a statistic of a population, such as the mean value of a variable or the correlation coefficient of two variables. This problem is called Sample Size Determination (SSD). The frequentist method for a random variable with a known type of distribution is formulated as follows: given a required absolute error δ and probability $1-\alpha$, find sample size n to ensure that the mean value of the variable is within a given interval with a high probability, such as 0.95.

METHODOLOGY

Bayesian Inference - Single Dimension

The technical approach is based on the application of the Bayesian probability model. Kelly and Smith (2011) presented the process of Bayesian Inference as:

1. Specify the *prior density* that represents the knowledge, as well as the uncertainty, in the parameter of interest, e.g. failure probability.
2. Specify the *stochastic model* which represents the process of observing the random events; i.e. likelihood.
3. Collect *data*
4. Calculate *posterior density* by updating the prior based on the observations
5. *Validate* the model

The above process is supported with the application of Bayes' formula:

$$f(\theta|\mathbf{x}) = \frac{f(\theta) \cdot L(\mathbf{x}|\theta)}{\int_0^1 f(\theta) \cdot L(\mathbf{x}|\theta) d\theta} \quad (1)$$

where, $f(\theta)$ is the the Prior Density and $L(\mathbf{x}|\theta)$ the Likelihood and θ and $\mathbf{x}$ represent the parameter of interest and the data, respectively.

Note: The above process shows an update of the Prior based on the collection of data. The intent of the methods presented in this paper are based on random realizations of data coming from sampling the *prior density* and the suggested *stochastic model*. The process depends on performing several thousand simulations to ensure appropriate coverage on the potential range of realizations.

Bayesian Inference - Multiple Dimensions

In the proposed problem, the above can be expanded to address a joint posterior, given a joint prior to describe dependency of the initial *beliefs* on two or more parameters (θ). As an example, a bivariate density may be considered that is comprised of multiple

failure probabilities and that the probability model parameters (θ) are correlated with Correlation Matrix (Ψ). The resulting Bayesian posterior density would be:

$$f_{\Theta}(\theta|\mathbf{x}) = \frac{f_{\Theta}(\theta) \cdot f_{\mathbf{X}}(\mathbf{x}|\theta)}{\int_{\Theta} f_{\Theta}(\theta) \cdot f_{\mathbf{X}}(\mathbf{x}|\theta) d\theta} \quad (2)$$

Note: **bold** symbols represent vectors.

Joint Prior $f_{\Theta}(\theta)$

Just as with the single dimensional case, each design alternative is given a marginal density that describes the location and scale of the parameter of interest (θ). For a multiple dimension decision problem, the prior becomes the joint density of these marginals. In this model, $f_{\Theta}(\theta)$ describes the joint density for the parameters of the multiple decision alternatives, i.e. $\theta_1, \dots, \theta_n$.

If the decision alternatives are considered independent, then the joint density simply becomes the product of the respective marginal densities of the parameters for each decision alternative. However, if the decision alternatives are considered correlated among one another, then further consideration must be given as to how the joint density on the prior will be constructed. Sklar's Theorem (1959) offers the suggestion on the use of Copulas to support such a construction.

Clemen and Reilly (1997) suggested the application of Sklar's theorem for the construction of a joint density with correlated marginals as:

$$f(x_1, \dots, x_n) = f_1(x_1) \times \dots \times f_n(x_n) \cdot c[F_1(x_1), \dots, F_n(x_n)] \quad (3)$$

where,

$$c = \frac{\partial^n C}{(\partial F_1, \dots, \partial F_n)}$$

is referred to as the *copula density*, and C is the *Copula Function*.

There are numerous options for copulas (C) (Nelson (2006)); however, for the problem discussed in this paper, literature has demonstrated frequent application of a *Gaussian Copula*.

Gaussian Copula - c_{Ψ}^{Gauss}

For this paper the correlation among the prior marginal densities is modeled by using a Gaussian Copula Density (c_{Ψ}^{Gauss}). Using Equation 3, the joint prior is defined as the product of the Gaussian Copula density and multiple marginal densities:

$$f_{\Theta}(\theta) = (c_{\Psi}^{Gauss}) \cdot \prod_{i=1}^n f_{\Theta_i}(\theta_i)$$

The Gaussian Copula is defined as:

$$c_{\Psi}^{\text{Gauss}}(\mathbf{u}) = \frac{1}{\sqrt{\det \Psi}} \exp \left(-\frac{1}{2} \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_n) \end{pmatrix}^T \cdot (\Psi^{-1} - \mathbf{I}) \cdot \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_n) \end{pmatrix} \right) \quad (5)$$

where,

- Ψ is the Correlation Matrix, based on *Pearson's Product-Moment Coefficient* (ρ). The Correlation Matrix is given as:

$$\Psi = \begin{bmatrix} 1 & \cdots & \rho_{1,n} \\ \vdots & \ddots & \vdots \\ \rho_{n,1} & \cdots & 1 \end{bmatrix}$$

- u_i is the cumulative probability from the marginal distribution of the i^{th} parameter

Likelihood

The Likelihood portion of Equation 2 is conditional on the data. It is the product of the densities used to describe the stochastic behavior of making the observations, specifically

$$f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n f_{X_i}(x_i|\theta_i) \quad (6)$$

where, x_i are the observed failure counts on the respective design alternatives.

Utility and Expected Utility

The risk tolerance of the decision maker is described by a utility function. The utility function is used to translate the values of the parameters from the different decision alternatives into a value that represents the risk tolerance of the decision maker. A utility function is a transfer function between the decision parameters and the risk tolerance of the decision maker.

The utility ranges between 0 and 1; where, a value of 1 represents the target value of the parameter of interest, e.g. with regards to failure probability a utility of 1 would be associated with 0 failure probability.

There is no specific form for a utility function, but a typical form presented in literature (Raiffa (1970)) is an Exponential utility function and is described in Equation 7.

$$U(x; T) = 1 - e^{-\frac{x}{T}} \quad (7)$$

where, T scales the decision maker's risk tolerance. Figure 1 compares two exponential utility functions for varying levels of risk tolerance - risk averse vs. risk seeking; such that risk averse behavior will always have a concave shape.

Expected Utility with No Information

The expected utility with no information is based purely on the *a priori* knowledge of the decision maker and does not consider any opportunity for additional data or "information" (Equation 8). As a result, the value of information is only the maximum of the expected utility of the individual design alternatives.

$$EUNI = \max \left(\int_{\Theta} f_{\Theta}(\boldsymbol{\theta}) \cdot U(\boldsymbol{\theta}) d\boldsymbol{\theta} \right) \quad (8)$$

Expected Utility with Perfect Information

Expected Utility with Perfect Information is based on knowing the value of θ_i and then calculating the Expected Utility based on the marginal, posterior density (Equation 9). The value represents the highest achievable utility that would be obtained with having complete knowledge, i.e. no uncertainty, of the design alternatives.

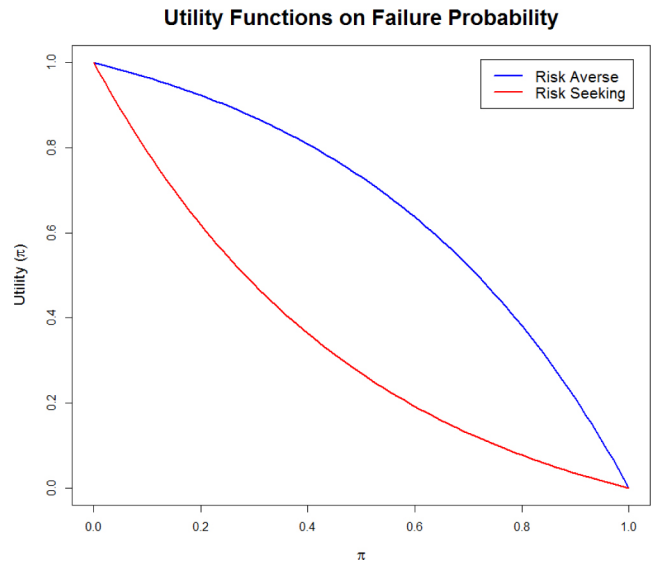


Figure 1. Example of Utility Functions based on Risk Tolerance

$$EUPI = \int_{\Theta_1} \cdots \int_{\Theta_n} f_{\Theta}(\boldsymbol{\theta}) \cdot \max(U(\boldsymbol{\theta})) d\theta_1 \cdots d\theta_n \quad (9)$$

The value of perfect information becomes the difference in expected utility with Perfect Information with that of the expected utility with no information.

Certainty Equivalent

The certainty equivalent is a payoff in the sure amount for a given decision scenario. This is the amount that the decision maker is willing to risk in order to undertake a decision with uncertainty. The certainty equivalent can be expressed in terms of the expected utility of the decision as:

$$CE = U^{-1}(E[U(\theta)]) \quad (10)$$

For the decision maker, the certainty equivalent may represent a better way to describe the value of the information that can be gained from the test or simulation.

Process of Determining Expected Utility

The approach being considered is depicted in Figure 2. The figure shows the procedure for a specific scenario of just two design alternatives. This process is repeated over multiple **chains**. A chain is a single iteration of the decision making process. It represents a single realization of an overall σ -field, or sample space, that exists when examining multiple observations of each design alternative. Typically, several thousand chains are simulated in order to best examine the overall σ -field that comprises all combinations of events coming from evaluating multiple design alternatives.

The process shows that the maximum expected utilities are calculated for each potential realization of r_1 and r_2 . The resultant expected utility of the Decision is then taken to be the average of the maximum utilities of each chain in the simulation.

The process consists of five steps:

1. Sample 'm' pairs of parameter θ from the joint prior; where, 'm' represents the number of chains used in the simulation, i.e. each pair is a potential realization of a data set. The process of sampling the correlated parameters and corresponding realizations consists of the steps below:
 - a. Draw a random sample from the Joint Standard Normal Distribution

$$\mathbf{y} \sim N(0, \Sigma)$$

- b. Determine Standard Normal Inverse of Joint Random Samples

$$\mathbf{u} = \Phi(\mathbf{y}, 0, 1)$$

- c. Calculate quantiles on model parameters from Cumulative Probabilities in (3)

$$\theta = F_{\Theta}^{-1}(\mathbf{u})$$

2. For each set of parameters, draw random samples of possible observations based on draws from the stochastic probabilistic model based on θ as:

$$\mathbf{r} \sim f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)$$

3. Calculate the expected utilities for each design alternative based on the sampled values for the failure probabilities and respective failure counts. The calculations for the expected utilities are based on the updated priors after the realizations of each r_p , i.e. the posterior densities. The expected utilities for the multiple decision alternatives are then calculated as follows:

$$EU_i = \int_{\Theta_1} \cdots \int_{\Theta_n} f_{\Theta}(\theta) \cdot f_{\mathbf{R}|\Theta}(\mathbf{r}|\theta) \cdot U(\theta_i) d\theta_1 \cdots d\theta_n \quad (11)$$

4. Calculate the maximum of the expected utilities in the j^{th} state of the simulation

$$EU_{max_j} = \text{maximum}(EU)$$

5. Take the average of the maximum expected utilities across all states

$$\overline{EU}_{Decision} = \frac{\sum_{j=1}^m EU_{max_j}}{m}$$

The above procedure can be managed using standard software packages that support numerical integration, such as MathCad, Matlab, etc.; but for this paper all calculations were carried out in R - a free and publically available programming language and environment for statistical computing (R (2016)).

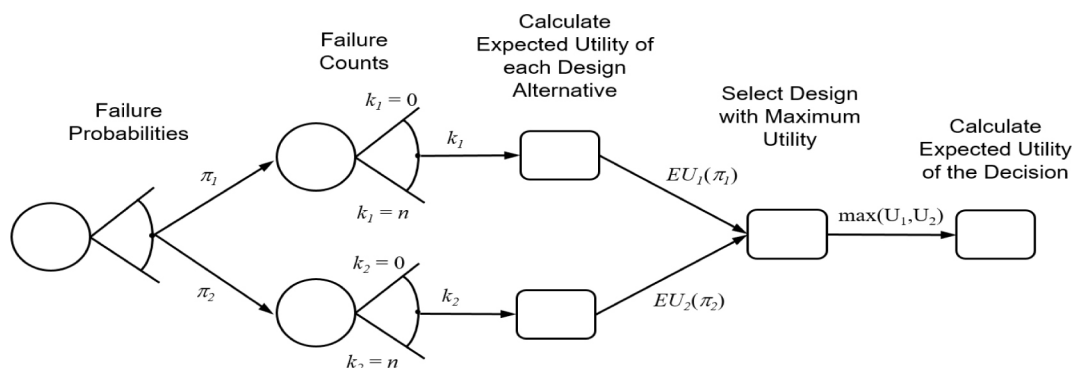


Figure 2. Flow Diagram for Simulation and Calculation of Utility Values