On the Efficient Allocation of Resources for Hypothesis Evaluation in Machine Learning: A Statistical Approach

Steve A. Chien[a], Jonathan M. Gratch[b], and Michael C. Burl[a,c]

[a]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

[b]Beckman Institute and Dept. of Computer Science, University of Illinois, Urbana, IL

[c]Department of Electrical Engineering, California Institute of Technology, Pasadena, CA

Keywords: machine learning, the utility problem, statistics, planning, scheduling, parameter estimation, adaptive problem–solving

Abstract

In this report we consider a decision–making problem of selecting a strategy from a set of alternatives on the basis of incomplete information (e.g., a finite number of observations); the system can, however, gather additional information at some cost. Balancing the cost of acquiring additional information against the expected utility of the information to be acquired is the central problem we address.

In our approach, the cost and utility of applying a particular strategy to a given problem are represented as random variables from a parametric distribution. By observing the performance of each strategy on a randomly selected sample of problems, we can use parameter estimation techniques to infer statistical models of performance on the general population of problems. These models can then be used to estimate: (1) the utility and cost of acquiring additional information; and (2) the desirability of selecting a particular strategy from a set of choices.

The techniques we have developed have been applied to adaptive problem–solving, in which a decision–making system automatically tunes various control parameters to improve performance. Empirical results are presented that compare the effectiveness of the hypothesis evaluation techniques applied to speedup learning for a NASA antenna scheduling application.

1. Introduction

In machine learning and basic decision making in AI, a system must often reason about alternative courses of action in the absence of perfect information – frequently balancing the cost of acquiring additional information against the expected utility of the information to be acquired. When one wishes some sort of statistical guarantees on the (local) optimality of the choice and/or the guarantee of rationality, a statistical decision theoretic framework is useful. This problem of decision–making with incomplete information and information costs can be analyzed in two parts:

> A1: How much information is enough? At what point do we have adequate information to select one of the alternatives?
>
> A2: If one wishes to acquire more information, which information will allow us to make the best possible decision at hand while minimizing information costs?

Possible solutions to this decision–making quandary depend on the context in which the decision is being made. This paper focuses upon the decision–making problems involved in adaptive problem–solving. Adaptive problem–solving occurs when a system has a number of control parameters which affect its performance over a distribution of problems. When the system solves a problem with a given set of control parameters, it produces a result which has a corresponding utility. The goal of adaptive problem–solving is: given a problem distribution, find the setting of control parameters which maximizes the expected utility of the result of applying the system to problems in the distribution.

Adaptive problem–solving is an important application for hypothesis evaluation techniques. Frequently, while a general class of problems may be intractable, effective domain specific (e.g., heuristic) solution strategies may exist. Unfortunately, determining these solution methods is a time consuming process which requires significant knowledge about both the the application domain and problem–solver. Hypothesis evaluation techniques offer significant promise in allowing an adaptive problem–solver to automatically search the space of control strategies to find an effective solution strategy for a specific applications, thus avoiding this bottleneck.

More rigorously, the adaptive problem–solving problem can be describes as follows. Given a flexible performance element PE with control points $CP_1 \ldots CP_n$, where each control point CP_i corresponds to a particular control decision and for which there is a set of alternative decision methods $M_{i,1} \ldots M_{i,k}$,[1] a control strategy is a selection of a specific method for every control point (e.g., STRAT = $<M_{1,3}, M_{2,6}, M_{3,1}, \ldots>$). A control strategy determines the overall behavior of the problem–solver. It may effect properties like computational efficiency or the quality of its solutions. Let PE(STRAT) be the problem solver operating under a particular control strategy. The function U(PE(STRAT), d) is a real valued utility function that is a measure of the goodness of the behavior of the scheduler over problem d. The goal of learning can

1. Note that a method may consist of smaller elements so that a method may be a set of control rules or a combination of heuristics. Note also that a method may also involve real–valued parameters. Hence, the number of methods for a control point may be infinite, and there may be an infinite number of strategies.

be expressed as: given a problem distribution D, find STRAT so as to maximize the expected utility of PE. Expected utility is defined formally as:

$$\sum_{d \in D} U(PE(STRAT), d) \times probability(d)$$

For example, in a planning system such as PRODIGY [Minton88], when planning to achieve a goal, control points would be: how to select an operator to use to achieve the goal; how to select variable bindings to instantiate the operator; etc. A method for the operator choice control point might be a set of control rules to determine which operators to use to achieve various goals plus a default operator choice method. A strategy would be a set of control rules and default methods for every control point (e.g., one for operator choice, one for binding choice, etc.). Utility might be defined as a function of the time to construct a plan, cost to execute the plan, or some overall measure of the quality of the plan produced.

This paper describes the application of two hypothesis evaluation methods to adaptive problem–solving: interval–based selection and expected loss selection. Specifically, our approach draws upon techniques from statistics in the of area of parametric statistical models to model the uncertainty in utility estimates. In parametric statistical models, one presumes that data is distributed according to some form of model (e.g., the normal distribution, the poisson distribution, etc.). This distribution can be described in terms of a fixed set of parameters (such as mean, variance, etc.). If one can infer the relevant parameters for the distribution underlying the data (so–called parameter estimation), then because the uncertainty in utility estimates is explicitly modelled in the statistics, two questions regarding utility can be answered: (1) which alternative is likely to have the highest expected utility; and (2) how certain are we of this ranking of the alternatives.

Of course, the accuracy of all of these estimates is dependent upon the goodness of fit of the parametric model used to model the utility estimates. Generally, we use the normal (gaussian) distribution model as our parametric model which, as shown by the Central Limit Theorem, is a good approximation to the underlying distribution of many real–world problems.

The first method, called interval–based selection, involves quantifying the uncertainty in competing hypotheses by using the statistical confidence that one hypothesis is better than another hypothesis. In this approach the system allocates examples to show that one hypothesis dominates all the other hypotheses with the specified confidence. These methods also rely upon an indifference parameter – if two hypotheses differ in performance by less than this amount, either is acceptable.[2]

The second method uses the decision theoretic concept of expected loss [Russell & Wefald 89, Russell & Wefald 91], which measures the probability of making a less preferable decision weighted by the lost utility with respect to the alternative choice. In the expected loss approach, the system acquires information until the expected loss is reduced below some specified threshold. This approach has the added benefit

2. This formalism is analogous to the PAC [Valiant84] framework – "probably" "approximately" "correct" maps onto "probably" "close to" "highest utility".

of not attempting to distinguish among two hypotheses with similar means and low variances (e.g., it recognizes indifference without a separate indifference parameter).

For both the interval-based and expected loss approaches, when selecting a best hypothesis, one must base this selection upon comparisons of the utility of the "best" hypothesis to the other possible hypotheses. Because there are multiple comparisons, the estimate for the overall error (or confidence) in a conclusion of selection of a best hypothesis, is based upon the errors associated with multiple smaller conclusions. For example, if we wish to show that H3 is the best choice among H1, H2, H3, H4, and H5, in the dominance–indifference approach, we might show that H1 and H3 are indifferent, that H3 dominates H2, H3 dominates H4, and H3 dominates H5. Thus if we wish a confidence level of 95%, with a straight sum error model and equal allocation of error, each of the individual hypotheses would need a 98.75% confidence level (since 4 x 1.25% = 5%).

The exact form of the error relationship between depends upon the particular error model used. However, taking additional examples can have varying effect on the reduction of error in these smaller conclusions. Also, there may be widely varying cost in applying additional examples to the various smaller conclusions. Thus, in many cases, it may be desirable to allocate the error estimates unequally. While we have implemented default strategies of equal allocation of errors, we have also developed a more sophisticated approach for allocating error levels to these smaller conclusions. In this approach, the system estimates the marginal benefit and marginal cost of sampling to extract another data point to compare two competing hypotheses. The marginal benefit is estimated by computing the decrease in the error estimate (or increase in confidence estimate) due to acquiring another sample presuming that the statistical parameters remain constant (e.g., in the case of the normal distribution, that the sample mean and variance do not change). The marginal cost is estimated using estimated parameters on the cost distribution for the relevant hypotheses. The system then allocates additional examples preferring the highest ratio of marginal benefit to marginal cost. We apply this strategy to both the interval-based and expected loss algorithms. Thus, in all, there are four new algorithms, interval-based equal error allocation, interval-based unequal error allocation, expected loss equal allocation, and expected loss unequal allocation.

The rest of this report is organized as follows. Section 2 describes the general hypothesis evaluation problem and frames the problem as Statistical Parameter Estimation. Section 3 describes the confidence interval approach and includes a discussions of it's strengths and weaknesses. Section 4 describes the expected loss approach and includes a discussion of its strengths and weaknesses. Section 5 describes and empirical evaluation of these techniques using synthetic and real–world scheduling data. Section 7 summarizes the principal points of this note.

2. The Hypothesis Evaluation Problem

We adopt a parametric statistical approach to the hypothesis evaluation problem. We begin by defining the adaptive problem–solving version of the hypothesis evaluation problem more concretely. Typically we have a set of problems D. Any particular problem d is selected from this set with probability $P_D(d)$.

We also have available a set of k potential alternative *strategies* $H_1 ... H_k$, for solving problems. Each hypothesis strategy H_i has associated with it a *utility distribution* U_i and a *cost distribution* C_i[3] which are induced by the probability distribution over D and the utility (or cost) of applying H_i to any particular problem.

Presumably the distributions Ui and Ci are unknown. However, the decision–making system can infer information about these distributions by observing a strategy Hi's behavior on problems drawn from D. Thus, the system can choose from among the following actions:

1. acquire more information; this action has cost drawn from C_i and provides another example from U_i; or
2. adopt a hypothesis strategy H_i.

Indeed the decision between 1 and 2 is exactly question A1 from Section 1: "How much information is enough? At what point do we have adequate information to select one of the alternatives?" If the system decides to acquire more information, then question is exactly A2: "which information will facilitate making the best possible decision at hand while minimizing information costs"

Our general approach to this problem consists of two parts: parameter estimation and hypothesis evaluation. In parameter estimation the underlying distributions of utility and cost are assumed to be of a particular form (e.g., normal, student T, etc.) reducing the problem to one of estimating parameters such as the mean and variance (for a normal distribution) from behavior on sample problems. In hypothesis evaluation, decision rules (to answer questions A1 and A2) are formulated based upon estimated parameters. As the result of applying these decision rules, the system may decide to gather additional information (samples), in which case it faces the decision between options 1 and 2 again. This process continues until option 2 is selected.

A common and reasonable assumption is to presume that the distribution of utility values and costs across different problems are normally distributed (also called a gaussian distribution). From this we can conclude that the average of values drawn from these distribution, the *sample mean*, is normally distributed about the true mean and confidence intervals regarding the true mean can be computed from the sample mean, sample variance, and number of samples. More concretely, one can show that the difference between the observed sample mean and true mean is normally distributed with 0 mean and with one *n*th the variance of the initial distribution, e.g. $\hat{\mu}-\mu \sim N(0,\frac{\sigma^2}{n})$.

Given the assumption of normality we can also conclude the *differential distribution* (the distribution of the difference in utility between any two strategies) is also normally distributed. This property is important in that it allows us to compute estimates that one strategy is better than (or roughly equivalent to) another strategy in expected utility by only maintaining information about the differential distributions. This simplifies the some of the mathematics. For example, for many applications the performance of strategies

3. An interesting special case to this problem is the *speed-up learning* problem where U_i and C_i are inversely related. Frequently in speed-up learning $U_i = -C_i$.

will likely be highly correlated (e.g. when strategies are small modifications to some common ancestor). Using the differential distributions encodes this correlational information without the need for explicitly computing covariance estimates[4].

The hypothesis evaluation problem that we pose is nearly identical to a problem frequently posed in the statistical literature, *the ranking and selection problem* [Bechhoffer54]. In the ranking and selection problem, one samples from various distributions of a known parameterized model (such as normal/gaussian), and the goal is to select the distribution with the highest mean. Ranking and selection strategies also allow for indifference zones, which specify how close to the highest the selected strategy must be.

Unfortunately, the methods developed in the statistics literature suffer from two drawbacks. First, because they have strong guarantees on the correctness of the decisions, they frequently require strong information on the distributions, such as presuming equal variances. Second, ranking and selection strategies typically attempt to minimize the overall number of examples. In the adaptive problem–solving special case, different examples have very different information costs (in particular consider the speed-up learning problem). Because statistical ranking and selection methods do not account for variable information cost, they are unlikely to perform as well when these costs vary greatly.

2.1 Notation

Throughout this paper we use the following notation:

U_i means the utility distribution for the hypothesis strategy H_i

Ci means the cost distribution for the hypothesis strategy Hi

μ_i is the true mean for the variable U_i

$\overline{U}_i$ is the sample mean for the variable U_i

S_{ui} is the sample standard deviation for U_i.

$\overline{C}_i$ is the sample mean for the variable C_i

U_{i-j} is the variable for the distribution computed by taking the utility of H_i minus the utility of H_j both solving the same problem. Note that this distribution is Gaussian (normal) if U_i and U_j are jointly gaussian even if U_i and U_j are not independent. $\overline{U}_{i-j}$ is the sample mean of this distribution and μ_{ui-uj} is the true mean of this distribution.

S_{ui-uj} is the sample standard deviation for the difference distributions for the utilities of H_i and H_j. Again, this difference distribution is created by taking values for H_i and H_j solving the same problem. Additionally, this distribution is gaussian (normal) if U_i and U_j are joint gaussian even if U_i and U_j are not independent.

We also define functions to allow computation of probabilities of normally distributed variables. The probability that a random variable y has a value in the interval (a,b) given that the variable is normally distributed with mean μ and standard deviation σ is:

4. This is a statistical technique known as blocking, see p. 299–300 of [Buringer80].

$$\Phi(a,b,\mu,\sigma) = \int_a^b \left(\frac{1}{\sqrt{2\pi}\sigma}\right) e^{-0.5\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

We further specialize for the standard normal distribution, with mean $\mu=0$ and standard deviation $\sigma=1$ as the following:

$$\Phi(a,b) = \Phi(a,b,0,1) = \int_a^b \left(\frac{1}{\sqrt{2\pi}}\right) e^{-0.5y^2} dy$$

3. The Interval–based Approach

The confidence interval–based approach depends on confidence parameter γ and an indifference parameter ε. The confidence interval approach attempts to show that with confidence γ there is a hypothesis strategy H_i for which every other hypothesis strategy H_j either: a) $E(U_{i-j}) > 0$; or b) $|E(U_{i-j})| < \varepsilon$. Intuitively, if such an H_i can be found it should be adopted because for every other hypothesis strategy H_j, with confidence γ, either H_i is either better than H_j (dominance) or H_i and H_j are close enough so that we do not care (indifference). This intuitive description will be further elaborated in the following paragraphs.

Consider two of the hypothesis strategies being evaluated H_i and H_j. Under the assumption that U_i and U_j are jointly normally distributed, the difference U_{i-j} is normally distributed. Hence, analyzing the difference U_{i-j} and computing the confidence that $U_{i-j} > 0$ gives the confidence that H_i dominates H_j. To represent the confidence in this pair-wise comparison of U_i and U_j we use the variable γ^*.

To compute the confidence that $U_{i-j} > 0$ we adapt a method for computing confidence intervals for the mean of a normal distribution with unknown variance from [Kreysig70]. However, our application differs from the standard confidence interval calculation as follows. In the standard problem, one is given a confidence level γ^*, and the task is to compute an interval such that the true mean lies in the interval with confidence γ^*. In our case, we are given the interval, and we wish to compute the confidence that the mean lies within the interval. We also use the normal distribution, rather than the t–distribution to model the distribution of the sample mean minus true mean. Further details of our approach and on computing confidence intervals are described in Appendix A. These assumptions result in the following formula (shown graphically in Figure 1):

$$\gamma^* = \Phi\left(0,\ \infty,\ \overline{U}_{i-j},\ \frac{S^2_{i-j}}{n}\right) \qquad \text{(dominance equation 1)}$$

$$= \Phi(-c,\ \infty) \text{ where } c = \overline{U}_{i-j}\frac{\sqrt{n}}{S_{ui-uj}}.$$

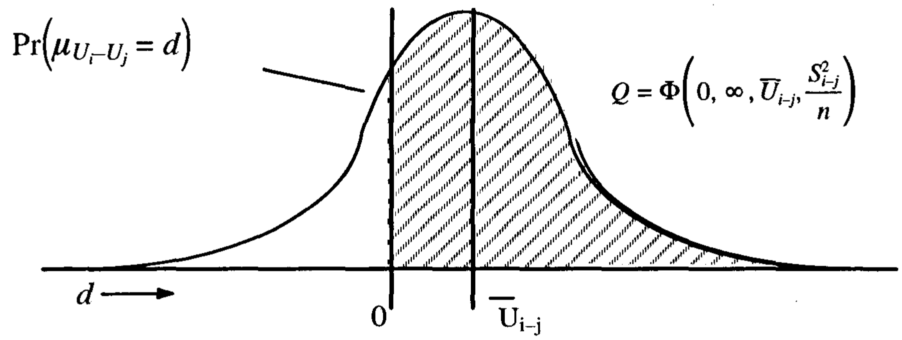


Figure 1: Probability distribution of μ_{ui-uj} : the difference between the utility means (expected utilities) U_i and U_j. Q is the probability that $\mu_{ui-uj} > 0$ (i.e. that $\mu_i > \mu_j$ which is the confidence that S_i dominates S_j).

To handle the case of indifference pruning, the confidence that the true mean of the difference utility distribution for the strategies is in the interval $\{-\varepsilon, \varepsilon\}$ can be computed similarly to the method described above, which results in the following (shown graphically in Figure 2):

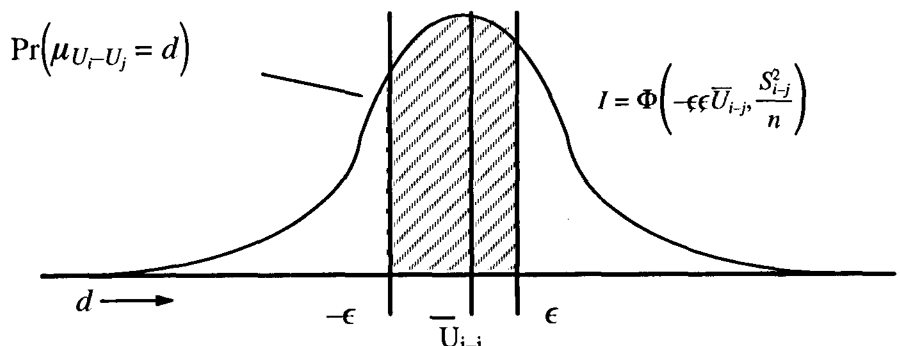


Figure 2: Probability distribution of μ_{ui-uj} : the difference between the utility means (expected utilities) U_i and U_j. I is the probability that $\varepsilon < \mu_{ui-uj} < \varepsilon$ (i.e. the confidence that S_i and S_j are indifferent).

$$\gamma^* = \Phi\left(-\epsilon, \epsilon, \overline{U}_{i-j}, \frac{S^2_{i-j}}{n}\right) \qquad \text{(indifference equation 1)}$$

$$= \Phi(a, b) \text{ where } a = \frac{(\overline{U}_{i-j} - \epsilon)\sqrt{n}}{S_{ui-uj}} \text{ and } b = \frac{(\overline{U}_{i-j} + \epsilon)\sqrt{n}}{S_{ui-uj}}$$

This can be interpreted using the confidence interval stopping criterion as follows. In the first case γ^* indicates our confidence in the hypothesis that the mean of the distribution U_i is greater than the mean of the distribution U_j, thus we prefer H_i over H_j (dominance). In the second case the difference between the means of U_i and U_j is less than ε with confidence γ^*, thus H_i and H_j are not worth distinguishing (indifference). If $\overline{U}_{i-j} < 0$, then H_j appears to be superior to H_i and we should be focussing on H_j and not H_i.

One complication is that the confidence term γ of the overall decision depends upon the number of hypotheses being considered. Suppose the system makes k conclusions, each with confidence γ^*. If we presume a pessimistic accumulation of error, we might project that the errors would add. In this case, to ensure that the total error is no more than $1 - \gamma$, we require that the sum of all of the k errors be less than $1 - \gamma$. This translates to the requirement that $k - \Sigma\gamma$'s $\leq 1 - \gamma$. The most straightforward way to achieve this is to force each of the errors to be less than $(1 - \gamma)/k$, however, this does not take advantage of the fact that reducing the error in some of the terms may be easier than in others. Pertaining to this issue we first outline an algorithm called STOP1 which distributes the error evenly, then show a variation on this basic algorithm STOP2 which accounts for the varying difficulty in reducing the error in each of the terms and takes into account the varying cost of sampling from each of the distributions.

3.1 The STOP1 Algorithm

The STOP1 algorithm can be described as follows. Let T be the set of hypothesis strategies $H_1 \ldots H_k$. Sample from each of the utility distributions $U_1 \ldots U_k$ some default number of samples n_0. Let H_i be the strategy in T which has the highest sample mean for U_i so far (hereafter called *the focus strategy*). For each strategy $H_j \in T$, if $\overline{U}_{j-i}$ is in the interval $\{-\varepsilon, \varepsilon\}$, attempt to show indifference. If not, attempt to show that H_i dominates H_j. Because the overall confidence must be γ, and we are drawing $k-1$ conclusions, evenly distributing the error indicates that the individual confidences must be:

$$\gamma^* = 1 - \frac{1-\gamma}{k-1} \qquad \text{(confidence equation 1)}$$

Unfortunately, in the worst case, for k strategies, the choice of the final selection may depend upon more than k–1 pair-wise comparisons. Consider the case where the focus strategy changes frequently while attempting to find a best strategy. Indeed, in the worst case, the final selection would depend upon all of the pair-wise combinations of selections of two of the k strategies (due to shifting of the focus hypothesis strategy). This is simply the sum of the integers from 1 to k–1, or k(k–1)/2. Thus, in the worst case, for the equal distribution of errors premise. the individual confidences must be:

$$\gamma^* = 1 - \frac{2(1-\gamma)}{k(k-1)} \qquad \text{(confidence equation 2)}$$

However, if n_0 is made large enough such that the focus strategy H_{high} changes rarely, the overall confidence will more closely resemble the linear relationship described in confidence equation 1. Indeed, if the errors tend to cancel each other, even this linear summation of errors will be an overestimate of the actual error.

Indifference is shown as follows. Compute the confidence that the true mean μ_{uj-ui} of U_{j-i} lies within the interval $\{-\varepsilon, \varepsilon\}$. If this confidence is greater than γ^* then indifference has been shown. Else, sample from

U_i and U_j as necessary until either: (1) the confidence that μ_{uj-ui} is within the interval $\{-\varepsilon, \varepsilon\}$ is greater than γ^* or (2) $\overline{U}_{j-i}$ goes above ε or below $-\varepsilon$. If $\overline{U}_{j-i}$ goes above ε, U_j now has a higher sample mean than U_i by a significant amount so that we should make H_j the target hypothesis and proceed. If $\overline{U}_{j-i}$ goes below $-\varepsilon$, H_j looks significantly worse than H_i so that we should attempt to show that H_i dominates H_j.

Dominance is shown similarly. Compute the confidence that the true mean μ_{ui-uj} of U_{i-j} is greater than 0. If this confidence is greater than γ^* we have shown dominance; otherwise sample from U_i and U_j as necessary until the either the confidence becomes greater than γ^* or $\overline{U}_{j-i}$ goes below ε. In this case, we might attempt to show indifference among H_i and H_j.

It is worth noting that sometimes when $\overline{U}_{j-i}$ is in the interval $\{-\varepsilon, \varepsilon\}$, there is more confidence in the claim that H_i dominates H_j than in the claim that H_i and H_j are indifferent. It is unclear whether a closed form exists that can be used to determine whether dominance or indifference has higher confidence. Thus, we avoid this problem by computing both the dominance and indifference and using the higher of the two confidences.

STOP1 ALGORITHM

let $T = \{H_1 \ldots H_k\}$
let $\gamma^* = \gamma/(k-1)$
solve n_0 problems with each strategy in T
let H_{high} be the strategy in T with the highest sample mean for U_{high}
compute utility comparison statistics
loop1
    let H_{high} be the strategy in T with the highest sample mean for U_{high}
    if for every $H_j \in T$ one of the following conditions holds
        U_{high} dominates U_j with confidence γ^*
        U_{high} and U_j are ambivalent with confidence γ^*
    THEN RETURN H_{high}
    ELSE
    select a strategy H_j such that neither of the following conditions holds
        U_{high} dominates U_j with confidence γ^*
        U_{high} and U_j are ambivalent with confidence γ^*
    generate data for the distribution U_{high-j}
    recompute utility comparison statistics
    continue with loop1

Note that the algorithm has been simplified for purposes of clarity. A realistic implementation would temporarily classify the strategies into indifference and dominance classes when confidence has been shown.

When S_{high} changes, these strategies must be returned to the unknown pool because they must be compared to the new S_{high}.

3.2 The STOP2 Algorithm

The STOP2 algorithm differs from the STOP1 algorithm in that it accounts for two factors ignored in the STOP1 approach. First, depending upon the sample variances and sample means of the individual U_{j-i} distributions, examples allocated to the distributions will have different effects on improving the confidence in a pair-wise dominance or indifference relation. Second, the cost of acquiring information (examples) varies from distribution to distribution. Because of these varying benefits and costs sometimes significant benefits can be derived from not bounding the statistical error equally across each of the pair-wise comparisons. The STOP1 algorithm, which does not account for these varying benefits and costs, uses equal bounds across the pair-wise comparisons. The STOP2 algorithm estimates the likely cost and benefit for each new example and allocates examples to the comparison with the highest estimated benefit divided by cost. This can result in a situation where each comparison is estimated to a different level of statistical error, although the sum of these errors still must remain below the overall bound of $1-\gamma$. As the individual pair-wise confidences may vary, we introduce the new notation γ^*_{ij} to signify the confidence that strategy S_i dominates or is indifferent with Strategy S_j as appropriate.

For example, as shown in Figure 3, if the uncertainty in determining the dominance of H_3 over H_5 has al-

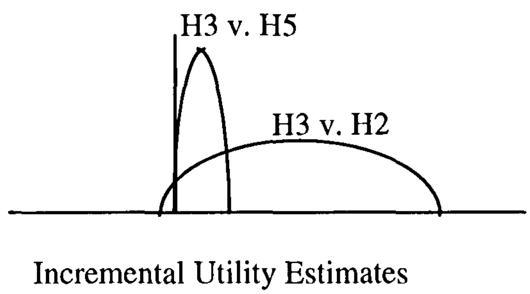


Figure 3: Varying Effects of Another Example

ready been reduced significantly, and the uncertainty in showing the dominance of H_3 over H_2 has not, additional examples to H_3 v. H_2 are likely to have greater effect on reducing the overall error than examples from H_3 v. H_5. More concretely, an additional sample from U_{32} will likely reduce γ^*_{32} by more than the amount that another sample from U_{35} is likely to reduce γ^*_{35}. Thus one can estimate the *marginal benefit* of allocating additional samples, the reduction in statistical error resulting from an additional example, by assuming that the mean and variance of U_{j-i} will change little and computing the increase in certainty.

The second factor considered by STOP2 and not by STOP1 is the varying cost of acquiring a sample. If acquiring an additional sample has an extremely high cost, it may not be worth the effort, even if the expected information gain is large. Likewise, a low information cost may make a lesser information gain look more attractive. To decide how best to allocate learning resources, STOP2 estimates *marginal cost.* This is the cost of acquiring another sample for a given pair-wise comparison and it consists of the cost of determining a utility value for each member of the pair. As each comparison shares the same hypothesis H_{high}, at least part of this cost may already have been incurred. Thus estimating the marginal cost involves two parts. First, determine which utility values must be determined (U_i, U_j, or both). Second, use the estimated means for C_i and C_j to estimate the cost of acquiring another sample U_j and U_i as appropriate. We then use the common greedy approach of selecting the course of action which has the highest ratio of marginal return to marginal cost. This process continues until a strategy emerges which can be shown with overall confidence γ to be dominant or indifferent with all other strategies.

Thus to estimate the marginal increase in dominance confidence from acquiring an additional example of U_{j-i} we use the following computation:

$$\Delta\gamma_{ji} \text{ for } U_{j-i} = \Phi\left(0, \infty, \overline{U}_{i-j}, \frac{\overline{S}^2_{i-j}}{n+1}\right) - \Phi\left(0, \infty, \overline{U}_{i-j}, \frac{\overline{S}^2_{i-j}}{n}\right)$$

$$= \Phi(-c1, -c2) \quad \text{where } c1 = \overline{U}_{i-j}\frac{\sqrt{n+1}}{S_{i-j}} \text{ and } c2 = \overline{U}_{i-j}\frac{\sqrt{n}}{S_{i-j}}$$

Of course, since the system computed the confidence for n samples when the nth sample was taken, there is no need to evaluate the second part of the integral. Thus the operational formula to compute the marginal increase in confidence is:

$$\Delta\gamma_{ji} \text{ for } U_{j-i} = \Phi(-c1, \infty) - previous\ \gamma^{*ji} \qquad \text{(marginal confidence dominance equation)}$$

where c1 is as defined above.

Similarly, we estimate the marginal increase in indifference confidence from acquiring an additional example of U_{j-i} as follows:

$$\Delta\gamma_{ji} \text{ for } U_{j-i} = \Phi(a1, b1) - \Phi(a2, b2)$$

$$\text{where } a1 = \frac{(\overline{U}_{i-j} - \epsilon)\sqrt{n+1}}{S_{ui-uj}} \text{ and } b1 = \frac{(\overline{U}_{i-j} + \epsilon)\sqrt{n+1}}{S_{ui-uj}}$$

$$\text{and} \quad a2 = \frac{(\overline{U}_{i-j} - \epsilon)\sqrt{n}}{S_{ui-uj}} \quad \text{and} \quad b2 = \frac{(\overline{U}_{i-j} + \epsilon)\sqrt{n}}{S_{ui-uj}}$$

Again, since the system has already computed the indifference confidence for n samples, there is no need to evaluate the second integral from a2 to b2. Thus the operational form of this integral is:

$$\Delta\gamma_{ji} \text{ for } U_{j-i} = \Phi(a1, b1) - previous\ \gamma^{*ji} \quad \text{(marginal confidence indifference equation)}$$

where a1 and b1 are as defined above.

The expected marginal cost of sampling from hypothesis strategy H_i which provides added information on U_i and C_i is simply the sample mean of C_i so far: $\overline{C}_i$. This relates to the marginal cost of determining another point from U_{j-i} as follows. Let Sa_i indicate the number of samples drawn from the strategy H_i so far. When we draw a problem from the distribution, we store it so that if we wish to sample p times from the distribution U_i, and p times from distribution U_j, we have the same p problems from the problem distribution. Furthermore, when we compute differences in utility from the distribution U_{j-i} these are computed by using the competing strategies on the same problem (as in COMPOSER). Thus if we wish to get the pth sample from the distribution U_{j-i}, (assuming p–1 samples have already been computed), Sa_i and Sa_j must each be at least p–1. The cost can be expressed as follows:

If both Sa_i and Sa_j are p or greater: the cost of computing the pth sample is 0.

If $Sa_i = p-1$ and $Sa_j \geq p$ then the expected cost is $\overline{C_i}$ (the sample mean of C_i) .

If $Sa_j = p-1$ and $Sa_i \geq p$ then the expected cost is $\overline{C_j}$.

If $Sa_j = p-1$ and $Sa_i = p-1$ then the expected cost is $\overline{C_i} + \overline{C_j}$.

Note that in general the system will be attempting to show that a specific strategy H_i dominates or is ambivalent with all the others. This means that Sa_i will be consistently ≥ to all other Sa_j. Anytime is Sa_i incremented to find out more information regarding H_i, this immediately reduces the cost of acquiring information for other H_j's, as they no longer need to pay the cost of sampling H_i. This will tend to mitigate the effects of different means and variances for U_{j-i} distributions. However, in cases where the focus strategy H_i changes, other more complex phenomena will occur.

STOP2 ALGORITHM

let $T = \{H_1 \ldots H_k\}$
solve n_0 problems with each strategy in T
let H_{high} be the strategy in T with the highest sample mean for U_{high}
compute utility comparison statistics
loop1

> let H_{high} be the strategy in T with the highest sample mean for U_{high}
> if for every $H_j \in$ T one of the following conditions holds
>
> > U_{high} dominates U_j with confidence γ_{highj}*
> > U_{high} and U_j are ambivalent with confidence γ_{highj}*
> >
> > such that $\sum_{j=1}^{k} \gamma_{highj} \leq \gamma$ where $\gamma_{nn} = 0$
>
> THEN RETURN H_{high}
> ELSE
> for each strategy $H_i \in$ T
>
> > Compute the marginal benefit MB_i and marginal cost MC_i
> > of acquiring another sample from U_{high-i}
>
> for the H_i with the highest MB_i / MC_i

generate data for the distribution U_{high-i}

recompute utility comparison statistics, reselecting H_{high} if necessary

continue with loop1

Again, the algorithm has been simplified to ease understanding. In fact, the marginal cost and utility of acquiring another sample need only be updated when relevant samples are taken. Additionally, acquiring a sample for H_{high} to acquire a sample for U_{high-i} may allow another U_{high-j} to be computed at zero cost (due to changes in H_{high}) and hence should be included in the relevant marginal benefit calculation.

4. The Expected Loss Approach

A commonly used measure in valuing information in game theory applications is the concept of *expected loss*. Put simply, expected loss is the chance that one makes the wrong decision, weighted by how wrong the decision turns out to be. The expected loss measure can be computed for any pair of alternatives. These computed values can then be used to answer both the question of "is the current information enough" and if additional information is needed "which information at which cost should we get". The former question can be answered by putting a bound on the expected loss that one is willing to tolerate, and making a decision when an alternative is found to have an expected loss of less than the bound. In our case of hypothesis evaluation, one can select a hypothesis strategy H_i when:

$$\sum_{j=1}^{k} \textit{Expected Utility Loss of selecting } i \textit{ over } j \ \leq L$$

for all j, the sum for the expected loss of all of the expected losses for e The latter question can be answered by acquiring the information which is expected to reduce the above sum by the greatest amount relative to the cost of acquiring the information.

More rigorously, we define the expected loss of utility from adopting H_i rather than H_j to be the integral of the joint utility of Hi and H_j over the regions where H_i has lower utility weighted by the difference in utility:

$$E(L(H_i, H_j)) = \int\int_{u_i<u_j} P_{U_i,U_j}(u_i, u_j)(u_j - u_i)du_i du_j$$

However, because U_i and U_j are jointly gaussian, and a linear combination of two jointly gaussian random variables is gaussian, we can use the differential distribution U_{i-j} to compute the expected loss directly.

Thus we simply estimate the mean and variance for our best guess at the true mean of the differential distribution U_{i-j}.[5] We compute the integral over the region where $U_{i-j} > 0$ of the term $\mathrm{Prob}(U_{i-j}=u)u$. To do this, we first compute the sample mean and variance for the differential distribution, and then apply a formula almost identical to that used in the dominance confidence interval calculation. The sole difference is that the the integral includes a multiplication times the difference of the values (utilities). This formula is shown below and it's derivation is shown in Appendix C.

5. An alternative approach would be to estimate the parameters for each of the individual utility distributions, then use these parameters to compute the mean and variances for the estimates of the differential distributions. This would result in the same parameters as our approach of computing the parameters of the differential distributions directly from the data.

$$E(L(H_i, H_j)) = \frac{S_{ui-uj} e^{-0.5n\left(\frac{\overline{U}_{i-j}}{S_{ui-uj}}\right)^2}}{\sqrt{2\pi n}} + \frac{\overline{U}_{i-j}}{\sqrt{2\pi}} \int_{-\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2} dz \qquad \text{(expected loss equation)}$$

One can compute the expected_utility_loss of the alternative choice (e.g., of choosing H_j over H_i) by taking the difference of the utilities in the opposite direction and integrating over the complementary region (i.e. where u < 0 of the term –uProb(U_{i-j}=u)).

4.1 The EL1 Algorithm

Given this definition of expected loss, we can define two algorithms EL1 and EL2. As with STOP1 and STOP2, EL1 and EL2 differ in that EL2 accounts for differences in marginal return from an additional sample for different distributions. EL2 also estimates the different costs for sampling from the different distributions. As with STOP2, EL2 samples from distribution of U_{i-j} so as to maximize the marginal benefit divided by the marginal cost.

EL1 ALGORITHM

let T = {H_1...H_k} and L be the expected loss threshold.
let L* be L/k
solve n_0 problems with each strategy in T
let H_{high} be the strategy in T with the highest sample mean for U_{high}
compute the expected_utility_loss statistic of selecting H_{high} over each other strategy H_i
loop1

    select a strategy H_i whose pair-wise such that the expected utility loss from selecting
      H_{high} over H_i is greater than L*
    If there is no such strategy,
    then   return H_{high}
    else   generate sample from H_i and H_{high}
            recompute expected utility losses
    continue loop1

4.2 The EL2 Algorithm

The EL2 algorithm extends the EL1 algorithm to account for differing gains from acquiring another sample for H_{high}, and H_i's based upon varying variances, relative means, and number of examples acquired so far. EL2 also accounts for the varying costs of acquiring these additional examples. Thus, the EL2 algorithm extends EL1 in exactly the same way that STOP2 extends STOP1.

The marginal decrease in expected_utility_loss (MDEUL) is computed by recomputing the integral for expected_loss, assuming that the variances and means will remain the same but incrementing n by 1 and subtracting the current expected_utility_loss. The resulting formula is shown below

$$\Delta E(L(H_i, H_j)) = \frac{S_{ui-uj} e^{-0.5(n+1)\left(\frac{\overline{U}_{i-j}}{S_{ui-uj}}\right)^2}}{\sqrt{2\pi(n+1)}} + \frac{\overline{U}_{i-j}}{\sqrt{2\pi}} \int_{\frac{\overline{U}_{i-j}\sqrt{(n+1)}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2} dz - \text{current } E(L(H_i, H_j))$$

(MDEUL equation)

The expected_marginal_cost_of_sampling is computed by using the mean of the expected cost distribution exactly as in STOP2. The EL2 algorithm is shown below.

EL2 ALGORITHM

let T = $\{H_1 \ldots H_k\}$ and L be the expected loss threshold.
solve n_0 problems with each strategy in T
let H_{high} be the strategy in T with the highest sample mean for U_{high}
compute expected_utility_loss statistic of selecting H_{high} over each other strategy H_i
  and let this be L^*_i (enforce that the expected utility loss of selecting H_i over H_i is 0)
loop1

if $\sum_{i=1}^{k} L^*_i \leq L$

then   return H_{high}
else   compute the marginal decrease in expected loss by sampling from each of the H_i's (including H_{high})
        compute the marginal cost of sampling each strategy using the C distributions
        sample from the distribution with the highest

$$\frac{MDEUL}{expected\_marginal\_cost\_of\_sampling}$$

        recompute L^*_i's as necessary
continue loop1

5. EMPIRICAL PERFORMANCE EVALUATION

We now turn to an empirical evaluation of the hypothesis selection techniques. This evaluation lends support to the techniques by addressing three key issues. First it demonstrates that the techniques perform as predicted. Second, the evaluation demonstrates the benefits of rational example allocation (as performed by STOP2 and EL2). Finally, it illustrates the applicability of the approach to a real-world hypothesis selection problem. Where possible, we contrast performance with that of other relevant approaches in the statistical literature.

5.1 Other Relevant Approaches

There exists a body of standard approaches for the interval-based formulation of the hypothesis evaluation problem. To demonstrate the power of our interval-based approaches we contrast them with two existing approaches. The first is a statistical approach proposed by Turnbull and Weiss [Turnbull87]. The second is a machine learning technique proposed by Gratch and DeJong [Gratch92].

The Turnbull and Weiss approach comes closest among statistical ranking and selection procedures to the generality of the STOP1 and STOP2 approaches. Most standard statistical approaches make strong assumptions about the form of the hypothesis evaluation problem; for instance that the variances associated with hypotheses are known or equal. As in our interval-based approaches, the Turnbull and Weiss treat hypotheses as normal variables with unknown mean, and unknown and unequal variance, however they make the additional assumption that hypotheses are independent. It can still be reasonable to use this approach when the hypotheses are not independent, but this can lead to excessive statistical error or unnecessarily large training set sizes under certain circumstances. The Turnbull technique is described in Appendix E.

The COMPOSER technique was proposed to solve hypothesis evaluation problems as they arise in the context of adaptive problem solving. COMPOSER treats hypotheses as dependent normal variables with unknown mean, and unknown and unequal variance. COMPOSER, however, does not implement the notion of an indifference interval. Rather it is trying to adopt the first hypothesis that can be demonstrated to be significantly better than a default hypothesis. When the best hypotheses are all close to each other in utility, COMPOSER will take an excessive number of training examples. COMPOSER is described in Appendix F.

5.2 Methodology

First we discuss some methodological issues. The interval-based and expected loss approaches embody different criteria for selecting hypotheses in that they use parameters which it is difficult to compare. Thus we first test the interval–based and expected loss approaches separately. Interval–based approaches have been investigated extensively in the statistical ranking and selection literature (see [Haseeb85] for a review of the recent literature). This affords us the opportunity to compare STOP1 and STOP2 against a standard statistical approach.

Techniques are evaluated on synthetic and real-world data sets. Synthetic data allows a systematic test of the formal properties of each technique while real data sets test the appropriateness of statistical assumptions – such as the normal approximation – and assess the practicality of each approaches on real-world problems.

Finally, in a comprehensive real–world test on scheduling data, we compare the interval based and expected loss approaches, using a wide range of parameter settings. This test reports on the bottom–line effectiveness of the competing techniques in a pragmatic problem–solving setting.

5.2.1 Synthetic Data

Synthetic data is used to show that: (1) the techniques perform as expected when the underlying assumptions are valid and (2) the use of rational example allocation exhibits substantial improvement when there is unequal cost or variance among the distributions. For interval-based approaches we show that the technique will choose the best hypotheses, or one ε–close to the best, with the requested probability. When all hypotheses are within ε of each other, the indifference–based technique should quickly terminate, returning any hypotheses. For the expected loss approaches the claim is that the technique will exhibit no more that the requested level of expected loss. One set of evaluations is devised to test this claim.

The second claim is that the techniques that use rational example allocation will exhibit substantial performance improvement when there is unequal cost or variance among the hypotheses. A second set of evaluations is devised to test this claim

For the synthetic data problems, hypotheses are modeled as random variables with parameterized properties. A specific hypothesis evaluation problem is constructed by fixing the values of each of these parameters. In the course of solving a specific problem, values for the utility and cost of each hypothesis on each example are assigned randomly according to the parameterized distribution functions. For a given problem let k define the number of hypotheses. For all synthetic evaluations the hypothesis utilities and costs are treated as independent normal random variables with some parameterized mean and variance. Each hypothesis is described by four parameters – its expected utility, the utility variance, its expected cost, and its cost variance. Thus a hypothesis evaluation problem is specified by $4k$ parameters.

The hypothesis evaluation techniques have additional parameters that govern how they attack the problem. To distinguish these we refer to *problem parameters* and *control parameters.* The interval–based techniques have three control parameters: an initial sample size n_0, a confidence setting γ^* and an indifference setting ε. The expected loss techniques have two control parameters: an initial sample size n_0 and a loss threshold H^*.

An experimental trial consists of solving a hypothesis evaluation problem with a given technique, where all problem and control parameters have been fixed. The performance on any single trial provides little information given the random nature of the task. To assess the average characteristics of the technique a trial is repeated multiple times and the results are averaged across trials. All experimental trials are repeated 5000 times.

An interval-based technique processes examples until it has identified a hypothesis that with probability γ^* is within ε of optimal. STOP1 attempts to ensure this property with the minimum number of training examples possible. STOP2 attempts to ensure this property with the minimum cost possible. To assess the competence of these techniques we track three quantities: the number of examples required to choose a hypothesis, the cost of the examples required to choose a hypothesis, and the observed probability that the expected utility of the chosen hypothesis is in fact within ε of the utility of the optimal hypothesis. For the expected loss techniques we track the analogous three quantities: the number of examples to choose a hypothesis, the cost of the examples, and the average loss (the average loss in utility when the technique chooses the non-optimal hypothesis weighted by the probability of choosing the non-optimal hypothesis).

Unless otherwise stated, each training example on any hypothesis is given equal cost. This means that the overall cost of a technique is directly proportional to the expected number of examples required to select a hypothesis. Thus, when each training example is given equal cost only the number of examples will be reported. One set of synthetic evaluations highlights the benefits of rational example allocation. In these evaluations we create a significant discrepancy in the cost of evaluating alternative hypotheses.

5.2.2 Scheduling Data

The test of real-world applicability is based on data drawn from an actual NASA scheduling application. This data provides a strong test of the applicability of the techniques. All of the statistical techniques make some form of normality assumption. However the data in this application is highly non-normal – in fact most of the distributions are bi-modal. This characteristic provides a rather severe test of the robustness of the approaches.

In this application a heuristic system was developed to schedule communication events between earth-orbiting satellites and ground-based radio antennas. In the course of development, extensive evaluations were performed with various scheduling heuristics. The goal of these evaluations was to choose a heuristic that solved scheduling problems quickly on average. This is easily seen as a hypothesis evaluation problem. Each of the heuristics corresponds to a hypothesis. The cost of evaluating a hypothesis over a training example is the cost of solving the scheduling problem with the given heuristic. The utility of the training example is simply the negation of its cost. In that way, choosing a hypothesis with maximal expected utility corresponds to choosing a scheduling heuristic with minimal average cost.

Using the data from the heuristic evaluations we derived four data sets. Each data set corresponds to a comparison of some set of scheduling heuristics, and contains data on the heuristics' performance over about one thousand scheduling problems. An experimental trial consists of executing a technique over one of these data sets. Each time a training example is to be processed, some problem is drawn randomly from the data set with replacement. The actual utility and cost values associated with this scheduling problem is then used. As in the synthetic data, each experimental trial is repeated 5000 times and all reported results are the average of these trials.

5.3 The Interval-Based Approach

The interval-based approaches, STOP1 and STOP2, are evaluated on both synthetic and scheduling data sets. Synthetic problems were constructed to answer the following three questions: 1) do the techniques select ε–close hypotheses with the specified probability, 2) do the techniques terminate quickly when all hypotheses are ε–close, and 3) does STOP2 outperform STOP1 when there is significant cost or variance differences between hypotheses. We also contrast the performance of our techniques with COMPOSER and the technique of Turnbull and Weiss.

5.3.1 Confidence Test

The statistical ranking and selection literature uses a standard methodology for evaluating the statistical error of hypothesis evaluation techniques. We adopt this methodology here. Robert Bechhofer introduced the concept of the *least favorable configuration of the population means* [Bechhofer54]. This is a parameter configuration that is most likely to cause a technique to choose a wrong hypothesis (one that is not ε–close) and thus provides the most severe test of the technique's abilities. Under this configuration, $k-1$ of the hypotheses have identical expected utilities, μ, and the remaining hypothesis has expected utility μ+ε. The last hypothesis has the highest expected utility and should be chosen by the technique. The costs and variances of all hypotheses are equal.[6]

Parameters			STOP1	STOP2	TURNBULL	COMPOSER
k	γ^*	σ/ε				
3	0.75	2	38 (0.85)	34 (0.83)	27 (0.75)	61 (0.96)
3	0.75	3	58 (0.08)	52 (0.78)	50 (0.72)	103 (0.90)
3	0.90	2	64 (0.92)	65 (0.92)	54 (0.86)	91 (0.98)
3	0.90	3	121 (0.91)	123 (0.91)	127 (0.87)	170 (0.95)
3	0.95	2	93 (0.95)	96 (0.97)	81 (0.92)	115 (0.99)
3	0.95	3	183 (0.94)	193 (0.95)	192 (0.93)	238 (0.97)
5	0.75	2	98 (0.86)	94 (0.86)	63 (0.71)	139 (0.96)
5	0.75	3	177 (0.83)	179 (0.81)	141 (0.71)	250 (0.89)
5	0.90	2	159 (0.93)	170 (0.94)	123 (0.84)	195 (0.97)
5	0.90	3	310 (0.92)	349 (0.93)	294 (0.88)	389 (0.94)
5	0.95	2	212 (0.96)	234 (0.97)	175 (0.91)	237 (0.98)
5	0.95	3	427 (0.95)	483 (0.96)	411 (0.94)	501 (0.97)
10	0.75	2	298 (0.89)	330 (0.90)	185 (0.66)	353 (0.95)
10	0.75	3	584 (0.87)	688 (0.87)	438 (0.70)	677 (0.89)
10	0.90	2	430 (0.95)	508 (0.95)	331 (0.83)	469 (0.97)
10	0.90	3	892 (0.93)	1,066 (0.95)	783 (0.89)	958 (0.93)
10	0.95	2	545 (0.97)	661 (0.97)	443 (0.91)	574 (0.98)
10	0.95	3	1,136 (0.95)	1,435 (0.97)	1,037 (0.94)	1,175 (0.95)

Table 1
Estimated expected total number of observations in the least favorable configuration. Achieved probability of a correct selection is shown in parenthesis.

We test each technique on the least favorable configuration under a variety of control parameter settings. The least favorable configuration becomes more difficult (requires more examples) as the confidence γ^*, the number of hypotheses k, or common utility variance σ^2, increases. It becomes easier as the indifference interval ε, increases. In the standard methodology a technique is evaluated varying values for k, γ^*, and

6. Note that in this evaluation ε acts as a problem parameter in addition to its role as a control parameter.

σ/ε. The last term combines the variance and indifference interval size into a single quantity which as it increases makes the problem more difficult. For our experiments, n_0 is set to seven, μ is fifty, σ^2 is sixty-four, and all other parameters are varied as indicated in the results. The sample size results and observed confidence levels are summarized in Table 1.

The results indicate that all systems are roughly comparable in the number of examples required to choose a hypotheses. As expected, the number of examples increases with k, γ^*, and σ/ε. The technique of Turnbull and Weiss tended to be the most efficient, COMPOSER the least. In terms of statistical error, all of the algorithms except Turnbull and Weiss' were correct at least as often as requested. The technique of Turnbull and Weiss often provided less than the requested confidence. However, since their technique only guarantees that the confidence will approach γ^* as ε/σ tends to zero, these results are consistent with their claim.

5.3.2 Indifference Test

The indifference interval approaches should terminate quickly when all hypotheses are indifferent to each other. To test this claim we repeated the least favorable configuration evaluations except that all hypotheses were assigned the same expected utility μ. Results are summarized in Table 2. Error rate results are not shown since any hypothesis is a correct selection in this configuration.

The key result to notice is that COMPOSER failed to terminate on any of the trials. This highlights the potential difficulties with COMPOSER that STOP1 and STOP2 were designed to correct. Again, the technique of Turnbull and Weiss slightly outperforms the other approaches.

5.3.3 Rational Allocation Test

STOP2 is designed to perform well when the cost of processing examples or the utility variance differs widely across hypotheses. The preceding evaluations did not contrast the two approaches under these conditions as both the cost and variances were equal. Consequently STOP1 and STOP2 was also approximately equal efficient in these tests. This evaluation contrasts the approaches by providing problem configurations with highly unequal costs.

Problem configurations are defined as follows. One hypothesis (the correct selection) is assigned a high mean μ_{best}. A second hypothesis is assigned a mean slightly below ε of the best, $\mu_{\text{best-1}}$. All remaining hypotheses are assigned a low mean, μ_{worst}. The second hypothesis is given a high cost c_{high} and all other hypotheses are given low cost c_{low}. All hypotheses are assigned a common variance of fifty, μ_{best} is seventy–four, $\mu_{\text{best-1}}$ is seventy–two, μ_{worst} is five, ε is one, and n_0 is seven. The rationale behind this configuration is given in Appendix D. Various confidence settings were evaluated. The results are summarized in Table 3.

The results illustrate the clear dominance of STOP2 under this configuration – up to seven times more efficient on one of the trials. An interesting question is whether there is a limit to how much better STOP2

Parameters			STOP1	STOP2	TURNBULL	COMPOSER
k	γ^*	σ/ε				
3	0.75	2	48	44	27	***
3	0.75	3	75	68	50	***
3	0.90	2	96	100	54	***
3	0.90	3	181	194	127	***
3	0.95	2	142	151	81	***
3	0.95	3	291	312	192	***
5	0.75	2	134	143	63	***
5	0.75	3	249	276	141	***
5	0.90	2	235	267	123	***
5	0.90	3	474	568	294	***
5	0.95	2	325	360	174	***
5	0.95	3	672	768	411	***
10	0.75	2	421	525	185	***
10	0.75	3	833	1104	438	***
10	0.90	2	649	772	331	***
10	0.90	3	1348	1667	782	***
10	0.95	2	835	975	444	***
10	0.95	3	1776	2100	1037	***

Table 2

Estimated expected total number of observations in the indifference configuration. Note that COMPOSER failed to terminate on any of the trials.

can be. In fact there is an upper bound on this difference that is proven in Appendix Z. This upper bound increases as the number of hypotheses increases or as the confidence level decreases.

5.3.4 Scheduling Test

We ran all four algorithms over the four scheduling data sets. In each case the confidence level was set at 95%, n_0 set to fifteen, and ε set to 4.0. Table 4 summarizes the results along with the number of hypotheses and the relative difficulty (σ/ε) of each data set.

The principle result is that STOP1 and STOP2 substantially exceeded the performance of the other algorithms except on one case. The one exception is an artifact of COMPOSER solving a slightly different task. Rather than choosing the hypothesis that is ε–close to optimal, COMPOSER chooses the first hypothesis to dominate a default hypothesis (the first hypothesis was arbitrarily defined to be the default in

Parameters		STOP1	STOP2	$\frac{\text{STOP1}}{\text{STOP2}}$
k	γ^*			
3	0.75	12,034	5,241	2.3
3	0.80	14,890	6,790	2.2
3	0.85	20,119	10,030	2.0
3	0.90	26,340	15,040	1.8
5	0.75	22,081	5,216	4.2
5	0.80	27,375	6,947	3.9
5	0.85	31,203	9,817	3.2
5	0.90	39,305	14,859	2.7
10	0.75	36,768	5,154	7.1
10	0.80	42,202	6,753	6.3
10	0.85	47,167	10,086	4.7
10	0.90	56,183	15,004	3.8

Table 3
Estimated expected total cost for rational allocation configuration.

	Parameters			STOP1	STOP2	TURNBULL	COMPOSER
	k	γ^*	σ/ε				
D1	3	0.95	34	908 (1.00)	648 (1.00)	26,691 (1.00)	78 (1.00)
D2	2	0.95	34	74 (1.00)	76 (1.00)	13,066 (1.00)	346 (1.00)
D3	7	0.95	14	2,371 (0.94)	2,153 (0.93)	94,308 (1.00)	2,456 (0.97)
D4	7	0.95	11	7,972 (0.96)	7,621 (0.94)	87,357 (1.00)	21,312 (0.89)

Table 4
Estimated expected total number of observations for scheduling data.
Achieved probability of a correct selection is shown in parenthesis.

these trials). In data set D1 the default is significantly worse than the other two hypotheses, which in turn are indifferent to each other. STOP1 and STOP2 take longer because they must verify this indifference.

Note that unlike the synthetic data where STOP1 was slightly more efficient than STOP2, in the scheduling data STOP2 was slightly more efficient. In fact, in the scheduling data there is some disparity between hypotheses in their utility variance. STOP2 is able to account for these factors when allocating examples, and thus exhibits greater efficiency.

5.4 Discussion of Interval-Based Evaluation

Taken together, the evaluation provides clear evidence for the effectiveness of STOP1 and STOP2 and demonstrates their superiority to alternative techniques. The techniques performed as predicted, guaranteeing the requested confidence level under a variety of configurations. In comparison to other approaches, they did perform the best on every configuration, however when they were out performed it was not by much and they often substantially outperformed the alternative techniques. For example, COMPOSER fails to terminate when multiple hypotheses are close to optimal. The technique of Turnbull and Weiss performed poorly on the real-world data sets. The scheduling evaluation demonstrates that STOP1 and STOP2's normal approximation allows effective performance on real-world hypotheses selection problems, even when the underlying distributions are not normal.

The rational allocation test illustrates that STOP2 can substantially outperform STOP1 when there are marked differences across heuristics in the cost of processing examples or in the variance of expected utility values. STOP2 should be used if the hypothesis evaluation problem has this characteristic. It appears that STOP1 is slightly more efficient when the cost and utilities are close to equal. Under these circumstances we recommend the use of STOP1.

5.5 The Expected Loss Approach

The expected loss approaches, EL1 and EL2, are evaluated on both synthetic and scheduling data sets. Synthetic problems are constructed to answer the following two questions: 1) do the techniques properly bound the expected loss, and 2) does EL2 outperform EL1 when there is significant cost or variance differences between hypotheses.

5.5.1 Expected Loss Test

The techniques are tested on a least favorable configuration with some number of hypotheses, k. The means of $k-1$ hypotheses are assigned the value μ and the remaining hypothesis is assigned mean $\mu+\varepsilon$. Each technique is then tested on various loss thresholds, H^*, over this problem. For this evaluation, μ is fifty, all hypotheses share a common utility variance of sixty–four, and ε is two. All other parameters are varied as indicated in the results. The sample size results and observed loss values are summarized in Table 5.

The results illustrate that the techniques perform as predicted. As the loss threshold is lowered the techniques take more training examples to ensure the expected loss remains below the threshold.

5.5.2 Rational Allocation Test

EL1 and EL2 greatly exceed the performance of Turnbull and Weiss' technique on all data sets. The poor performance of the latter algorithm is due to two factors. First, the technique is unable to quickly discard hypotheses which are clearly dominated by other hypotheses. Second, the technique's independence assumption appears to have degraded performance on these data sets where hypotheses are positively correlated. A technique that views two hypotheses as independent will tend to overestimate (underestimate)

Parameters			EL1		EL2	
k	ε	H*	Observations	Loss	Observations	Loss
3	2	1.0	33	0.5	26	0.8
3	2	0.75	38	0.4	29	0.7
3	2	0.5	46	0.2	35	0.5
3	2	0.25	58	0.1	48	0.3
5	2	1.0	73	0.4	54	0.9
5	2	0.75	83	0.3	62	0.7
5	2	0.5	98	0.2	78	0.5
5	2	0.25	127	0.1	114	0.2
10	2	1.0	201	0.2	157	0.8
10	2	0.75	221	0.2	182	0.6
10	2	0.5	255	0.1	220	0.4
10	2	0.25	312	0.0	269	0.2

Table 6
Estimated expected total number of observations and expected loss of an incorrect selection for the least favorable configuration.

their joint variance when the hypotheses are positively (negatively) correlated. Overestimating the variance, in turn, leads to higher sample sizes.

EL2 is designed to perform well when the cost of processing examples or the utility variance differs widely across hypotheses. The preceding evaluations did not contrast the two techniques as the cost and variances were equal across hypotheses. This evaluation contrasts the approaches using unequal costs across the hypotheses. The configuration used is identical to the one described in Section 5.3.3. The difference in expected costs between solving problems with EL1 and EL2 is summarized in Table 7.

The results indicate that EL2 substantially outperformed EL1 – in one trial solving the configuration four times more efficiently. EL2 achieves greater efficiency as the number of hypotheses increases. As with STOP2 we suspect that the potential for greater efficiency is not unbounded, but we have not as yet obtained an upper bound on the relative efficiency of EL2.

5.5.3 Scheduling Test

We ran the two techniques over the four scheduling data sets. In each case the loss threshold was set at three and n_0 was fifteen. Table 8 summarizes the results.

The main result is that the algorithms correctly bounded the expected loss with one exception – EL2 gave greater than expected loss on data set D3. It appears that this exception arose from a significant departure from normality in the distributions comprising the data set. Additional trials demonstrated this discrepancy goes away if the initial sample size is increased, thereby improving the normal approximation.

Parameters		STOP1	STOP2	STOP1 / STOP2
k	H^*			
3	1.00	5,757	3,733	1.5
3	0.75	6,980	3,992	1.8
3	0.50	8,899	4,636	1.9
3	0.25	14,102	6,847	2.1
5	1.00	8,070	3,737	2.2
5	0.75	9,688	3,985	2.5
5	0.50	12,807	4,664	2.8
5	0.25	19,525	6,873	2.9
10	1.00	12,745	3,740	3.2
10	0.75	15,035	4,037	3.7
10	0.50	19,144	4,718	4.1
10	0.25	26,901	6,861	3.9

Table 7
Estimated expected total cost for rational allocation configuration.

	Parameters		EL1		EL2	
	k	H^*	Observations	Loss	Observations	Loss
D1	3	3.0	78	0.1	49	1.0
D2	2	3.0	30	1.8	30	1.8
D3	7	3.0	335	3.0	177	3.9
D4	7	3.0	735	1.7	283	2.2

Table 8
Estimated expected total number of observations and expected loss of an incorrect selection for the scheduling data.

5.6 Discussion of Expected Loss Evaluation

The three evaluations of EL1 and EL2 give clear support for the effectiveness of these algorithms. The techniques performed as predicted, properly bounding the expected loss under a variety of parameter configurations. We did observe that under some of the configurations EL2 gave slightly larger than requested loss. More generally, it appears that the expected loss approach will be more susceptible to departures from normality in the utility distributions, when compared with interval-based approach. Both approaches use a normal distribution to approximate the distribution of a sample mean. However the interval-based approach is only sensitive to the area under parts of the normal curve. The expected loss computation makes use of both the area and the shape of certain parts of the normal curve. Thus the expected loss approach demands more fidelity from its approximation, and this fidelity is degraded when the underlying distribution is not normal. This effect can be compensated by using greater initial sizes for the expected loss technique.

5.7 Comparing Interval-based to Expected Loss

One cannot state that interval-base techniques are better or worse than expected loss approaches – each is solving a slightly different problem. Interval-based approaches are attempting to identify a nearly optimal hypothesis with high confidence while expected loss approaches are attempting to minimize the cost of a mistaken selection. If the goal of the task is to identify the best hypothesis then clearly an interval-based approach should be use. If the goal is to simply improve expected utility as much as possible, either could be used and it is unclear which is to be preferred.

Our original motivation in developing these approaches was to develop effective techniques for adaptive problem solving. In this section we attempts to assess how the various approaches perform on this task. In particular we consider how the approaches perform in the problem of learning a *set* of problem solving heuristics for the NASA scheduling domain. In this test the algorithms were given the task of optimizing four control parameters of the adaptive scheduler, with the goal of speeding up the schedule generation process. The solution to this consists of identifying a good heuristic for each of the four control parameters, where the best choice for a particular parameter depends on the heuristics chosen for the other control parameters. We implement a hill-climbing strategy for finding a good combination of heuristics. For more details on this application domain see [Gratch et al. 93].

We run each algorithm under a variety of parameter settings and compare the best performance of each algorithm (i.e., the lowest cost setting that resulted in a high expected utility on average). In this test the interval-based algorithms are run with confidence levels γ^*=0.75,0.90,0.95 and indifference levels ε=1.0, 4.0, 7.0. The expected loss algorithms are run with loss bound L=5, 1, 0.5. For each setting 1000 runs are conducted, we then determined the best settings as the lowest cost solution within 1.0 utility of the average best solution found per algorithm (effectively enforcing a minimum utility of 16.5). These best settings found and their averaged results (from 1000 runs each) are shown below in Table 4.

	Cost 100s of CPU seconds	Examples	Utility
COMPOSER (0.90)	6128	4075	17.3
STOP1 (0.75,1.0)	4199	2785	17.1
STOP2 (0.75,1.0)	3140	1924	16.6
EL1(1.0)	2347	1557	16.8
EL2(0.5)	2211	1454	16.4

Table 4: Direct Comparison of all four algorithms

These results show that the algorithms produce roughly comparable utilities, the difference in utilities is smaller than the smallest indifference interval specified to the interval-based algorithms.

From this comparison we must conclude that, at least in the case of this NASA scheduling application, there is little difference between the interval-based and expected loss approaches, both in terms of expected improvement and in terms of sample complexity. As expected, the unequal allocation approaches performed better in terms of learning cost. Finally, all of the improved algorithms outperformed the benchmark COMPOSER algorithm in terms of learning cost.

6. Discussion and Conclusions

There are many relevant issues pertaining to the topic of hypothesis evaluation which have not been covered in this technical report. This section briefly discusses a number of these issues.

One issue is modelling the computational cost of inferring (parameter estimation) and applying the statistical models. In some applications, one might imagine that these costs would play a significant role in determining the usefulness of our hypothesis evaluation mode. However, in our target application of learning for scheduling, the cost of gathering further information heavily outweighs the cost of inferring and applying the statistical models. However, for other domains we concede that this may not be the case. A second related issue is to estimate and tradeoff this cost of applying the statistics and decision theory relative to the cost of additional examples.

Another issue is to better understand the qualitative conditions under which the cost sensitive measures (STOP2 and EL2) will outperform the equal error distribution models (STOP1 and EL1). Generally speaking, if the means and variances vary significantly, the cost sensitive measures should perform better. Additionally, if the marginal computations are reasonable projections, the cost sensitive measures should also outperform the other measures.

An important issue is the use of the O(k) error function. Further empirical evaluation needs to be performed to better understand the relationship between n_0 and the number of S_{high} switches during hypothesis evaluation, and exactly how this relates to the error models (e.g., vs. the $O(k^2)$ error model and to the required confidence parameter γ. As a further subtlety, one might consider removing strategies which become dominated at any point in the evaluation (in contrast with the current approach which requires all strategies to be compared against the final S_{high}).

A better understanding of how the estimates of the means and variances vary with more examples would also be useful. Confidence estimates on next hypotheses to evaluate might be useful. Analytical estimates of how many examples are required as a function of the size of the hypothesis set would clarify this matter. Additionally, determining the exact impact of the dual example phenomenon (where two examples are needed to compute each data point for the differential distribution) would be desirable.

Finally, how will the cost of sampling vary, and how will the incremental increase in confidence (or decrease in expected loss) vary? Preliminary evidence in the scheduling CPU case indicates that the cost varies considerably, and that quickly pruning bad hypotheses is of significant importance. Additionally, if we had a method of estimating a utility difference with unequal numbers of examples that would be very helpful, but since the utilities are covarying it seems unlikely that such a technique will be found.

Another interesting problem is that of hybrid utility functions where the value of a solution is inversely related to the amount of time needed to create it. In the speed-up case the exact relationship between the cost of sampling and utility is known, however, one might imagine time–sensitive applications where the time to a solution drastically impacts its utility. In general, there is a continuum of utility functions, involving time and final schedule quality. This strongly relates to the concept of "anytime algorithms" which can be interrupted at anytime, returning increasingly good solutions.

This report has described techniques for choosing among a set of alternatives in the presence of incomplete information and varying costs of acquiring information. In our approach, the cost and utility of various alternatives are represented using parameterized statistical models. Using techniques from an area of statistics called parameter estimation, models can be inferred from performance on sample problems. These statistical models can then be used to estimate the utility and cost of acquiring additional information and the utility of selecting specific alternatives from the possible choices at hand. These techniques have been applied to adaptive problem–solving, a technique in which a system automatically tunes various control parameters on a performance element to improve performance in a given domain. Empirical results were presented comparing the effectiveness of these techniques on artificially generated data and speed-up learning from a real–world NASA scheduling domain.

Acknowledgements

Portions of this work were performed by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration and portions at the Beckman Institute, University of Illinois under National Science Foundation Grant NSF–IRI–92–09394.

References

[Bechhoffer54] R.E. Bechhofer, "A Single–sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances," **Annals of Mathematical Statistics (25) 1**,1954 pp. 16–39.

[Bell93] C. E. Bell, "Scheduling Deep Space Network Data Transmissions: A Lagrangian Relaxation Approach," **Proceedings of the SPIE Conference on Applications of Artificial Intelligence 1993: Knowledge–based Systems in Aerospace and Industry**, Orlando, FL, April 1993, pp. 330–340. (An extended version of this paper appears as **Unpublished JPL Manuscript**, July 1992.)

[Buringer80] H. Buringer, H. Martin, K. Schriever, Nonparametric Sequential Selection Procedures, Birkhauser,Boston, MA, 1980.

[Fisher81] M. Fisher, "The Lagrangian Relaxation Method for Solving Integer Programming Problems," ***Management Science 27***, 1 (1981), pp. 1–18.

[Gratch92] J. Gratch and G. DeJong, "COMPOSER: A Probabilistic Solution to the Utility Problem in Speed-up Learning," **Proceedings of the National Conference on Artificial Intelligence,** San Jose, CA, July 1992, pp. 235–240.

[Gratch et al. 93a] J. Gratch, S. Chien, and G. DeJong, "Learning Search Control Knowledge for Deep Space Network Scheduling," **Proceedings of the Tenth International Conference on Machine Learning**, Amherst, MA, June 1993, pp. 135–142.

[Gratch et al. 93b] J. Gratch, S. Chien, and G. DeJong, "Learning Search Control Knowledge to Improve Schedule Quality," **Proceedings of the of the IJCAI93 Workshop on Production Planning, Scheduling, and Control**, Chamberry, France, August 1993.

[Gratch & Chien 93] J. Gratch and S. Chien, "Learning Search Control Knowledge for the Deep Space Network Scheduling Problem: Extended Report and Guide to Software," Technical Report UIUCDCS–R–93–1801, Department of Computer Science, University of Illinois, Urbana, IL, January 1993.

[Gratch & DeJong 93] J. Gratch and G. DeJong, "Rational Learning: A Principled Approach to Balancing Learning and Action," Technical Report UIUCDCS–R–93–1801, Department of Computer Science, University of Illinois, Urbana, IL, April 1993.

[Haseeb85] R. M. Haseeb, "Modern Statistical Selection," American Sciences Press, Columbus, OH, 1985.

[Hunter86]D. Hunter, "Uncertain reasoning using maximum entropy inference," *Uncertainty in Artificial Intelligence*, 1986, pp. 203–209.

[Kreysig70] E. Kreysig, Introductory Mathematical Statistics: Principles and Methods, John Wiley & Sons, New York, USA.

[Russell & Wefald 89] S. Russell and E. Wefald, "On Optimal Game Tree Search using Rational Meta-Reasoning," **Proceedings of the Eleventh International Joint Conference on Artificial Intelligence,** Detroit, MI, August 1989, pp. 334–340.

[Russell & Wefald 91] S. Russell and E. Wefald, "Do the Right Thing: Studies in Limited Rationality," MIT Press, Cambridge, MA 1991.

[Turnbull & Weiss 87] A class of sequential procedures for k–sample problems concerning normal means with unknown unequal variances, in **Design of Experiments: ranking and selection,** T. J. Santner & A. C. Tamhane (eds.), Marcel Dekker, 1984.

[Valiant84] L. G. Valiant, "A Theory of the Learnable," *Communications of the Association for Computing Machinery 27,* (1984), pp. 1134–1142.

Appendix A: Confidence Calculation

To compute the confidence that $U_{i-j} > 0$ we adapt a method for computing confidence intervals for the mean of a normal distribution with unknown variance from [Kreysig70]. Computing the confidence interval for a specific confidence γ^* can be done as follows:

Assume that we have n samples.

Find c such that F(c)=0.5 * (1 + γ^*) where F(c) is the t–distribution with n–1 degrees of freedom.

Compute the sample mean $\overline{U}_{i-j}$ and variance S^2_{iui-uj} for the sample.

Compute $k = S_{ui-uj}\frac{c}{\sqrt{n}}$

The confidence interval is: $\overline{U}_{i-j}-k \le \mu_{i-j} <= \overline{U}_{i-j}+k$ with confidence γ^*.

We now adapt this to work from $\overline{U}_{i-j}$, n, and S_{ui-uj}, using the normal distribution rather than the t distribution.. We also modify the integral account for the fact that we are interested in the confidence that $0 \le \mu_{i-j} \le \infty$, rather than $\overline{U}_{i-j}-k \le \mu_{i-j} \le \overline{U}_{i-j}+k$.

Substituting $\overline{U}_{i-j}$ in for k, $\overline{U}_{i-j} = S_{ui-uj}\frac{c}{\sqrt{n}}$, thus $c = \overline{U}_{ui-uj}\frac{\sqrt{n}}{S_{ui-uj}}$.

As with the default settings for the COMPOSER approach, we use the standard normal distribution model rather than the student t distribution. This results in the following:

$$\gamma^* = \frac{1}{\sqrt{2\pi}}\int_{-c}^{\infty} e^{-0.5y^2}dy$$

where again $c = \overline{U}_{i-j}\frac{\sqrt{n}}{S_{ui-uj}}$.

To handle the case of indifference pruning, the confidence that the true mean is in the interval $\{-\varepsilon, \varepsilon\}$ can be computed similarly to the method described above. We want to compute the confidence that μ_{i-j} (the true mean of U_{i-j}) is in the interval $\{-\varepsilon, \varepsilon\}$. This results in the following derivation.

$$\Pr(-\epsilon \le \mu_{i-j} \le \epsilon)$$

$$= \Pr(\epsilon \ge -\mu_{i-j} \ge -\epsilon)$$

$$= \Pr(\overline{U}_{i-j}+\epsilon \ge \overline{U}_{i-j}+\mu_{i-j} \ge \overline{U}_{i-j}-\epsilon)$$

$$= \Pr\left(\frac{(\overline{U}_{i-j}+\epsilon)\sqrt{n}}{S_{ui-uj}} \ge \frac{(\overline{U}_{i-j}+\mu_{i-j})\sqrt{n}}{S_{ui-uj}} \ge \frac{(\overline{U}_{i-j}-\epsilon)\sqrt{n}}{S_{ui-uj}}\right)$$

Therefore we can compute the confidence that the true mean μ_{i-j} is in the interval $\{-\varepsilon, \varepsilon\}$ as:

$$\gamma^* = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-0.5y^2} dy$$

where $a = \dfrac{(\overline{U}_{i-j} + \varepsilon)\sqrt{n}}{S_{ui-uj}}$ and $b = \dfrac{(\overline{U}_{i-j} - \varepsilon)\sqrt{n}}{S_{ui-uj}}$

Appendix B: The Applicability of Dominance

One interesting question is "in what cases will it be easier to prove dominance than to compare against a default strategy?". In this appendix, we examine precisely this question, and derive interesting qualitative information as to when these cases occur. Consider the case where S_i looks like a good strategy, better than S_0 with medium confidence, and S_j is worse than S_0 with medium confidence. This results in the following confidences:

$$\gamma^* (\mu_{ui-u0} < 0) = \frac{1}{\sqrt{2\pi}} \int_{c_{i0}}^{\infty} e^{-0.5y^2} dy \text{ where } c_{i0} = \overline{U}_{i0} \frac{\sqrt{n}}{S_{i0}}.$$

$$\gamma^* (\mu_{uj-u0} < 0) = \frac{1}{\sqrt{2\pi}} \int_{-c_{j0}}^{\infty} e^{-0.5y^2} dy \text{ where } c_{j0} = \overline{U}_{j0} \frac{\sqrt{n}}{S_{j0}}.$$

$$\gamma^* (\mu_{uj-ui} < 0) = \frac{1}{\sqrt{2\pi}} \int_{-c_{ji}}^{\infty} e^{-0.5y^2} dy \text{ where } c_{ji} = \overline{U}_{ji} \frac{\sqrt{n}}{S_{ji}}.$$

Given this context, we ask the question: under what conditions are both $\gamma^* (\mu_{ui-u0} < 0)$ and $\gamma^* (\mu_{uj-u0} < 0)$ less than $\gamma^* (\mu_{uj-ui} < 0)$. This can be reduced to the inequalities: $-c_{ji} < -c_{j0}$ and $-c_{ji} < c_{io}$. These inequalities can be reduced further as follows.

$$-\overline{U}_{ji} \frac{\sqrt{n}}{S_{ji}} < -\overline{U}_{j0} \frac{\sqrt{n}}{S_{j0}} \qquad\qquad -\overline{U}_{ji} \frac{\sqrt{n}}{S_{ji}} < -\overline{U}_{i0} \frac{\sqrt{n}}{S_{i0}} \quad \text{(since } \sqrt{n} \text{ must be positive)}$$

$$\frac{\overline{U}_{ji}}{S_{ji}} > \frac{\overline{U}_{j0}}{S_{j0}} \qquad\qquad \frac{\overline{U}_{ji}}{S_{ji}} > -\frac{\overline{U}_{i0}}{S_{i0}} \quad \text{(multiply both sides by } -1\text{, flip inequality)}$$

and because for covarying normal distributions $S_{x-y} = \sqrt{S_x^2 + S_y^2 - 2S_{x-y}}$ this can be reduced to:

$$\frac{\overline{U}_j - \overline{U}_i}{\sqrt{S_j^2 + S_i^2 - 2S_{ij}}} > \frac{\overline{U}_j - \overline{U}_0}{\sqrt{S_j^2 + S_0^2 - 2S_{j0}}} \qquad\qquad \frac{\overline{U}_j - \overline{U}_i}{\sqrt{S_j^2 + S_i^2 - 2S_{ij}}} > \frac{\overline{U}_0 - \overline{U}_i}{\sqrt{S_i^2 + S_0^2 - 2S_{i0}}} \qquad (1)$$

Note that in the case that where i, j, and 0 are independent, the covariances are zero, resulting in: $s_{x-y} = \sqrt{s_x^2 + s_y^2}$

Looking at the equations marked (1), we can make the following observations. First, since we chose an example where $\mu_j < \mu_0 < \mu_i$, clearly the numerators are satisfied in the inequalities in (1) (i.e., necessarily, is is always the case that $\mu_j - \mu_i > \mu_j - \mu_0$ and $\mu_j - \mu_i > \mu_0 - \mu_i$. Thus as the sample means converge on the true means and the differences between the means of U_i and U_j relative to U_0 grow we are more likely to reach confidence quicker. What remains to be shown is that the denominators have the right relative magnitudes. Consider the cases that:

$$\sqrt{S_j^2 + S_i^2 - 2S_{ij}} \le \sqrt{S_j^2 + S_0^2 - 2S_{j0}} \qquad\qquad \sqrt{S_j^2 + S_i^2 - 2S_{ij}} \le \sqrt{S_i^2 + S_0^2 - 2S_{i0}}$$

This tells us that when the variances of U_i and U_j: S_i^2 and S_j^2 are less than the variance of U_0: S_0^2 the dominance will be easier to show. This also tells us that covariance helps, when i and j covary more than j and 0 and i and 0, we are more likely to gain an advantage.

As a concrete example, consider $\overline{U}_i = 5, \overline{U}_j = -5, \overline{U}_0 = 0, S_i^2 = S_j^2 = S_0^2 = 10.$ This results in the following.

$\frac{10}{\sqrt{10^2 + 10^2}} > \frac{5}{\sqrt{10^2 + 10^2}}$ $\frac{10}{\sqrt{10^2 + 10^2}} > \frac{5}{\sqrt{10^2 + 10^2}}$ Which can be easily shown to be true.

Appendix C: The Expected Loss Calculation

We begin by noting that we want to integrate over the difference between the two utilities, over the region in which the unselected hypothesis strategy has a higher utility. Consider the expected loss for the selection of hypothesis strategy H_j over H_i. In order to compute this, we need to examine the differential distribution U_{i-j}, and integrate from zero to infinity.

$$E(L(H_i, H_j)) = \frac{1}{S_{ui-uj}\sqrt{2\pi}} \int_0^{\infty} e^{-0.5\left(\frac{(\overline{U}_{ui-uj}-l)\sqrt{n}}{S_{ui-uj}}\right)^2} l\,dl$$

we then make the substitution of $z = \frac{(\overline{U}_{i-j}-l)\sqrt{n}}{S_{ui-uj}}$ which results in the following implied substitutions

$l = \frac{S_{ui-uj}z}{\sqrt{n}} + \overline{U}_{i-j}$, $dz = \frac{\sqrt{n}}{S_{ui-uj}}dl$ and $dl = \frac{S_{ui-uj}}{\sqrt{n}}dz$ furthermore, to compute the limits of integration, we note that

when $l = 0$ then $z = -\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}$ and when $l = \infty$ then $z = \frac{(\infty - \overline{U}_{i-j})\sqrt{n}}{S_{ui-uj}} = \infty$ this resulting in the formula

$$\text{expected loss} = \frac{1}{S_{ui-uj}\sqrt{2\pi}} \int_{-\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2}\left(\frac{zS_{ui-uj}}{\sqrt{n}} + \overline{U}_{i-j}\right)S_{ui-uj}dz$$

$$\text{expected loss} = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2}\left(\frac{zS_{ui-uj}}{\sqrt{n}} + \overline{U}_{i-j}\right)dz$$

$$\text{expected loss} = \frac{1}{\sqrt{2\pi}}\left[\int_{-\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2}\left(\frac{zS_{ui-uj}}{\sqrt{n}}\right)dz + \int_{-\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2}(\overline{U}_{i-j})dz\right]$$

$$\text{expected loss} = \frac{S_{ui-uj}}{\sqrt{2\pi n}}\left[\int_{-\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2}z\,dz\right] + \frac{\overline{U}_{i-j}}{\sqrt{2\pi}}\left[\int_{-\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2}dz\right]$$

we now note that the first integral has an analytic solution, namely that

$\int e^{-0.5x^2}x\,dx = e^{-0.5x^2}$, solution of this integral leaves us with the following:

$$\text{expected loss} = \frac{S_{ui-uj}e^{-0.5n\left(\frac{\overline{U}_{i-j}}{S_{ui-uj}}\right)^2}}{\sqrt{2\pi n}} + \frac{\overline{U}_{i-j}}{\sqrt{2\pi}} \int_{-\frac{\overline{U}_{i-j}\sqrt{n}}{S_{ui-uj}}}^{\infty} e^{-0.5z^2}dz \qquad \text{(expected loss formula 1)}$$

Appendix D: An Analysis of Possible Performance Differences between Cost–sensitive and non Cost–sensitive Approaches

The PAC requirement constrains but does not completely determine the behavior of a hypothesis selection algorithm. We would like an algorithm to satisfy the requirement with the minimum cost possible. Several of the factors that contribute to the cost are unknown before learning begins. For this reason standard (non-rational) hypothesis selection algorithms ignore these factors when making their selection. This section discusses the relevant factors and shows that they can be folded into a single value, the *disparity index*. We show that in theory an algorithm can achieve large performance improvements by exploiting this information, if only it were available. In fact, comparable performance improvements can be achieved in practice using sequential techniques, as we show in the next section.

First, in our analysis we use a simplified dominance indifference criterion. In this criterion, the PAC requirement is that H_{sel} must be within some constant ε of the best hypothesis with probability $1-\delta$. It suffices to show:[7]

$$\left(\sum_{i=1}^{k-1} \Pr\left[H_i > H_{sel} + \epsilon\right]\right) \leq \delta \qquad (1.)$$

That is, there is some probability that one or more hypotheses are ε-greater than the selected hypothesis. The hypothesis selection algorithm must insure that the probability that one or more of these events occurs is less than δ. With this equation we see that the problem of bounding the probability of error reduces to bounding the probability of error of each of the $k-1$ comparisons of H_{sel} to H_i.

With the normality assumption the probability that H_i is greater than H_{sel} is a function of the estimates, the number of examples, n, used for each estimate, the closeness parameter ε, and an unknown variance term, σ^2. Variance measures how much each observation can differ from its expected value, which can be estimated from the data.[8] To simplify the presentation we ignore the ε parameter in the discussion that follows. For a given pair-wise comparison the (simplified) probability of incorrect selection is:

$$\Phi\left(-(H_{sel}-H_i)\frac{\sqrt{n}}{\sqrt{\sigma^2_{sel,i}}}\right) \qquad (2.)$$

7. This is a worst case bound. If the hypotheses are independent then Equation 1 is overly conservative. However often there will be some dependency between hypotheses.

8. We assume that the learning system can evaluate multiple hypotheses over any given example (i.e. it can determine the utility of any and all hypotheses on that example). The estimates for the difference in expected utility, $H_{sel} - H_i$, can then be constructed by averaging the difference in utility between the two hypotheses on each observed example. Variance is also based on these differences. This "sharing" of examples means that utility estimates will be correlated with each other. This can be a disadvantage but in the current context it is an advantage as the hypotheses are likely to be positively correlated (e.g., when each hypothesis is a slight variant on a basic search control strategy). When the hypotheses are positively correlated the variance in the estimates will be less when sharing examples than if each hypothesis had to be evaluated on a separate example (the standard way to avoid correlations). It is trivial to modify the algorithm to work for the case where it is not possible to share examples.

Given this pairwise comparison information, Equation 3. illuminates the factors that effect the cost of selecting a hypothesis.

$$n_{sel,i} = \frac{\sigma^2_{sel,i}}{(H_{sel}-H_i)^2}[\Phi^{-1}(\delta)]^2 \quad (3.)$$

In order to satisfy the PAC requirement we must, for each non-selected hypothesis, bound the probability that it is better than the selected hypothesis. The total cost is is the sum of the cost of processing each training example. Equation 3. shows that the number of examples allocated to the two hypotheses increases as the variance increases, as the difference in utility between the hypotheses decreases, or as the acceptable probability of making a mistake decreases.

The first two factors are determined by the environment, but the last, the probability threshold associated with each comparison, can conceivably vary and thus be placed under the control of the hypothesis selection algorithm. The algorithm must only ensure that the *sum* of these probabilities remain less than δ (Equation 1.).

If one comparison requires a great many examples and another very few, it seems possible that allowing greater error for the first and less for the second might reduce the total cost. In fact, allowing the algorithm to judiciously allocate error to each comparison can result in a substantial reduction in overall cost.

Minimizing the cost of selection can be cast as constrained optimization problem. Total cost is the sum of the number of examples allocated to each comparison multiplied by the average cost to process an example. Let $c_{sel,i}$ denote the average cost per example to compare the selected hypothesis with hypothesis i. Let α_i be the error level allocated to the comparison between the selected hypothesis and hypothesis i. Then the optimal allocation of error can be determined by solving the following optimization problem:

Resource Optimization Problem

Choose α_i to minimize the cost $\sum_{i=1}^{k-1} c_{sel,i} \frac{\sigma^2_{sel,i}}{(H_{sel}-H_i)^2}[\Phi^{-1}(\alpha_i)]^2$

Subject to the constraint that $\sum_{i=1}^{k-1} \alpha_i \leq \delta$

Of course in an actual hypothesis selection problem the expected utility of the hypotheses, and perhaps the variance and cost will be unknown before learning begins. Without considering such information the only reasonable policy is to assign an equal error level to each comparison (i.e. $\alpha_i=\delta/[k-1]$).[9] However, comparing this *equal allocation* policy with the optimal solution shows that equal allocation can be highly sub-optimal. To see this, consider the case with three hypotheses, $k=3$, which results in two comparisons with error α_1 and $\delta-\alpha_1$. The cost is:

$$\frac{c_{sel,1}\sigma^2_{sel,1}}{(H_{sel}-H_1)^2}[\Phi^{-1}(\alpha_1)]^2+\frac{c_{sel,2}\sigma^2_{sel,2}}{(H_{sel}-H_2)^2}[\Phi^{-1}(\delta-\alpha_1)]^2$$

which is proportionate to:

$$[\Phi^{-1}(\alpha_1)]^2 + D \times [\Phi(\delta-\alpha_1)]^2$$

To be optimal, α_1 must be chosen so as to minimize this cost. The equal allocation policy assigns α_1 equal to $\delta/2$. The proportionate equation shows that the characteristics of the two comparisons depend on a single value D called the *disparity index*. The disparity index is the normalized difference in evaluation cost, variance, and expected utility of the comparisons. The equal allocation solution is optimal *only* when the disparity index is equal to one, an unlikely event. This is illustrated in Figure 3, which shows the cost equation as a function of α_1 for two different values of the disparity index. The minimum under this curve is the optimal cost and the value of α_1 at this point determines the optimal error allocation. In contrast, the equal allocation policy yields a cost that may differ significantly from this minimum.

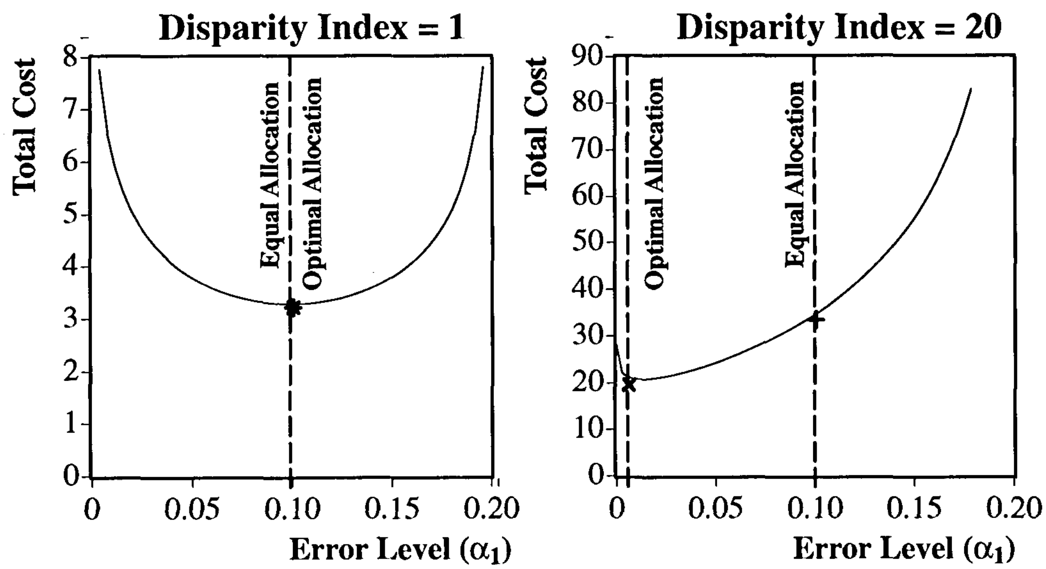


Figure 3. An illustration of the difference between equal and optimal allocation with and without disparity between the comparisons.

In practice it is unlikely that the disparity index will be close to one for all comparisons. Even if the example cost is similar for every hypothesis, the variance and expected utilities of hypotheses will almost certainly differ. The inefficiency of the equal allocation solution increases as these disparities increases. The

9. This corresponds to the maximum entropy solution [Hunter86].

inefficiency also increases as the number of hypotheses increases. It can be shown that for k hypotheses the ratio of equal allocation cost to the optimal cost can be up to $[\Phi^{-1}(\delta/[k-1])]^2 / [\Phi^{-1}(\delta)]^2$. The ratio can be quite large as illustrated in Figure 4. Ignoring disparity information can result in costs up to an order of magnitude greater with as few as ten hypotheses under consideration.

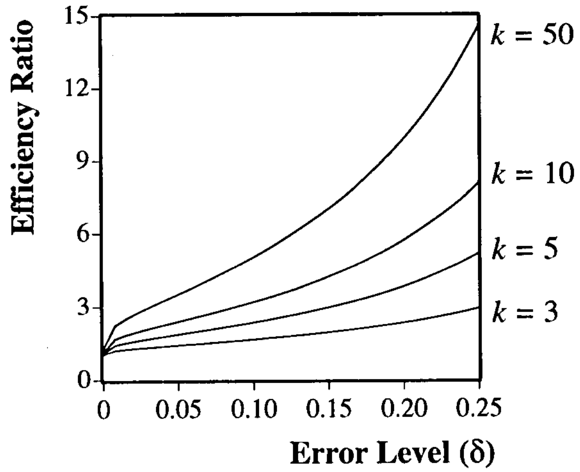


Figure 4. This illustrates the potential disparity between the equal allocation cost and the optimal cost. Results plot the ratio of equal allocation cost to optimal cost for several error levels and number of hypotheses.

Appendix E: The Turnbull and Weiss Algorithm

Turnbull and Weiss have proposed a sequential interval-based procedure for selecting the member of a population with largest mean. Members are considered normal variables with unknown mean and unknown variance. The procedure is as follows. For each hypotheses take an initial sample of n_o observations, then take observations sequentially. Stop sampling from a hypothesis when:

$$S_i^2/n_i \leq 1/n*$$

Where S_i^2 is the sample variance and n_i is the number of examples taken for hypothesis i. The value n^* will be defined momentarily. When sampling has stopped on all hypotheses, select the hypothesis with the highest sample mean. The value n^* is defined as d^2/ε^2 where d is chosen to satisfy:

$$\int_{-\infty}^{\infty} [F(y+d)]^{k-1} f(y) dy = \gamma *$$

where $F(y)$ and $f(y)$ are the cumulative distribution function and probability density function of the standard normal distribution, ε is the indifference interval, and γ^* is the confidence level. Bechhoffer provides extensive tables to determine d [Bechhoffer54].

Turnbull and Weiss provide a proof that their algorithm asymptotically exhibits the requested confidence as the average variance of the hypotheses divided by the indifference interval converges to zero.

Appendix F: The COMPOSER System

The COMPOSER system [Gratch93] uses a statistical approach to select a good strategy from the available alternatives. Because COMPOSER performs hill–climbing, it is always comparing the current strategy, to a set of alternative strategies. COMPOSER designates the current strategy the default strategy, and estimates the marginal utility of each of the other strategies with respect to the default choice. If the confidence interval for one of the marginal utilities moves entirely *below* 0, it is discarded. If one or more marginal utility confidence intervals moves entirely *above* 0, that alternative strategy with the highest sample mean is adopted. Thus, COMPOSER performs quickest ascent hill–climbing, i.e. it adopts the first alternative which is a step uphill, breaking ties by taking the step with the greatest increase.

More specifically, the COMPOSER algorithm can be described as follows[10]. Let H denote a strategy. COMPOSER takes an initial strategy, H_0, and identifies a sequence, H_0, H_1, ... where each subsequent H has higher expected utility with probability $1 - \delta$. Let $T = T_0, T_1, \ldots$ be a set of candidate transformed strategies. *The incremental utility* of adopting a strategy T_i over H_i for a single problem is the utility derived from applying T_i to d minus the utility derived from applying H_i to d. COMPOSER finds a H_i with high expected utility by identifying transformations with positive expected incremental utility. The expected incremental utility is estimated by averaging a sample of randomly drawn incremental utility values. Given a sample of n values, the average of that sample is denoted by $\overline{U}_{ti-hi}$. The likely difference between the average and the true expected incremental utility depends on the variance of the distribution, estimated from a sample by the *sample variance* $S_{uti-uhi}$, and the size of the sample, n. COMPOSER provides a statistical technique for determining when sufficient examples have been gathered to decide, with error δ, that the expected incremental utility of a transformation is positive or negative. The COMPOSER algorithm is summarized in Figure 1.

Let T = candidate strategies $j = 0$ $\delta^* = \delta/(2|T|)$

While more examples and $T \neq \emptyset$ do

$j = j + 1$

$\forall T_i \in T$: Get $U_{T_i - H_i}$ /* Gather statistics and find transformations that have reached significance */

$$significant = \left\{ T_k \in T : j \geq n_0 \text{ and } \frac{S^2_{T_k - h_i}}{(\overline{U}_{T_k - H_i})^2} < \frac{j}{a^2} \right\} \quad \text{where } \Phi(a) = \int_{-\infty}^{a} \left(1/\sqrt{2\pi}\right) exp\{-0.5y^2\} dy = \delta^*$$

$T = T - \{T_k \in significant : \overline{U}_{T_k - H_i} < 0\}$ /* Discard transformations that decrease expected utility */

If $\exists T_k \in stopped : \overline{U}_{T_k - H_i} > 0$ Then /* Adopt transformation that most increases expected utility */

$H_{i+1} = (T_k \in significant : \forall T_l \in significant \left[\overline{U}_{T_k - H_i} > \overline{U}_{T_l - H_i}\right]$

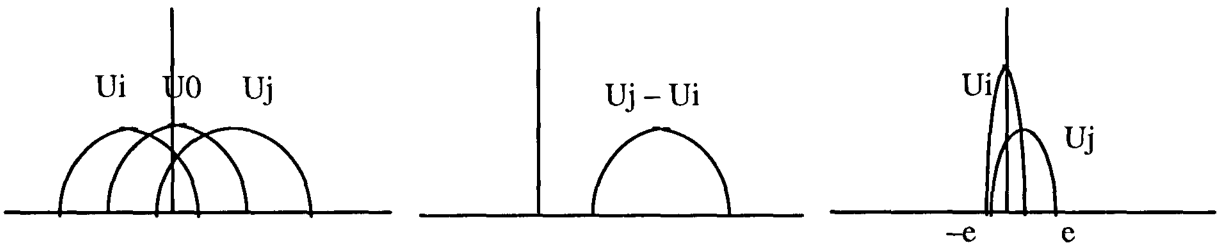
$j = 0$ $\delta^* = \delta/(2|T|)$

Return H_{i+1}

Figure 1: The COMPOSER algorithm

10. This paragraph and figure are taken from [Gratch et al. 93a]. For further details on the COMPOSER algorithm, see [Gratch93].

There are two difficulties with this approach. First, because each strategy is compared to the default, the presence of an extremely good hypothesis strategy cannot be used to prune other hypothesis strategies. This problem, shown graphically in Figure 2, occurs where a good hypothesis strategy (e.g. better than



Dominance　　Dominance Test　　Ambivalence

Figure 2: Problem Cases

the current strategy) can be shown to dominate a poor hypothesis more easily (faster) than the poor hypothesis can be shown to be dominated by the current strategy. This case can be most easily handled by simply analyzing the distribution of values for the difference in utility between the strategies as shown in Figure 2b. The second difficulty is that of irrelevant distinctions. In some cases, one or more hypotheses will have approximately the same utility as the current strategy. Thus it may take many samples to determine which strategy is better, but the overall gain or loss is insignificant. This is a poor expenditure of sampling resources.

Two portions of the confidence interval approach can be added directly to COMPOSER: dominance pruning and indifference pruning. COMPOSER currently only tracks the utility difference between the default strategy and each other strategy. In dominance pruning, COMPOSER tracks the incremental utility comparing all of the pairs of strategies, and allows pruning of the worst strategy from the hypothesis set if any other strategy in the hypothesis set dominates it. In the case where there is a Strategy H_i which is worse than the default H_0 with medium confidence (e.g., less than γ^*), and there is is a strategy H_j which is better than the default strategy H_0 with medium confidence (e.g., less than γ^*), the current COMPOSER strategy would not allow pruning of H_i, but if H_j is better than H_i with high confidence (e.g., greater than γ^*) dominance pruning would allow faster pruning of H_i. In particular dominance pruning would result in significant performance improvement if C_i has a high mean (and hence expected value) as in the speedup learning special case. This issue of cases in which using dominance is expected to improve performance is discussed in more detail in Appendix B.

In indifference pruning, COMPOSER would require an indifference threshold ε. In the case where there is a strategy H_i that can be shown with confidence γ^* that U_{i-0} has true mean μ_{i-0} in the interval $\{-\varepsilon, \varepsilon\}$, the strategy would be pruned from the candidate set. This would significantly improve performance of the COMPOSER algorithm if the system was wasting resources investigating candidate strategies with low expected improvement.