Sample size determination with imprecise risk aversion

Malcolm Farrow
Newcastle University, UK
malcolm.farrow@newcastle.ac.uk

Abstract

We consider multi-attribute utility functions, particularly applied to the choice of a design and sample sizes for an experiment. We extend earlier work, which allowed imprecision in the trade-offs between attributes, to allow imprecision also in the shape of marginal utility functions. The method is illustrated with a simple example involving a two-group binomial experiment.

Keywords. Design of experiments, imprecise utility, risk aversion, sample size.

1 Introduction

In earlier work [8, 9, 10] a method for decision analysis with multiattribute utilities has been developed which does not require the specification of precise trade-offs between different risks. The original motivation for this work was the design of experiments [7, 8]. Multiattribute utilities may be imprecisely specified, due to an unwillingness or inability on the part of the client to specify fixed trade-offs or precise marginal utility functions or because of disagreement within a group with responsibility for the decision. In particular this may be so when the decision is the choice of a design or sample size for an experiment. For example, in the design of a medical experiment, participants in the decision-making process may have different viewpoints, may put different weights on such attributes as the information gain and the risks to trial subjects and may be more or less risk averse in terms of these attributes.

An approach to constructing imprecise multiattribute utility hierarchies and finding the Pareto optimal rules was introduced in [8]. The structure used was based on a utility hierarchy with utility independence at each node and used the notion of imprecise utility trade-offs within such a hierarchy, based on limited collections of stated preferences between outcomes. Pareto optimality, over the set of possible trade-off specifications, was used to reduce the set of alternatives.

Many real decision problems, for example in experimental design, have very large spaces of possible choices. Relaxing the requirement for precise utility specification reduces our ability to eliminate choices by dominance and can leave us with a large class of choices, none of which is dominated by any other over the whole range of possible utility functions allowed by the imprecise specification. Methods were described in [9] to reduce the class of alternatives that must be considered, by eliminating choices which are "ε-dominated" and combining choices which are "ε-equivalent." The effects of different values of ε and of different parts of the hierarchy were explored to see when and why choices were eliminated.

To choose a single alternative d^* from our reduced list, we can use the boundary linear utility approach described in [8], or select the choice which is the last to be eliminated as we increase the value of our ε criterion as described in [9]. We can then find the set D^* of choices which are "almost equivalent" to d^* and perhaps use secondary considerations to choose among them. In [10] methods based on the boundary linear utility for exploring the sensitivity of possible choices to variation in the utility trade-offs were described. This helps us to find a decision which, as far as possible, is a good choice over the whole range of possible trade-offs.

For some other approaches to imprecise utility, see, for example, [12, 2, 13, 16, 17, 5]. A particular feature of the approach used in [8, 9, 10] and this paper is the generality of the form of the utility hierarchy and of the shape of the feasible region.

The purpose of this paper is twofold. Firstly we show how the imprecise utility structure can be extended in a simple way to include imprecision in the shape of the marginal utility functions for attributes, and

therefore in the degree of risk aversion, and that this extension preserves all of the results derived for the structure in previous work. Secondly, we return to the original motivation of the work by applying the methods to the choice of design and determination of sample size for experiments.

In Section 2 we briefly outline the Bayesian approach to experimental design, viewing it as a multi-attribute decision problem. In Section 3 we review the earlier work on decisions with imprecise utility trade-offs. In Section 4 we introduce the extension to include imprecision in the shape of the marginal utility functions. Finally, in Section 5, we apply the ideas to sample-size determination for a simple two-group experiment.

2 Bayesian Experimental Design

2.1 Introduction

The problem of experimental design is essentially that of choosing a design for an experiment from a, possibly infinite, set of possibilities. In simple cases this might just be a matter of choosing a sample size. In more complicated cases it may involve choosing several sample sizes, for observations of different types, or even of selecting types of observations to make, for example determining the values of covariates to use. In any case, this is clearly a decision and, usually, the values of various attributes, typically more than one, which are relevant to us, are unknown before the experiment and our distributions for them depend on the choice of design. We therefore formulate experimental design as a multi-attribute decision problem and choose the design which maximises our expectation of a multi-attribute utility function.

A recent, brief, introduction to this view of experimental design is given by [6]. For a more technical introduction to the field of Bayesian experimental design see, for example, [3]. A discussion of sample-size determination in clinical trials is given in Chapter 6 of [19]. See also, for example, [15, 20].

In much published work on Bayesian experimental design, a fixed total number of observations N is assumed. The problem is then to allocate these observations to design points (*ie* types of observation) while keeping the total fixed (sometimes allowing non-integer allocations on the grounds that it is the proportions of the total sample size which are being determined). Often some measure of information gain is used to provide a utility function and costs are assumed to depend only on the total sample size and therefore need not be considered. This is described as the "design problem" (although, perhaps, "allocation problem" might be a better name).

In contrast, in the " sample size problem", the trade-off between costs and benefits is explicitly considered so a utility function is required which involves both, *eg* [20]. Usually, relatively simple designs are considered.

In many real practical problems we need both to determine a total sample size and how the observations should be allocated to different design points. In this paper we do not distinguish between these two types of problem.

Typically, in experimental design we require a multi-attribute utility function where the attributes include costs and benefits. Each of these may be of more than one kind.

In some cases we might represent the "benefit" from an experiment in terms of some measure of information. For example we might use the posterior precision for some quantity of interest. We may, of course, be interested in several different unknown quatities so each would have its respective marginal utility and these utilities need to be combined. In other cases we might base our benefit utility directly on the pay-off from some *terminal decision*, in which our choice is informed by the result of the experiment. In fact the information-measure approach is (usually, at least) a special case of the terminal-decision approach, in which the terminal decision is to declare a value for some unknown (vector) quantity. The benefit utility is then based on the difference between our declared value and the true value.

Figure 1 shows an influence diagram for a typical problem in experimental design. For example this could refer to the design of a clinical trial in which we wish to compare two or more treatments. There could also be several groups of patients, for example divided by age-group, severity-group, sex *etc.* The initial decision D_X consists of the choice of design d_X. Often the set of possible choices will include the option of no experiment at all. In the experiment, we observe data X. The distribution of X depends on d_X and on unknown quantities (parameters) θ. A vector of pay-offs C_X refers to various attributes, for example financial costs or effects on subjects. The distribution of these depends on d_X and X. Having seen the data X we make a terminal decision D_Y. This may well be the choice of treatment for future patients. We choose d_Y. The outcomes Y of this terminal decision may be, for example, the clinical outcomes for some future patients but may also include other attributes such as costs of future treatments. The distribution of these depends on d_Y and on the unknown θ . These outcomes lead to rewards (pay-offs) C_Y which depend on d_Y and Y. (More generally, they may also depend on θ). There may, of course, be a potentially unbounded

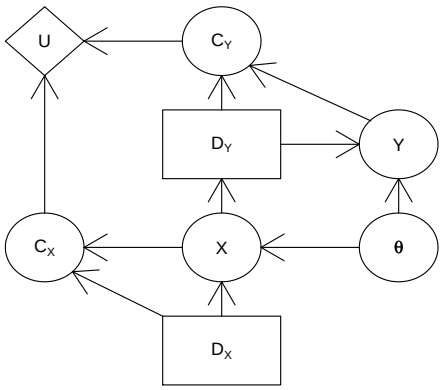

Figure 1: Influence diagram for a typical problem in experimental design.

number of future patients. However, in our utility function, we might discount outcomes as we look further into the future. This might be justified on the grounds that, further into the future, it becomes less likely that our choice of treatment will still be dictated by this experiment. Finally, our overall utility $U = U(C_X, C_Y)$ depends on C_X and on C_Y.

To determine our choice of design $d(X)$ we work our way backwards through the influence diagram. After observing the data $X = x$ in our experiment, we choose

$$\begin{aligned} d_Y &= \arg \max_{d_Y \in D_Y} [\mathrm{E}_{d_Y}\{U(C_X, C_Y) \mid x\}] \\ &= \arg \max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y \mid x)]. \end{aligned}$$

Our expected utility at this stage is

$$\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y) \mid x].$$

Before observing the data, we choose the design

$$d_X = \arg \max_{d_X \in D_X} \mathrm{E}_{d_X}\{\max_{d_Y \in D_Y} [U(d_Y; C_X, C_Y) \mid X]\}.$$

A useful variation on this is to use two different prior distributions, an *inference* or *fitting* or *terminal* prior, which is used for choosing d_Y, and a *design* or *sampling* prior which is used for choosing d_X. This approach was suggested by [21]. Similarly we can have different utility functions for the two decisions.

2.2 Risks in Experimental Design

Since we are concerned in this paper with degrees of risk aversion, let us briefly consider some of the many risks associated with experimental design.

We have already mentioned the financial cost of the experiment, which may not be known in advance with certainty, and the effects on experimental subjects. Particularly in the cases of human and animal subjects we are likely to be concerned about the possibility of adverse reactions but, even in other experiments, there might be other costs concerned with effects on valuable material or equipment. We may come to a conclusion, based on our experiment, which is far from the truth. This could lead to a bad choice in a terminal decision and therefore to a bad pay-off. A type of risk which seems to have had little formal consideration is that something may go wrong with the experiment and that this leads to less useful information than expected or perhaps to none at all. In particular we may suffer from missing observations. Some designs, for example those for microarray experiments, could be very sensitive to missingness. See *eg* [1].

In choosing an experimental design we will be seeking to optimise our expectation of a utility function which involves some or all of these risks. Our choice will therefore depend on how we are willing to trade these risks against each other and this, in turn, depends on our attitudes to these risks, including the shapes of our marginal utility functions since these shapes describe our degrees of risk aversion with respect to the various attributes.

2.3 Utility Hierarchy

A hierarchical structure for utilities in a multi-attribute problem was suggested by [14] and [8] adopted such a structure. In [8], an example was used in which there were financial costs of the experiment and also "ethical costs" which related to possible effects on the experimental subjects. The marginal utilities of these are combined into a Cost utility. In an experiment we potentially learn about a number of quantities and, in their example, [8] represented this collection in four groups, each of which had a marginal utility based on the distance of our posterior expectation from the true value. These were combined into a Benefit utility. Finally the Cost and Benefit utilities were combined in an overall utility for the chosen design.

3 Imprecision in utility trade-offs

3.1 Mutually utility independent hierarchies

In order to introduce imprecision into the trade-offs between attributes, [8] proposed a general class of multi-attribute utility functions which uses the concept of mutual utility independence among sets of attributes in order to impose a structure on the utility function. Attributes $\underline{Y} = (Y_1, ..., Y_k)$ are *utility independent* of the attributes $\underline{Z} = (Z_1, ..., Z_r)$ if condi-

tional preferences over lotteries with differing values of $\underline{Y}$ but fixed values, $\underline{z}$, of $\underline{Z}$, do not depend on the particular choice of $\underline{z}$. Attributes $\underline{X} = (X_1, ..., X_s)$ are *mutually utility independent* if every subset of $\underline{X}$ is utility independent of its complement. If attributes $\underline{X}$ are mutually utility independent, then [14] showed that the utility function for $\underline{X}$ must be given by the *multiplicative form*

$$U(\underline{X}) = B^{-1}\left\{\prod_{i=1}^{s}[1 + ka_iU_i(X_i)] - 1\right\}, \quad (1)$$

where B does not depend on $U_1(X_1), \ldots, U_s(X_s)$, or the *additive form*

$$U(\underline{X}) = \sum_{i=1}^{s} a_iU_i(X_i), \quad (2)$$

where $\mathrm{U}_i(\mathrm{X}_i)$ is a conditional utility function for attribute X_i, namely an evaluation of the utility of X_i for fixed values of the other attributes. The coefficients in (1) and (2) are the *trade-off parameters*; the a_i reflect the relative importance of the attributes and k reflects the degree to which rewards may be regarded as complementary, if $k > 0$, or as substitutes, if $k < 0$.

The assumption of mutual utility independence is enough in itself to reduce the problem to one of considering a finite number of parameters.

The next step is to form a hierarchical structure, in which, at each node, several utilities are merged into a combined utility. This combined utility is merged with others at a node in the next level until, finally, one overall utility function is formed. If, at each node, we have mutual utility independence for the utilities combined at that node, then we term such a utility function a *Mutually Utility Independent Hierarchic (MUIH)* utility. Thus, in a MUIH utility, at each node we combine utilities using either (1) or (2).

This hierarchical structure allows us to relax the requirement for overall mutual utility independence by allowing the user to specify utility independence just at the nodes of the hierarchy and, of course, the user can choose this structure.

Nodes in the hierarchy, other than the marginal nodes, are termed *child nodes* and classified by [8] into the following three types:

1. an *additive node*, where utilities are combined as in (2) with $\sum_{i=1}^{s} a_i \equiv 1$ and $a_i > 0$ for $i = 1, \ldots, s$;

2. a *binary node*, where precisely two utilities are combined, where we rescale the combined utility as
$$U = a_1U_1 + a_2U_2 + hU_1U_2 \quad (3)$$
where $0 < a_i < 1$ and $-a_i \leq h \leq 1 - a_i$, for $i = 1, 2$, and $a_1 + a_2 + h \equiv 1$. Note that (3) is derived by setting $s = 2$ and $h = ka_1a_2$ in (1).

3. a *multiplicative node*, where more than two utilities are combined and the parameter k in (1) may be nonzero. We scale the utility using
$$B = \prod_{i=1}^{s}(1 + ka_i) - 1 \quad (4)$$
with $a_1 \equiv 1, k > -1$ and, for $i = 1, \ldots, s$, we have $a_i > 0$ and $ka_i > -1$. When $k = 0$ we obtain (2).

At each child node n, we have a collection $\underline{\phi}_n = (\phi_{n,1}, \ldots, \phi_{n,r_n})$ of trade-off parameters which determine how the parent utilities at node n are combined to give the value at the child node. If there are N child nodes, then we denote by $\underline{\theta} = (\underline{\phi}_1, \ldots, \underline{\phi}_N)$ the collection of all the trade-off parameters in the hierarchy. A hierarchy in which imprecision is allowed in some of the elements of $\underline{\theta}$ is called an *imprecise independence hierarchy (IIH)*. If the hierarchy contains only additive and binary nodes, then the specification is a *simple imprecise independence hierarchy (SIIH)*

So that the interpretation of utility values does not depend on the choice of trade-off parameters, we place all utilities in the hierarchy on a *standard scale*. Each marginal utility is normed to lie between 0, the worst outcome that we shall consider for the problem, and 1, the best outcome. The relative weights of the marginal utilities are governed by the trade-off parameters at the nodes of the hierarchy and these are chosen to reflect this norming. Consider a child node n. Let C_n be an outcome such that all marginal predecessor nodes have utility 1, and c_n be an outcome such that all marginal predecessor nodes have utility 0. The scalings described above for additive, binary and multiplicative nodes ensure that, at n, the utilities of C_n and c_n are 1 and 0 respectively. Therefore, a utility value of u at node n may always be interpreted as the utility of a gamble giving C_n with probability u and c_n with probability $1 - u$, irrespective of the chain of trade-off parameters in the hierarchy.

3.2 Specification of imprecise utility trade-offs

In standard utility theory, the decision maker must make statements which define the preferences between all combinations of outcomes. In the case of imprecise utility, the decision maker may state preferences

just for some, but not all, choices of outcome combinations. Imprecise utility is defined by obeying all of the constraints implied by the stated preferences. In [8, 9, 10] it was supposed that the decision maker could make preference statements over all outcomes of each individual attribute, and so could specify precise marginal utilities, but could only make preference statements for some, but not all, combinations of the various attributes. Each such preference statement imposed constraints on the tradeoff parameters which are used to combine the individual attributes into an imprecise multi-attribute utility. These constraints together specify a feasible region R for $\underline{\theta}$. Comments on the process of elicitation are made in [8, 9, 10].

In Section 4 below we will drop the assumption that the decision maker has to specify precise marginal utilities.

3.3 Analysis with imprecise utility trade-offs

In earlier work [8, 9, 10], methods have been developed which exploit the IIH structure to reduce the number of choices to be considered and select choices and to explore the sensitivity of choices. Our aim in Section 4 below will be to extend the structure to allow imprecision in the marginal utility functions while preserving the various results derived and retaining our ability to carry out these analyses. In this section we briefly summarise these results and methods.

Having obtained our imprecise specification for the parameters of our multi-attribute utility function we can reduce the number of possible choices, that is designs, by retaining only choices which are Pareto optimal (non-dominated) with respect to the range R of the parameters $\underline{\theta}$.

We have to choose from a set $\mathcal{D}$ of choices. We denote the utility of a particular choice $A \in \mathcal{D}$, evaluated with trade-off parameters $\underline{\theta}$ as $U_{A\underline{\theta}}$. This is evaluated as the expected value of U_θ, with respect to the probability distribution, induced by the choice A, over the marginal attributes involved in U. For two alternatives, $A,\ B$, let $d_{AB}(\underline{\theta}) = U_{A\underline{\theta}} - U_{B\underline{\theta}}$.

We write $A \succeq B$, if $U_{A\underline{\theta}} \geq U_{B\underline{\theta}}\ \forall \underline{\theta} \in R$. We say that A is preferred to B over R, written $A \succ B$, if $A \succeq B$ and $U_{A\underline{\theta}} > U_{B\underline{\theta}}$ for some $\underline{\theta} \in R$, and that A is equivalent to B, written $A \simeq B$, if $U_{A\underline{\theta}} = U_{B\underline{\theta}} \forall \underline{\theta} \in R$. We call alternative A *Pareto optimal* for R if there is no other allowable alternative B for which $B \succ A$ over R. We restrict attention to Pareto optimal alternatives. Furthermore, if we form equivalence classes of equivalent decisions $A_1 \simeq A_2 \simeq ... \simeq A_r$, then it is reasonable to restrict attention to only one representative member of each equivalence class.

To reduce the number of choices further, [9] introduced the concept of *ε-preference* as follows. Let $\varepsilon \geq 0$ be a value chosen to indicate a practical indifference between utility values. For two alternatives A and B, we say that A is almost-preferable with tolerance ε, or, more concisely, "*ε-preferable*" to B, written $A \succeq_\varepsilon B$, over the set R of parameter specifications if $\inf_R(d_{AB}(\underline{\theta})) \geq -\varepsilon$. Two alternatives A, B are said to be almost-equivalent with tolerance ε, or, more concisely, "*ε-equivalent*", written $A \simeq_\varepsilon B$, if both $A \succeq_\varepsilon B$ and $B \succeq_\varepsilon A$. Note that ε-preference does not define a complete ordering of the alternatives and nor does ε-equivalence define an equivalence relation. Alternative A is said to *ε-dominate* alternative B, written $A \succ_\varepsilon B$, if $A \succeq_\varepsilon B$ but $B \not\succeq_\varepsilon A$, where the negation of the relationship is indicated in the usual way. Setting $\varepsilon = 0$, an alternative which is not 0-dominated by any other is Pareto optimal. The notation is extended to collections of alternatives as follows. The collection $\mathcal{A}$ is ε-preferable to the collection $\mathcal{B}$ of alternatives, written $\mathcal{A} \succeq_\varepsilon \mathcal{B}$ if, for each $B \in \mathcal{B}$, there is at least one $A \in \mathcal{A}$ for which $A \succeq_\varepsilon B$.

In [9] a number of results are derived concerning the properties and uses of ε-preference in IIH utilities. In particular, an algorithm is presented for gradually reducing the number of choices by increasing ε from zero and eliminating choices while our retained list remains an ε-Pareto set. Eventually we are left with a single choice d^*. Notice that this choice is made without having to specify a value for ε in advance.

In [10] methods for exploring the sensitivity of choices are presented. In particular the *boundary linear utility*, which had been introduced in [8], is described and results concerning its properties and uses with IIH utilities are given. Let P be the set of vertices of R. In [8] it is shown that, for a SIIH utility, Pareto optimal alternatives for R are the same as Pareto optimal alternatives for P. This forms part of the motivation for the boundary linear utility

$$\bar{U}_\lambda = \sum_{i=1}^{s} \lambda_i U_i$$

where U_i is the utility function determined by the choice of trade-offs $\underline{\theta}_i \in P = \{\underline{\theta}_1, \ldots, \underline{\theta}_s\}$ and $\lambda_1, \ldots, \lambda_s$ are nonnegative constants with $\sum_{i=1}^{s} \lambda_i = 1$.

The results and methods which are developed, some of which may be extended to the case of general IIH utilities, allow us to exploit the idea that, by varying the λ weights, we can change the emphasis which is placed on different parts of the feasible region.

4 Imprecise risk aversion

4.1 Use of basis functions

Now we consider dropping the assumption that the decision maker can give a precise specification of each marginal utility function. Recall that two utility functions, U_A and U_B, are strategically equivalent if $U_B = c + dU_A$ where c and d are constants with $d > 0$. Therefore, without loss of generality we can rescale a marginal utility function to be on the standard scale, as in [8, 9, 10]. Without loss of generality we can also rescale a scalar attribute Z so that the "worst value" is $z = 0$ and the "best value" is $z = 1$. All that is left is to determine the shape of the utility curve between the points $(0, 0)$ and $(1, 1)$. The shape will typically reflect the degree of risk aversion, with a concave curve representing a risk-averse utility function and a convex curve representing a risk-seeking utility function, with respect to the (rescaled) attribute Z. See, for example, Section 4.4.1 of [14].

We could introduce imprecision into the shape of a marginal utility function $U(z)$ by introducing a collection of basis functions $U_1(z), \ldots, U_s(z)$ so that $U(z) = \sum_{i=1}^{s} b_i U_i(z)$ with $b_i \geq 0$ for all i and $\sum_{i=1}^{s} b_i = 1$. We would then elicit a feasible region for the weights $b_1, \ldots, b_s$. An important feature of this approach is that, in effect, we are simply adding an extra layer to the utility hierarchy by making each marginal utility an additive node and introducing the basis functions as new marginal quantities which are parents to the previously marginal nodes. Therefore all of the theory and methods developed previously for the case where imprecision applied only to the trade-offs extends to cover imprecision in the marginal utility functions as well.

A simple example of basis functions is given by quadratic functions. Consider $U_i(z) = c_0 + c_1 z + c_2 z^2$. The constraints $U(0) = 0$ and $U(1) = 1$ simplify this to $U(z) = cz + (1-c)z^2$. The constraints $U'(0) \geq 0$ and $U'(1) \geq 0$, where $U'(z) = dU(z)/dz$, imply $0 \leq c \leq 2$. With $c = 0$, we obtain $U_1(z) = z^2$ and, with $c = 2$, we obtain $U_2(z) = 2z - z^2$. Let $b = c/2$. Then

$$U(z) = (1 - b)U_1(z) + bU_2(z)$$

with $0 \leq b \leq 1$. If $b > 1/2$ we have a risk averse utility function, with $b = 1/2$ it is risk neutral and with $b < 1/2$ it is risk seeking. Curves with $b = 0,\ 0.25,\ 0.5,\ 0.75,\ 1$ are shown in Figure 2.

Note that we can rewrite the basis functions as $U_1(z) = z - h(z)$ and $U_2(z) = z + h(z)$ with, in this case, $h(z) = z - z^2$. It can be seen from Figure 2 that this offers a rather limited range of shapes. While restricting ourselves to monotonic functions, the greatest range that we can obtain in this form is with $h(z) = z$ $(0 \leq z \leq 1/2)$ and $h(z) = 1 - z$ $(1/2 < z \leq 1)$. Even this is somewhat restricted in range and certainly in shape. We can obtain greater flexibility with a more direct approach to eliciting the utility function.

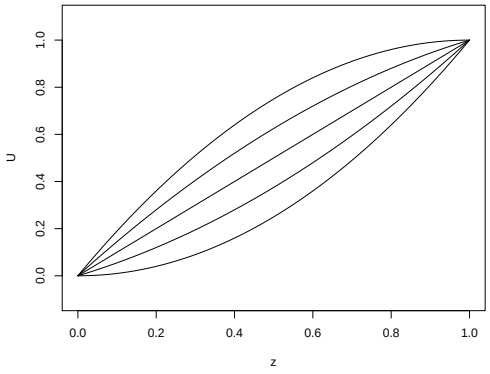

Figure 2: Quadratic utility curves with $b = 0.0, 0.25, 0.5, 0.75, 1$.

While an elicitation procedure for use in practice might involve more refined questions, in principle we can use the probability-equivalent method. In its simplest form, to determine a range for $U(z^*)$ where $0 < z^* < 1$, we offer the decision maker a choice between d_A : the attribute value corresponding to $z = z^*$, with certainty, and d_B : with probability α, the attribute value corresponding to $z = 1$ and, with probability $1 - \alpha$, the attribute value corresponding to $z = 0$. For large α the decision maker will choose d_B, for small α the decision maker will choose d_A but for an intermediate range the decision maker may express no clear preference. The *lower utility* for z^*, $U_1(z^*)$ is the largest value of α at which the decision maker would choose d_A and the *upper utility* for z^*, $U_2(z^*)$ is the smallest value of α at which the decision maker would choose d_B. By repeating this process at a range of values z^* and using suitable interpolation, we obtain lower and upper utility functions, $U_1(z)$ and $U_2(z)$. These can then be our two basis functions. Linear interpolation may well be adequate.

With two basis functions, all allowable utility functions are weighted averages of these two. We could obtain more degrees of flexibility in the shape by adding additional basis functions, for example one which is closer to $U_1(z)$ for some of the range of z and otherwise closer to $U_2(z)$. This would, of course, require more sophisticated elicitation procedures.

4.2 Effect on trade-offs

While the standard scale ensures that all utilities are in $[0, 1]$, where in that range they are likely to be will be different for the lower and upper utility functions. In itself this does not cause a problem. Of more concern is the fact that $U'(z)$ may be different between the lower and upper marginal utility functions. This could affect our consideration of the trade-off at the immediate successor node in the hierarchy. For example, suppose that our marginal utility is U_z and, at the child node, this is combined with another utility U_x to give $U_n = a_n U_z + (1 - a_n)U_x$. Then, if $U_z = (1-b)U_1(z) + bU_2(z)$, the effect on U_n of a fixed change in z may depend on the choice of b. This may be acceptable. After all, the *average* gradient, given a uniform distribution for Z, will remain 1. However the decision maker, with the help of the analyst, needs to consider this consequence of allowing imprecision in the shape of $U_z(z)$. A possible solution would be to elicit a joint feasible region for a and b (or, more generally, for all of the parameters involved at the marginal and child nodes) so that the range of a can depend on the choice of b. If the child node is an additive node it can be extended straightforwardly to include all the basis functions at its parent (marginal) nodes as separate parents. If the child node is a binary node then it can similarly be extended although its new form will not imply mutual utility independence between all of its new parents.

5 Sample size example

To illustrate the method we consider a simple example. Suppose we wish to design a trial, for example a clinical trial, with two treatments and binary outcomes (*eg* cure/not cure). For $g = 1, 2$, we will give treatment g to n_g subjects and observe the number X_g of successes. Using these data, a choice will be made between these treatments for use with future cases.

Suppose that the unknown success rate with treatment g is θ_g. For simplicity assume that our terminal prior gives a Beta$(a_{t,g}, b_{t,g})$ distribution to θ_g with θ_1 and θ_2 independent and that our terminal utility is such that we will choose whichever treatment has the greater posterior probability of success. That is we choose treatment g if the posterior expectation of θ_g is greater than that of $\theta_{g'}$. We set $a_{t,1} = a_{t,2} = b_{t,1} = b_{t,2} = 1.5$.

In our design prior, θ_1 and θ_2 are not independent. A number of methods are available for constructing this joint distribution. For example we could use a bivariate normal distribution for the logits or probits of θ_1 and θ_2 or we could link beta marginal distributions using a copula. However, in this example, the prior is constructed using a mixture distribution. In each component, c, we give θ_1 and θ_2 independent beta distributions, Beta$(a_{c,g}, b_{c,g})$, $g = 1, 2$. The effect of the mixture is to induce correlation between θ_1 and θ_2. A three component mixture is used, with parameters as given in Table 1. The advantage of this form of prior distribution is that prior predictive distributions for the observations can be calculated analytically within each component leading to simple calculations of expected utilities. The results can then be averaged over components.

Component	Probability	Parameters			
c		$a_{c,1}$	$b_{c,1}$	$a_{c,2}$	$b_{c,2}$
1	0.25	7.5	3.0	4.5	4.5
2	0.50	4.5	3.0	3.0	4.5
3	0.25	4.5	6.0	3.0	6.0

Table 1: Parameters of design prior mixture distribution. Within each component $\theta_g \sim$ Beta(a_g, b_g).

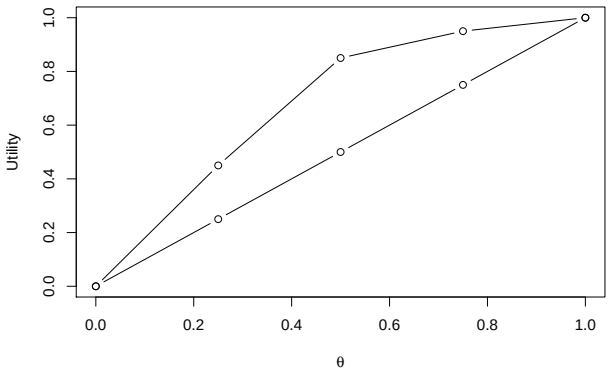

Figure 3: Lower and upper benefit utility functions.

For simplicity in this example we use a simple (precise) form for the marginal cost design utility. Let $n_{\max,1}$ and $n_{\max,2}$ be the largest sample sizes which we would consider. Let

$$Z_{C,g} = \begin{cases} 1 & (n_g = 0) \\ 1 - \frac{h_{0,g} + h_{1,g} n_g}{h_{0,g} + h_{1,g} n_{\max,g}} & (n_g > 0) \end{cases} . \qquad (5)$$

Then the marginal cost utility is $U_C = a_{c,1} Z_{C,1} + a_{c,2} Z_{C,2}$. We use $a_{c,1} = a_{c,2} = 0.5$, $h_{0,1} = h_{0,2} = 10$, $h_{1,1} = h_{1,2} = 1$, $n_{\max,1} = 100$ and $n_{\max,2} = 60$.

The overall design utility is $U = b_C U_C + b_B U_B$. We use $0.03 \le b_C \le 0.07$ and $b_B = 1 - b_C$.

Order	n_1	n_2	ε	Order	n_1	n_2	ε	Order	n_1	n_2	ε
	17	13		25	19	15	0.000084	12	20	15	0.000022
37	0	0	0.004334	24	16	12	0.000067	11	25	19	0.000018
36	19	16	0.000724	23	16	10	0.000048	10	25	16	0.000018
35	14	12	0.000571	22	15	11	0.000048	9	22	19	0.000013
34	18	15	0.000295	21	22	18	0.000048	8	21	17	0.000010
33	21	18	0.000271	20	18	14	0.000044	7	23	17	0.000009
32	13	10	0.000220	19	16	15	0.000043	6	16	16	0.000008
31	15	12	0.000134	18	18	16	0.000043	5	23	19	0.000008
30	21	16	0.000126	17	17	15	0.000040	4	13	13	0.000007
29	17	14	0.000114	16	16	11	0.000037	3	19	17	0.000002
28	13	11	0.000095	15	15	15	0.000033	2	24	18	0.000001
27	24	19	0.000092	14	15	13	0.000023	1	20	16	0.000001
26	16	13	0.000088	13	12	12	0.000022				

Table 2: Results of selection by ε-preference. The order of dropping is shown. The last-retained design is $n_1 = 17,\ n_2 = 13$.

The benefit utility depends on the outcomes for future patients. For a future patient i, let Z_i be 1 or 0 depending on the success or failure of the treatment. This suggests an attribute of the form $Z_B = \sum_{i=1}^{\infty} k_i Z_i$ with $\sum_{i=1}^{\infty} k_i = 1$. For example, we could use $k_i = (1-\lambda)\lambda^{i-1}$ with $0 < \lambda < 1$. Another possibility is $k_i = m^{-1}$ for $i = 1, \ldots, m$ and $k_i = 0$ for $i > n$. For simplicity in this example we adopt the second form and furthermore let $m \to \infty$ so that, given a value of θ, $Z_B \to \theta$.

Using the probability-equivalent method we elicit a lower and an upper utility function $U_{B,L}(\theta)$ and $U_{B,U}(\theta)$ with evaluations at a range of values of θ and linear interpolation. At $\theta = 0, 0.25, 0.5, 0.75, 1$, the lower values are chosen to be $U_{B,L}(\theta) = \theta$, giving risk neutrality. The upper values are $U_{B,L}(\theta) = 0.00, 0.45, 0.85, 0.95, 1.00$, giving risk aversion. These two functions are shown in Figure 3.

Let $\underline{\theta} = (\theta_1, \theta_2)^T$ and $\underline{x} = (x_1, x_2)^T$. We can write the joint probability density of component c, parameters θ_1, θ_2, observations X_1, X_2, and the benefit utility U_B, given sample sizes n_1, n_2, as

$$P = \Pr(c) f_{c,\theta,X}(\underline{\theta}, \underline{x} \mid c) f_U(U_B \mid \underline{x}, \underline{\theta}, c) \qquad (6)$$

where

$$\begin{aligned} f_{c,\theta,X}(\underline{\theta}, \underline{x} \mid c) &= \prod_{g=1}^{2} f_{c,g}(\theta_g \mid c) f_{X|\theta,n_1}(x_g \mid \theta_g) \\ &= \prod_{g=1}^{2} f_{X|n_g}(x_g \mid c) f_{c,g|x}(\theta_g \mid x_g, c) \end{aligned}$$

where $f_{X|n_g}(x_g \mid c)$ is the prior predictive probability function of X_g, given c, and $f_{c,g|x}(\theta_g \mid x_g, c)$ is the conditional posterior density, using the design prior, given c, of θ_g after observing the data $X_g = x_g$. The density of U_B depends on x_1 and x_2 both because we use the posterior density of θ_1 and θ_2 and because the choice of treatment (and hence θ_1 or θ_2) for future cases depends on the posterior distributions, given x_1 and x_2, using the terminal prior. From (6) we can see that we can evaluate conditional expectations within each component of the mixture straightforwardly and then average over the mixture components. The conditional posteriors are beta distributions and the conditional prior predictive distributions for X_g can be evaluated analytically.

With $0 \le n_1 \le 100$ and $0 \le n_2 \le 60$, there are 6161 potential designs. Of these, 38 are non-dominated. With the exception of $(0, 0)$, all of the non-dominated designs have $12 \le n_1 \le 25$, all have $0.6n_1 < n_2 \le n_1$ and all but three have $0.7n_1 < n_2 \le n_1$. Applying the ε-preference algorithm described in Section 5.2 of [9], we obtain the results shown in Table 2. Designs are eliminated one by one as we increase the value of the tolerance ε. Finally one design, $n_1 = 17,\ n_2 = 13$, is left. Interestingly, the last eliminated design is the null experiment, reflecting the fixed cost of any non-null experiment given in (5).

6 Concluding comments

Imprecision in the shape of the marginal utility functions is a natural extension of the earlier work on imprecision in utility trade-offs. In this paper this extension has been made in a way which preserves the results from the earlier work.

The remaining extension to give a fully imprecise analysis would be to allow imprecision in the probability distributions for outcomes given choices. In fact,

if our utility hierarchy is fully additive then we can work directly in terms of previsions of marginal utilities and thus deal with this imprecision in the same way as we have done in this paper. When our multiattribute utility involves products of marginal utilities then incorporation of imprecision in our beliefs in this way would still be possible if we were prepared to regard all of the marginal utilities as uncorrelated. The generalisation to the case without this assumption awaits further work. See, for example, [4] for a different approach.

The simple example in this paper presented no serious computational difficulty. However more complicated experimental design problems will often present computational challenges, both because of the number of potential designs to be compared and, particularly in cases where computationally intensive methods would normally be used to evaluate posterior distributions, the difficulty of evaluating the expected utility for any proposed design. These difficulties apply even without the introduction of imprecision. One possible approach in such cases is to use a simulation-based method, as in [18]. Another possibility is to use a method which does not require such intensive computation, such as Bayes linear methods [8] or Bayes linear kinematics [11, 22] and such an approach, using Bayes linear kinematics is under investigation.

Acknowledgement

The author is grateful for the helpful comments of three reviewers on an earlier version of this paper.

References

[1] R. A. Bailey. Designs for two-colour microarray experiments. *Applied Statistics*, 56:365–394, 2007.

[2] J. Butler, J. Jia and J. Dyer. Simulation techniques for the sensitivity analysis of multi-criteria decision models. *European Journal of Operational Research*, 103:531–546.

[3] K. Chaloner and I. Verdinelli. Bayesian experimental design – a review. *Statistical Science*, 10:273–304, 1995.

[4] M. Danielson, L. Ekenberg and A. Larsson. Distribution of expected utility in decision trees. *International Journal of Approximate Reasoning*, 46:387–407, 2007.

[5] M. Danielson, L. Ekenberg and A. Riabacke. A prescriptive approach to elicitation of decision data. *Journal of Statistical Theory and Practice*, 3:157–168, 2009.

[6] M. Farrow. Optimal Experiment Design, Bayesian. In *Encyclopedia of Systems Biology* (W. Dubitzky, O. Wolkenhauer, K-H. Cho and H. Yokota, Eds). Springer, 2013.

[7] M. Farrow and M. Goldstein. Reconciling costs and benefits in experimental design. In *Bayesian Statistics 4* (J.M.Bernardo, J.O.Berger, A.P.Dawid and A.F.M.Smith, Eds). Oxford University Press, 607-615, 1992.

[8] M. Farrow and M. Goldstein. Trade-off sensitive experimental design: a multicriterion, decision theoretic, Bayes linear approach. *Journal of Statistical Planning and Inference*, 136:498–526, 2006.

[9] M. Farrow and M. Goldstein. Almost-Pareto decision sets in imprecise utility hierarchies. *Journal of Statistical Theory and Practice*, 3:137–155, 2009.

[10] M. Farrow and M. Goldstein. Sensitivity of decisions with imprecise utility trade-off parameters using boundary linear utility. *International Journal of Approximate Reasoning*, 51:1100-1113, 2010.

[11] M. Goldstein and S. Shaw. Bayes linear kinematics and Bayes linear Bayes graphical models. *Biometrika*, 91:425–446, 2004.

[12] G. B. Hazen. Partial information, dominance and potential optimality in multiattribute utility theory. *Operations Research*, 34:296-310.

[13] A. Jiménez, S. Ríos-Insua and A. Mateos. A decision support system for multiattribute utility evaluation based on imprecise assignments. *Decision Support Systems*, 36:65-79, 2003.

[14] R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. John Wiley & Sons, 1976.

[15] D. V. Lindley. The choice of sample size. *The Statistician*, 46:129–138, 1997.

[16] A. Mateos, A. Jiménez and S. Ríos-Insua. Modelling individual and global comparisons for multiattribute preferences. *Journal of Multi-Criteria Decision Analysis*, 12:177–190, 2003.

[17] A. Mateos, A. Jiménez and S. Ríos-Insua. Dominance, potential optimality and alternative ranking in imprecise multi-attribute decision making. *Journal of the Operational Research Society*, 58:326–336, 2007.

[18] P. Müller. Simulation-based optimal design. In *Bayesian Statistics 6* (J.M.Bernardo, J.O.Berger, A.P.Dawid and A.F.M.Sith Eds.) Oxford University Press, 459–474, 1999.

[19] D. J. Spiegelhalter, K. R. Abrams and J. P. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, 2004.

[20] S. B. Tan and A. F. M. Smith. Exploratory thoughts on clinical trials with utilities. *Statistics in Medicine*, 17:2771–2791, 1998.

[21] F. Wang and A. E. Gelfand. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*, 17:193–208, 2002.

[22] K. J. Wilson and M. Farrow. Bayes linear kinematics in the analysis of failure rates and failure time distributions. *Journal of Risk and Reliability*, 224:309–321, 2010.