Bayesian D–Optimal Choice Designs for Mixtures

Aiste Ruseckaite[1,4]
Peter Goos[2,3]
Dennis Fok[1,4]

[1] Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands;
[2] Universiteit Antwerpen, Belgium;
[3] Katholieke Universiteit Leuven, Belgium;
[4] Tinbergen Institute, the Netherlands.

Bayesian D-Optimal Choice Designs for Mixtures

Aiste Ruseckaite[1,4], Peter Goos[1,2,3] and Dennis Fok[1,4]

[1]Erasmus School of Economics, Erasmus Universiteit Rotterdam, the Netherlands
[2]Faculty of Applied Economics & StatUa Center for Statistics, Universiteit Antwerpen, Belgium
[3]Faculty of Bioscience Engineering, Katholieke Universiteit Leuven, Belgium
[4]Tinbergen Institute, the Netherlands

May 7, 2014

Abstract

Consumer products and services can often be described as mixtures of ingredients. Examples are the mixture of ingredients in a cocktail and the mixture of different components of waiting time (e.g., in-vehicle and out-of-vehicle travel time) in a transportation setting. Choice experiments may help to determine how the respondents' choice of a product or service is affected by the combination of ingredients. In such studies, individuals are confronted with sets of hypothetical products or services and they are asked to choose the most preferred product or service from each set.

However, there are no studies on the optimal design of choice experiments involving mixtures. We propose a method for generating an optimal design for such choice experiments. To this end, we first introduce mixture models in the choice context and next present an algorithm to construct optimal experimental designs, assuming the multinomial logit model is used to analyze the choice data. To overcome the problem that the optimal designs depend on the unknown parameter values, we adopt a Bayesian D-optimal design approach. We also consider locally D-optimal designs and compare the performance of the resulting designs to those produced by a utility-neutral (UN) approach in which designs are based on the assumption that individuals are indifferent between all choice alternatives. We demonstrate that our designs are quite different and in general perform better than the UN designs.

Keywords: Bayesian design; Choice experiments; D-optimality; Experimental design; Mixture coordinate-exchange algorithm; Mixture experiment; Multinomial logit model; Optimal design

1 Introduction

One way to obtain information on consumer preferences is to employ choice experiments. In a choice experiment, products or services are characterized by a combination of attribute levels called a profile or an alternative. Each respondent then chooses one alternative from a group of alternatives called a choice set. The respondents repeat this task for several other choice sets presented to them. All choice sets that respondents face make up an experimental design (Kessels et al., 2009; Rose and Bliemer, 2009). The preferences recorded in the course of a choice experiment allow us to estimate the importance of each attribute and its levels. Such estimates are crucial to design new successful products and services, to predict market shares and to determine willingnesses to pay.

Alternative ways to quantify consumer preferences utilize rating and ranking approaches. These approaches are, however, not very appealing if the number of alternatives that has to be rated or ranked is large and when the differences between alternatives are small and subjective. In such scenarios, many researchers in sensometrics, marketing, transportation, environmental and health economics and psychology have resorted to pairwise comparisons (Agresti, 2002; David, 1963). In pairwise comparison studies, the respondents repeatedly evaluate pairs of alternatives, which is cognitively less demanding and considered more reliable than rating or ranking large sets of alternatives. Paired comparison studies can be viewed as choice experiments with choice sets of two alternatives.

While the design of choice experiments involving categorical and (general) quantitative attributes has received substantial attention in the literature, the design of choice experiments where the attributes are proportions of ingredients in a mixture has not been studied at all. This is surprising given the fact that many consumer products are in fact mixtures and given the fact that Cornell (2002) describes many experiments where respondents have to rate or rank mixtures. The examples in Cornell's textbook deal with sports drinks (with various kinds of sweeteners as ingredients), fish patties (with mullet, sheepshead and croaker as ingredients) and tobacco blends (with flue-cured tobacco, burley, turkish blend and processed tobacco as ingredients), among other things.

In this article, we show how to optimally design choice experiments involving mixtures. To this end, we bring together the large body of work of the design and analysis of mixture experiments in general and the work on the optimal design of discrete choice experiments, also known as stated preference studies. The focus in this article is on applications in sensometrics. However, our work might be useful for other application areas, too. For instance, in transportation, the total traveling time can be viewed as a mixture of congested traveling time and free-flow traveling time and the total traveling cost may be a mixture of fuel cost and tolls. In the RGB color model, colors are represented as mixtures of the ingredients red, green and blue.

Mixture experiments have a rich history in agriculture and industry, which started with the seminal papers of Scheffé (1958, 1963). These papers presented linear regression models suitable for analyzing data from mixture experiments, as well as experimental designs that allowed estimation of these models. The initial models treated the responses as quantitative variables and

assumed that the responses depend only on the proportions of the mixture. However, extensions of these models to cope with situations in which the total amount of the mixture also influences the responses, or in which the ingredients are mixtures themselves have been presented as well. An overview of the most commonly used models can be found in Cornell (2002) and Smith (2005).

In this article, we embed the Scheffé model in a multinomial logit framework to link the choices made by respondents in a choice experiment to the mixture of ingredients. This is the article's first novel aspect. Next, we adapt the mixture coordinate-exchange algorithm of Piepel et al. (2005) to compute optimal designs for choice experiments. This is the second novel aspect of the article. Finally, to speed up the search for an optimal design, we use a high-quality starting design as input to the mixture coordinate-exchange algorithm, based on a theoretical result on minimum support designs in Goos and Donev (2007). This is the third innovative aspect of the paper. We adopt a Bayesian approach and evaluate the Bayesian D-optimality criterion using a computationally efficient quasi-Monte Carlo approach.

In the next section, we review the relevant literature on the design of choice experiments and on mixture experiments. In Section 3, we discuss how we combine the mixture and the multinomial logit models and we present the design criterion. We present our algorithm for obtaining the optimal design involving mixtures in Section 4. We show a selection of our computational results in Section 5, starting with locally optimal designs and ending with a Bayesian optimal design for a choice experiment involving a cocktail. We pay special attention to the performance of the newly developed designs relative to utility-neutral designs, a class of locally optimal choice designs that has received much attention in the literature. Section 6 contains a conclusion.

2 Literature review

This article combines two streams of literature, namely, that on the optimal design of choice experiments and that on the design of mixture experiments. In optimal design of experiments, the information contained within an experiment is quantified by means of the information matrix on the estimated model's unknown parameters. An optimal design optimizes some function of this information matrix. For instance, the design which maximizes the determinant of the information matrix is called D-optimal. D-optimal designs are the most commonly used types of optimal designs, both in the choice experiments literature and in the mixture experiments literature. Therefore, we also focus on D-optimal designs. In this section, we first describe the state of the art when it comes to designing choice experiments. Next, we turn our attention to the literature on mixture experiments.

The first paper to discuss the D-optimal design of choice experiments was Huber and Zwerina (1996), who assumed a multinomial logit model and showed that taking into account prior knowledge on the respondents' preferences had a beneficial impact on the quality of the experimental design and the subsequent analysis. This is possible because the information matrix for the multinomial logit model, the most commonly used model for analyzing choice data, depends on the unknown model parameters. Therefore, any D-optimal design also depends on the parameter values. A key feature of the approach taken by Huber and Zwerina (1996) is that it uses

a degenerate prior distribution, meaning that the optimal design is based on a single a-priori chosen value of the model parameters. This approach, which is referred to as the locally optimal design approach in the optimal experimental design literature, does not allow for any uncertainty concerning the parameters' values.

To overcome this weakness, Sándor and Wedel (2001) used a Bayesian optimal design approach involving a proper prior distribution, which summarizes the prior beliefs about the model parameters, including the uncertainty. They showed that, even if the prior distribution is misspecified (or not well specified), more efficient experimental designs are obtained, due to the fact that the prior uncertainty is taken into account. Since the appearance of Sándor and Wedel's work, the Bayesian optimal design approach has received much attention in the choice design literature. Kessels et al. (2006, 2009) introduced better algorithms for constructing optimal designs for the multinomial logit model and studied Bayesian A-, G- and V-optimal designs, in addition to Bayesian D-optimal designs. The best-performing algorithm, in terms of speed and solution quality, is the coordinate-exchange algorithm (Meyer and Nachtsheim, 1995) as proposed by Kessels et al. (2009).

More recent work on the optimal design of choice experiments has focused on other application areas such as transportation (Hensher and Rose, 2009), environmental economics (Vermeulen et al., 2011), on fast approaches to evaluate the Bayesian optimality criteria (Bliemer et al., 2008; Yu et al., 2010b), on models other than the multinomial logit model (Bliemer and Rose, 2010; Bliemer et al., 2009a; Goos et al., 2010; Yu et al., 2009, 2011), or on design approaches that take into account the complexity of choice tasks (Danthurebandara et al., 2011; Kessels et al., 2011a).

Some authors use a very specific locally optimal design approach to set up choice experiments, i.e., they assume that all model parameters are zero. We refer to Burgess and Street (2005); Grasshoff et al. (2003, 2004); Grossmann et al. (2006, 2009); and Street and Burgess (2007) for D-optimal designs derived under that assumption. Under the assumption that the model parameters are zero, the information matrix for the multinomial logit model is proportional to the information matrix for a linear regression model, in case the experiment is run in blocks (Kessels et al., 2011b). Blocked experiments are used whenever not all experimental tests can be performed under homogeneous circumstances. The equivalence of the problem of finding an optimal design for a blocked experiment and that of finding an optimal design for a choice experiment under the assumption of zero model parameters implies that standard software for the optimal design of experiments serves as a useful benchmark for the designs we propose in this paper. Designs derived under the assumption that model parameters are zero are referred to as utility-neutral designs, because, under that assumption, all alternatives have the same utility and the respondents are indifferent between all alternatives.

Much research has also been done on the optimal design of mixture experiments. The research focused on industrial and bioscience applications, where linear regression models are used and experiments are usually completely randomized or blocked (Cornell, 2002; Smith, 2005). D-optimal completely randomized designs for the most commonly used models, i.e., the linear, quadratic and special cubic Scheffé models, in scenarios with simplex-shaped experimental regions have been known for many years (Kiefer, 1961; Uranisi, 1964). For more general scenarios and other model

types, Piepel et al. (2005) introduced a coordinate-exchange algorithm for mixture experiments. Goos and Donev (2006, 2007) describe an algorithmic approach for designing blocked mixture experiments. The latter work is particularly relevant here because it also produces utility-neutral choice designs in case the alternatives are mixture of ingredients. We will use these D-optimal blocked mixture experiments as benchmarks in this article.

In spite of all published work on the design of choice experiments and the design of mixture experiments, the construction of optimal choice designs for mixtures is still an unexplored research area. In this article, we fill this void in the literature.

3   Models and Design Criterion

In this section, we first describe general models for data from mixture experiments. Next, we discuss these models in combination with the multinomial logit model. We then discuss the D-optimality criterion which we use as the design selection criterion.

First, however, we need to introduce some notation. In a mixture choice experiment, a product or service is represented by a combination of q *ingredient proportions* $x_1, x_2, \ldots, x_q$ that take values between zero and one, sum up to one and make up an *alternative*. A group of alternatives presented to a respondent is called a *choice set*. The collection of all choice sets is called a *design*. We denote the total number of alternatives in an experimental design by n, the number of choice sets by S and the number of alternatives within a choice set by J.

3.1   Models for data from mixture experiments

In order to model choice as a function of mixtures of ingredients, a suitable model is required, i.e., a model that can handle proportions as explanatory variables. Such models have been developed for situations involving continuous responses. As a result, they have only been used in the linear regression context. The *mixture constraint*

$$\sum_{i=1}^{q} x_i = x_1 + x_2 + \cdots + x_q = 1 \tag{1}$$

has a substantial impact on the models that can be considered. The first major consequence of this mixture constraint is that a linear regression model for mixture data cannot contain an intercept. Otherwise, the model's parameters cannot be estimated uniquely. Moreover, the cross-products $x_i x_j$ (generally required for quantifying interaction effects) and the squares x_i^2 (generally required for quantifying quadratic effects) cannot be included simultaneously in the model, as this also leads to perfect collinearity. To see this, note that

$$x_i^2 = x_i (1 - \sum_{\substack{j=1 \\ j \neq i}}^{q} x_j) = x_i - \sum_{\substack{j=1 \\ j \neq i}}^{q} x_i x_j,$$

4

for every proportion x_i. Thus, the square of a proportion is a linear combination of that proportion and its cross-products with all other proportions composing the mixture. As a result, the mixture constraint naturally leads to the family of models proposed by Scheffé (1958, 1963). The first-order Scheffé model for a continuous dependent variable Y is given by

$$Y = \sum_{i=1}^{q} \beta_i x_i + \varepsilon, \tag{2}$$

whereas the second-order Scheffé model is given by

$$Y = \sum_{i=1}^{q} \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} x_i x_j + \varepsilon. \tag{3}$$

The so-called special-cubic model can be written as

$$Y = \sum_{i=1}^{q} \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} x_i x_j + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{k=j+1}^{q} \beta_{ijk} x_i x_j x_k + \varepsilon. \tag{4}$$

The interpretation of the coefficient β_i in Equation (2) is the expected response if x_i is 100%, i.e., if a product consists of ingredient i only. We cannot interpret it as the partial effect of ingredient i, since changing the proportion x_i requires at least one other proportion to be changed as well. Otherwise, the mixture constraint in Equation (1) is violated. It is, therefore, relatively difficult to interpret parameters in the Scheffé models.

If we expect interaction effects like synergism (interaction of ingredients such that the total effect is greater than the sum of the individual effects) or antagonism (interaction of ingredients such that the total effect is smaller than the sum of the individual effects), we should use the second-order or special-cubic model in Equation (3) or (4). As the number of terms in the second-order or special-cubic Scheffé models increases rapidly with the number of ingredients q, estimating them may require a large number of observations.

In a mixture experiment, the experimental region is the set of all possible combinations of the proportions $x_1, x_2, \ldots, x_q$ that satisfy the mixture constraint in Equation (1). That region is a $(q-1)$-dimensional regular simplex, which is an equilateral triangle when $q = 3$. If we impose constraints on the ingredient proportions, such as lower and/or upper bounds, we obtain a constrained experimental region. When only lower bounds are used, the experimental region remains simplex-shaped and it is in this case common to redefine the proportions in terms of so-called L-pseudocomponents, each of which again takes values on the interval $(0, 1)$ and the sum of which is one. If we denote the lower bound for ingredient i as L_i, where $0 \leq L_i \leq x_i$, then the ith pseudocomponent x_i' is defined as

$$x_i' = \frac{x_i - L_i}{1 - L},$$

where $L = \sum_{i=1}^{q} L_i < 1$. The original reason for using L-pseudocomponents is to enhance the numerical stability in the data analysis. We use these pseudocomponents to be able to generate

optimal designs subject to lower bound restrictions on the proportions.

3.2 Multinomial logit model

To capture an individual's preferences and to model the choices made in mixture choice experiments, we build on the multinomial logit (MNL) model. The model is based on random utility theory and expresses the utility of alternative j in choice set s as a function of the observed alternative specific characteristics plus an error term:

$$U_{js} = \mathbf{f}'(\mathbf{x}_{js})\boldsymbol{\beta} + \epsilon_{js}, \tag{5}$$

where $\mathbf{f}'(\mathbf{x}_{js})$ represents the model expansion of the characteristics of alternative j in choice set s and $\boldsymbol{\beta}$ is the corresponding parameter vector. Note that, because the same parameter vector $\boldsymbol{\beta}$ is attached to every respondent in the MNL model, the model is suitable whenever people's preferences are homogeneous across the population.

In the mixture context, the alternatives' characteristics are ingredient proportions. Therefore, the model expansions we use in this paper are based on, but not identical to, the basic Scheffé models in Equations (2) - (4). When the Scheffé mixture models are directly embedded in a MNL framework, not all parameters can be estimated independently. For the model parameters to be identified, we have to drop one of the q proportions x_i from the model. Which proportion is dropped is arbitrary. We choose to drop the linear term corresponding to the last ingredient proportion x_q, which comes down to substituting $1 - x_1 - \ldots - x_{q-1}$ for x_q and next dropping the resulting constant term. Assuming, for instance, that one starts with the special cubic model, the utility for the jth alternative in choice set s then becomes

$$U_{js} = \sum_{i=1}^{q-1}(\beta_i - \beta_q)x_{ijs} + \sum_{i=1}^{q-1}\sum_{k=i+1}^{q}\beta_{ik}x_{ijs}x_{kjs} + \sum_{i=1}^{q-2}\sum_{k=i+1}^{q-1}\sum_{l=k+1}^{q}\beta_{ikl}x_{ijs}x_{kjs}x_{ljs} + \varepsilon_{js},$$

$$= \sum_{i=1}^{q-1}\beta_i^* x_{ijs} + \sum_{i=1}^{q-1}\sum_{k=i+1}^{q}\beta_{ik}x_{ijs}x_{kjs} + \sum_{i=1}^{q-2}\sum_{k=i+1}^{q-1}\sum_{l=k+1}^{q}\beta_{ikl}x_{ijs}x_{kjs}x_{ljs} + \varepsilon_{js},$$

where $\beta_i^* = \beta_i - \beta_q$.

In each choice set, the respondent chooses the alternative that has the largest utility. The probability that a respondent chooses alternative j in choice set s is, therefore,

$$p_{js} = P\{U_{js} > \max\{U_{1s}, \ldots, U_{j-1,s}, U_{j+1,s}, \ldots, U_{Js}\}\}.$$

In the MNL model, the stochastic components ϵ_{js} are assumed to be mutually independent and to follow the so-called log Weibull distribution (also known as type-I extreme value distribution). Their cumulative distribution function is given by

$$F(\epsilon_{js}) = \exp(-\exp(-\sigma\epsilon_{js})),$$

where σ is a positive scale parameter. The probability that a respondent chooses alternative j in

choice set s can now be written as

$$p_{js} = \frac{\exp(\sigma \mathbf{f}'(\mathbf{x}_{js})\boldsymbol{\beta})}{\sum\limits_{t=1}^{J} \exp(\sigma \mathbf{f}'(\mathbf{x}_{ts})\boldsymbol{\beta})}. \tag{6}$$

It is clear from Equation (6) that the scale parameter σ and the vector $\boldsymbol{\beta}$ cannot be estimated independently. In some sense, the parameter σ defines the scale of the parameters contained within $\boldsymbol{\beta}$. The larger the value of σ for a given $\boldsymbol{\beta}$, the larger the degree of certainty with which we can predict the choices respondents make. The smaller the value of σ, the more randomness is present in the respondents' choices. As the parameter σ is not identified, we set it to one as is customary in the literature. In Section 5.2, however, we explicitly consider the impact of the scale of the $\boldsymbol{\beta}$ vector. This is equivalent to studying the impact of σ.

3.3 Design optimality criterion

The goal when constructing an optimal experimental design is to obtain precisely estimated parameter values. This goal is formalized in an objective function. In the optimal experimental design literature, this objective function is called the design optimality criterion. Different design optimality criteria exist (for instance, the D-, A-, G- and V-optimality criteria), all of which are functions of the Fisher information matrix on the parameters (Atkinson et al., 2007). We use the D-optimality criterion, which is the most commonly used one in the literature on the design of mixture experiments and in the literature on the design of choice experiments. Designs that are optimal with respect to this criterion in general also perform well in terms of other criteria. A D-optimal design maximizes the determinant of the information matrix, or minimizes its inverse, the determinant of the variance-covariance matrix of the parameter estimator.

For the MNL model, the total information matrix $\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})$ is obtained as the sum of the information matrices of the S choice sets, $\mathbf{I}_s(\mathbf{X}_s, \boldsymbol{\beta})$ and given by

$$\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}) = \sum_{s=1}^{S} \mathbf{I}_s(\mathbf{X}_s, \boldsymbol{\beta}) = \sum_{s=1}^{S} \mathbf{X}_s'(\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s')\mathbf{X}_s, \tag{7}$$

with $\mathbf{X}_s = [\mathbf{f}(\mathbf{x}_{js})]'_{j=1,\ldots,J}$, $\mathbf{X} = [\mathbf{X}_s']'_{s=1,\ldots,S} = [\mathbf{f}(\mathbf{x}_{js})]'_{j=1,\ldots,J;s=1,\ldots,S}$, $\mathbf{p}_s = (p_{1s},\ldots,p_{Js})'$ and $\mathbf{P}_s = \mathrm{diag}(p_{1s},\ldots,p_{Js})$. From Equation (7), it is clear that the information matrix depends on the choice probabilities p_{js} and, therefore, also on the unknown parameter vector $\boldsymbol{\beta}$. As a result, in order to optimize the D-optimality criterion value when constructing an optimal design, we need to assume (prior) values for the model parameters.

Generally, in optimal experimental design, a prior *distribution* is specified for the unknown parameters and a design is sought that maximizes the D-optimality criterion averaged over that prior distribution. The resulting design is referred to as the Bayesian optimal experimental design (Atkinson et al., 2007). We denote the prior distribution for the parameter vector $\boldsymbol{\beta}$ by $\pi(\boldsymbol{\beta})$ and use the following Bayesian D-optimality criterion for mixture choice experiments for estimating

the MNL model:

$$D = \log \Big(\int_{\mathbf{R}^{\mathbf{p}}} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}))\}^{1/p} \pi(\boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta} \Big), \tag{8}$$

where p is the number of parameters in $\boldsymbol{\beta}$. The exponent $1/p$ makes the D-criterion comparable across models with different numbers of parameters. Note that the log transformation has no impact on the optimal design. The Bayesian D-optimality criterion was introduced in the context of choice experiments by Sándor and Wedel (2001) and it has been used by many other authors studying the D-optimal design of choice experiments. This criterion corresponds to Criterion II in Atkinson et al. (2007) and is appropriate when the variance of the parameter estimates is of the highest importance.

Through the prior distribution, we can express different beliefs about the parameters as well as our uncertainty concerning these prior beliefs. It is possible to use a degenerate prior distribution putting all weight on a single value for the parameter vector, say $\tilde{\boldsymbol{\beta}}$. The Bayesian D-optimality criterion then simplifies to the local D-optimal design criterion

$$\log(\det(\mathbf{I}^{-1}(\mathbf{X}, \tilde{\boldsymbol{\beta}}))^{1/p}). \tag{9}$$

When $\tilde{\boldsymbol{\beta}} = \mathbf{0}_p$, the locally D-optimal design is called the utility-neutral design. We discuss the choice of a proper prior distribution in Section 4.3.

It is interesting to note that there may be more than one D-optimal design in a given scenario. This is especially true in the utility-neutral case. However, designs that are equivalent in the utility-neutral case are generally not equivalent when evaluated using a nonzero prior parameter vector $\tilde{\boldsymbol{\beta}}$ or using a non-degenerate prior distribution $\pi(\boldsymbol{\beta})$.

4   Design Construction Algorithm

In this section, we present an algorithm that minimizes the D-optimality criterion value in Equation (8) through a local search procedure. The algorithm involves two main steps: (1) the generation of a starting design and (2) the use of a mixture coordinate-exchange algorithm to improve the starting design, proportion by proportion. This algorithm is a heuristic and, therefore, there is no guarantee that it will produce the globally Bayesian D-optimal design. To increase the likelihood of finding the global optimum, we repeat the two steps a number of times. This approach is common in algorithms for constructing optimal experimental designs (Atkinson et al., 2007).

Our algorithmic approach differs from the mixture coordinate-exchange algorithm of Piepel et al. (2005) in two important ways. First, we use a better procedure for generating a starting design, which takes into account the fact that a design for a mixture choice experiment requires the mixtures to appear in groups. Second, we have adapted the algorithm so that it optimizes the Bayesian D-optimality criterion for the MNL model.

4.1 Generating a starting design

Various implementations of the coordinate-exchange algorithm for constructing optimal experimental designs, including the mixture coordinate-exchange algorithm of Piepel et al. (2005) for linear regression models, use completely random starting designs. Other algorithms for constructing D-optimal designs generate starting designs partly at random and partly using a greedy heuristic (Atkinson and Donev, 1989). The random part ensures that the starting designs are different. The greedy heuristic ensures that the starting design is reasonably good, which speeds up the search for the optimal design considerably.

To obtain a starting design, we use similar ideas. We randomly generate starting designs that are already optimal in some sense. More specifically, we use a method introduced by Goos and Donev (2007) to generate D-optimal minimum support mixture designs in blocks. Due to the equivalence of the problem of finding a D-optimal design for a blocked experiment and the D-optimal choice design problem in case the parameters are zero, the resulting starting designs are automatically D-optimal utility-neutral choice designs within the class of minimum support designs. In the context of mixture experiments, a minimum support design is a design which has as many distinct mixtures as there are parameters in the model.

Since our starting design is D-optimal for a problem that approximates the Bayesian D-optimal mixture choice design problem, we can trust that the starting design is of high-quality. Our computational experiments showed that using these minimum support designs rather than completely random starting designs substantially decreases the required computation time of the algorithm[1].

The approach we take to construct the starting design involving S choice sets of J mixtures works as follows (Goos and Donev, 2007):

- Choose the p distinct mixtures which are D-optimal for the Scheffé model under consideration in case a completely randomized experiment would be performed. For the first-order Scheffé model, the required mixtures correspond to the q vertices of the simplex. For the second-order model, the q vertices and $q(q-1)/2$ edge midpoints are required. For the special-cubic model, the same mixtures are required as for the second-order model, plus all $q(q-1)(q-2)/3$ mixtures of three ingredients, each with proportion $1/3$.

- Replicate the p distinct mixtures as evenly as possible, so that the total number of mixtures is $n = SJ$. Each of the mixtures then appears $\text{int}(n/p)$ or $\text{int}(n/p) + 1$ times in the design. When p is not a divisor of n, the mixtures that are replicated most frequently are randomly chosen. The resulting design is a D-optimal completely randomized design for the model under consideration.

- Randomly allocate the n mixtures to the S choice sets, subject to the constraint that a given mixture should not appear more than once in any choice set. The resulting design is now D-optimal in the class of minimum support designs for the model under consideration in case there are S blocks of size J.

---

[1]To generate random starting designs uniformly on the simplex, the proportions can be drawn from a standard exponential distribution and then normalized.

It is clear that the procedure to create the starting design involves a large number of random components. This procedure is, therefore, capable of producing a large number of different starting designs.

4.2 Improving the starting design

To improve the starting design, we use a mixture coordinate-exchange algorithm (Meyer and Nachtsheim, 1995; Piepel et al., 2005). The algorithm starts by optimizing the first ingredient proportion of the first alternative in the first choice set of the starting design using the method of Brent (1973). Brent's method is a one-dimensional optimization algorithm, based on a combination of golden section search and successive parabolic interpolations. The mixture coordinate-exchange algorithm then successively optimizes the second ingredient proportion of the first alternative in the first choice set, the third ingredient proportion, ..., and the qth ingredient proportion. This process is repeated for each alternative in each choice set of the design. Whenever at least one proportion has been improved in a complete pass of the coordinate-exchange algorithm through the entire design, the optimization procedure is restarted from the first ingredient proportion. The algorithm stops when no improvements have been performed in a complete pass through all the proportions.

This mixture coordinate-exchange algorithm differs in a substantial way from the original coordinate-exchange algorithm described in Meyer and Nachtsheim (1995). In our case, the proportions cannot be changed independently: if one proportion changes, then at least one other proportion has to change as well. Otherwise, the mixture constraint would be violated. To deal with this complication, we adopt the approach by Piepel et al. (2005) and make sure that, when we optimize a proportion x_i, the pairwise ratios of the other ingredient proportions remain fixed. This approach is known as exchanges along Cox-effect directions (Cornell, 2002; Piepel, 1982). The general expression for recomputing the proportions $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_q$ after a proportion x_i is changed to $x_i + \Delta$ is

$$x_j - \frac{\Delta x_j}{1 - x_i}.$$

In case $x_i = 1$ before the change, we set the proportions $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_q$ to $\Delta/(q-1)$.

4.3 Prior Distribution

Specifying a prior distribution for the parameters in the identified utility model, involving the differences $\beta_1 - \beta_q, \ldots, \beta_{q-1} - \beta_q$, is not straightforward. We recommend first specifying a prior distribution for the full (unidentified) Scheffé model (as in Equations (2)-(4)) and transforming it into a prior for the identified model. This also makes the resulting prior specification invariant to the chosen reference proportion.

As a first example, consider the parameter vector $(\beta_1, \beta_2, \ldots, \beta_q)'$ for the first-order Scheffé model and assume a normal prior distribution with mean vector $\boldsymbol{\beta}_0$ and variance-covariance matrix $\kappa \mathbf{I}_q$, with $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \ldots, \beta_{0q})'$ and $\mathbf{I}_q$ the identity matrix of dimension q. This distribution

translates into the prior distribution

$$
\begin{pmatrix} \beta_1 - \beta_q \\ \beta_2 - \beta_q \\ \vdots \\ \beta_{q-1} - \beta_q \end{pmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \beta_{01} - \beta_{0q} \\ \beta_{02} - \beta_{0q} \\ \vdots \\ \beta_{0,q-1} - \beta_{0q} \end{bmatrix}, \begin{bmatrix} 2\kappa & \kappa & \kappa & \ldots & \kappa \\ \kappa & 2\kappa & \kappa & \ldots & \kappa \\ & & & \vdots & \\ \kappa & \kappa & \kappa & \ldots & 2\kappa \end{bmatrix} \right) \tag{10}
$$

for the differences $(\beta_1 - \beta_q, \beta_2 - \beta_q, \ldots, \beta_{q-1} - \beta_q)'$. Note that a very specific correlation structure results from this transformation. Applying the same approach to a special-cubic model for three ingredients, we obtain

$$
\begin{pmatrix} \beta_1 - \beta_3 \\ \beta_2 - \beta_3 \\ \beta_{12} \\ \beta_{13} \\ \beta_{23} \\ \beta_{123} \end{pmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \beta_{01} - \beta_{03} \\ \beta_{02} - \beta_{03} \\ \beta_{0,12} \\ \beta_{0,13} \\ \beta_{0,23} \\ \beta_{0,123} \end{bmatrix}, \begin{bmatrix} 2\kappa & \kappa & 0 & 0 & 0 & 0 \\ \kappa & 2\kappa & 0 & 0 & 0 & 0 \\ 0 & 0 & \kappa & 0 & 0 & 0 \\ 0 & 0 & 0 & \kappa & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa & 0 \\ 0 & 0 & 0 & 0 & 0 & \kappa \end{bmatrix} \right). \tag{11}
$$

In this paper, we either start from independent normal prior distributions for the parameters of the full Scheffé model, or base the distribution on prior research. There are also other ways to obtain sensible prior distributions. Huber and Zwerina (1996) suggest using a small pilot test to obtain a-priori estimates of the model parameters. Sándor and Wedel (2001) use managers' beliefs as a starting point for constructing a prior distribution.

4.4 Computational issues

The most computationally intensive part in our work is the evaluation of the p-dimensional integral in Equation (8). As there is no analytical expression for it, that integral has to be evaluated numerically for each design considered in the course of our mixture coordinate-exchange algorithm. Since the number of designs we need to consider during the optimization procedure is large, it is of crucial importance to use a computationally efficient numerical integration approach.

A simple but computationally demanding approach would be to use pseudo Monte-Carlo sampling. In that approach, R pseudo random draws for $\boldsymbol{\beta}$, say $\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(R)}$, are taken from the prior distribution and the Bayesian D-optimality criterion is approximated as

$$
\mathrm{D} = \log \Big(\frac{1}{R} \sum_{i=1}^{R} \{\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}_{(i)}))\}^{1/p} \Big). \tag{12}
$$

The required number of determinant evaluations in the numerical integration increases linearly with the number of draws, R. A weakness of the pseudo Monte-Carlo approach is that draws are made independently. This may result in a large variability of the integral approximation and a slow convergence to the true value of the integral as R increases. The use of pseudo Monte-Carlo sampling, therefore, requires a large value for R (Bliemer et al., 2009b; Yu et al., 2010a). This

results in long computation times for the mixture coordinate-exchange algorithm.

In our work, we use a systematic sample from the prior distribution, based on Halton sequences (Halton, 1960). The fact that the resulting draws $\boldsymbol{\beta}_i$ are systematically spread over the domain, improves the accuracy of the integral approximation. This means that fewer draws and, hence, much shorter computing times are required to evaluate the Bayesian D-optimality criterion (Bliemer et al., 2009b; Yu et al., 2010a).

In order to speed up the computations even further, we use the Cholesky decomposition to calculate the determinant of the information matrix. This is faster by a factor of two for symmetric, positive definite matrices, such as the information matrix.

5 Illustrations

In this section, we provide illustrations of the designs that our algorithm produces. First, we consider utility-neutral designs, i.e., setting $\boldsymbol{\beta} = \mathbf{0}_p$. We demonstrate that, in this case, the performance of our algorithm matches that of the mixture coordinate-exchange algorithm of Piepel et al. (2005), as implemented in the software package JMP10. The excellent performance of of our algorithm in this case provides confidence that the algorithm is also capable of finding good designs in other, more complex, cases that cannot be tackled using standard software. Next, we consider locally optimal designs. We obtain designs for fixed, non-zero, values of $\boldsymbol{\beta}$ and show that these designs outperform the utility-neutral design when both designs are evaluated at $\boldsymbol{\beta}$. We also investigate the impact of the value of $\boldsymbol{\beta}$, where we separately study the impact of changing the shape of the implied utility function and the impact of changing the implied range of the utilities. Finally, we study Bayesian D-optimal designs. Such designs account for uncertainty in $\boldsymbol{\beta}$ through the use of a normal prior distribution with a non-zero variance-covariance matrix.

While our algorithm is capable of handling any number of ingredients and any kind of Scheffé model, in this section, we only consider experimental designs for the special-cubic model for three ingredients with 14 alternatives in seven choice sets of size two. This allows for a straightforward comparison of the designs under different scenarios. All reported designs are obtained using 1,000 different starting designs. For the Bayesian D-optimal design, we used 128 Halton draws to approximate the Bayesian D-optimality criterion.

5.1 Utility-neutral designs

First, we construct choice designs assuming that $\boldsymbol{\beta} = \mathbf{0}_p$. We do so for two reasons. First, as pointed out in Section 2, several authors in the choice design literature focus on this scenario. Second, studying the case where $\boldsymbol{\beta} = \mathbf{0}_p$ allows us to compare the performance of our algorithm to that of commercial software.

In Figure 1, we show the utility-neutral designs generated both by our algorithm and by JMP10 using a ternary plot. We provide the exact ingredient proportions for each of the designs in Tables A.1 and A.2 in Appendix A. The vertices of the triangle in Figure 1 correspond to the three alternatives where one proportion equals 100%. The interior points represent mixtures in which none of the three ingredients is absent, i.e., where x_1, x_2 and x_3 are all strictly positive.

The centroid corresponds to the mixture with equal proportions for each of the three ingredients (Cornell, 2002). Each symbol in the ternary plot corresponds to a different choice set. As certain mixtures in Figure 1 appear in more than one choice set, symbols may overlap. In total, there are seven symbols in Figure 1, each of which occurs twice, as we consider designs with seven choice sets of two alternatives.

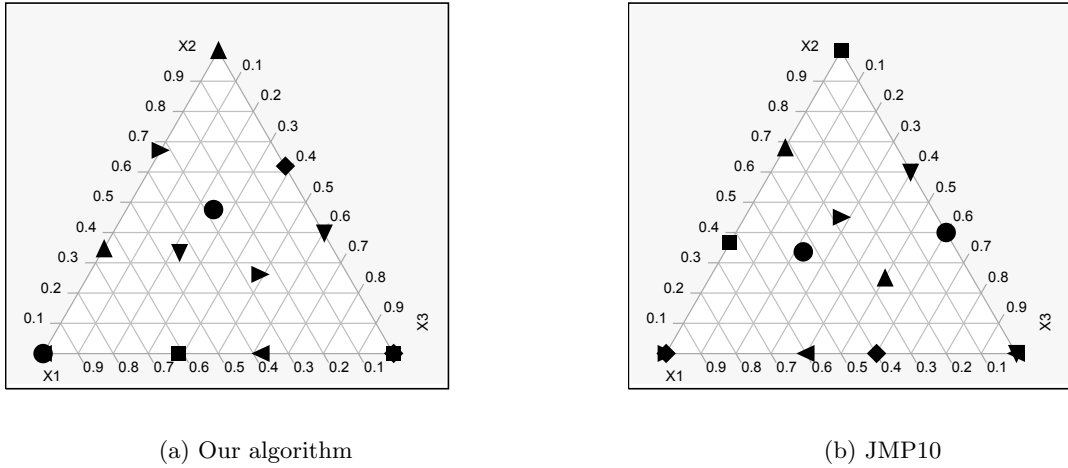(a) Our algorithm   (b) JMP10

Figure 1: Utility-neutral designs

The locations of the mixtures in the ternary plots for the two designs in Figure 1 are remarkably similar, as is the pairwise grouping. The D-optimality criterion value equals 2.9397 for the design constructed using the algorithm presented in Section 4 and 2.9410 for the design constructed using JMP10. Thus, our design performs slightly better and we conclude that our algorithm works well in this scenario. Evaluating the performance of our algorithm in several other utility-neutral scenarios yields similar results.

In case $\boldsymbol{\beta} = \mathbf{0}_p$, then, for any given design, it is possible to construct $q!$ different, but equivalent designs by permuting the labels of all ingredients. For instance, for the design problem in Figure 1, it is possible to construct $3! = 6$ different designs, all of which have the same D-optimality criterion value when $\boldsymbol{\beta} = \mathbf{0}_p$, by permuting the labels x_1, x_2 and x_3. In the next sections, we will, however, see that the $q!$ different utility-neutral designs may have quite different performances in case $\boldsymbol{\beta} \neq \mathbf{0}_p$.

5.2   Locally optimal design approach

Using parameter values equal to zero might seem attractive, as it is convenient and helps to reduce the computation time required to generate an optimal design. However, the corresponding preference assumption is unrealistic, since it is hard to believe that respondents have equal preferences over alternatives. For instance, consumers generally prefer low prices to high ones and certain brands are consistently preferred over others. In a mixture context, not all ingredients might be equally important to the respondents. For example, if we consider a fruit punch consisting of

13

watermelon, pineapple and orange juice concentrates, respondents might prefer mixtures which contain more of the orange juice concentrate and less of the other two ingredients.

In this section, we demonstrate how we use this kind of consumer preference information to construct locally optimal experimental designs assuming a particular value for $\boldsymbol{\beta}$. We focus on locally optimal experimental designs and investigate the impact of changing the shape and range of the implied utility function. The shape of the utility function determines the relative importance that respondents attribute to certain ingredients, or combinations of ingredients, comprising a mixture. The range influences the magnitude of the odds ratios within a choice set.

5.2.1 A first example

In order to use realistic parameter values, we analyzed the data from a three-ingredient mixture experiment described in Cornell (2002). The experiment was performed to determine whether an artificial sweetener could be used in a popular athletic-sports drink. The three sweeteners considered are glycine (x_1), saccharin (x_2) and an enhancer (x_3). The response of interest was "intensity of sweetness aftertaste". The special-cubic model estimated based on rank-order data from 380 individuals is given by

$$\hat{y} = 11.25x_1 + 5.54x_2 + 3.73x_3 + 26.93x_1x_2 + 20.52x_1x_3 + 28.44x_2x_3 - 180.68x_1x_2x_3.$$

We interpret the intensity of sweetness aftertaste as being proportional to the utility and use the above estimates as input for the parameters in the MNL choice model. In the utility model, we need to account for identification restrictions, therefore, the assumed utility model becomes

$$\hat{u} = (11.25 - 3.73)x_1 + (5.54 - 3.73)x_2 + 26.93x_1x_2 + 20.52x_1x_3 + 28.44x_2x_3 - 180.68x_1x_2x_3,$$
$$= 7.52x_1 + 1.81x_2 + 26.93x_1x_2 + 20.52x_1x_3 + 28.44x_2x_3 - 180.68x_1x_2x_3.$$

We refer to the parameters in this model as $\tilde{\boldsymbol{\beta}}$.

The locally optimal design for this parameter setting is shown in Figure 2a and given in Table A.3 of Appendix A. For comparison, we present the corresponding utility-neutral design in Figure 2b. The contour lines in the plot represent the utility function, where the color roughly represents the utility value for a particular mixture.

A key feature of the design presented in Figure 2a is that, within a choice set, the two alternatives have similar utilities. Hence, in the optimal choice design we ask respondents to compare mixtures that have similar a-priori utilities. The locally optimal design is, therefore, approximately utility balanced. Figure 2b clearly shows that the utility-neutral design involves only one choice set where mixtures have similar utilities. The utility-neutral design is, therefore, certainly not utility balanced.

5.2.2 Changing the scale of utility

For the parameter values $\tilde{\boldsymbol{\beta}}$, the utility ranges from 0 to 11.70, which results in a relatively extreme odds ratio of $\exp(11.70)/\exp(0) > 1.2 \times 10^5$ for choosing the best available alternative
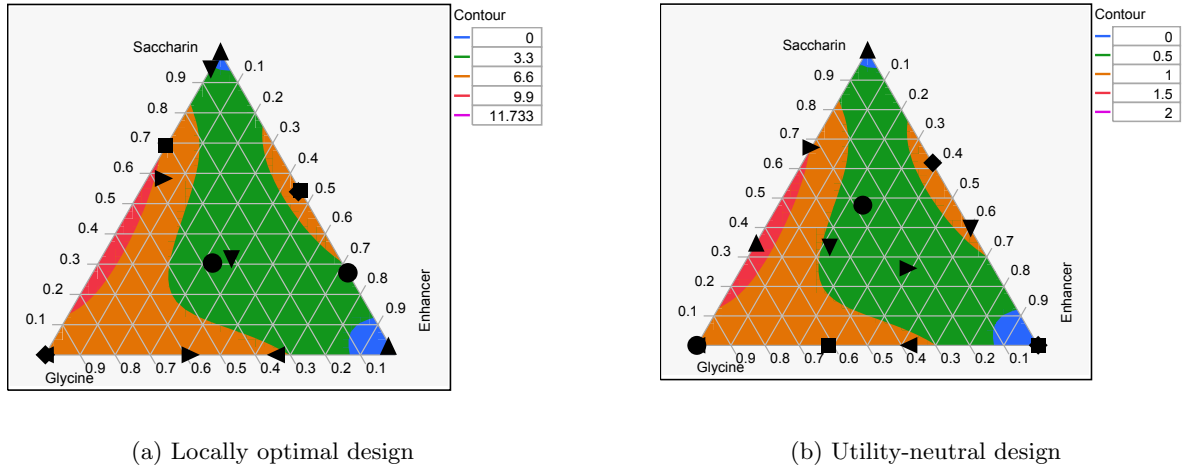
(a) Locally optimal design   (b) Utility-neutral design

Figure 2: Locally optimal design for parameter vector $\tilde{\boldsymbol{\beta}}$ versus the utility-neutral design for the intensity of sweetness aftertaste example

compared to the worst. It is interesting to investigate the impact of reducing this odds ratio, where we make sure not to change the shape of the utility function. We first rescale the parameter values such that the maximum odds ratio decreases to 6. Next, we further rescale the parameters to make the maximum odds ratio decrease to 3.8. In these two new scenarios, the utility values are in the interval $[0, 1.79]$ and $[0, 1.33]$, respectively. The rescaled parameters are $\boldsymbol{\beta}_1 = (1.15, 0.28, 4.12, 3.14, 4.36, -27.67)'$ and $\boldsymbol{\beta}_2 = (0.86, 0.21, 3.07, 2.34, 3.24, -20.59)'$.

We graphically show the locally optimal designs for these two scenarios in Figure 3 and we provide the designs in tabular format in Tables A.4 and A.5 in Appendix A. Note that when the utility range narrows, we approach the utility-neutral case, for which the utility range is $[0, 0]$ and the parameter vector equals the zero vector.

Figure 3 has contour lines with the same shape as those in Figure 2, but, as the legend shows, the utility values corresponding to the contours differ markedly. The designs in Figure 3a and 3b differ substantially from the one in Figure 2a, showing that the a-priori information provided has a substantial impact on the design. As expected, the design constructed for the narrowest utility range, i.e., the design in Figure 3b is most similar to the utility-neutral design in Figure 2b.

5.2.3   Comparison to all utility-neutral designs in terms of D-efficiency

The optimal solution to the utility-neutral design problem is not unique. Under the utility-neutral assumption, all column permutations of any given utility-neutral design matrix result in a design with the same D-value as the original. In Appendix B, we list all permutations of the utility-neutral design in Figure 1a.

However, the permuted designs are not equivalent when evaluated under a non-zero parameter vector. Since the choice of the permutation to eventually use in an experiment is arbitrary, all permutations should be considered when comparing the locally optimal design approach to the
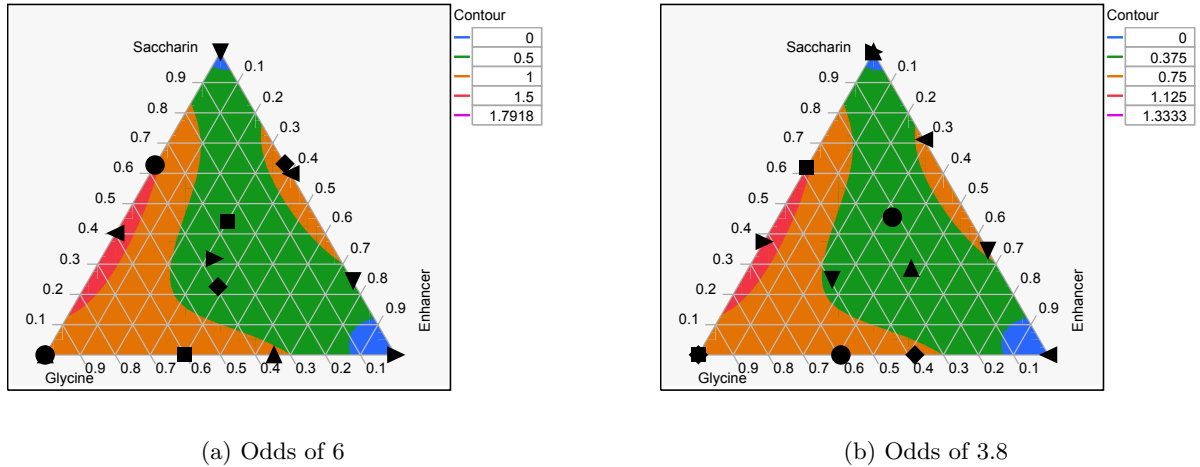
(a) Odds of 6   (b) Odds of 3.8

Figure 3: Locally optimal experimental designs for rescaled parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in the intensity of sweetness aftertaste example

utility-neutral approach. For each non-zero parameter vector we considered above, i.e., $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, we identify the best possible utility-neutral design (having the smallest D-optimality criterion value) as well as the worst utility-neutral design (having the largest D-optimality criterion value). For each of the three parameter vectors, the best utility-neutral design turns out to be second permutation listed in Table B.2. For prior parameter vector $\tilde{\boldsymbol{\beta}}$, the third permutation (see Table B.3) is worst, whereas, for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, it is the first permutation (Table B.1) which performs worst.

The D-optimality criterion values for the locally optimal designs and the best and worst utility-neutral designs are given in Table 1. The difference in D-optimality values is largest for the original parameter vector $\tilde{\boldsymbol{\beta}}$. In order to quantify the performance of the best and worst utility-neutral designs relative to the locally optimal design, we compute the local D-efficiencies $(\det(\mathbf{I}(\mathbf{X}_{\text{best}}, \boldsymbol{\beta})) / \det(\mathbf{I}(\mathbf{X}_{\text{loc.opt.}}, \boldsymbol{\beta})))^{-1/p}$ and $(\det(\mathbf{I}(\mathbf{X}_{\text{worst}}, \boldsymbol{\beta})) / \det(\mathbf{I}(\mathbf{X}_{\text{loc.opt.}}, \boldsymbol{\beta})))^{-1/p}$, for $\boldsymbol{\beta}$ equal to $\tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_1$, or $\boldsymbol{\beta}_2$, where $\mathbf{X}_{\text{best}}$, $\mathbf{X}_{\text{worst}}$ and $\mathbf{X}_{\text{loc.opt.}}$ represent the design matrices of the best and worst utility-neutral design and of the locally optimal design. A D-efficiency of, say, 50% of one design relative to another means that the former design requires twice as many respondents as the latter to achieve the same precision of parameter estimates. The percentage difference in the required number of respondents for the former design is, thus, 100%. In Table 2, we show the D-efficiencies for the best and worst utility-neutral designs relative to the locally optimal designs for each of the parameter vectors.

Table 1 shows that the locally optimal designs perform better in all three scenarios. However, the difference in performance decreases as the prior parameter vector approaches the zero vector. For parameter vector $\tilde{\boldsymbol{\beta}}$, the locally optimal design is nearly 20% more efficient than the best and 33% more efficient than the worst utility-neutral design. The difference in efficiency drops to 0.02% and 0.5% for the best and worst utility-neutral designs, respectively, for $\boldsymbol{\beta}_2$, i.e., for the parameter vector closest to zero. In any case, our results show that the conclusion of Huber and

16

	$\tilde{\boldsymbol{\beta}}$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$
Locally optimal	3.5380	3.0474	3.0132
Best utility-neutral	4.9880	3.0704	3.0147
Worst utility-neutral	6.3447	3.1284	3.0482

Table 1: D-optimality criterion values for locally optimal and utility-neutral designs for three different prior parameter vectors

	$\tilde{\boldsymbol{\beta}}$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$
Best utility-neutral	81.29%	99.67%	99.98%
Worst utility-neutral	66.97%	98.85%	99.50%

Table 2: D-efficiencies for utility-neutral designs relative to the locally optimal design for three different prior parameter vectors

Zwerina (1996) that non-zero priors can be used to generate statistically more efficient choice designs extends from choice experiments involving categorical attributes to choice experiments involving mixtures.

5.2.4 Comparison to all utility-neutral designs in terms of utility balance

All locally optimal designs we constructed are approximately utility balanced. The alternatives within any given choice set have similar utilities and, hence, similar choice probabilities. This is not always the case for the utility-neutral designs. Formally, for a utility-balanced design with choice sets of two alternatives, the product of the two choice probabilities in any choice set equals 0.25. If the alternatives differ in utility, this product is smaller than 0.25.

Consider again the three locally optimal designs, constructed assuming the parameter vectors $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, along with the best and worst utility-neutral designs. For each of these designs, we provide in Table 3 the minimum, mean and maximum values of the product of the choice probabilities across choice sets. The table shows that the locally optimal design always scores very well in terms of utility balance, whereas the utility-neutral designs exhibit a complete lack of utility balance in various choice sets under $\tilde{\boldsymbol{\beta}}$. Each of the possible utility-neutral designs has at least one choice set for which the product of the choice probabilities is zero in that case and the average of the choice probabilities' product is at most 0.08. Under the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ (which are closer to the zero vector), the lack of utility balance is less extreme for the utility-neutral designs. Nevertheless, under $\boldsymbol{\beta}_1$, the utility-neutral designs have choice sets for which the choice probabilities' product equals 0.15, in which case the choice probabilities are as unequal as 0.2 and 0.8. None of the three locally optimal designs results in choice probabilities that differ to this extent.

5.2.5 Changing the shape of the utility function

We next study the impact of changing the shape of the utility function, keeping the range or scale of the utilities constant. We start with the parameter vector $\boldsymbol{\beta}_1$, make the first param-

		$\tilde{\boldsymbol{\beta}}$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$
Min	Locally optimal	0.17	0.22	0.18
	Best utility-neutral	0.00	0.15	0.18
	Worst utility-neutral	0.00	0.15	0.19
Mean	Locally optimal	0.21	0.23	0.23
	Best utility-neutral	0.08	0.22	0.23
	Worst utility-neutral	0.04	0.21	0.22
Max	Locally optimal	0.24	0.25	0.25
	Best utility-neutral	0.25	0.25	0.25
	Worst utility-neutral	0.25	0.25	0.25

Table 3: Measures of utility balance for the locally optimal and utility-neutral designs

eter (corresponding to the first ingredient) ten times as large as the second parameter (corresponding to the second ingredient) and rescale the resulting parameter vector to ensure the range of the utility function remains unchanged. The resulting new parameter vector is then $\boldsymbol{\beta}_1^{\text{new}} = (1.69, 0.17, 2.52, 1.92, 2.66, -16.90)'$ and the odds for choosing the best alternative instead of the worst are still equal to 6. The corresponding locally optimal design and the contour lines corresponding to the new utility function are shown in Figure 4. The design is given in tabular format in Table A.6 in Appendix A.
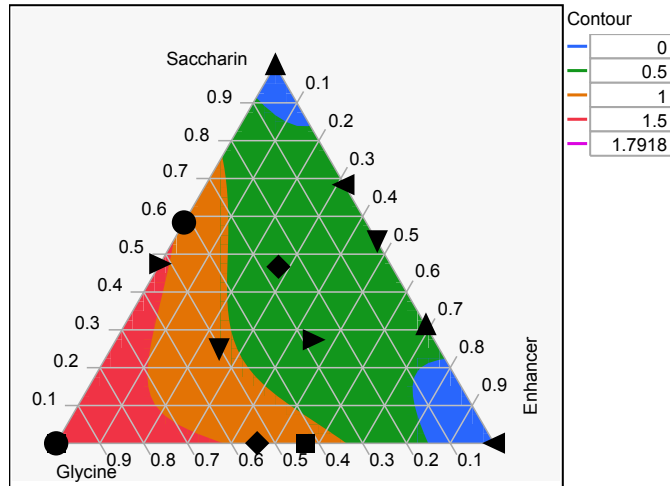

Figure 4: Locally optimal design for a different shape of the utility function, corresponding to parameter vector $\boldsymbol{\beta}_1^{\text{new}}$

It is interesting to investigate the performance of the new design in Figure 4 in case the true model parameter vector differs from $\boldsymbol{\beta}_1^{\text{new}}$. To this end, we compute its D-efficiency relative to the locally optimal designs in Figures 2a, 3a and 3b. Since the D-optimality criterion values of the new design under $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ equal 5.3156, 3.0883 and 3.0391, respectively, the relative D-efficiencies equal 77.57%, 99.42% and 99.63%. We observe that misspecifying the prior parameter vector can have a drastic impact on the performance of the design.

To reduce the dependence of the design quality on the set value of $\boldsymbol{\beta}$, we advocate using a

non-degenerate prior distribution which incorporates the uncertainty about the unknown model parameters. The performance of the resulting designs will be more stable across different values of the model parameter vector. We consider such designs next.

5.3 Bayesian approach

Bayesian D-optimal designs, constructed using a proper prior distribution that incorporates the a-priori uncertainty about the model parameter vector, are known to perform well under a broad range of scenarios. In this section, we first investigate the impact of the uncertainty, using independent prior distributions for the parameters. Next, we show how to deal with a multivariate prior distribution involving correlations between the different parameters. We do so using a practical example taken from the literature.

5.3.1 The impact of the parameter uncertainty

As prior distribution, we choose a normal distribution with prior mean equal to $\boldsymbol{\beta}_2$ and the variance-covariance matrix as given in Equation (11). The a-priori uncertainty is controlled by the parameter κ. We use κ values of 0.5, 5, 10 and 30 in this section. We show the Bayesian D-optimal designs corresponding to the different κ values in Figure 5 and provide their design points in tabular format in Tables A.7 to A.10 in Appendix A. The contours in Figure 5 correspond to the mean parameter vector ($\boldsymbol{\beta}_2$). When the uncertainty increases, the design becomes less symmetric and the design points tend to be spread over the entire simplex. We also find that the design for $\kappa = 0.5$ is almost identical to the locally optimal design in Figure 3b, which implicitly assumes a κ value of zero.

A remarkable feature of the Bayesian D-optimal designs is that, for larger values of κ, the mixtures within a choice set tend to be close to each other. To investigate this issue, we computed the minimum, mean and maximum distances between the two alternatives within a choice set for every design. The results are shown in Table 4. It is clear that the distances decrease substantially when the prior uncertainty, as measured by κ, is increased. The distances for the Bayesian D-optimal design when $\kappa = 30$ are only half as large as for the locally optimal design, for which $\kappa = 0$. The explanation for this phenomenon lies in the fact that a prior distribution with a large variance results in parameter draws $\boldsymbol{\beta}_{(i)}$ with large absolute values when evaluating the Bayesian D-optimality criterion using Equation (12). If the absolute values of the parameters in the utility function are large, then substantially different alternatives within a choice set have substantially different utilities, which results in a complete lack of utility balance and extreme choice probabilities. The Bayesian D-optimal design prevents this from happening by making the alternatives within a choice set resemble each other, which guarantees similar utilities and a certain degree of utility balance.

To show that the Bayesian D-optimal designs indeed exhibit approximate utility balance, we provide the minimum, mean and maximum values of the product of choice probabilities in a choice set in Table 5, along with the corresponding values for the locally optimal design for $\boldsymbol{\beta}_2$ (for which $\kappa = 0$). The mean values are all very close to 0.25, while the maximum values are

19
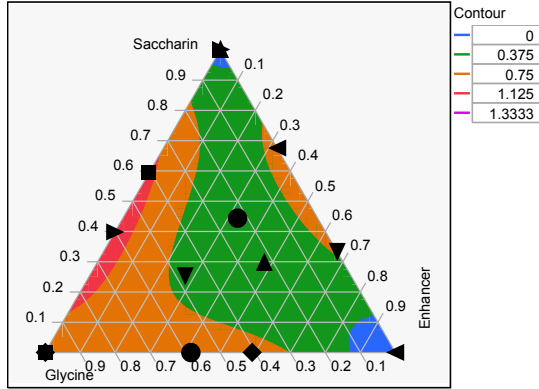
(a) $\kappa = 0.5$
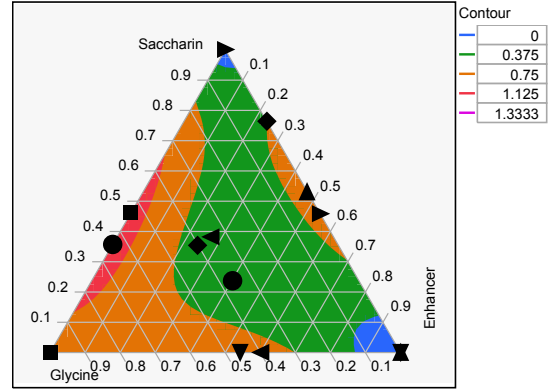
(b) $\kappa = 5$
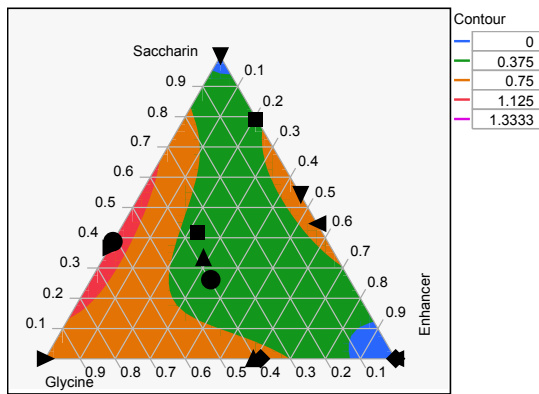
(c) $\kappa = 10$
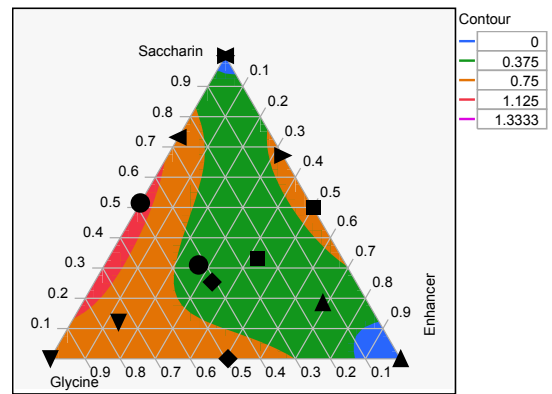
(d) $\kappa = 30$

Figure 5: Bayesian optimal designs for $\boldsymbol{\beta}_2$ and different levels of uncertainty

	Locally Optimal	Bayesian D-Optimal			
	$\kappa = 0$	$\kappa = 0.5$	$\kappa = 5$	$\kappa = 10$	$\kappa = 30$
min	0.59	0.58	0.51	0.42	0.31
mean	0.82	0.79	0.63	0.53	0.36
max	1.00	0.96	0.76	0.65	0.46

Table 4: Distances between alternatives within a choice set for the locally optimal design and its Bayesian counterparts

	Locally Optimal	Bayesian D-Optimal			
	$\kappa = 0$	$\kappa = 0.5$	$\kappa = 5$	$\kappa = 10$	$\kappa = 30$
min	0.18	0.18	0.20	0.21	0.22
mean	0.23	0.23	0.22	0.23	0.23
max	0.25	0.25	0.25	0.25	0.25

Table 5: Utility balance measures for the designs with uncertainty

0.25 for all the designs. Because the locally optimal design and the Bayesian D-optimal design for $\kappa = 0.5$ are very similar, the utility-balance measures for the two designs are identical.

### 5.3.2   An example with dependent prior distributions

As a final example, we consider a case where the prior distribution for the utility parameters has non-zero correlations. In order to obtain realistic values for the prior mean and covariance, we base our prior on an experiment for quantifying preferences for cocktails described in Courcoux and Séménou (1997). The cocktails were tasted by a panel of consumers and involved three ingredients: mango juice (x_1), blackcurrant syrup (x_2) and lemon juice (x_3).

Since Courcoux and Séménou (1997) imposed lower bounds of 0.3, 0.15 and 0.1 on the three ingredient proportions, we redefine the coordinates of the allowed subregion of the simplex in terms of pseudocomponents (see Section 3.1). We use a mixed logit model (Train, 2009) to obtain estimates of the parameter means and variances, starting from the choice probabilities reported in Courcoux and Séménou (1997) and assuming a special-cubic Scheffé model. The preference distribution we obtain is $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta_0}, \Sigma_0)$, with mean vector $\boldsymbol{\beta_0} = (1.36, 1.57, 2.47, -0.43, 0.50, 1.09)'$ and variance-covariance matrix

$$\Sigma_0 = \begin{pmatrix} 6.14 & 5.00 & 2.74 & -0.43 & -2.81 & -3.33 \\ 5.00 & 6.76 & 4.47 & -1.79 & -6.13 & -3.51 \\ 2.74 & 4.47 & 3.45 & -1.38 & -4.71 & -2.17 \\ -0.43 & -1.79 & -1.38 & 1.18 & 2.39 & 0.71 \\ -2.81 & -6.13 & -4.71 & 2.39 & 7.43 & 2.71 \\ -3.33 & -3.51 & -2.17 & 0.71 & 2.71 & 2.49 \end{pmatrix}.$$

We next use this distribution as prior distribution for computing a Bayesian D-optimal design with seven choice sets of two alternatives each. The design is presented in two ways in Figure 6, once in terms of the true ingredients (left) and once in terms of the pseudocomponents (right). The figure also shows the contours of the expected utility function. The optimal design in terms of the pseudocomponents is given in tabular format in Table A.11 in Appendix A.

We compare the Bayesian design's performance to that of the utility-neutral designs that result in the best and worst Bayesian D-optimality criterion values. Interestingly, the worst utility-neutral design is obtained by a clock-wise rotation of the best utility-neutral design. The best and worst utility-neutral designs are shown in Figure 7. These designs, expressed in terms of the pseudocomponents, also appear in Tables B.2 and B.3 in Appendix B. We also compare the performance of the Bayesian design to the locally optimal design which we constructed for
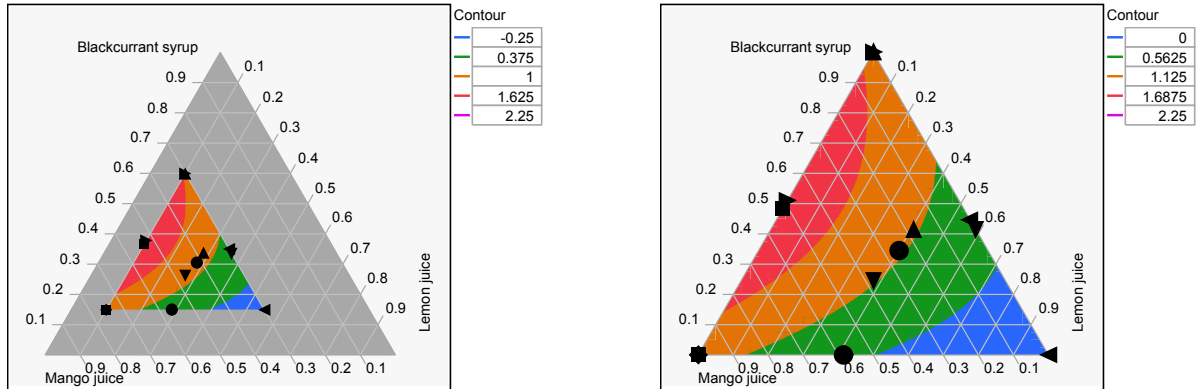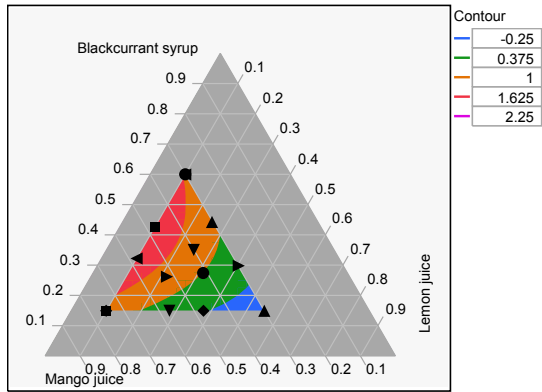
(a) Design in the original components   (b) Design in the pseudocomponents

Figure 6: Bayesian optimal design for the experiment of Courcoux and Séménou (1997)

$\boldsymbol{\beta}_0$ assuming no uncertainty. The locally optimal design in the original as well as the pseudocomponents is depicted in Figure 8 and tabulated in terms of the pseudocomponents in Table A.12 in Appendix A.

The utility-neutral designs cover the design region more uniformly than the locally optimal design, which in turn covers the design region more uniformly than the Bayesian design. One result of this is that the Bayesian design is approximately utility balanced. None of the choice sets involves alternatives that are more than one color zone apart in the graphical representation in Figure 6. This is not the case for the two utility-neutral designs in Figure 7, where, for some of the choice sets, the alternatives are two color zones apart. There is also one choice set for the locally optimal design in Figure 8 in which the alternatives are two color zones apart. To further quantify the utility balance, in the Bayesian design, the mean of the product of the choice probabilities of the different choice sets is 0.23, while its maximum and minimum are equal to 0.24 and 0.21, respectively. The mean, maximum and minimum products equal 0.22, 0.24, 0.18, respectively, for the best utility-neutral design and 0.20, 0.24, 0.16 for the worst utility-neutral design. For the locally optimal design, the same measures are 0.23, 0.25 and 0.21. In sum, the utility-neutral designs score less well in terms of utility balance under $\boldsymbol{\beta}_0$, while the locally optimal design scores a bit better.

We next study how the Bayesian D-optimal design performs in comparison to the utility-neutral and locally optimal designs for different potential true parameter values. To this end, we randomly draw 10,000 parameter vectors from the normal prior distribution with mean $\boldsymbol{\beta}_0$ and variance-covariance matrix Σ_0 and, for each draw, calculate D-optimality criterion values for all four designs. We visualize the resulting D-optimality criterion values in three ways. First, we plot their distribution for each of the four designs. Second, we create histograms of the D-optimality criterion values of the Bayesian D-optimal design relative to those of the two utility-neutral and the locally optimal designs. Finally, we produce scatter plots of the D-optimality criterion values for the Bayesian design versus those of the utility-neutral and the locally optimal designs.

(a) Best utility-neutral design in terms of the original components
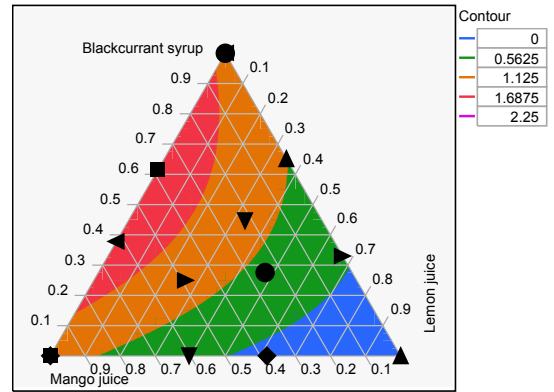
(b) Best utility-neutral design in terms of the pseudocomponents
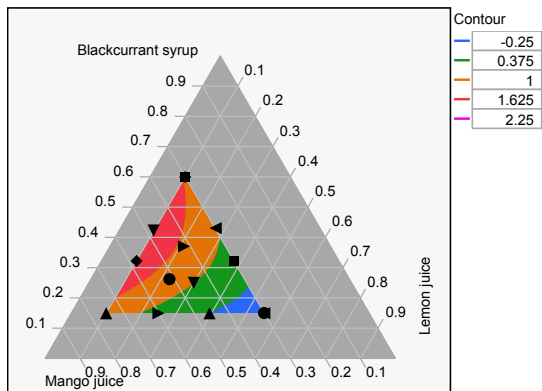
(c) Worst utility-neutral design in terms of the original components

(d) Worst utility-neutral design in terms of the pseudocomponents

Figure 7: Utility-neutral designs for the experiment of Courcoux and Séménou (1997)

<table>
<tr><td>(a) Design in the original components</td><td>(b) Design in the pseudocomponents</td></tr>
</table>

Figure 8: Locally optimal design for the experiment of Courcoux and Séménou (1997)

Figure 9 shows the estimated densities of the D-optimality criterion values for the Bayesian, the two utility-neutral and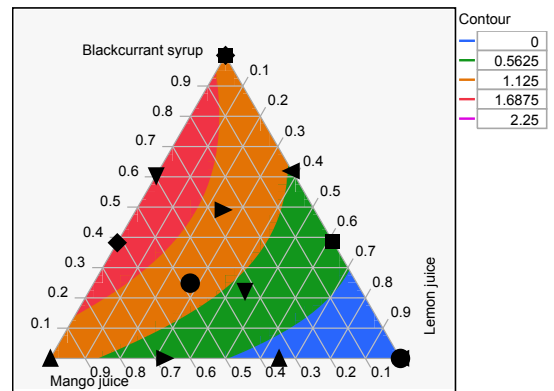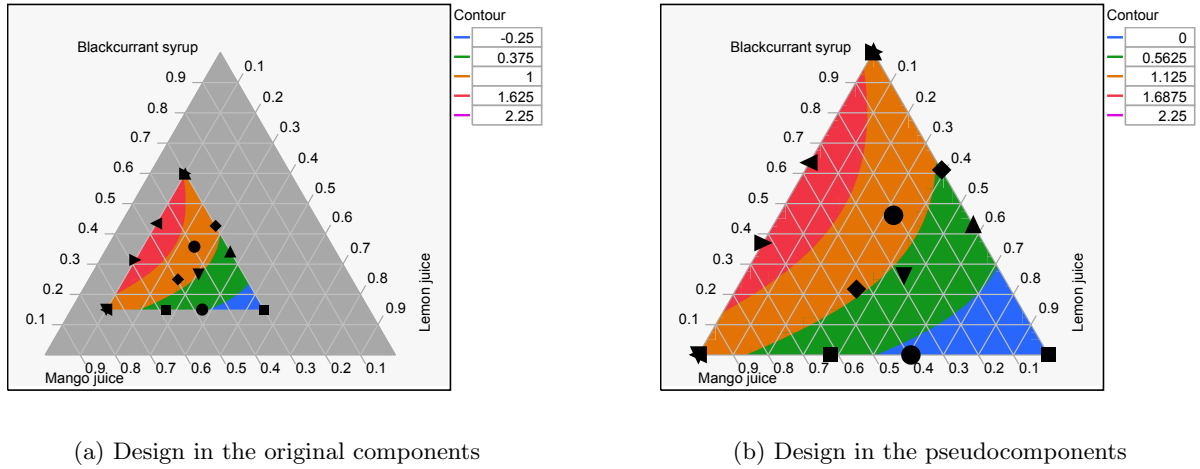 the locally optimal design. The densities for the locally optimal and both utility-neutral designs have substantially fatter and longer right tails compared to the Bayesian design. This indicates that the locally optimal and the utility-neutral designs can result in much higher D-optimality criterion values. The densities also show that, for some parameter values, the locally optimal and the utility-neutral designs sometimes yield smaller D-optimality values than the Bayesian design. However, these cases do not compensate the impact of the long and fat right tails. This is reflected in the means of the four distributions, which equal 3.47, 3.68, 3.58 and 3.56 for the Bayesian, worst and best utility-neutral and locally optimal designs, respectively. The locally optimal design performs better than both utility-neutral designs, in the sense that the distribution's right tail is lighter and its mean D-optimality criterion value is smaller than the corresponding means of the utility-neutral designs.

Figure 10 shows the ratios of the D-optimality criterion values of the Bayesian design to the D-optimality criterion values of the utility-neutral and the locally optimal designs computed for the 10,000 draws. We clearly see that the histograms are left skewed, meaning that there are many cases where the ratio is substantially smaller than one. In other words, the utility-neutral and the locally optimal designs are often substantially worse than the Bayesian design. This is especially so for the worst utility-neutral design. These histograms also show that there are also scenarios in which the locally optimal or the utility-neutral designs outperform the Bayesian design, but that the extent to which this happens is almost never larger than 5%. Moreover, these scenarios are not as numerous as the scenarios where the locally optimal or the utility-neutral designs are inferior. The means of the ratios are 0.97 and 0.95 for the best and worst utility-neutral designs, respectively, and 0.98 for the locally optimal design.

Figure 11 shows scatter plots for (3,000 out of 10,000) D-optimality criterion values computed for the four designs. In each plot, we compare the Bayesian design to an alternative design. Each plot shows the D-optimality criterion value for the alternative design on the vertical axis and

24
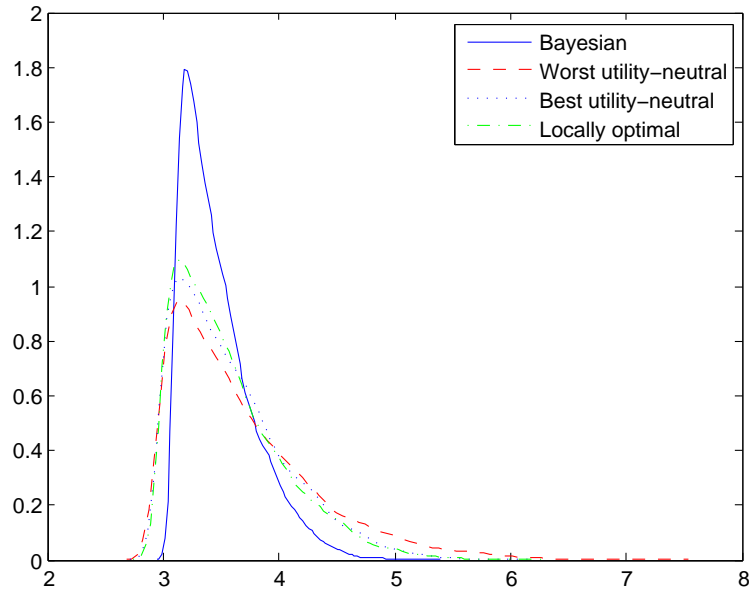
Figure 9: Distribution of D-optimality criterion values for Bayesian, utility-neutral and locally optimal designs

the criterion value for the Bayesian design on the horizontal axis. Every point above the 45-degree line corresponds to a parameter vector for which the Bayesian design produces a smaller D-optimality criterion value than the locally optimal or utility-neutral design. Hence, the larger the number of points appearing above the 45-degree line, the better the relative performance of the Bayesian design. The large green circle denotes the scenario corresponding to the zero parameter vector. The scatter plots show that the Bayesian design is substantially more robust than the locally optimal or utility-neutral design. As a matter of fact, whenever a point is below the 45-degree line, it is never very far below. Conversely, some of the points are well above the line. The Bayesian design outperforms the locally optimal design in 58% of the cases, the best utility-neutral design in 68% of the cases and the worst utility-neutral design in 60% of the cases. Again, the locally optimal design performs worse than the Bayesian design but better than both utility-neutral designs.
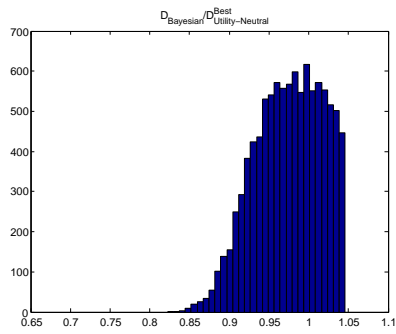
(a) Best utility-neutral design
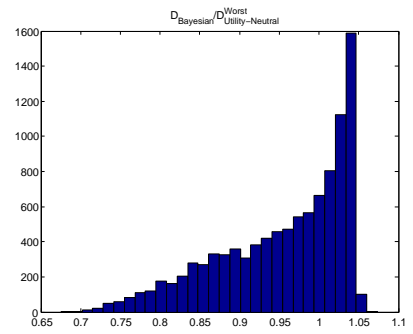
(b) Worst utility-neutral design
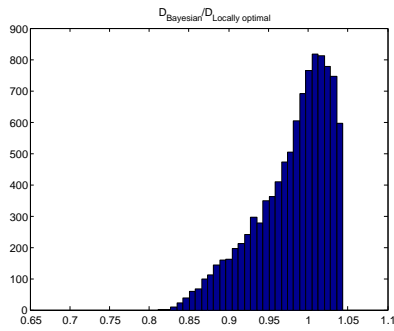
(c) Locally optimal design

Figure 10: Histogram of ratios of the D-optimality criterion values of the Bayesian design to the D-optimality criterion values of the utility-neutral and the locally optimal designs
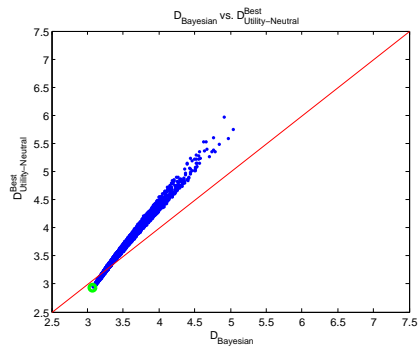
(a) Best utility-neutral design
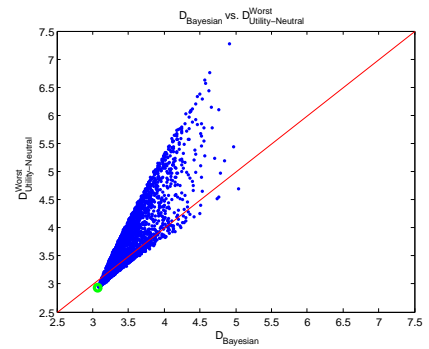
(b) Worst utility-neutral design
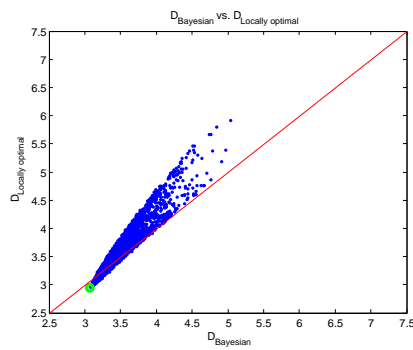
(c) Locally optimal design

Figure 11: Scatter plots of D-optimality criterion values for the Bayesian design and the utility-neutral or locally optimal designs

6 Conclusion

In this paper, we studied the problem of constructing D-optimal experimental designs for the multinomial logit model when the alternatives are mixtures of ingredients. We developed a mixture coordinate-exchange algorithm that is capable of constructing designs for any number of ingredients and any type of Scheffé model. Because of the nonlinearity of the multinomial logit model, we adopt a Bayesian approach involving a prior distribution for the model parameters.

We showed that the Bayesian and locally D-optimal designs differ both in look and in performance from the utility-neutral ones. Bayesian designs are robust in the sense that they perform well in a broad range of scenarios and that they never seem to perform very poorly. The Bayesian and locally D-optimal designs we construct are close to being utility balanced, which is considered a desirable feature in the literature on choice experiments.

Like all algorithms for constructing Bayesian optimal choice designs, our algorithm is a heuristic. Therefore, our algorithm might not yield the globally optimal design. The extensive use of coordinate-exchange algorithms in the optimal experimental design literature, however, indicates that the resulting designs are highly efficient and practically useful. Another issue worth mentioning is that, while we focused on the D-optimality criterion, our algorithm can be used for generating designs that are optimal with respect to other criteria as well, such as the A-, G- and V-optimality criteria.

Finally, it is our hope that our work will have a beneficial impact on the use of choice experiments for mixtures. The examples in the textbook by Cornell (2002) suggest that ranking data are very commonly used, but it is known that ranking large numbers of products is tough and usually requires many pairwise comparisons. It is, therefore, better to use choice sets of size two (i.e., paired comparisons) or three and ask the respondents to pick the mixture of their choice in each set (David, 1963).

Appendices

A   Experimental Designs

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	0.62	0.38
0.28	0.48	0.25	0.00	0.00	1.00
0.65	0.35	0.00	0.38	0.00	0.62
0.00	1.00	0.00	1.00	0.00	0.00
0.00	0.40	0.60	0.25	0.26	0.49
0.45	0.33	0.22	0.33	0.67	0.00
0.61	0.00	0.39			
0.00	0.00	1.00			

Table A.1: Optimal utility-neutral design with 14 alternatives in 7 choice sets of size 2, produced by our algorithm, for the special-cubic model for 3 ingredients. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.40	0.60	0.40	0.00	0.60
0.44	0.34	0.22	1.00	0.00	0.00
0.32	0.68	0.00	0.60	0.00	0.40
0.25	0.25	0.50	0.00	0.00	1.00
0.00	0.60	0.40	1.00	0.00	0.00
0.00	0.00	1.00	0.27	0.45	0.27
0.63	0.37	0.00			
0.00	1.00	0.00			

Table A.2: Optimal utility-neutral design with 14 alternatives in 7 choice sets of size 2, produced by JMP10, for the special-cubic model for 3 ingredients. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.27	0.73	0.01	0.54	0.45
0.37	0.30	0.33	1.00	0.00	0.00
0.00	0.03	0.96	0.34	0.00	0.66
0.00	1.00	0.00	1.00	0.00	0.00
0.31	0.32	0.37	0.37	0.58	0.05
0.06	0.94	0.00	0.59	0.00	0.41
0.31	0.69	0.00			
0.00	0.54	0.46			

Table A.3: Locally optimal design with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the original parameter vector $\tilde{\beta}$ in the artificial sweetener example in Cornell (2002). Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	0.63	0.37
0.37	0.63	0.00	0.39	0.22	0.38
1.00	0.00	0.00	0.60	0.40	0.00
0.35	0.00	0.65	0.00	0.60	0.40
0.00	0.24	0.76	0.00	0.00	1.00
0.00	1.00	0.00	0.36	0.32	0.33
0.26	0.44	0.30			
0.60	0.00	0.40			

Table A.4: Locally optimal design with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_1$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.22	0.46	0.33	1.00	0.00	0.00
0.59	0.00	0.41	0.38	0.00	0.62
0.25	0.29	0.47	0.00	0.71	0.29
0.00	1.00	0.00	0.00	0.00	1.00
0.00	0.35	0.65	0.00	1.00	0.00
0.49	0.25	0.26	0.63	0.37	0.00
1.00	0.00	0.00			
0.38	0.62	0.00			

Table A.5: Locally optimal design with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_2$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.26	0.47	0.27
0.42	0.58	0.00	0.54	0.00	0.46
0.00	1.00	0.00	0.00	0.69	0.31
0.00	0.31	0.69	0.00	0.00	1.00
0.50	0.25	0.25	0.53	0.47	0.00
0.00	0.53	0.47	0.27	0.28	0.45
0.43	0.00	0.57			
1.00	0.00	0.00			

Table A.6: Locally optimal design with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the original parameter vector $\boldsymbol{\beta}_1^{\text{new}}$. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.23	0.44	0.33	1.00	0.00	0.00
0.58	0.00	0.42	0.41	0.00	0.59
0.23	0.30	0.48	0.00	0.68	0.32
0.00	1.00	0.00	0.00	0.00	1.00
0.00	0.33	0.67	0.00	1.00	0.00
0.47	0.26	0.27	0.60	0.40	0.00
1.00	0.00	0.00			
0.41	0.59	0.00			

Table A.7: Bayesian optimal design with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_2$ and a κ value of 0.5. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.36	0.24	0.40	0.40	0.35	0.24
0.64	0.36	0.00	0.00	0.76	0.24
0.00	0.00	1.00	0.40	0.00	0.60
0.00	0.53	0.47	0.35	0.38	0.27
0.00	0.00	1.00	0.00	1.00	0.00
0.46	0.00	0.54	0.00	0.46	0.54
1.00	0.00	0.00			
0.54	0.46	0.00			

Table A.8: Bayesian optimal design with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_2$ and a κ value of 5. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.40	0.26	0.34	0.39	0.00	0.61
0.61	0.39	0.00	0.00	0.00	1.00
0.41	0.00	0.59	0.00	0.44	0.56
0.38	0.33	0.28	0.00	0.00	1.00
0.00	0.54	0.46	0.63	0.37	0.00
0.00	1.00	0.00	1.00	0.00	0.00
0.36	0.42	0.22			
0.00	0.79	0.21			

Table A.9: Bayesian optimal design with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_2$ and a κ value of 10. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.42	0.31	0.27	0.49	0.00	0.51
0.49	0.51	0.00	0.41	0.25	0.34
0.13	0.18	0.69	0.27	0.73	0.00
0.00	0.00	1.00	0.00	1.00	0.00
0.74	0.12	0.13	0.00	0.67	0.33
1.00	0.00	0.00	0.00	1.00	0.00
0.25	0.33	0.42			
0.00	0.50	0.50			

Table A.10: Bayesian optimal design with 14 alternatives in 7 choice sets of size 2 for the special-cubic model for 3 ingredients, for the parameter vector $\boldsymbol{\beta}_2$ and a κ value of 30. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.25	0.34	0.40	1.00	0.00	0.00
0.59	0.00	0.41	0.58	0.00	0.42
0.18	0.41	0.41	0.00	0.45	0.55
0.00	1.00	0.00	0.00	0.00	1.00
0.00	0.41	0.59	0.00	1.00	0.00
0.38	0.24	0.38	0.49	0.51	0.00
1.00	0.00	0.00			
0.52	0.48	0.00			

Table A.11: Bayesian optimal design with 14 alternatives in 7 choice sets of size 2 for the example from Courcoux and Séménou (1997), expressed in terms of the pseudocomponents. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.39	0.00	0.61	0.00	0.61	0.39
0.21	0.46	0.33	0.44	0.22	0.34
0.00	1.00	0.00	1.00	0.00	0.00
0.00	0.43	0.57	0.37	0.63	0.00
1.00	0.00	0.00	0.00	1.00	0.00
0.28	0.26	0.46	0.63	0.37	0.00
0.00	0.00	1.00			
0.62	0.00	0.38			

Table A.12: Locally optimal design with 14 alternatives in 7 choice sets of size 2 for the example from Courcoux and Séménou (1997), expressed in terms of the pseudocomponents. Choice sets are separated by dashed lines

B   Permutations of the Utility-Neutral Design

In this appendix, we list all permutations of the columns of the design matrix of the utility-neutral design introduced in Section 5.1 and given in Figure 1a that lead to the same D-optimality criterion value of 2.94. As we consider an example with three ingredient proportions, the number of different permutations is equal to $3! = 6$.

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	0.62	0.38
0.28	0.48	0.25	0.00	0.00	1.00
0.65	0.35	0.00	0.38	0.00	0.62
0.00	1.00	0.00	1.00	0.00	0.00
0.00	0.40	0.60	0.25	0.26	0.49
0.45	0.33	0.22	0.33	0.67	0.00
0.61	0.00	0.39			
0.00	0.00	1.00			

Table B.1: Permutation N⁰ 1 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	1.00	0.00	0.38	0.00	0.62
0.25	0.28	0.48	1.00	0.00	0.00
0.00	0.65	0.35	0.62	0.38	0.00
0.00	0.00	1.00	0.00	1.00	0.00
0.60	0.00	0.40	0.49	0.25	0.26
0.22	0.45	0.33	0.00	0.33	0.67
0.39	0.61	0.00			
1.00	0.00	0.00			

Table B.2: Permutation N⁰ 2 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.00	1.00	0.62	0.38	0.00
0.48	0.25	0.28	0.00	1.00	0.00
0.35	0.00	0.65	0.00	0.62	0.38
1.00	0.00	0.00	0.00	0.00	1.00
0.40	0.60	0.00	0.26	0.49	0.25
0.33	0.22	0.45	0.67	0.00	0.33
0.00	0.39	0.61			
0.00	1.00	0.00			

Table B.3: Permutation N⁰ 3 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	0.00	1.00	0.38	0.62	0.00
0.25	0.48	0.28	1.00	0.00	0.00
0.00	0.35	0.65	0.62	0.00	0.38
0.00	1.00	0.00	0.00	0.00	1.00
0.60	0.40	0.00	0.49	0.26	0.25
0.22	0.33	0.45	0.00	0.67	0.33
0.39	0.00	0.61			
1.00	0.00	0.00			

Table B.4: Permutation Nº 4 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
0.00	1.00	0.00	0.62	0.00	0.38
0.48	0.28	0.25	0.00	0.00	1.00
0.35	0.65	0.00	0.00	0.38	0.62
1.00	0.00	0.00	0.00	1.00	0.00
0.40	0.00	0.60	0.26	0.25	0.49
0.33	0.45	0.22	0.67	0.33	0.00
0.00	0.61	0.39			
0.00	0.00	1.00			

Table B.5: Permutation Nº 5 of the utility neutral design. Choice sets are separated by dashed lines

x_1	x_2	x_3	x_1	x_2	x_3
1.00	0.00	0.00	0.00	0.38	0.62
0.28	0.25	0.48	0.00	1.00	0.00
0.65	0.00	0.35	0.38	0.62	0.00
0.00	0.00	1.00	1.00	0.00	0.00
0.00	0.60	0.40	0.25	0.49	0.26
0.45	0.22	0.33	0.33	0.00	0.67
0.61	0.39	0.00			
0.00	1.00	0.00			

Table B.6: Permutation Nº 6 of the utility neutral design. Choice sets are separated by dashed lines

C Pseudo Code

Notation:

$\mathbf{X} :=$ the design matrix,
$n :=$ the number of alternatives in the experimental design,
$k :=$ the number of ingredient proportions,

$p :=$ the number of the parameters in the model,

$S :=$ the number of choice sets,

$\mathbf{S} :=$ the set of starting designs,

$\mathbf{D}^S :=$ the set of D-values of starting designs in $\mathbf{S}$,

$\mathcal{S} :=$ a particular starting design with the D-value D^S,

$\mathbf{O} :=$ the set of the constructed optimal designs for each of the starting design,

$\mathbf{D}^{Opt} :=$ the set of D-values of starting designs in $\mathbf{O}$,

$\mathcal{O} :=$ a particular optimal experimental design with the D-value D^{Opt}.

Then, the algorithm for constructing the D-optimal experimental design can be described as follows:

> **parfor** $i \leftarrow 1$, starts **do**
> $\mathbf{S}(i) \leftarrow$ OBTAINSTARTINGDESIGN(n, k, p, S)
> $[\mathbf{O}(i), \mathbf{D}^{Opt}(i)] \leftarrow$ OPTIMIZESTARTINGDESIGN$(\mathbf{S}(i))$
> **end parfor**
> $i \leftarrow \arg\min_k \mathbf{D}^{Opt}(k)$
> **return** $\mathbf{O}(i)$

> **procedure** OBTAINSTARTINGDESIGN(n, k, p, S)
>   1. Choose p distinct design points that maximize the information matrix for a minimum support design with p observations (see Section 4.1 for a discussion)
>   2. Replicate them as evenly as possible to obtain n observations for the design matrix $\mathbf{X}$
>   3. Spread the replicated design points as evenly as possible over the choice sets. Avoid replicating points within a choice set
>   4. Assign the remaining non-replicated points to choice sets.
> **end procedure**

> **procedure** OPTIMIZESTARTINGDESIGN$(\mathcal{S})$
> $D^S \leftarrow$ D$(\mathcal{S})$
> $D^{\text{Init}} \leftarrow 10 + D^S$
> $D^{\text{Opt}} \leftarrow D^{\mathcal{S}}$
> $\mathbf{X} \leftarrow \mathcal{S}$
>   **while** abs$(D^{\text{Init}} - D^{\text{Opt}}) > 0.0001$ **do**
> $D^{\text{Init}} \leftarrow D^{\text{Opt}}$
>   **for** $i \leftarrow 1,\ n$ **do**
>   **for** $j \leftarrow 1,\ k$ **do**
> $\mathbf{X}^* \leftarrow$ BRENT$(\mathbf{X}, i, j)$
> $D^{\text{New}} \leftarrow$ D$(\mathbf{X}^*)$
>   **if** $D^{\text{New}} < D^{\text{Opt}}$ **then**
> $D^{\text{Opt}} \leftarrow D^{\text{New}}$

$$\mathbf{X} \leftarrow \mathbf{X}^*$$
               end if
            end for
         end for
      end while
      return $\mathbf{X}, D^{\mathrm{Opt}}$
  end procedure

  procedure $\mathrm{D}(\cdot)$
      Compute the D-value (see Equation (8))
  end procedure

  procedure $\textsc{Brent}(\mathbf{X}, i, j)$
      Optimize a coordinate (i, j) of a design matrix $\mathbf{X}$ by using the method of Brent (1973) (see Section 4.2)
      return optimal design $\mathbf{X}^*$
  end procedure

References

A. Agresti. *Categorical Data Analysis*. John Wiley & Sons, Inc, 2002.

A. Atkinson, A. Donev, and R. Tobias. *Optimum Experimental Designs, with SAS*. Oxford Statistical Science Series, 2007.

A. C. Atkinson and A. N. Donev. The construction of exact D-optimum experimental designs with application to blocking response surface designs. *Biometrika*, 76:515–526, 1989.

M. C. J. Bliemer and J. M. Rose. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, 44:720–734, 2010.

M. C. J. Bliemer, J. M. Rose, and S. Hass. Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modelling*, 1:98–127, 2008.

M. C. J. Bliemer, J. M. Rose, and D. A. Hensher. Efficient stated choice experiments for estimating nested logit models. *Transportation Research Part B: Methodological*, 43:19–35, 2009a.

M.C.J. Bliemer, J.M. Rose, and S. Hess. Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modeling*, 1(1):98–127, 2009b.

R.P. Brent. *Algorithms for Minimization without Derivatives*. Prentice-Hall, 1973.

L. Burgess and D. Street. Optimal designs for choice experiments with asymmetric attributes. *Journal of Statistical Planning and Inference*, 134(1):288–301, 2005.

J.A. Cornell. *Experiments with Mixtures*. John Wiley & Sons, Inc, 2002.

Ph. Courcoux and M. Séménou. Une méthode de segmentation pour l'analyse de données issues de comparaisons par paires. *Revue de Statistique Appliquée*, 45(2):59–69, 1997.

V. M. Danthurebandara, J. Yu, and M. Vandebroek. Effect of choice complexity on design efficiency in conjoint choice experiments. *Journal of Statistical Planning and Inference*, 141: 2276–2286, 2011.

H. David. *The Method of Paired Comparisons*. Charles Griffin & Co., 1963.

P. Goos and A.N. Donev. The D-optimal design of blocked experiments with mixture components. *Journal of Quality Technology*, 38(4):319–332, 2006.

P. Goos and A.N. Donev. D-optimal minimum support mixture designs in blocks. *Metrika*, 65 (1):53–68, 2007.

P. Goos, B. Vermeulen, and M. Vandebroek. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Journal of Statistical Planning and Inference*, 140:851–861, 2010.

U. Grasshoff, H. Grossmann, H. Holling, and R. Schwabe. Optimal paired comparison designs for first-order interactions. *Statistics*, 37:373–386, 2003.

U. Grasshoff, H. Grossmann, H. Holling, and R. Schwabe. Optimal designs for main effects in linear paired comparison models. *Journal of Statistical Planning and Inference*, 126:361–376, 2004.

H. Grossmann, H. Holling, U. Grasshoff, and R. Schwabe. Optimal designs for asymmetric linear paired comparisons with a profile strength constraint. *Metrika*, 64:109–119, 2006.

H. Grossmann, U. Grasshoff, and R. Schwabe. Approximate and exact optimal designs for paired comparisons of partial profiles when there are two groups of factors. *Journal of Statistical Planning and Inference*, 139:1171–1179, 2009.

J. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Matematik*, 2(1):84–90, 1960.

D. A. Hensher and J. M. Rose. Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay. *Transportation Research Part E: Logistics and Transportation Review*, 45:583–590, 2009.

J. Huber and K. Zwerina. The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, 33(3):307–317, 1996.

R. Kessels, P. Goos, and M. Vandebroek. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, 43(3):409–419, 2006.

R. Kessels, B. Jones, P. Goos, and M. Vandebroek. An efficient algorithm for constructing Bayesian optimal choice designs. *Journal of Business & Economic Statistics*, 27(2):279–291, 2009.

R. Kessels, B. Jones, and P. Goos. Bayesian optimal designs for discrete choice experiments with partial profiles. *Journal of Choice Modelling*, 4:52–74, 2011a.

R. Kessels, B. Jones, P. Goos, and M. Vandebroek. The usefulness of Bayesian optimal designs for discrete choice experiments. *Applied Stochastic Models in Business and Industry*, 27:173–188, 2011b.

J. Kiefer. Optimum design in regression problems II. *Annals of Mathematical Statistics*, 32: 298–325, 1961.

R.K. Meyer and C.J. Nachtsheim. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1):60–69, 1995.

G.F. Piepel. Measuring component effects in constrained mixture experiments. *Technometrics*, 24(1):29–39, 1982.

G.F. Piepel, S.K. Cooley, and B. Jones. Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm. *Quality Engineering*, 17(4): 579–594, 2005.

J. Rose and M. Bliemer. Constructing efficient stated choice experimental designs. *Transport Reviews: A Transnational Transdisciplinary Journal*, 29(5):587–617, 2009.

Z. Sándor and M. Wedel. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*, 38(4):430–444, 2001.

H. Scheffé. Experiments with mixtures. *Journal of the Royal Statistical Society, Ser. B*, (20): 344–360, 1958.

H. Scheffé. The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society, Ser. B*, (25):235–263, 1963.

W.F. Smith. *Experimental Design for Formulation.* Siam, 2005.

D. J. Street and L. Burgess. *The Construction of Optimal Stated Choice Experiments: Theory and Methods.* Wiley, 2007.

K. Train. *Discrete choice methods with simulation.* Cambridge University Press, 2009.

H. Uranisi. Optimal design for the special cubic regression model on the q-simplex. Mathematical Report 1, Kyushu University, General Education Department, 1964.

Bart Vermeulen, Peter Goos, Riccardo Scarpa, and Martina Vandebroek. Bayesian conjoint choice designs for measuring willingness to pay. *Environmental and Resource Economics*, 48: 129–149, 2011.

J. Yu, P. Goos, and M. Vandebroek. Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science*, 28:122–135, 2009.

J. Yu, P. Goos, and M. Vandebroek. Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments. *Transportation Research Part B: Methodological*, 44(10):1268–1289, 2010a.

J. Yu, P. Goos, and M. Vandebroek. Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments. *Transportation Research Part B: Methodological*, 44:1268–1289, 2010b.

J. Yu, P. Goos, and M. Vandebroek. Individually adapted sequential Bayesian designs for conjoint choice experiments. *International Journal of Research in Marketing*, 28:378–388, 2011.