Statistics and Probability Letters 145 (2019) 127–132

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On three-level D-optimal paired choice designs

Rakhi Singh

IITB-Monash Research Academy, Mumbai, India

ARTICLE INFO

Article history:
Received 5 February 2018
Received in revised form 4 July 2018
Accepted 11 September 2018
Available online 24 September 2018

Keywords:
Choice experiment
Generators
Linear paired comparison model
Multinomial logit model
Utility-neutral setup

ABSTRACT

Considering three-level D-optimal paired choice designs for estimating all the main effects and two-factor interaction effects under the utility-neutral multinomial logit model, we provide general characteristics (and examples) required in generators allowing significant reduction in the number of choice pairs.

1. Introduction

Choice experiments are often used by industry and government alike to assess the importance of certain characteristics to their users. Typically, a design of choice experiment involves designing multiple choice sets with multiple options. Each of the respondents is then asked to choose their preferred option from each of the choice sets. Each option in a choice set is described by a set of k factors and each factor can have $v(\geq 2)$ levels. We consider the choice experiment where N choice sets of size two (or, choice pairs) are shown to every respondent and each of the option is described by k three-level factors. In the setup, each respondent has to choose one of the options in each choice pair. Thus, a paired choice design, say d, is a collection of N choice pairs with k three-level factors, employed in a choice experiment.

Paired choice designs are studied under the multinomial logit model (see, Huber and Zwerina (1996), Street and Burgess (2007)). The multinomial logit model supposes that the probability of preferring option 1 over option 2 in the ith choice pair is $\pi_{12i} = e^{u_{1i}}/(e^{u_{1i}} + e^{u_{2i}})$, where u_{1i} and u_{2i} represent the systematic part of the utilities attached to the two options in the ith choice pair. Similarly $\pi_{21i} = 1 - \pi_{12i}$ is the probability that option 2 is preferred over option 1. For a paired choice design d with N choice pairs, following Huber and Zwerina (1996), the utilities u_j are modeled as $u_j = P_j\theta$, where θ is a $\left(2k+4\binom{k}{2}\right) \times 1$ vector representing the main and two-factor interaction effects, P_j is an $N \times \left(2k+4\binom{k}{2}\right)$ effects-coded matrix for the jth option, and $u_j = (u_{ji})$ is an $N \times 1$ utility vector for the jth option, $j = 1, 2$; $i = 1, \ldots, N$. We also define $P = P_1 - P_2$ and refer to it as the design matrix of design d. For attaining theoretically optimal designs under the multinomial logit model, a utility-neutral approach (that is, taking $\theta = 0$) is in practice for finding the information matrix. Under such a utility-neutral multinomial logit model, the Fisher information matrix for a design d reduces to $(1/4)M_d$, where $M_d = P^TP$.

Graßhoff et al. (2003) and Graßhoff et al. (2004) have studied linear paired comparison designs. These are analyzed under the linear paired comparison model. It has been shown by Großmann and Schwabe (2015) that the optimal designs under linear paired comparison models are also optimal under the utility-neutral multinomial logit model for paired choice designs. Optimal designs under these two models have been obtained by several authors and we refer the reader to comprehensive reviews provided by Street and Burgess (2007) and Großmann and Schwabe (2015).

E-mail address: agrakhi@gmail.com.

https://doi.org/10.1016/j.spl.2018.09.005
0167-7152/

In this paper, we are interested in the estimation of all the main effects and all two-factor interaction effects. As an example, interest on such main-effects and all two-factors interaction effects may arise when say, a fast-food joint wants to assess the effect of four factors (food, drinks, sides, and price) and their interactions on its marketing strategies. These four factors are say at 3 levels each: food (vegetarian, egg, and chicken), drinks (hot coffee, fruit juice, and soft-drinks), sides (fries, onion rings, and popcorn), and price (3, 5, 7). In such situation, when the fast-food joint wants to assess not just the impact of these factors as main-effects but also the impact of interaction effects of each the two factors (interaction of price and food, interaction of sides and drinks, etc.) on their marketing strategies, the designs in this paper will be useful. Designs for such estimation problems in choice experiments are studied by several authors (see, Street and Burgess, 2007, Großmann and Schwabe, 2015, etc.).

Graßhoff et al. (2003) provided D-optimal designs for estimating main effects and two-factor interaction effects with total number of choice pairs $N = g3^k$, where $g = \binom{k}{t^*}2^{t^*}$ when 3 does not divide $k-2$ and $g = \binom{k}{t^*}2^{t^*} + \binom{k}{t^*+1}2^{t^*+1}$, otherwise. Here, $t^* = k - 1 - [\frac{k-2}{3}]$. Street and Burgess (2007) reduced the total number of choice pairs to $N = gn$, where g is same as Graßhoff et al. (2003) and n is the size of a strength four orthogonal array on k three-level factors. Thus, Street and Burgess (2007) reduced the number of choice pairs by using an orthogonal array instead of a complete factorial design. In this paper, we further provide a significant reduction in the number of choice pairs for such an optimal design by reducing the number of generators g. For example, for $k = 4$, currently a D-optimal design would need $N = 32n$ choice pairs, whereas we provide construction of D-optimal design in $N = 4n$ choice pairs, implying a reduction of 88% in the number of choice pairs. We provide construction of such designs for $k = 3, 4, 5, 6$ factors after obtaining generators with much reduced values of g. Using the approach of Singh et al. (2018), we also provide a way to further reduce the number of choice pairs by using orthogonal blocking methodology.

2. Preliminaries

In this section, we introduce some notations and discuss the existing work done in details. Let P_j for main effects and two-factor interaction effects be denoted by X_j and Y_j respectively, $j = 1, 2$. Also, let $X = X_1 - X_2$ and $Y = Y_1 - Y_2$. When our interest lies in the estimation of both the main effects and the two-factor interaction effects, the corresponding information matrix M_d under the linear paired comparison model (Graßhoff et al., 2003) is

$$M_d = P^T P = \begin{bmatrix} X^T X & X^T Y \\ Y^T X & Y^T Y \end{bmatrix}. \tag{1}$$

For main effects, the effects-coded vectors for levels 0, 1 and 2 are (1 0), (0 1) and (−1 −1), respectively. Let $X^i_{j\ell}$ represent X_j corresponding to the ith choice pair and ℓth factor. Then, ith row of X_j is $(X^i_{j1}\ X^i_{j2}\ \dots\ X^i_{jk})$. Also, ith row of Y_j is defined as $(X^i_{j1} \otimes X^i_{j2}, X^i_{j1} \otimes X^i_{j3}, \dots, X^i_{j(k-1)} \otimes X^i_{jk})$.

In our context, a choice design d is connected if each of the main effects and the two-factor interaction effects are estimable, and this happens if and only if M_d has rank $2k + 4\binom{k}{2} = 2k^2$. In what follows, the class of all connected paired choice designs with k three-level factors and N choice pairs is denoted by $\mathcal{D}_{k,N}$. We make use of the standard D-optimality criteria. A design that minimizes $det(M_d^{-1})$ among all designs in $\mathcal{D}_{k,N}$ is said to be D-optimal.

A design is said to be a uniform design (Graßhoff et al., 2003) if it assigns equal weight to all choice pairs with meaningful comparisons, that is, for each factor, equal weight is given to each of six choice pairs (s, t) of distinct levels, $s \neq t$. The comparison depth t in a design d is an integer such that exactly t of the k factors have different levels in both the options and in each of the choice pairs. For estimating main effects and two-factor interaction effects, Graßhoff et al. (2003) showed that the information matrix M_d in (1) for any uniform design d with comparison depth t can be written as

$$M_{d(t)} = \begin{pmatrix} h_1(t) I_k \otimes M_2 & 0 \\ 0 & h_2(t) I_{k(k+1)/2} \otimes M_2 \otimes M_2 \end{pmatrix} \tag{2}$$

where $M_2 = (I_2 + J_2)$, I_ℓ denotes the identity matrix of order ℓ and J_ℓ denotes the $\ell \times \ell$ matrix of all ones, and $\otimes$ denotes the Kronecker product. Also, $h_1(t) = Nt/k$ and $h_2(t) = N\frac{t}{k}(\frac{2}{3} - \frac{t-1}{2(k-1)})$, where N is the total number of choice pairs in a choice design d.

Let $t^* = k - 1 - [\frac{k-2}{3}]$ and $w^* = (t^* + 1)/(3t^* + 1)$, where $[x]$ denotes the largest integer less than or equal to x. Graßhoff et al. (2003) showed that if 3 does not divide $k - 2$, then a uniform design $d(t^*)$, which gives equal weight to all $N = g3^k = \binom{k}{t^*}2^{t^*}3^k$ choice pairs with comparison depth t^*, is D-optimal in $\mathcal{D}(k, N)$. These N choice pairs are formed by pairing each of the 3^k options to 2^{t^*} options obtained such that t^* positions in second option is different than the corresponding positions in the first option and this needs to be done for each of the $\binom{k}{t^*}$ possibilities. The information matrix for such an optimal design $d(t^*)$ is then given by $M_{d(t^*)}$. Furthermore, if 3 divides $k - 2$, then an optimal design d is a combination of two uniform designs d_1 and d_2 with weights w^* and $1 - w^*$ respectively. Here, d_1 and d_2 are paired choice designs with all the choice pairs having comparison depths of t^* and $t^* + 1$, respectively. The information matrix for such an optimal design $d(t^*)$ is then given by $w^* M_{d(t^*)} + (1 - w^*) M_{d(t^*+1)}$. In this case, total number of choice pairs in an optimal design are $N = g3^k = \{\binom{k}{t^*}2^{t^*} + \binom{k}{t^*+1}2^{t^*+1}\}3^k$.

An orthogonal array $OA(n, 3^k, t)$, of strength t, is an $n \times k$ array with elements from a set of 3 distinct symbols $\{0, 1, 2\}$, such that all possible combinations of symbols appear equally often as rows in every $n \times t$ subarray. Street and Burgess