An Introduction to Constructing Discrete Choice Experiments (DCEs)

Deborah Street

Centre for Health Economics Research and Evaluation,
University of Technology Sydney

Outline

- A simple DCE
- Design Requirements
- MNL Model and Design Comparison
- Construction of Generator-Developed Designs for Main Effects
- Extension to Interactions
- Concluding comments

Basic Features of a DCE

- Screening Question(s)
- Background Information
- Choice Context
- Choice Set(s)
- Follow-up questions about the task (possibly)

Lightbulbs - Screening Questions

- Demographics
  - Age
  - Gender
  - State of residence
- Have you ever bought a lightbulb?

Lightbulbs - Background Information

Australia is about to ban the sale of incandescent light bulbs. Once this happens all light bulbs on sale in Australia will be energy efficient light bulbs. This type of bulb typically **lasts for a long time** (from 3 to 7.5 years under standard domestic conditions) and can **cost** rather more initially than a standard bulb. The light that such bulbs give can have more blue in it, although bulbs with the same **sort of light** as an incandescent gives are also available. The bulbs themselves have a higher concentration of mercury and so they need to be **recycled** to avoid environmental damage near rubbish dumps. Energy efficient bulbs use about 25% of the power of an incandescent bulb of an equivalent brightness and so over the lifetime of the bulb the total amount of mercury used to make the bulb and the power to run it is less than for an equivalent incandescent bulb. Some **bulbs can take up to a minute to reach full brightness**.

Lightbulbs - Typical survey page

Suppose that you need to replace the lightbulb in your living room. This room is currently used for talking, reading books and watching TV (at various times). Because of the new regulations about incandescent bulbs, you need to replace the bulb with an energy efficient bulb which will give the same level of light as you currently have in your living room. Which of the three lightbulbs shown below would you choose?

Attribute	**Option A**	**Option B**	**Option C**
Quality of light	warm white	daylight white	cool white
Lifetime of bulb	15,000 hours	6,000 hours	12,000 hours
Recycling	Shopping mall	kerbside	Shopping mall
Dimmable	no	yes	no
Time to full brightness	60 secs	5 secs	60 secs
Cost	$14	$3.50	$7
YOUR CHOICE			

Lightbulbs - Typical survey page

Choice Context - *Suppose that you need to replace the lightbulb in your living room. This room is currently used for talking, reading books and watching TV (at various times). Because of the new regulations about incandescent bulbs, you need to replace the bulb with an energy efficient bulb which will give the same level of light as you currently have in your living room.*

Which of the three lightbulbs shown below would you choose?

Choice set - block in a blocked factorial experiment

Attribute	**Option A**	**Option B**	**Option C**
Quality of light	warm white	daylight white	cool white
Lifetime of bulb	15,000 hours	6,000 hours	12,000 hours
Recycling	Shopping mall	kerbside	Shopping mall
Dimmable	no	yes	no
Time to full brightness	60 secs	5 secs	60 secs
Cost	$14	$3.50	$7
YOUR CHOICE			

Goal of the Choice Experiment

- To estimate how the attribute levels affect the utility of the individual

Assumptions

- Individuals make utility-maximising decisions
- Utility has two components
  - Systematic component, relating to the attributes and levels
  - Random component, may relate to
    - unobserved determinants of utility,
    - individual characteristics,
    - random variation in choices.

CHERE CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

Notation

- A choice experiment consists of N choice sets.
- Each choice set consists of m options which are generic (unlabelled).
- Respondents are asked which option they would choose from each choice set.
- Each option is described by k attributes

  attribute q has ℓ_q levels $0, 1, \ldots, \ell_q - 1$, $q = 1, \ldots, k$, so

  $$\begin{array}{l} \text{attribute 1 has } \ell_1 \text{ levels } 0, 1, \ldots, \ell_1 - 1 \\ \qquad\qquad\qquad \vdots \\ \text{attribute } k \text{ has } \ell_k \text{ levels } 0, 1, \ldots, \ell_k - 1 \end{array}$$

- Typical option is a k-tuple, $(2, 1, \ell_3 - 2, \ldots, \ell_k - 1)$

Desirable Properties of a DCE

1. A manageable number of choice sets for the respondents;
2. Choice sets which are statistically efficient;
3. Parameters are
   - identifiable and
   - independently estimated;
4. No repeated choice sets;
5. All options in a choice set are distinct;
6. No option that will always be best or worst in any choice set.

Example - Choice sets that satisfy all of the properties if at least one attribute does not have ordered levels.

Here $k = 3$, $\ell_1 = \ell_2 = \ell_3 = 2$.

So the options are triples with all entries equal to either 0 or $\ell_q - 1 = 2 - 1 = 1$.

Choice set	Option 1	Option 2	Option 3
1:	0 0 0	0 0 1	1 1 0
2:	0 1 1	0 1 0	1 0 1
3:	1 0 1	1 0 0	0 1 1
4:	1 1 0	1 1 1	0 0 0

CHERE CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

Dominance Issues

Consider plane tickets.

Attribute	Attribute Levels
cost	\$199 \$299
check-in time	Must check in 30 minutes before flight or ticket forfeit Check-in up to 15 minutes before departure

A \$199 fare with relaxed check-in conditions is going to be preferred by all respondents, all else being equal.

Such an option is said to *dominate*.

The design question for choice experiments

Given that we want to estimate the impact of the attributes on the utility,

- Which level combinations should be placed in the same choice set?
- Which choice sets should be placed in the same design?

Form of the utility function for main effects MNL model

$U_{ijc} = V_{ijc} + e_{ijc}, 1 \leq i \leq I; 1 \leq j \leq m; 1 \leq c \leq N$

i - individual
j - alternative
c - choice set
V_{ijc} - systematic component of utility
e_{ijc} - unobserved error term

- Assume the e_{ijc} are independent extreme value Type 1. Then
$$P(T_a > \{T_1, T_2, \ldots, T_m\} \backslash \{T_a\}) = \frac{\exp(\lambda V_a)}{\sum_{j=1}^{m} \exp(\lambda V_j)}.$$
- λ is positive scale parameter,
  - inversely proportional to the variance of the random component,
  - commonly assumed to be 1.

Calculating V_{ijc} as a function of the attribute levels

- For main effects model, replace a level of attribute q by a vector of length $\ell_q - 1$.

Dummy-coding

For each attribute

- One level is chosen as the reference level
- Each of the other ℓ_q levels are replaced by a row of the identity matrix of order $\ell_q - 1$
- The reference level is replaced by a row of $(\ell_q - 1)$ 0s.

Effects-coding

For each attribute

- One level is chosen as the reference level
- Each of the other ℓ_q levels are replaced by a row of the identity matrix of order $\ell_q - 1$
- The reference level is replaced by a row of $(\ell_q - 1)$ -1s.

The EQ-5D-3L Attributes and Levels

The five dimensions (attributes)

- Mobility
- Self-Care
- Usual Activities
- Pain and Discomfort
- Anxiety and Depression

The three levels of each dimension

- No problems - level 1
- Moderate problems - level 2
- Extreme problems / Unable to - level 3

Dummy coding an option for the main effects MNL model

- Use 1 as the reference level for all 5 dimensions.
- For the EQ-5D-3L health state 22331
- Dummy-coding gives (1,0,1,0,0,1,0,1,0,0)=X_{ijc}
- The $\sum(\ell_q - 1) = 10$ entries of $\boldsymbol{\beta}$ are
  - β_1 corresponds to moderate problems with mobility
  - β_2 corresponds to extreme problems with mobility
  - β_3 corresponds to moderate problems with self-care
  - and so on.
- $V_{ijc} = X'_{ijc}\boldsymbol{\beta} = \beta_1 + \beta_3 + \beta_6 + \beta_8$

CHERE CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

The information matrix and design comparison

For choice sets of size 2,

$$P(A \text{ is preferred to } B) = \frac{\exp(V_A)}{\exp(V_A) + \exp(V_B)}.$$

The information matrix for the design is the sum of the information matrices for the N choice sets.

$$M_{d,\boldsymbol{\beta}} = \sum_{c=1}^{N} X_c'(\text{diag}(p_{c,\boldsymbol{\beta}}) - p_{c,\boldsymbol{\beta}}p_{c,\boldsymbol{\beta}}')X_c,$$

$p_{c,\boldsymbol{\beta}}$ - column vector of probabilities associated with choice set c, given the column of parameters $\boldsymbol{\beta}$,
and
X_c - matrix of order 2×10 (in general, $m \times \sum_q(\ell_q - 1)$) with the dummy-coded options in choice set c as rows.

CHERE
CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

Calculations for one choice set

Choice set has health states (1,2,1,2,2) and (2,1,2,1,2) as options.

$$X_c = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\beta = (-0.2, -0.9, -0.1, -0.6, -0.2, -0.3, -0.1, -0.9, -0.3, -0.8)$$

$$V_c = X_c.\beta' = \begin{pmatrix} -0.5 \\ -0.7 \end{pmatrix}$$

$$p_{c,\beta} = \begin{pmatrix} e^{-0.5}/(e^{-0.5} + e^{-0.7}) \\ e^{-0.7}/(e^{-0.5} + e^{-0.7}) \end{pmatrix} = \begin{pmatrix} 0.55 \\ 0.45 \end{pmatrix}$$

$$\text{diag}(p_{c,\beta}) - p_{c,\beta}p'_{c,\beta} = \begin{pmatrix} 0.248 & -0.248 \\ -0.248 & 0.248 \end{pmatrix}$$

Information matrix for choice set with (1,2,1,2,2) and (2,1,2,1,2)

$$X'(\mathrm{diag}(p_{c,\boldsymbol{\beta}}) - p_{c,\boldsymbol{\beta}}p'_{c,\boldsymbol{\beta}})X =$$

$$X' \begin{pmatrix} 0.248 & -0.248 \\ -0.248 & 0.248 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 0.248 & 0.0 & -0.248 & 0.0 & 0.248 & 0.0 & -0.248 & 0.0 & 0.0 & 0.0 \\ 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.0 & 0.0 \\ -0.248 & 0.0 & 0.248 & 0.0 & -0.248 & 0.0 & 0.248 & 0.0 & 0.0 & 0.0 \\ 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.0 & 0.0 \\ 0.248 & 0.0 & -0.248 & 0.0 & 0.248 & 0.0 & -0.248 & 0.0 & 0.0 & 0.0 \\ 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.0 & 0.0 \\ -0.248 & 0.0 & 0.248 & 0.0 & -0.248 & 0.0 & 0.248 & 0.0 & 0.0 & 0.0 \\ 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.0 & 0.0 \\ 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.0 & 0.0 \\ 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.000 & 0.0 & 0.0 & 0.0 \end{pmatrix}$$

Example - Two binary attributes in choice sets of size 2

- Represent the attributes by A and B
- Represent the levels by 0 and 1
- List all the possible options

CHERE CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

Example - Two binary attributes in choice sets of size 2

- Represent the attributes by A and B
- Represent the levels by 0 and 1
- List all the possible options
  - 00 and 01 and 10 and 11

Example - Two binary attributes in choice sets of size 2

- Represent the attributes by A and B
- Represent the levels by 0 and 1
- List all the possible options
  - 00 and 01 and 10 and 11
- List all the possible choice sets of size 2

Example - Two binary attributes in choice sets of size 2

- Represent the attributes by A and B
- Represent the levels by 0 and 1
- List all the possible options
  - 00 and 01 and 10 and 11
- List all the possible choice sets of size 2
  - 00 with 01
  - 00 with 10
  - 00 with 11
  - 01 with 10
  - 01 with 11
  - 10 with 11

Contribution to $M_{d,\beta}$ from each pair

00 with 01

$$\begin{pmatrix} 0 & 0 \\ 0 & \frac{e^{b2}}{\left(1+e^{b2}\right)^2} \end{pmatrix}$$

00 with 11

$$\begin{pmatrix} \frac{e^{b1+b2}}{\left(1+e^{b1+b2}\right)^2} & \frac{e^{b1+b2}}{\left(1+e^{b1+b2}\right)^2} \\ \frac{e^{b1+b2}}{\left(1+e^{b1+b2}\right)^2} & \frac{e^{b1+b2}}{\left(1+e^{b1+b2}\right)^2} \end{pmatrix}$$

01 with 11

$$\begin{pmatrix} \frac{e^{b1}}{\left(1+e^{b1}\right)^2} & 0 \\ 0 & 0 \end{pmatrix}$$

---

00 with 10

$$\begin{pmatrix} \frac{e^{b1}}{\left(1+e^{b1}\right)^2} & 0 \\ 0 & 0 \end{pmatrix}$$

01 with 10

$$\begin{pmatrix} \frac{e^{b1+b2}}{\left(e^{b1}+e^{b2}\right)^2} & -\frac{e^{b1+b2}}{\left(e^{b1}+e^{b2}\right)^2} \\ -\frac{e^{b1+b2}}{\left(e^{b1}+e^{b2}\right)^2} & \frac{e^{b1+b2}}{\left(e^{b1}+e^{b2}\right)^2} \end{pmatrix}$$

10 with 11

$$\begin{pmatrix} 0 & 0 \\ 0 & \frac{e^{b2}}{\left(1+e^{b2}\right)^2} \end{pmatrix}$$

List of all of the 63 possible designs

00, 01	00, 10	00, 11	01, 10	01, 11	10, 11	Design
0	0	0	0	0	0	0
0	0	0	0	0	1	1
0	0	0	0	1	0	2
0	0	0	0	1	1	3
0	0	0	1	0	0	4
0	0	0	1	0	1	5
0	0	0	1	1	0	6
0	0	0	1	1	1	7
0	0	1	0	0	0	8
0	0	1	0	0	1	9
0	0	1	0	1	0	10
0	0	1	0	1	1	11
0	0	1	1	0	0	12
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	1	1	1	1	63

Comparing information matrices

d_{12}: 00 with 11 and 01 with 10

$$M_{12} = e^{b1+b2}\begin{pmatrix} \left(\frac{1}{\left(1+e^{b1+b2}\right)^2}+\frac{1}{\left(e^{b1}+e^{b2}\right)^2}\right) & \left(\frac{1}{\left(1+e^{b1+b2}\right)^2}-\frac{1}{\left(e^{b1}+e^{b2}\right)^2}\right) \\ \left(\frac{1}{\left(1+e^{b1+b2}\right)^2}-\frac{1}{\left(e^{b1}+e^{b2}\right)^2}\right) & \left(\frac{1}{\left(1+e^{b1+b2}\right)^2}+\frac{1}{\left(e^{b1}+e^{b2}\right)^2}\right) \end{pmatrix}$$

d_{33}: 00 with 01 and 10 with 11

$$M_{33} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{2e^{b2}}{\left(1+e^{b2}\right)^2} \end{pmatrix}$$

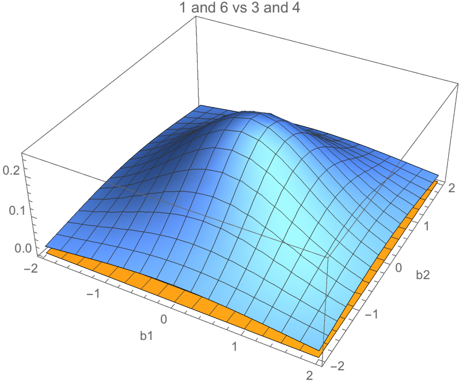
d_{10}: 00 with 11 and 01 with 11

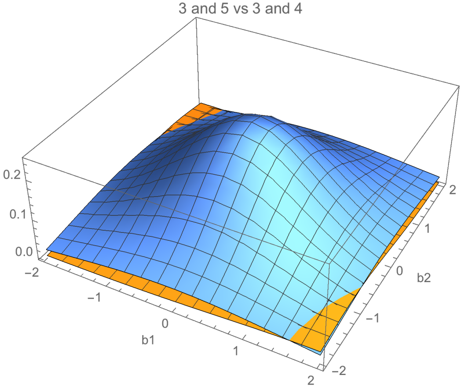
$$M_{10} = \begin{pmatrix} e^{b1}\left(\frac{e^{b2}}{\left(1+e^{b1+b2}\right)^2}+\frac{1}{\left(1+e^{b1}\right)^2}\right) & \frac{e^{b1+b2}}{\left(1+e^{b1+b2}\right)^2} \\ \frac{e^{b1+b2}}{\left(1+e^{b1+b2}\right)^2} & \frac{e^{b1+b2}}{\left(1+e^{b1+b2}\right)^2} \end{pmatrix}$$

The determinant is a good one number summary.

$\det(M_{33}) = 0$ and easy to see there is no information about one of the βs.

Plots of $\det(M_{d,\beta})$





d_1: 00 with 01 and 10 with 11

vs

d_2: 00 with 11 and 01 with 10

d_1: 00 with 11 and 01 with 11

vs

d_2: 00 with 11 and 01 with 10

Possible construction strategies

- Random allocation of treatment combinations to the choice sets.
- Generator-developed design - start with an orthogonal array (OA) for the first option and add suitable generators for subsequent options
- Computer generated designs (search algorithms)
  - Start with a set of choice sets
  - Calculate $\det(M_{d,\beta})$
  - Change one or more of the choice sets. See if the new design is better (has a larger $\det(M_{d,\beta})$).
  - Repeat until the design is “good enough".

CHERE
CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

Orthogonal array - OA$[R; \ell_1, \ell_2, \ldots, \ell_k]$

- $R \times k$ array
- Column q has ℓ_q levels
  - The levels are $0, 1, \ldots, \ell_q - 1$.
- Any two distinct columns, i and j, have each of the $\ell_i \ell_j$ distinct ordered pairs equally often.

Finding the OA

Websites with OAs include

- http://support.sas.com/techsup/technote/ts723_Designs.txt
- http://www.pietereendebak.nl/oapackage/series.html

OA[16;4,4,2,2,2,4]

0	0	0	0	0	0
0	1	0	1	1	3
0	2	1	0	1	1
0	3	1	1	0	2
1	0	0	1	1	2
1	1	0	0	0	1
1	2	1	1	0	3
1	3	1	0	1	0
2	0	1	0	0	3
2	1	1	1	1	0
2	2	0	0	1	2
2	3	0	1	0	1
3	0	1	1	1	1
3	1	1	0	0	2
3	2	0	1	0	0
3	3	0	0	1	3

OA[16;4,4,2,2,2,4]

0	0	0	0	0	0
0	1	0	1	1	3
0	2	1	0	1	1
0	3	1	1	0	2
1	0	0	1	1	2
1	1	0	0	0	1
1	2	1	1	0	3
1	3	1	0	1	0
2	0	1	0	0	3
2	1	1	1	1	0
2	2	0	0	1	2
2	3	0	1	0	1
3	0	1	1	1	1
3	1	1	0	0	2
3	2	0	1	0	0
3	3	0	0	1	3

OA[16;4,4,2,2,2,4]

0	0	0	0	0	0
0	1	0	1	1	3
0	2	1	0	1	1
0	3	1	1	0	2
1	0	0	1	1	2
1	1	0	0	0	1
1	2	1	1	0	3
1	3	1	0	1	0
2	0	1	0	0	3
2	1	1	1	1	0
2	2	0	0	1	2
2	3	0	1	0	1
3	0	1	1	1	1
3	1	1	0	0	2
3	2	0	1	0	0
3	3	0	0	1	3

Random allocation

- Given $k, \ell_1, \ldots, \ell_k, m, N$,
  - find an orthogonal array
  - with $R = Nm$ treatment combinations,
- Randomly group the R treatment combinations into $N = R/m$ choice sets.

Example - $k = 5$, $\ell_1 = \ell_2 = \ell_3 = \ell_4 = \ell_5 = 4$, $m = 2$, $N = 16$

So need an OA with $R = N \times m = 16 \times 2 = 32$ treatment combinations.

Random allocation: One possible set of choice sets

Option 1	Option 2
2 1 0 2 3	0 0 3 1 1
0 1 2 2 0	0 3 0 1 1
3 1 1 1 3	3 2 2 3 1
0 2 1 2 0	2 2 3 2 3
2 3 2 3 0	1 2 0 3 2
3 1 1 3 1	2 3 2 1 2
3 3 3 2 2	1 3 1 2 1
3 3 3 0 0	0 1 2 0 2
1 1 3 3 2	3 0 0 2 2
3 0 0 0 0	3 2 2 1 3
2 0 1 3 0	1 0 2 2 1
0 3 0 3 3	1 0 2 0 3
2 2 3 0 1	1 1 3 1 0
0 0 3 3 3	0 2 1 0 2
2 1 0 0 1	1 2 0 1 0
1 3 1 0 3	2 0 1 1 2

Constructed 100 different designs for $m = 2$.

12 had $\det(M) = 0$.

For the rest, efficiency ranged from 32.04% to 59.61%, with an average of 45.08% (relative to the best possible design when $\beta = \mathbf{0}$).

In none of the designs constructed was any pair of main effects uncorrelated.

Shows properties of designs depends on the allocation into choice sets.

Generator Developed DCEs

For estimation of main effects:

- Given k, $\ell_1, \ell_2, \ldots, \ell_k$ find an orthogonal array.
- Each treatment combination gives rise to one choice set.
- Obtain the rest of the items in each choice set by adding generators.
  - Addition is carried out component-wise mod ℓ_q in position q.
  - Need $m - 1$ generators for choice sets of size m.
- The relative performance of the DCEs depends on the combination of generators and OA.

DCE from OA[16;4,4,2,2,2,4] and generators g_2=(1,1,1,1,1,1) and g_3=(2,2,0,0,0,2)

Option 1	Option 2	Option 3	
0 0 0 0 0 0	1 1 1 1 1 1	2 2 0 0 0 2	0+1 (mod 4)=1
0 1 0 1 1 3	1 2 1 0 0 0	2 3 0 1 1 1	1+1 (mod 4)=2
0 2 1 0 1 1	1 3 0 1 0 2	2 0 1 0 1 3	2+1 (mod 4)=3
0 3 1 1 0 2	1 0 0 0 1 3	2 1 1 1 0 0	3+1 (mod 4)=0
1 0 0 1 1 2	2 1 1 0 0 3	3 2 0 1 1 0	
1 1 0 0 0 1	2 2 1 1 1 2	3 3 0 0 0 3	0+2 (mod 4)=2
1 2 1 1 0 3	2 3 0 0 1 0	3 0 1 1 0 1	1+2 (mod 4)=3
1 3 1 0 1 0	2 0 0 1 0 1	3 1 1 0 1 2	2+2 (mod 4)=0
2 0 1 0 0 3	3 1 0 1 1 0	0 2 1 0 0 1	3+2 (mod 4)=1
2 1 1 1 1 0	3 2 0 0 0 1	0 3 1 1 1 2	
2 2 0 0 1 2	3 3 1 1 0 3	0 0 0 0 1 0	0+3 (mod 4)=3
2 3 0 1 0 1	3 0 1 0 1 2	0 1 0 1 0 3	1+3 (mod 4)=0
3 0 1 1 1 1	0 1 0 0 0 2	1 2 1 1 1 3	2+3 (mod 4)=1
3 1 1 0 0 2	0 2 0 1 1 3	1 3 1 0 0 0	3+3 (mod 4)=2
3 2 0 1 0 0	0 3 1 0 1 1	1 0 0 1 0 2	
3 3 0 0 1 3	0 0 1 1 0 0	1 1 0 0 1 1	0+1 (mod 2)=1
			1+1 (mod 2)=0

Generator entries for attribute q

		m		
		2	3	4
	2	$g_{2q}=1$	$g_{2q}=1, g_{3q}=0$	$g_{2q}=1, g_{3q}=0, g_{4q}=1$
	3	$g_{2q}=1$	$g_{2q}=1, g_{3q}=2$	$g_{2q}=1, g_{3q}=2, g_{4q}=0$
	4	$g_{2q}=1$	$g_{2q}=1, g_{3q}=2$	$g_{2q}=1, g_{3q}=2, g_{4q}=3$
ℓ	5	$g_{2q}=1$	$g_{2q}=1, g_{3q}=2$	$g_{2q}=1, g_{3q}=2, g_{4q}=3$
	6	$g_{2q}=1$	$g_{2q}=1, g_{3q}=3$	$g_{2q}=1, g_{3q}=3, g_{4q}=2$
	7	$g_{2q}=1$	$g_{2q}=1, g_{3q}=3$	$g_{2q}=1, g_{3q}=2, g_{4q}=4$
	8	$g_{2q}=1$	$g_{2q}=1, g_{3q}=3$	$g_{2q}=1, g_{3q}=2, g_{4q}=4$

What's the most information possible when $\beta = 0$?

Caution: This uses orthogonal polynomial coding.

The *D*-optimal design, for estimating main effects only, has a maximum value of the determinant of the information matrix C given by

$$\det(C_{\text{opt}}) \quad = \quad \prod_{q=1}^{k} \left[\frac{2\ell_q S_q}{m^2 L(\ell_q - 1)} \right]^{\ell_q - 1}.$$

where

$$S_q = \begin{cases} (m^2 - (\ell_q x^2 + 2xy + y))/2, & 2 \leq \ell_q < m, \\ m(m-1)/2, & \ell_q \geq m, \end{cases}$$

and

- $m = \ell_q x + y$ for $x \geq 0$ and $0 \leq y < \ell_q$,
- $L = \prod_q \ell_q$,
- $C = M/N$.

CHERE CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

Orthogonal polynomial coding

$\ell_q = 2$

$$\begin{array}{ccr} 0 & \mapsto & -1/\sqrt{2} \\ 1 & \mapsto & 1/\sqrt{2} \end{array}$$

$\ell_q = 3$

$$\begin{array}{ccrr} 0 & \mapsto & -1/\sqrt{2}, & 1/\sqrt{6} \\ 1 & \mapsto & 0, & -2/\sqrt{6} \\ 2 & \mapsto & 1/\sqrt{2}, & 1/\sqrt{6} \end{array}$$

$\ell_q = 4$

$$\begin{array}{ccrrr} 0 & \mapsto & -1/2, & -1/2, & 1/2 \\ 1 & \mapsto & -1/2, & 1/2, & -1/2 \\ 2 & \mapsto & 1/2, & -1/2, & -1/2 \\ 3 & \mapsto & 1/2, & 1/2, & 1/2 \end{array}$$

CHERE CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

Do we need to avoid dominated options?

- Need to be sure that everyone will order the levels of each attribute in the same way.
- A few choice sets with dominated options provide a chance to check that respondents are making rational choices.
- But some respondents may choose an option that is “good enough” without looking at all the possibilities.
- To avoid dominated options, it often helps to have a mix of levels in the generators, some high, some low and some in the middle.

The EQ-5D-3L Attributes and Levels

The five dimensions (attributes)

- Mobility
- Self-Care
- Usual Activities
- Pain and Discomfort
- Anxiety and Depression

The three levels of each dimension

- No problems - level 0
- Moderate problems - level 1
- Extreme problems / Unable to - level 2

Getting the initial OA

Websites with OAs include

- http://support.sas.com/techsup/technote/ts723_Designs.txt
- http://www.pietereendebak.nl/oapackage/series.html

From the SAS website, for example, we find

```
3^6 6^1      n=18
0000000
0011221
0102212
0120123
0212104
0221015
1002125
1020214
1111110
1122001
1201203
1210022
2012013
2021102
2101024
2110205
2200111
2222220
```

The first 6 columns each have 3 levels.

Any 5 of the first 6 columns give the required OA.

You must not remove rows as that destroys the properties of the OA.

18 run OA for EQ-5D-3L

0	0	0	0	0
0	0	1	1	2
0	1	0	2	2
0	1	2	0	1
0	2	1	2	1
0	2	2	1	0
1	0	0	2	1
1	0	2	0	2
1	1	1	1	1
1	1	2	2	0
1	2	0	1	2
1	2	1	0	0
2	0	1	2	0
2	0	2	1	1
2	1	0	1	0
2	1	1	0	2
2	2	0	0	1
2	2	2	2	2

Use the first 5 columns.

18 run OA for EQ-5D-3L

0	0	1	0	1
0	0	2	1	0
0	1	1	2	0
0	1	0	0	2
0	2	2	2	2
0	2	0	1	1
1	0	1	2	2
1	0	0	0	0
1	1	2	1	2
1	1	0	2	1
1	2	1	1	0
1	2	2	0	1
2	0	2	2	1
2	0	0	1	2
2	1	1	1	1
2	1	2	0	0
2	2	1	0	2
2	2	0	2	0

Use the first 5 columns.

Add 1 (mod 3) to columns 3 and 5.

18 run OA for EQ-5D-3L

0	0	1	2	1
0	0	2	0	0
0	1	1	1	0
0	1	0	2	2
0	2	2	1	2
0	2	0	0	1
1	0	1	1	2
1	0	0	2	0
1	1	2	0	2
1	1	0	1	1
1	2	1	0	0
1	2	2	2	1
2	0	2	1	1
2	0	0	0	2
2	1	1	0	1
2	1	2	2	0
2	2	1	2	2
2	2	0	1	0

Use the first 5 columns.

Add 1 (mod 3) to columns 3 and 5.

Add 2 (mod 3) to column 4.

Using more than one set of generators

We might do this

- to get a more efficient design
  - For a design with 7 4-level attributes
    - one generator - {1,1,1,1,1,1,1}- 94.49% efficient
    - one generator - {2,2,2,2,2,2,2}- 0% efficient
    - two generators - {1,1,1,1,1,1,1} and {2,2,2,2,2,2,2} - 100%
- to allow for some overlap to make the designs easier for respondents
  - Using {1,1,1,0,0} and {0,0,1,1,1} in an EQ-5D design would mean only 3 of the 5 health state dimensions would vary between the two options in each pair.

Having different versions

- There is a limit to the number of tasks an individual can perform.
- For large OAs, or for designs with several generators, many choice sets are generated.
- Subdivide them into versions with fewer sets.
  - Use a spare attribute to determine the versions.
  - Allocate randomly.
- Both have been used and work well in practice.

What about including interactions?

- Used when we think that the response to one factor depends on the level of another factor.
  - Cost is a familiar example.
  - Accuracy and waiting time for test results.
  - Brand of medicine and pharmacist recommendation
- Model needs to be extended.
  - Include more elements in $\boldsymbol{\beta}$.
  - Extend the coding to allow for interaction terms.
- The initial OA needs different properties.
  - Need to think about the frequency with which level combinations across 4 columns appear.
  - If only some interactions are of interest then only some sets of 4 columns need to be considered.
- The generators need a different structure.
  - For each attribute have at least one generator with a non-zero entry.
  - For each pair of attributes have at least one generator with a zero entry for one of the attributes.

Empirical design comparison by Domínguez-Torreiro (2014)

- Compared a generator-developed design and a *D*-efficient design constructed using priors
- Each design had 24 choice sets.
- Used a split-sample so all respondents saw 6 choice sets from each design
- *D*-efficient design did not produce significantly lower standard errors
- Overall goodness-of-fit was not found to be superior in the model from the *D*-efficient design compared to that from the gen-dev design

Empirical design comparison by Walker, Wang, Thorhauge and Ben-Akiva (2018)

- Interested in robustness of efficient designs to misspecification of the assumed prior β
- Looked at 10 design strategies (none were gen-dev)
- “it is risky to use an efficient design with uncertain priors”
- “this paper serves as a reference to support the use of non-efficient [constructed without a prior] design methods in stated choice experiments”

Using non-zero priors

- Burgess et al (2015) used the same attributes and levels in two studies.
  - The first study provided priors for some designs used in the second study.
  - “using prior parameter estimates has resulted in a reduction in design performance ... for the main effects and two-factor interactions MNL model”
- Mulhern et al (2017) compared designs to value health states within the EQ-5D-5L
  - Design with non-zero prior produced more inconsistent coefficients than a design with a zero prior.

THANK YOU