Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables

Mario Becerra with Peter Goos May 25th, 2022 International Choice Modelling Conference Reykjavik, Iceland

# Outline

- 1. Choice modeling and choice experiments
- 2. Mixture experiments
- 3. Combining choice models and mixture models
- 4. Optimality criteria for choice experiments
- 5. Results

# Choice modeling and choice experiments







Quantify consumer preferences







- Quantify consumer preferences
- Preference data is collected





- Quantify consumer preferences
- Preference data is collected
- Respondents are presented sets of alternatives (choice sets) and asked to choose





- Quantify consumer preferences
- Preference data is collected
- Respondents are presented sets of alternatives (choice sets) and asked to choose
  - Example: choosing to buy product A, B or C





- Quantify consumer preferences
- Preference data is collected
- Respondents are presented sets of alternatives (choice sets) and asked to choose
  Example: choosing to buy product A, B or C
- Latent utility function -> probability of making each decision





# Mixture experiments





• Many products and services can be described as mixtures of ingredients

- Many products and services can be described as mixtures of ingredients
- Examples:
  - $\circ$  wheat varieties in bread





- Many products and services can be described as mixtures of ingredients
- Examples:
  - $\circ$  wheat varieties in bread
  - ingredients used to make a cocktail







- Many products and services can be described as mixtures of ingredients
- Examples:
  - $\circ$  wheat varieties in bread
  - o ingredients used to make a cocktail
  - $\circ$  types of fish used to make a fish patty









 In mixture experiments, products are expressed as combinations of proportions of ingredients

- In mixture experiments, products are expressed as combinations of proportions of ingredients
- The researchers' interest is generally in one or more characteristics of the mixture

- In mixture experiments, products are expressed as combinations of proportions of ingredients
- The researchers' interest is generally in one or more characteristics of the mixture
- In this work, the characteristic of interest is the **preference** of respondents

- In mixture experiments, products are expressed as combinations of proportions of ingredients
- The researchers' interest is generally in one or more characteristics of the mixture
- In this work, the characteristic of interest is the **preference** of respondents
- Choice experiments are ideal to collect data for quantifying and modeling preferences for mixtures

# Combining choice models and mixture models









• First example by Courcoux and Séménou (1997), preferences for cocktails



- First example by Courcoux and Séménou (1997), preferences for cocktails
  - mango juice
  - lemon juice 🧉
  - blackcurrant syrup





- First example by Courcoux and Séménou (1997), preferences for cocktails
  - mango juice 🧧
  - lemon juice 🧉
  - blackcurrant syrup 🚙
- 60 people, each making 8 pairwise comparisons



• Experiments are expensive, cumbersome and time-consuming



- Experiments are expensive, cumbersome and time-consuming
- Efficient experimental designs  $\rightarrow$  reliable information



- Experiments are expensive, cumbersome and time-consuming
- Efficient experimental designs  $\rightarrow$  reliable information
- Optimal design of experiments: the branch of statistics that deals with the construction of efficient experimental designs



# Optimality criteria for choice experiments



• **D-optimal** experimental designs  $\rightarrow$  low-variance estimators

- **D-optimal** experimental designs  $\rightarrow$  low-variance estimators
- We want to have a mixture that maximizes consumer preference



- **D-optimal** experimental designs  $\rightarrow$  low-variance estimators
- We want to have a mixture that maximizes consumer preference
- Precise predictions are crucial

- **D-optimal** experimental designs  $\rightarrow$  low-variance estimators
- We want to have a mixture that maximizes consumer preference
- Precise predictions are crucial
- I-optimal experimental designs  $\rightarrow$  low-variance prediction

### Models for data from mixture experiments



# Models for data from mixture experiments

 Mixture models assume two or more ingredients and a response variable that depends only on the relative proportions of the ingredients in the mixture

# Models for data from mixture experiments

- Mixture models assume two or more ingredients and a response variable that depends only on the relative proportions of the ingredients in the mixture
- Each mixture is described as a combination of *q* ingredient proportions (0 to 1)
# Models for data from mixture experiments

- Mixture models assume two or more ingredients and a response variable that depends only on the relative proportions of the ingredients in the mixture
- Each mixture is described as a combination of *q* ingredient proportions (0 to 1)
- Constraint: proportions sum up to one

# Models for data from mixture experiments

- Mixture models assume two or more ingredients and a response variable that depends only on the relative proportions of the ingredients in the mixture
- Each mixture is described as a combination of *q* ingredient proportions (0 to 1)
- Constraint: proportions sum up to one  $\rightarrow$  perfect collinearity

## Models for data from mixture experiments

- Mixture models assume two or more ingredients and a response variable that depends only on the relative proportions of the ingredients in the mixture
- Each mixture is described as a combination of *q* ingredient proportions (0 to 1)
- Constraint: proportions sum up to one  $\rightarrow$  perfect collinearity
- Special-cubic Scheffé model:

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

• The preference for a mixture may depend on other characteristics

- The preference for a mixture may depend on other characteristics
- Additional variables → **process variables**

- The preference for a mixture may depend on other characteristics
- Additional variables → **process variables**
- Second-order Scheffé model

- The preference for a mixture may depend on other characteristics
- Additional variables → process variables
- Second-order Scheffé model

$$Y = \sum_{k=1}^{q} \gamma_k^0 x_k + \sum_{k=1}^{q-1} \sum_{l=k+1}^{q} \gamma_{kl}^0 x_k x_l + \sum_{i=1}^{r} \sum_{k=1}^{q} \gamma_k^i x_k z_i + \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} \alpha_{ij} z_i z_j + \sum_{i=1}^{r} \alpha_i z_i^2 + \varepsilon$$



• A respondent faces S choice sets involving J alternatives each

- A respondent faces S choice sets involving J alternatives each
- Respondent chooses the alternative that has the highest perceived utility

- A respondent faces S choice sets involving J alternatives each
- · Respondent chooses the alternative that has the highest perceived utility
- The probability that a respondent chooses alternative j ∈ {1, ..., J} in choice set s is

$$p_{js} = rac{\exp\left[oldsymbol{f}^T(oldsymbol{x}_{js})oldsymbol{eta}
ight]}{\sum_{t=1}^J \exp\left[oldsymbol{f}^T(oldsymbol{x}_{ts})oldsymbol{eta}
ight]}$$

# Model for choice data concerning mixtures

• We assume vector  $x_{js}$  contains the q ingredient proportions and r process variables

# Model for choice data concerning mixtures

- We assume vector  $\boldsymbol{x}_{js}$  contains the q ingredient proportions and r process variables
- Perceived utility modeled as

$$u_{js} = \boldsymbol{f}(\boldsymbol{x}_{js})^{T} \boldsymbol{\beta}$$
  
=  $\sum_{i=1}^{q-1} \gamma_{i}^{0*} x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^{q} \gamma_{ik}^{0} x_{ijs} x_{kjs} + \sum_{i=1}^{r} \sum_{k=1}^{q} \gamma_{k}^{i} x_{kjs} z_{ijs} + \sum_{i=1}^{r} \sum_{k=i+1}^{r} \alpha_{ik} z_{ijs} z_{kjs} + \sum_{i=1}^{r} \alpha_{i} z_{ijs}^{2}$ 



11 / 23

- D-optimality criterion
  - $\mathcal{D} = \det \left( oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta}) 
    ight)$

D-optimality criterion

$$\mathcal{D} = \det \left( \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{\beta}) \right) \longrightarrow$$
 prior distribution  $\pi(\boldsymbol{\beta})$ 

- D-optimality criterion
  - $\mathcal{D} = \det \left( \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{eta}) 
    ight)$
- Bayesian D-optimality criterion  $\mathcal{D}_B = \int_{\mathbb{R}^m} \det \left( \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{\beta}) \right) \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$

- D-optimality criterion
  - $\mathcal{D} = \det \left( \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{eta}) 
    ight)$
- Bayesian D-optimality criterion  $\mathcal{D}_B = \int_{\mathbb{R}^m} \det \left( \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{\beta}) \right) \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$
- Numerical approximation to Bayesian D-optimality criterion

$$\mathcal{D}_B \approx \frac{1}{R} \sum_{i=1}^R \det \left( \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{\beta}^{(i)}) \right)$$





I-optimality criterion

$$\mathcal{I} = \int_{\chi} oldsymbol{f}^T(oldsymbol{x}_{js})oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta})oldsymbol{f}(oldsymbol{x}_{js})doldsymbol{x}_{js}$$

I-optimality criterion

$$egin{aligned} \mathcal{I} &= \int_{\chi} oldsymbol{f}^T(oldsymbol{x}_{js})oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta})oldsymbol{f}(oldsymbol{x}_{js})doldsymbol{x}_{js} \ &= ext{tr}\left[oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta})oldsymbol{W}_u
ight] \end{aligned}$$

$$oldsymbol{W}_u = \int_{\chi} oldsymbol{f}(oldsymbol{x}_{js}) oldsymbol{f}^T(oldsymbol{x}_{js}) doldsymbol{x}_{js}$$

I-optimality criterion

$$egin{aligned} \mathcal{I} &= \int_{\chi} oldsymbol{f}^T(oldsymbol{x}_{js})oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta})oldsymbol{f}(oldsymbol{x}_{js})doldsymbol{x}_{js} \ &= ext{tr}\left[oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta})oldsymbol{W}_u
ight] \end{aligned}$$

Bayesian I-optimality criterion

$$\mathcal{I}_B = \int_{\mathbb{R}^m} \operatorname{tr} \left[ \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{\beta}) \boldsymbol{W}_u \right] \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$$

$$oldsymbol{W}_u = \int_{\chi} oldsymbol{f}(oldsymbol{x}_{js}) oldsymbol{f}^T(oldsymbol{x}_{js}) doldsymbol{x}_{js}$$

**KU LEUVEN** 

I-optimality criterion

$$egin{aligned} \mathcal{I} &= \int_{\chi} oldsymbol{f}^T(oldsymbol{x}_{js})oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta})oldsymbol{f}(oldsymbol{x}_{js})doldsymbol{x}_{js} \ &= ext{tr}\left[oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta})oldsymbol{W}_u
ight] \end{aligned}$$

Bayesian I-optimality criterion

$$\mathcal{I}_B = \int_{\mathbb{R}^m} \operatorname{tr} \left[ \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{\beta}) \boldsymbol{W}_u \right] \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}$$

Numerical approximation to Bayesian I-optimality criterion

$$\mathcal{I}_B \approx \frac{1}{R} \sum_{i=1}^{R} \operatorname{tr} \left[ \boldsymbol{I}^{-1}(\boldsymbol{X}, \boldsymbol{\beta}^{(i)}) \boldsymbol{W}_u \right]$$

$$oldsymbol{W}_u = \int_{\chi} oldsymbol{f}(oldsymbol{x}_{js}) oldsymbol{f}^T(oldsymbol{x}_{js}) doldsymbol{x}_{js}$$

**KU LEUVEN** 

# Results



Original experiment by Courcoux and Semenou

- Original experiment by Courcoux and Semenou
- September 2019: students from KU Leuven replicated the experiment with 35 respondents

- Original experiment by Courcoux and Semenou
- September 2019: students from KU Leuven replicated the experiment with 35 respondents
- Each respondent tasted 4 choice sets of size 2

- Original experiment by Courcoux and Semenou
- September 2019: students from KU Leuven replicated the experiment with 35 respondents
- Each respondent tasted 4 choice sets of size 2
- Simulated responses for temperature (process variable)

- Original experiment by Courcoux and Semenou
- September 2019: students from KU Leuven replicated the experiment with 35 respondents
- Each respondent tasted 4 choice sets of size 2
- Simulated responses for temperature (process variable)
- Prior distribution for parameter vector β in a second-order Scheffé model and MNL model for Bayesian D- and I-optimal designs


































1.0

0.5

0.0

-0.5

-1.0









15/23









17 / 23

• Experiment from the 1980s by Cornell

- Experiment from the 1980s by Cornell
- Interest in firmness of patties

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - $\circ$  mullet

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet



- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet







- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - $\circ$  sheepshead

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - $\circ$  mullet
  - $\circ$  sheepshead



- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - $\circ$  sheepshead
  - $\circ$  croaker

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - $\circ$  sheepshead
  - $\circ$  croaker



- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - $\circ$  sheepshead
  - $\circ$  croaker

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - sheepshead
  - $\circ$  croaker
- Subjected to different processing conditions:

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - sheepshead
  - croaker
- Subjected to different processing conditions:
  - oven cooking temperature (375 or 425 degrees Fahrenheit)

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - sheepshead
  - croaker
- Subjected to different processing conditions:
  - oven cooking temperature (375 or 425 degrees Fahrenheit)
  - oven cooking time (25 or 40 minutes)

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - sheepshead
  - croaker
- Subjected to different processing conditions:
  - oven cooking temperature (375 or 425 degrees Fahrenheit)
  - oven cooking time (25 or 40 minutes)
  - deep fat frying time (25 or 40 seconds)

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - sheepshead
  - croaker
- Subjected to different processing conditions:
  - oven cooking temperature (375 or 425 degrees Fahrenheit)
  - oven cooking time (25 or 40 minutes)
  - deep fat frying time (25 or 40 seconds)
- Assuming firmness is proportional to utility



- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - sheepshead
  - croaker
- Subjected to different processing conditions:
  - oven cooking temperature (375 or 425 degrees Fahrenheit)
  - oven cooking time (25 or 40 minutes)
  - deep fat frying time (25 or 40 seconds)
- Assuming firmness is proportional to utility
- Point estimates with 3 levels of uncertainty controlled by κ parameter





-0.5 0 0.5

z1

0.5 0 z2







20 / 23

#### **Contributions**

• Developed the theory for choice experiments involving mixtures and process variables

#### **Contributions**

- Developed the theory for choice experiments involving mixtures and process variables
- Implemented the theory in a coordinate-exchange algorithm

#### **Contributions**

- Developed the theory for choice experiments involving mixtures and process variables
- Implemented the theory in a coordinate-exchange algorithm
- Created an open-source R package with our algorithms (<u>https://github.com/mariobecerra/opdesmixr</u>)

#### **Contributions**

- Developed the theory for choice experiments involving mixtures and process variables
- Implemented the theory in a coordinate-exchange algorithm
- Created an open-source R package with our algorithms (<u>https://github.com/mariobecerra/opdesmixr</u>)
- Demonstrated that the I-optimal designs perform better than their D-optimal counterparts

#### **Contributions**

- Developed the theory for choice experiments involving mixtures and process variables
- Implemented the theory in a coordinate-exchange algorithm
- Created an open-source R package with our algorithms (<u>https://github.com/mariobecerra/opdesmixr</u>)
- Demonstrated that the I-optimal designs perform better than their D-optimal counterparts

#### Future work

• Extending the algorithm to find designs with an upper bound on the number of distinct mixtures and/or an upper bound on the number of distinct choice sets







#### **Contributions**

- Developed the theory for choice experiments involving mixtures and process variables
- Implemented the theory in a coordinate-exchange algorithm
- Created an open-source R package with our algorithms (<u>https://github.com/mariobecerra/opdesmixr</u>)
- Demonstrated that the I-optimal designs perform better than their D-optimal counterparts

#### Future work

- Extending the algorithm to find designs with an upper bound on the number of distinct mixtures and/or an upper bound on the number of distinct choice sets
- Models that take into account possible presence of consumer heterogeneity



#### More information

- Mario Becerra and Peter Goos. Bayesian I-optimal designs for choice experiments with mixtures. Chemometrics and Intelligent Laboratory Systems 217 (2021): 104395. DOI: 10.1016/j.chemolab.2021.104395
- Mario Becerra's website (with links to paper, R package, and code to reproduce the paper): <u>mariobecerra.github.io/</u>





#### Extra: Cocktail preferences





#### Extra: Fish patty experiment


#### Extra: Optimal design criteria

- D-optimal designs: low-variance estimators
- I-optimal designs: low-variance predictions
- Information matrix of multinomial logit model:  $I(X, \beta) = \sum X_s^T (P_s p_s p_s^T) X_s$
- With
  - $$\begin{split} \boldsymbol{P}_{s} &= \operatorname{diag}(\boldsymbol{p}_{s}) \\ \boldsymbol{p}_{s} &= (p_{1s}, ..., p_{Js})^{T} \\ \boldsymbol{X}_{s}^{T} &= [\boldsymbol{f}(\boldsymbol{x}_{js})]_{j \in \{1, ..., J\}} \\ \boldsymbol{X} &= [\boldsymbol{X}_{1}, ..., \boldsymbol{X}_{S}] \\ \boldsymbol{p}_{js} &= \frac{\exp\left[\boldsymbol{f}^{T}(\boldsymbol{x}_{js})\boldsymbol{\beta}\right]}{\sum_{t=1}^{J} \exp\left[\boldsymbol{f}^{T}(\boldsymbol{x}_{ts})\boldsymbol{\beta}\right]} \end{split}$$



#### Extra: Model for choice data concerning mixtures

- The attributes of the alternatives in a choice experiment are the ingredients of a mixture
- Vector  $x_{js}$  contains the q ingredient proportions and that  $f(x_{js})$  represents the model expansion of these proportions
- Most natural thing to do:

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

• Rewrite 
$$x_{qjs}$$
 as  $1 - x_{1js} - \dots - x_{q-1,js}$   
 $U_{js} = \mathbf{f}^T(\mathbf{x}_{js})\mathbf{\beta} = \sum_{i=1}^{q-1} \beta_i^* x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_{ijs} x_{kjs} + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^q \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} + \varepsilon_{js}$ 

• With  

$$f(\boldsymbol{x}_{js}) = (x_{1js}, x_{2js}, \dots, x_{q-1,js}, x_{1js}x_{2js}, \dots, x_{q-1,js}x_{qjs}, x_{1js}x_{2js}x_{3js}, \dots, x_{q-2,js}x_{q-1,js}x_{qjs})^T$$

$$\beta_i^* = \beta_i - \beta_q \text{ for } i \in \{1, \dots, q-1\}$$

$$\boldsymbol{x}_{js} = (x_{1js}, x_{2js}, \dots, x_{qjs})^T \qquad \boldsymbol{\beta} = (\beta_1^*, \beta_2^*, \dots, \beta_{q-1}^*, \beta_{1,2}, \dots, \beta_{q-1,q}, \beta_{123}, \dots, \beta_{q-2,q-1,q})^T$$

# Fish patty experiment

**D-optimal** 



#### **I**-optimal



18 / 23

## **D-optimal designs**

• D-optimality criterion 
$$\mathcal{D} = \det \left( oldsymbol{I}^{-1}(oldsymbol{X},oldsymbol{eta}) 
ight)$$

#### **D-optimal designs**

• D-optimality criterion  

$$\mathcal{D} = \det \left( I^{-1}(X, \beta) \right)$$
  $\longrightarrow$  point estimate

## **D-optimal designs**



## Fish patty experiment

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - $\circ$  sheepshead





## Fish patty experiment

- Experiment from the 1980s by Cornell
- Interest in firmness of patties
- Three fish species:
  - mullet
  - $\circ$  sheepshead



