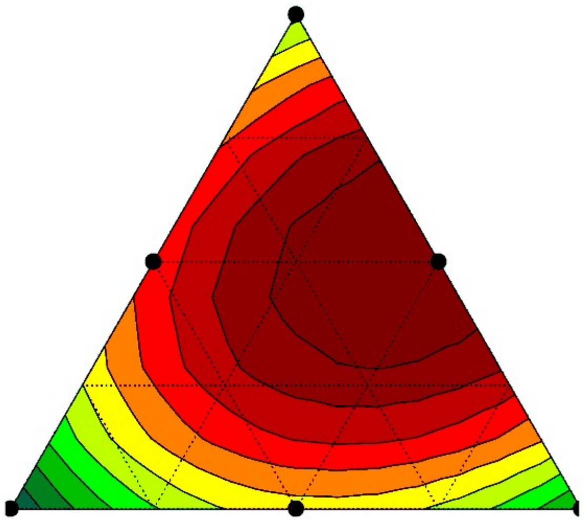
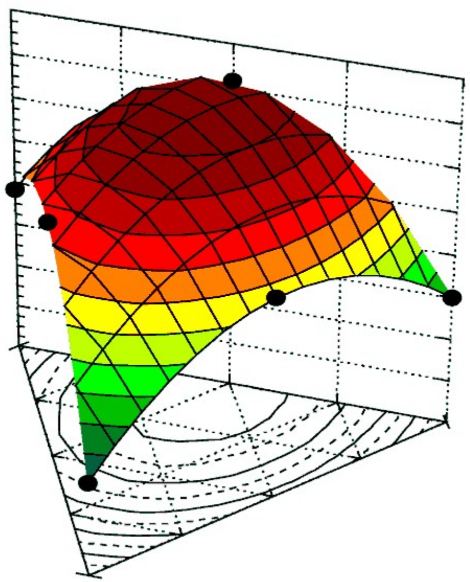
Cours sur les plans de mélange





F. Husson
Département de statistique & informatique
francois.husson@institut-agro.fr
https://husson.github.io/MOOC_plan/plan_melange.pdf

Exemple d'utilisation des plans de mélange

Objectif : une firme alimentaire produit des jus de fruit et désire mettre sur le marché un nouveau cocktail de fruits à base de :

- jus d'orange
- jus de banane
- jus de mangue
- et en ajoutant ou non un additif pour la couleur

- Evaluation du goût par un jury de consommateurs

Qu'est-ce qu'un plan de mélange ?

Mélange :

mixture obtenue en mélangeant différents ingrédients en certaines proportions

Caractéristiques d'un plan de mélange :

- Les facteurs mis en jeu sont les proportions des différents ingrédients (et non les quantités absolues)
- Le domaine des facteurs est contraint : $X_1+X_2+X_3 \ldots + X_k = 1$
- Les réponses ne sont pas influencées par les quantités absolues des facteurs mais par les proportions relatives de ceux-ci

Comment étudier un plan de mélange ?

Modèle pour les plans de mélange :

- Rappel du modèle de régression d'ordre 1 : $Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i$

  Mais il faut prendre en compte la contrainte : $X_1 + X_2 + ... + X_k = 1$

- Modèle d'ordre 1 : $Y_i = \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i$ k

  nb paramètres

- Modèle d'ordre 2 : $Y_i = \sum_{k=1}^{K} \beta_k x_{ik} + \sum_{k=2}^{K} \sum_{j<k}^{K-1} \delta_{jk} x_{ij} x_{ik} + \varepsilon_i$ k(k+1) / 2

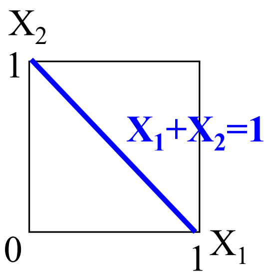
- Modèle d'ordre 3 : prend en compte des interations d'ordre 3 (utilité ?)

Domaine expérimental d’un mélange

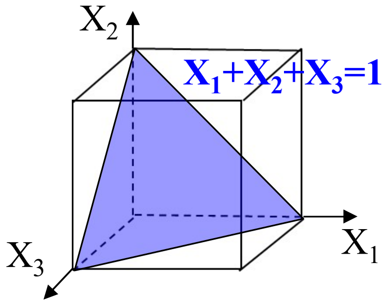
Le domaine expérimental d’un mélange est contraint par la relation :

$$X_1+X_2+X_3 \ldots + X_k = 1$$

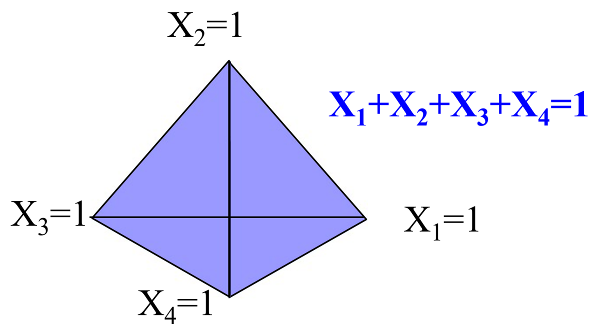
K=2



K=3

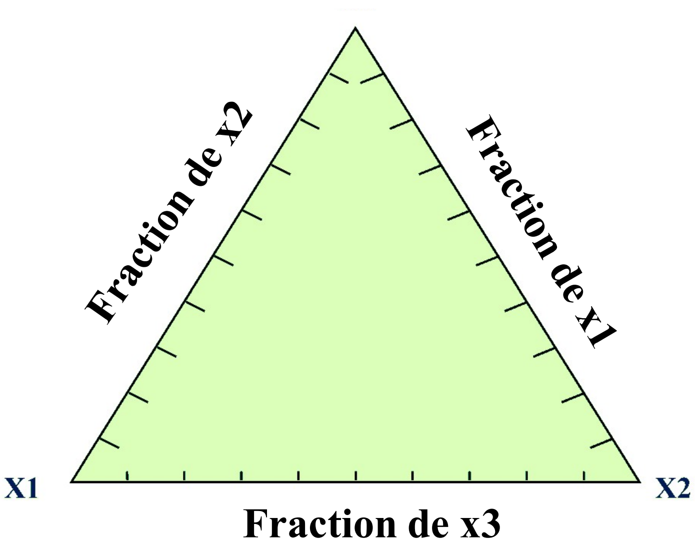


K=4

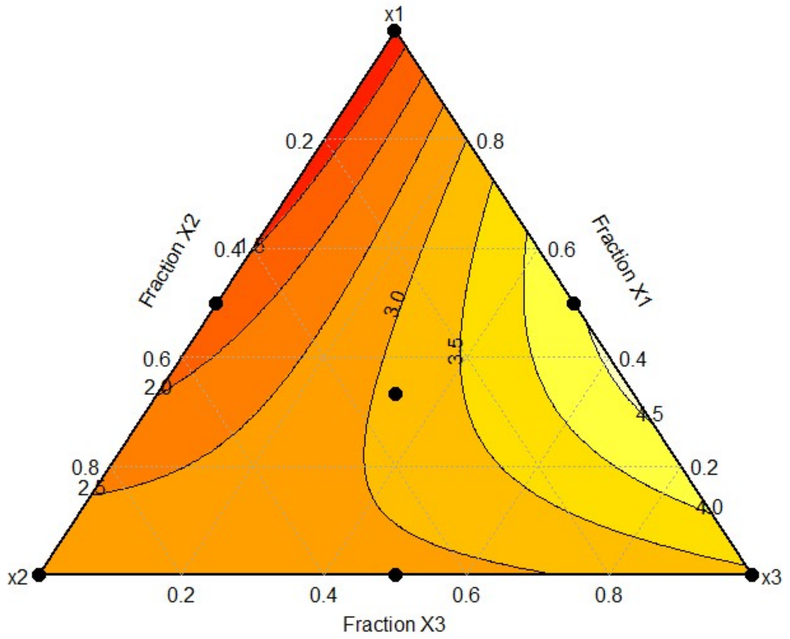


Représentation triangulaire d’un mélange

Représentation triangulaire
d’un domaine de mélange à
trois facteurs

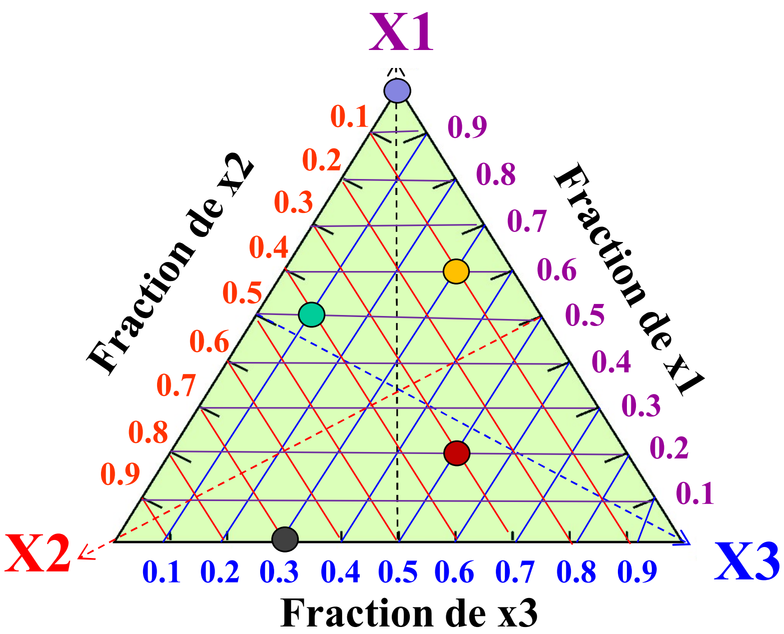


Surface de réponse dans
un domaine de mélange
à trois facteurs



Représentation triangulaire d’un mélange

Représentation triangulaire
d’un domaine de mélange à
trois facteurs



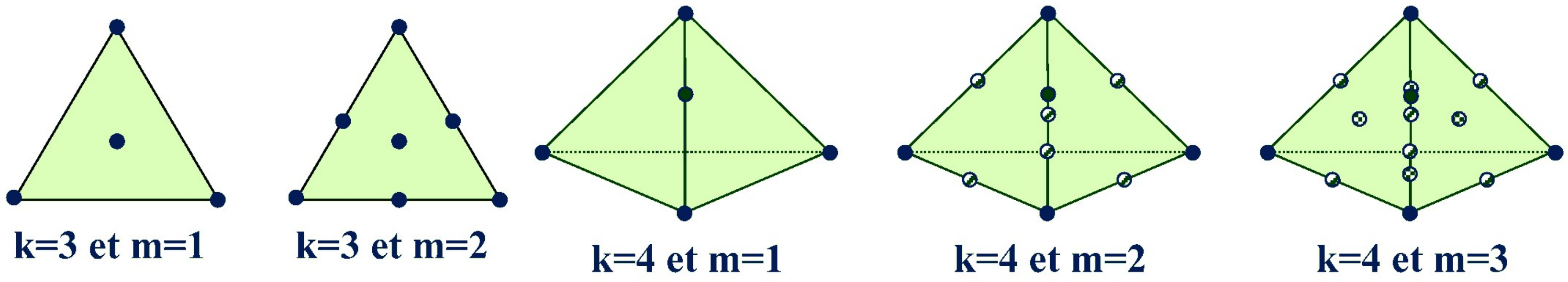
Quel mélange pour les
expériences suivantes ?

- (0.5, 0.4, 0.1)
- (0, 0,7, 0.3)
- (0.2, 0.3, 0.5)
- (0.6, 0.1, 0.3)
- (1, 0, 0)

Plan de mélange simplex centroïd

Un plan **simplex centroïd** à k facteurs étudié avec un modèle de degré m est constitué des mélanges suivants ($m < k$) :

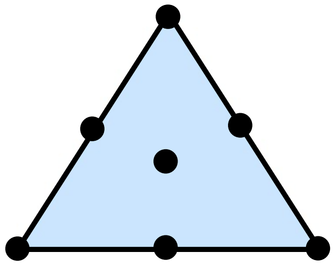
- ✓ chaque constituant pur
- ✓ mélange de 2 constituants en proportions égales
- ✓ …
- ✓ mélange de m constituants en proportions égales
- ✓ d'essais au centre du domaine (tous les constituants en proportions égales)



Bien souvent en pratique, on se contente de $m \leq 2$

Plan de mélange simplex centroïde

Exemple : plan **simplex centroïd** à 3 facteurs étudié avec un modèle de degré 2

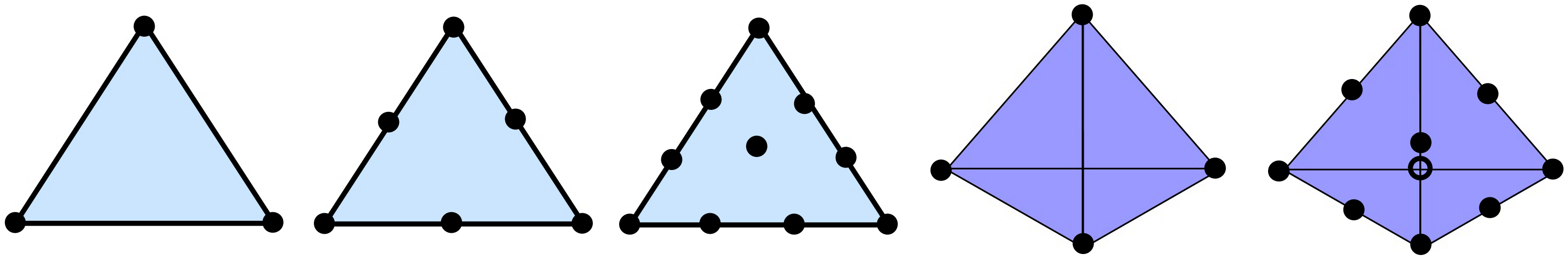


Essai	X1	X2	X3
1	1	0	0
2	0	1	0
3	0	0	1
4	0.5	0.5	0
5	0.5	0	0.5
6	0	0.5	0.5
7	1/3	1/3	1/3

Plan de mélange : réseau simplex (réseau de Scheffé)

Un **réseau simplex** pour étudier k facteurs avec un modèle de degré (au plus) m est constitué de toutes les combinaisons possibles de m+1 niveaux pour chaque constituant (m+1 niveaux sont testés pour chaque facteur) :

$$(0,\ 1/m,\ 2/m,\ 3/m,\ \ldots,\ 1)$$



k = 3 et m = 1 — k = 3 et m = 2 — k = 3 et m = 3 — k = 4 et m = 1 — k = 4 et m = 2

Grille régulière dans le domaine qui peut être considéré comme un plan factoriel complet dans le cas classique des plans

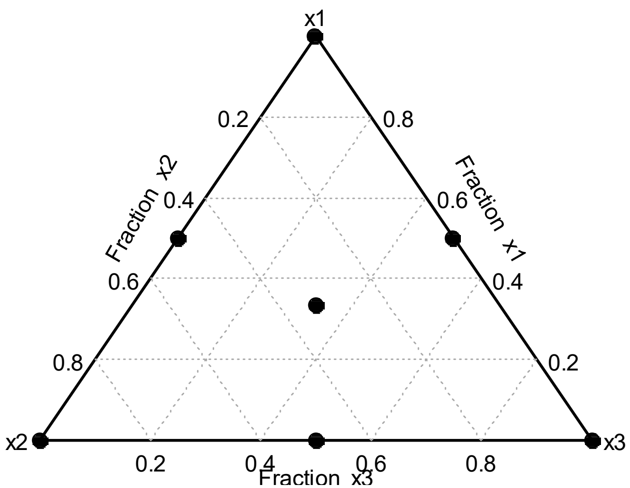
Construction de plans avec R

Construction du plan centroïd

```
library(mixexp)
plan <- SCD(fac = 3)
DesignPoints(plan)
```

Nb d'ingrédients

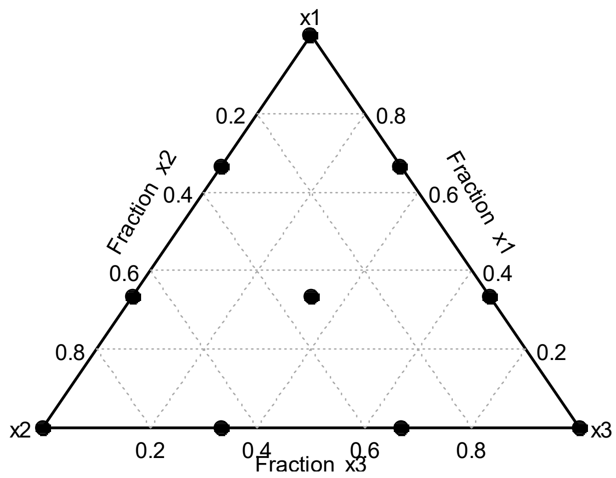
	x1	x2	x3
1	1.0000000	0.0000000	0.0000000
2	0.0000000	1.0000000	0.0000000
3	0.0000000	0.0000000	1.0000000
4	0.5000000	0.5000000	0.0000000
5	0.5000000	0.0000000	0.5000000
6	0.0000000	0.5000000	0.5000000
7	0.3333333	0.3333333	0.3333333



Construction d'un réseau de Scheffé

```
plan2 <- SLD(fac = 3, lev=3)
DesignPoints(plan2)
```

	x1	x2	x3
1	1.0000000	0.0000000	0.0000000
2	0.6666667	0.3333333	0.0000000
3	0.3333333	0.6666667	0.0000000
4	0.0000000	1.0000000	0.0000000
5	0.6666667	0.0000000	0.3333333
6	0.3333333	0.3333333	0.3333333
7	0.0000000	0.6666667	0.3333333
8	0.3333333	0.0000000	0.6666667
9	0.0000000	0.3333333	0.6666667
10	0.0000000	0.0000000	1.0000000



Plan de mélange de type II

Contraintes sur les facteurs : $I_i \leq X_i \leq 1$

Quelle est la forme du domaine et quel plan utiliser ?

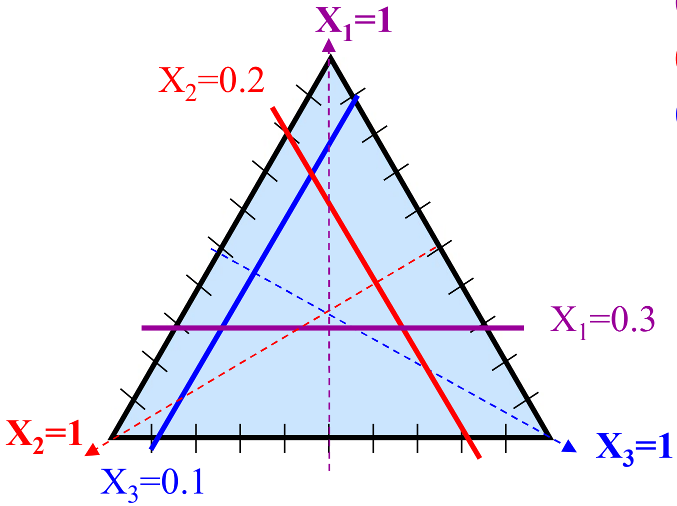
Exemple : $0.3 \leq$ orange ≤ 1; $0.2 \leq$ banane ≤ 1; $0.1 \leq$ mangue ≤ 1;

Contraintes :

$0.3 \leq$ orange ≤ 1

$0.2 \leq$ banane ≤ 1

$0.1 \leq$ mangue ≤ 1



Contraintes réelles :

$0.3 \leq$ orange $\leq 1\text{-}0.2\text{-}0.1=0.7$

$0.2 \leq$ banane $\leq 1\text{-}0.3\text{-}0.1=0.6$

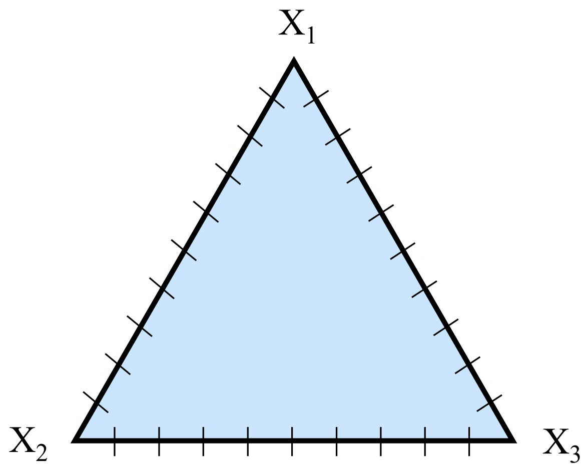
$0.1 \leq$ mangue $\leq 1\text{-}0.3\text{-}0.2=0.5$

```
library(mixexp)
plan <- Xvert(3, lc=c(.3,0.2,0.1), uc=c(1,1,1), ndm=1)
DesignPoints(plan)
```

Plan de mélange de type III

Contraintes sur les facteurs : $I_i \leq X_i \leq S_i$

Exemple : $0.3 \leq$ orange ≤ 0.6 ;
$0.2 \leq$ banane ≤ 0.5;
$0.1 \leq$ mangue ≤ 0.4;



Quelle est la forme du domaine et quel plan utiliser ?

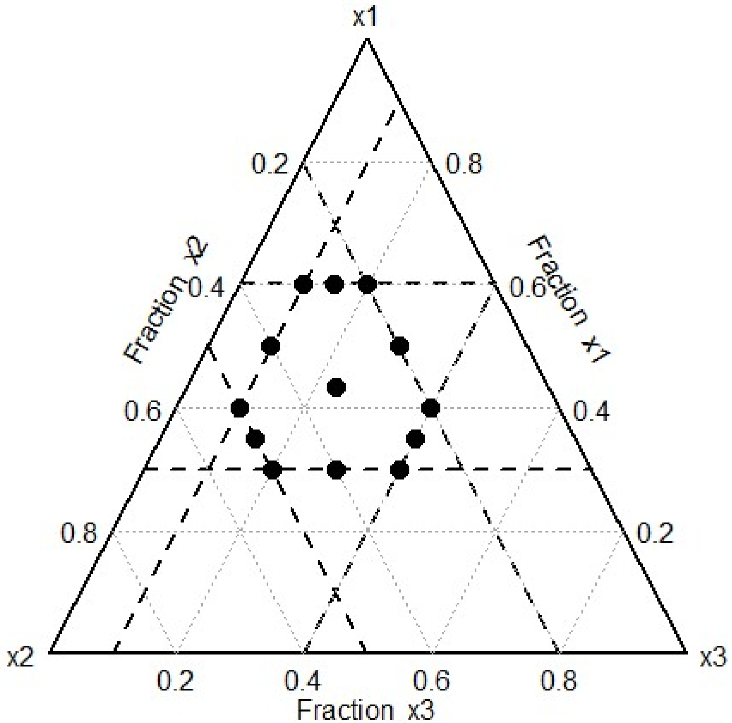
Plan de mélange de type III

Contraintes :

$0.3 \leq$ orange ≤ 0.6

$0.2 \leq$ banane ≤ 0.5

$0.1 \leq$ mangue ≤ 0.4




```
plan <- Xvert(3,lc=c(.3,0.2,0.1),uc=c(0.6,0.5,0.4),ndm=1)
DesignPoints(plan)
```

Quand les proportions des constituants sont soumises à des contraintes inférieures ET supérieures, le domaine expérimental est un polyèdre irrégulier

Plans D-optimaux

Un plan tel que l'estimation des effets de chaque variable est la plus précise possible, et qui prend en compte les contraintes pratiques

Rappel : $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$

Dépend uniquement du **choix** des expériences

Variabilité résiduelle : dépend des résultats des expériences

⇒ trouver les expériences telles que $(X'X)^{-1}$ soit minimale

Plan D-optimal minimise le déterminant de la matrice $(X'X)^{-1}$

Plans D-optimaux

Principe :

- Beaucoup d'essais candidats
- N essais choisis au hasard puis algorithme d'échange : échange entre un essai du plan et un essai hors plan effectué si cela diminue le déterminant de $(X'X)^{-1}$
- L'algorithme de Federov s'arrête quand aucun échange n'améliore le critère

Avantages :

- Plan très flexibles par rapport au nombre d'essais
- Possibilité d'imposer des essais
- Démarche séquentielle possible (modèle du 1er ordre, si le modèle n'est pas adapté, l'algorithme fournit les essais supplémentaires)

Inconvénient : un plan est toujours fourni. Est-il de bonne qualité ?

Plan de mélange de type III

Générer un ensemble de *m* essais candidats :

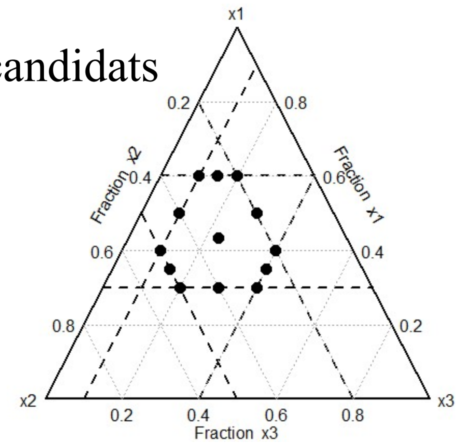
✓Sommets du polyèdre

✓Milieux des arêtes et des faces

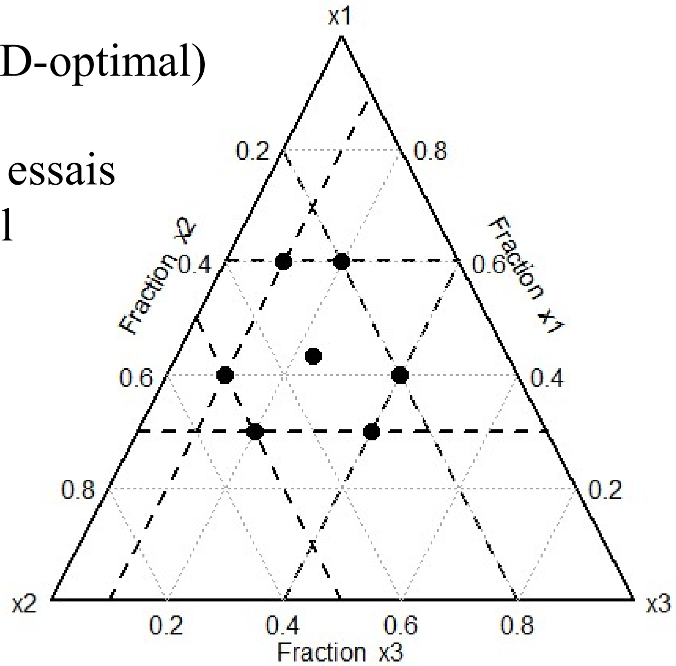
✓Centre du polyèdre

Puis choisir les essais à conserver (algorithme D-optimal)

Essais candidats



Plan en 7 essais D-optimal




```
library(AlgDesign)
planD=optFederov(~ -1 + x1+x2+x3+x1:x2+x1:x3+x2:x3, plan, nTrials=7)
DesignPoints(planD$design)
```

Plan de mélange de type IV et V

Mélange de type IV

Un des facteurs est très dominant, les autres en faible quantité.
Exemple : Eau entre 99% et 99.5%, acide citrique entre 0.5% et 1%, carraghénane entre 0.5% et 1%
Construire un plan d'expérience pour facteurs quantitatifs sur les additifs uniquement (et compléter à l'eau)

Mélange de type V

Problème avec des facteurs de mélange et des facteurs quantitatifs et/ou qualitatifs
Solution : Répéter le plan de mélange pour les différents niveaux des facteurs quantitatifs ou qualitatifs
Si cela engendre trop d'essais, extraire un plan D-optimal à partir de cette liste d'essais candidats

Plan de type V avec sélection d'essais

Construction d'un plan de mélange avec un facteur quantitatif

```
library(mixexp)
plan <- Xvert(3,lc=c(.3,.2,.1),uc=c(0.6,0.5,0.4),ndm=1)
plan2 <- rbind.data.frame(plan[,1:3], plan[,1:3], plan[,1:3])
names(plan2) <- c("orange", "banane", "mangue")
fac.quanti <- rep(c(5,10,15), each=nrow(plan))
plan3 <- cbind.data.frame(plan2,fac.quanti)
```

contraintes sur les var. de mélange

facteur quantitatif prenant 3 valeurs possibles : 5, 10, 15

Construction d'un plan optimal en 12 essais

```
library(AlgDesign)
planD=optFederov(~ -1 + orange+banane+mangue + orange:banane + orange:mangue +
    banane:mangue + fac.quanti + I(fac.quanti^2), plan3,nTrials=12)
```

	orange	banane	mangue	fac.quanti
1	0.6000000	0.2000000	0.2000000	5
4	0.4000000	0.5000000	0.1000000	5
6	0.3000000	0.3000000	0.4000000	5
13	0.4333333	0.3333333	0.2333333	5
15	0.3000000	0.5000000	0.2000000	10
16	0.6000000	0.3000000	0.1000000	10
18	0.4000000	0.2000000	0.4000000	10
26	0.4333333	0.3333333	0.2333333	10
27	0.6000000	0.2000000	0.2000000	15
31	0.4000000	0.2000000	0.4000000	15
33	0.3000000	0.4000000	0.3000000	15
37	0.5000000	0.4000000	0.1000000	15

Représentation de surfaces de réponse avec R

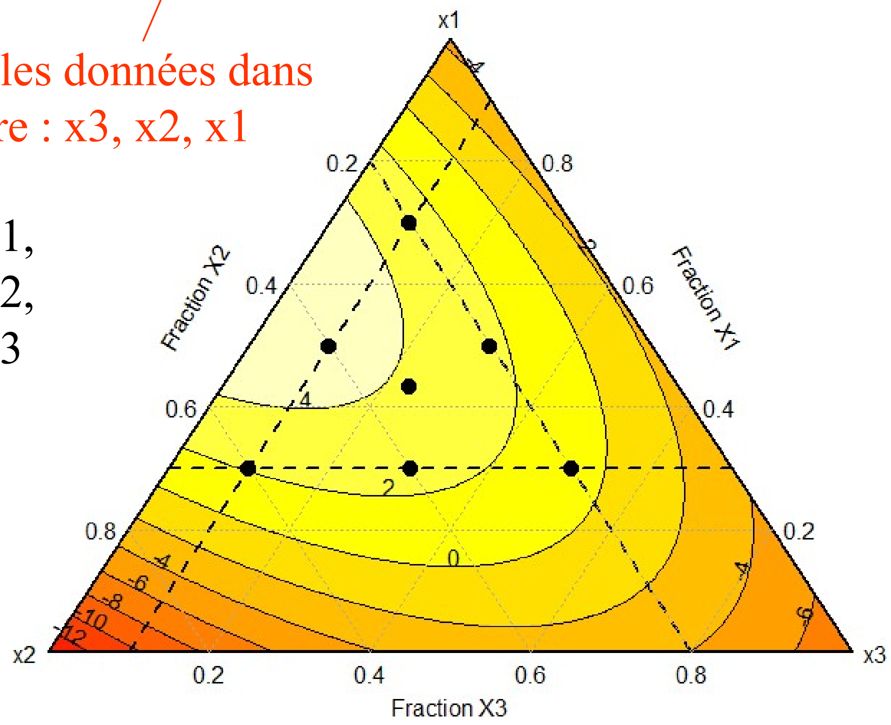
Dessin de la surface de réponse

mod=2 pour modèle quadratique

```
library(mixexp)
plan <- Xvert(3,lc=c(.3,.2,.1),uc=c(1,1,1),ndm=1)
Y <- c(1,2.4,3.2,1.4,1.2,3.6,6.4)
MixturePlot(x=plan$x3,y=plan$x2,z=plan$x1, w=Y, cols=TRUE, mod=2,
  lims=c(.3,1,.2,1,.1,1), constrts = TRUE)
```

Mettre les données dans cet ordre : x3, x2, x1

lim inf x1, lim sup x1,
lim inf x2, lim sup x2,
lim inf x3, lim sup x3



Détermination d'un optimum

Estimation des paramètres du modèle

```
mod=lm(Gout~ -1+(X1+X2+X3)^2+X4+ I(X4^2),data=don)
summary(mod)
           Estimate Std. Error t value Pr(>|t|)
X1        -19.91189    4.03501  -4.935 0.001685 **
X2        -17.96241    4.10318  -4.378 0.003244 **
X3        -11.54983    5.09151  -2.268 0.057609 .
X4          1.04690    0.13971   7.493 0.000138 ***
I(X4^2)    -0.05935    0.00708  -8.382 6.76e-05 ***
X1:X2      67.91227   14.70224   4.619 0.002429 **
X1:X3      52.47022   14.65999   3.579 0.008986 **
X2:X3      43.43361   11.18708   3.882 0.006034 **
```

Données

X1	X2	X3	X4	Gout
0.4	0.5	0.1	5	3.25
0.4	0.5	0.1	5	3.75
0.55	0.2	0.25	5	3
0.3	0.3	0.4	5	5.25
0.3	0.3	0.4	5	5.75
0.42	0.34	0.24	5	5
0.55	0.35	0.1	10	3.5
0.3	0.5	0.2	10	4.75
0.3	0.5	0.2	10	5.25
0.4	0.2	0.4	10	5.5
0.42	0.34	0.24	10	6
0.55	0.35	0.1	15	1.5
0.55	0.2	0.25	15	2
0.4	0.2	0.4	15	3.5
0.3	0.4	0.3	15	3.5

Détermination de l'optimum avec la fonction optim

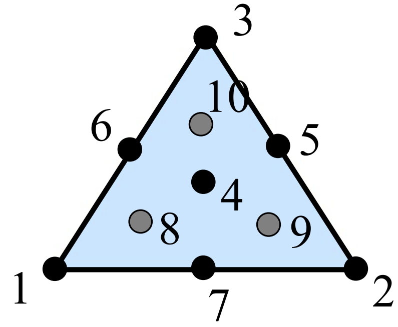
```
fct.a.opt = function(x){
  x1 = x[1]  ;  x2 = x[2]  ; x4 = x[3]
  x3 = 1-x[1]-x[2] ### On impose la contrainte sur un des paramètres
  Y = -19.9*x1-18*x2-11.5*x3+1.05*x4+67.9*x1*x2+52.47*x1*x3+43.4*x2*x3-0.059*x4^2
}
yy=optim(c(0.4,0.3,10),fct.a.opt,lower=c(0.3,0.2,5),upper=c(0.6, 0.5, 15), control$fnscale= -1)
```

Initialisation des paramètres — Bornes inf paramètres — Bornes sup paramètres — Mettre -1 pour chercher un maximum

```
$par
[1] 0.3354582 0.3170189 8.8983053
$value
[1] -6.412429
```

Démarche statistique

Construire un plan avec peu d’essais et estimer le modèle avec les effets linéaires.
Valider le modèle grâce au point au centre
si le modèle n’est pas validé, ajouter des essais et considérer
un modèle avec effets quadratiques
Valider le modèle avec les point au centre des domaines



Exemple pour un mélange de 3 ingrédients :

- Le réseau Scheffé {3;1} comporte 3 points pour un polynôme de degré 1
- Faire les essais pour les points 1, 2 et 3 ce qui permet d’estimer 3 paramètres a1, a2, a3
- Calculer la prévision pour le centre du domaine (essai 4) et faire l’essai de validation,
- Si le modèle du 1er degré est rejeté (à l’incertitude de mesure près), faire l’hypothèse d’un modèle quadratique :
- Le réseau Scheffé {3;2} comporte 6 points pour un polynôme de degré 2,
- Faire les essais supplémentaires 5, 6 et 7 et estimer les 6 paramètres a1, a2, a3, a12, a23, a13
- Calculer les prévisions pour les centres des domaines 8, 9, 10 et éventuellement 4.

Construction de plans avec R
package RcmdrPlugin.DoE

Package très complet MAIS il faut bien connaître les plans d'expériences pour pouvoir l'utiliser

Certaines procédures ne sont pas encore programmées, notamment avec les plans de mélange (construction + dépouillement)