



# Planification optimale des expériences

## Introduction et cas du modèle linéaire

Frédéric Bertrand<sup>1</sup>

<sup>1</sup>IRMA, Université de Strasbourg  
Strasbourg, France

ENSAI 3<sup>e</sup> Année

19-03-2012



## Référence

Ce cours s'appuie essentiellement sur

1. le livre de Hardeo Sahai et Mohammed I. Ageel, **The Analysis of Variance : Fixed, Random and Mixed Models**, aux éditions Birkhäuser, 2000.
2. le livre de Hardeo Sahai et Miguel Ojeda, **Analysis of Variance for Random Models, Volume I: Balanced data**, aux éditions Birkhäuser, 2004.
3. le livre de Hardeo Sahai et Miguel Ojeda, **Analysis of Variance for Random Models, Volume II: Unbalanced data**, aux éditions Birkhäuser, 2005.
4. le livre de Douglas C. Montgomery, **Design and Analysis of Experiments**, 7<sup>th</sup> Edition, aux éditions Wiley, 2009.
5. le livre de Samuel D. Silvey, **Optimal Design**, aux éditions Chapman and Hall, 1980.



# Sommaire

## Introduction

- Plans pour pesée
- Définition générale du problème
- Exemples

## Planification, modèle linéaire et symétries

- Définitions, admissibilité, optimalité
- Invariances
- Cas de l'isovariance

## Caractérisation différentielle de l'optimalité

- Définitions, première propriétés
- Théorèmes de caractérisation



# Sommaire

## Introduction

Plans pour pesée

Définition générale du problème

Exemples

## Planification, modèle linéaire et symétries

Définitions, admissibilité, optimalité

Invariances

Cas de l'isovariance

## Caractérisation différentielle de l'optimalité

Définitions, première propriétés

Théorèmes de caractérisation



## Introduction

La planification d'expérience traite, comme son nom l'indique, de l'organisation réfléchie des expériences de manière à pouvoir

1. analyser celles-ci avec des modèles permettant de reproduire fidèlement la, plus ou moins grande, complexité des processus expérimentaux à décrire ;
2. maximiser la quantité d'information (au sens de Fisher par exemple) que nous apporte les essais que nous choisirons de réaliser.



















## Plans pour pesée

### Comparaisons de deux plans

Ainsi pour le premier plan, nous obtenons avec le théorème de Gauss-Markov et la formule  $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = y$ , les BLUE

1.  $\widehat{\beta}_1 = y_1, \text{Var}[\widehat{\beta}_1] = \sigma_e^2;$
2.  $\widehat{\beta}_2 = y_2, \text{Var}[\widehat{\beta}_2] = \sigma_e^2;$
3.  $\widehat{\beta}_3 = y_3, \text{Var}[\widehat{\beta}_3] = \sigma_e^2;$
4.  $\widehat{\beta}_4 = y_4, \text{Var}[\widehat{\beta}_4] = \sigma_e^2.$



## Plans pour pesée

### Comparaisons de deux plans

Ainsi pour le premier plan, nous obtenons avec le théorème de Gauss-Markov et la formule  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{y}$ , les BLUE

1. Pesée 1 :  $\hat{\beta}_1 = \frac{1}{4} (y_1 + y_2 + y_3 + y_4)$ ,  $\text{Var}[\hat{\beta}_1] = \frac{\sigma_e^2}{4}$  ;
2. Pesée 2 :  $\hat{\beta}_2 = \frac{1}{4} (y_1 + y_2 - y_3 - y_4)$ ,  $\text{Var}[\hat{\beta}_2] = \frac{\sigma_e^2}{4}$  ;
3. Pesée 3 :  $\hat{\beta}_3 = \frac{1}{4} (y_1 - y_2 - y_3 + y_4)$ ,  $\text{Var}[\hat{\beta}_3] = \frac{\sigma_e^2}{4}$  ;
4. Pesée 4 :  $\hat{\beta}_4 = \frac{1}{4} (y_1 - y_2 + y_3 - y_4)$ ,  $\text{Var}[\hat{\beta}_4] = \frac{\sigma_e^2}{4}$  .



## En conclusion

Le second plan pour pesée est le plus intéressant puisque la variance des estimateurs des masses des objets est quatre fois plus petite.

Il vaut mieux combiner astucieusement les objets pour la balance pour les peser tous quatre fois, plutôt que de les peser chacun directement. En effet, dans le premier cas, l'estimateur est un moyenne de quatre termes non corrélés, ce qui réduit sa variance.





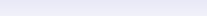


## Définition générale du problème

### Modélisation de la réponse $\tilde{y}$

Nous supposons que la loi de probabilité de la variable aléatoire  $\tilde{y}$  dépend de :

1. un vecteur colonne  $u = (u_1, u_2, \dots, u_r)'$  de variables réelles appelées variables **contrôlées** parce qu'elles ont été choisies par l'expérimentateur et puisque les valeurs qu'elles prennent sont parfaitement connues et peuvent être fixées par lui ;



## Définition générale du problème

### Modélisation de la réponse $\tilde{y}$

Nous supposons que la loi de probabilité de la variable aléatoire  $\tilde{y}$  dépend de :

2. un vecteur colonne  $\theta = (\theta_1, \theta_2, \dots, \theta_k)'$  de paramètres qui sont fixes mais dont les valeurs sont **inconnues** de l'expérimentateur ; ce sont les valeurs de ces paramètres, ou de fonctions de ces paramètres, que l'expérimentateur essaye de connaître ;
3. un vecteur colonne  $\tau = (\tau_1, \tau_2, \dots, \tau_r)'$  de paramètres de nuisance ; ceux-ci sont également fixes et **inconnus** mais l'expérimentateur n'est pas particulièrement intéressé par le fait de connaître leur valeur.

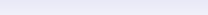


## Définition générale du problème

### Paramètres d'intérêts et de nuisance

La distinction entre les paramètres d'intérêt et les paramètres de nuisance peut être difficile à faire, en particulier lorsque notre attention porte sur l'estimation d'un nombre de fonctions de  $\theta$  strictement à inférieure au nombre  $k$  de paramètres.

Néanmoins, typiquement les vecteurs  $u$  et  $\theta$  apparaîtront dans  $\mathbb{E}[\tilde{y}]$  et  $\tau$  dans la « variance des erreurs ».



## Définition générale du problème

### Domaines d'appartenance des vecteurs

1. Les vecteurs  $u$  peut être choisis dans la partie  $\mathcal{U} \subset \mathbb{R}^r$  ;
2. Nous savons que le vrai vecteur de paramètres  $\theta$  appartient à la partie  $\Theta \subset \mathbb{R}^k$  ;
3. Nous savons que le vrai vecteur de paramètres  $\tau$  appartient à la partie  $\mathcal{T} \subset \mathbb{R}^l$  ;



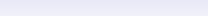
## Définition générale du problème

### Loi de $\tilde{y}$

Nous supposons que pour, pour des valeurs données de  $u$ ,  $\theta$  et  $\tau$ , la loi de  $\tilde{y}$  est donnée par une densité de probabilité par rapport à une mesure  $\sigma$ —finie.

Si la variable  $\tilde{y}$  est

- discrète cette mesure est la mesure de comptage ;
- continue cette mesure est la mesure de Lebesgue.



## Définition générale du problème

### Principe de la planification

L'expérimentateur est autorisé à réaliser  $N$  observations indépendantes de la réponse  $\tilde{y}$  aux points  $u_{(1)}, u_{(2)}, \dots, u_{(N)}$  qui auront été choisis par lui dans l'ensemble  $\mathcal{U}$ .

Ce choix des  $N$  vecteurs, s'appelle le choix du plan expérimental en  $N$  essais ou d'un  $N$ -plan.

La question classique est alors de savoir comment construire un  $N$ -plan « intéressant ».



## Définition générale du problème

### Principe de la planification

Pour évaluer la qualité d'un  $N$ -plan plusieurs approches ont été proposées :

1. Théorie de la décision, voir Brooks (1972), Bandemer (1979) pour les premières utilisations, en définissant une fonction d'utilité et une distribution a priori sur  $\theta$  et  $\tau$  et en calculant la richesse, au sens de Bayes, apportée par chaque  $N$ -plan. Il « suffit » alors de choisir l'un des plans qui maximise cette richesse.
2. Utilisation de l'information au sens de Fisher.

La seconde approche est la plus souvent utilisée car la première, bien que simple en apparence, pose de réels problèmes aussi bien pratiques que théoriques.



## Définition générale du problème

### Information de Fisher

Supposons que nous souhaitons estimer  $\theta$ .

Nous supposons que la famille

$$\{p(y|u, \theta, \tau) ; u \in \mathcal{U}, \theta \in \Theta, \tau \in \mathcal{T}\}$$

est suffisamment régulière pour que la matrice d'information de Fisher existe et qu'il soit possible d'utiliser la seconde caractérisation de l'information.





## Définition générale du problème

### Information de Fisher

Pour une observation de  $\tilde{y}$  au vecteur  $u$ , la matrice partitionnée d'information de Fisher adaptée à l'estimation de  $(\theta, \tau)$  est

$$J(u; \theta, \tau) = \begin{pmatrix} J_{\theta\theta}(u) & J_{\theta\tau}(u) \\ J'_{\theta\tau}(u) & J_{\tau\tau}(u) \end{pmatrix}.$$

où  $J_{\theta\theta}(u)$  est la matrice carrée d'ordre  $k$  dont la coordonnée  $(i, j)$  est

$$\mathbb{E} \left[ - \frac{\partial^2 \log p(y|u, \theta, \tau)}{\partial \theta_i \partial \theta_j} \right]$$

et  $J_{\theta\tau}(u)$  et  $J_{\tau\tau}(u)$  sont définies de manière similaire.



## Définition générale du problème

### Information de Fisher

Par conséquent la matrice d'information pour  $N$  observations indépendantes de  $\tilde{y}$  réalisées aux points  $u_{(1)}, u_{(2)}, \dots, u_{(N)}$  est

$$L(\mathbf{u}; \theta, \tau) = \sum_{i=1}^N J(u_{(i)}; \theta, \tau) = \begin{pmatrix} L_{\theta\theta}(\mathbf{u}) & L_{\theta\tau}(\mathbf{u}) \\ L_{\theta\tau}(\mathbf{u}) & L_{\tau\tau}(\mathbf{u}) \end{pmatrix}.$$

où  $\mathbf{u} = (u_{(1)}, \dots, u_{(N)})$ ,  $L_{\theta\theta}(\mathbf{u}) = \sum_i^N J_{\theta\theta}(u_{(i)})$  et  $L_{\theta\tau}(\mathbf{u})$  et  $L_{\tau\tau}(\mathbf{u})$  sont définies de manière similaire.



## Définition générale du problème

### Information de Fisher

Supposons maintenant que  $\mathbf{u}$  est tel que la matrice  $L(\mathbf{u}; \theta, \tau)$  est inversible<sup>1</sup>.

L'**inégalité de Fréchet-Darmonis-Cramer-Rao** fournit alors une borne inférieure :

1.  $L^{-1}(\mathbf{u}; \theta, \tau)$ , pour la variance d'un estimateur sans biais de  $(\theta, \tau)$  ;
2.  $V(\mathbf{u}; \theta, \tau) = (L_{\theta\theta}(\mathbf{u}) - L_{\theta\tau}(\mathbf{u})L_{\tau\tau}^{-1}(\mathbf{u})L'_{\theta\tau}(\mathbf{u}))^{-1}$ , pour la variance d'un estimateur sans biais de  $\theta$ .

$V(\mathbf{u}; \theta, \tau)$  est en fait la première matrice principale carré d'ordre  $k$  extraite de  $L^{-1}(\mathbf{u}; \theta, \tau)$ .

---

1. Nous verrons dans la suite que nous ne pourrons pas toujours faire cette hypothèse et comment l'affaiblir.



## Définition générale du problème

### Information de Fisher

Il est important de noter que

1. les bornes inférieures précédentes dépendent du plan  $\mathbf{u}$  qui a été choisi par l'expérimentateur. Comme nous l'avons vu, un choix judicieux peut donc améliorer la précision de l'estimation ;
2. dans le cadre du modèle linéaire gaussien,  $V(\mathbf{u}; \theta, \tau)$  est exactement la variance de l'estimateur de Gauss-Markov de  $\theta$ .
3. si  $N$  est grand, et sous des hypothèses de régularité suffisante, la matrice de variance-covariance de l'estimateur par maximum de vraisemblance de  $\theta$  est approximativement  $V(\mathbf{u}; \theta, \tau)$  puisque ces estimateurs sont asymptotiquement efficaces ;



## Définition générale du problème

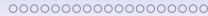
### Estimation d'une fonction du paramètre

Supposons que nous souhaitons estimer les fonctions à valeurs réelles  $g_1(\theta), \dots, g_s(\theta)$  à la place de  $\theta$ .

Posons  $g$  la fonction  $(g_1, \dots, g_s)'$  à valeurs dans  $\mathbb{R}^s$  et notons

$Dg(\theta) = \left( \left( \frac{\partial g_j(\theta)}{\partial \theta_i} \right) \right)_{1 \leq i, j \leq n}$  sa différentielle au point  $\theta$

lorsqu'elle existe.



## Définition générale du problème

### Estimation d'une fonction du paramètre

Sous des hypothèses de régularité suffisante et pour un plan  $\mathbf{u}$  donné, une borne inférieure pour la matrice de variance-covariance d'un estimateur sans biais de  $g(\theta)$  est

$$V_g(\mathbf{u}; \theta, \tau) = \{Dg(\theta)\}' V(\mathbf{u}; \theta, \tau) \{Dg(\theta)\}. \quad (1)$$

À nouveau, pour  $N$  suffisamment grand, la variance-covariance de l'estimateur par maximum de vraisemblance de  $g$  sera proche de cette borne inférieure.



## Définition générale du problème

### Estimation d'un fonction du paramètre

Le cas de l'estimation du paramètre  $\theta$  correspond au cas où  $g_i(\theta) = \theta$  pour tout  $i = 1, \dots, k$  et ainsi  $Dg(\theta) = I_k$ , la matrice identité d'ordre  $k$ .

Par conséquent, nous nous intéresserons uniquement au cas de l'estimation d'une fonction du paramètre  $\theta$ .



## Définition générale du problème

### Objectif de la planification

L'idée simple, que nous poursuivrons dans la plus grande partie de l'exposé qui va suivre, est de s'efforcer de choisir le plan  $\mathbf{u}$  de telle sorte que la quantité qui apparaît dans l'équation (1) soit la plus petite possible, ou, ce qui est similaire, de rendre son inverse le plus grand possible.

En procédant de la sorte, nous utilisons un plan  $\mathbf{u}$  qui nous apporte le plus d'information possible, au sens de Fisher, sur la fonction  $g(\theta)$ .





# Définition générale du problème

## Objectif de la planification

Nous pourrions utiliser effectivement toute cette information soit

- dans le cadre de l'estimation de Gauss-Markov du modèle linéaire gaussien ;
- soit avec des estimateurs du maximum de vraisemblance et un effectif  $N$  suffisamment grand.



## Définition générale du problème

### Objectif de la planification

Le problème ainsi posé semble considérable et deux questions se posent immédiatement à nous :

1. au sens de quel critère souhaitons-nous rendre la matrice  $V_g(\mathbf{u}; \theta, \tau)$  « petite » ?
2. est-il possible de trouver un plan  $\mathbf{u}$ , noté  $\mathbf{u}_*$ , indépendant de  $\theta$  et  $\tau$ , qui rende  $V_g(\mathbf{u}; \theta, \tau)$  « petite », au sens où nous le souhaitons, pour toutes les valeurs de  $\theta$  et  $\tau$  ?



## Définition générale du problème

### Objectif de la planification

Pour toute une classe de problèmes, le modèle linéaire par exemple, l'expression de  $V_g(\mathbf{u}; \theta, \tau)$  est si simple qu'il est possible de répondre affirmativement à la seconde question, ce qui motivera l'étude de plans classiques comme les plans factoriels, fractionnaires, en blocs, . . .

Pour les autres situations, pour lesquelles un tel plan  $\mathbf{u}_*$  n'existe pas, nous devons réfléchir à comment tenir compte de cette situation. Nous commençons par étudier la première classe de problèmes qui pourra nous fournir des pistes pour aborder la seconde, plus complexe.



# Sommaire

## Introduction

Plans pour pesée

Définition générale du problème

**Exemples**

## Planification, modèle linéaire et symétries

Définitions, admissibilité, optimalité

Invariances

Cas de l'isovariance

## Caractérisation différentielle de l'optimalité

Définitions, première propriétés

Théorèmes de caractérisation

# Exemple 1

## Régression linéaire simple, gaussienne et contrôlée

Dans cet exemple :

- $u$  est un nombre réel ;
- $\theta = (\theta_0, \theta_1)'$  ;
- $\tau$  est un nombre réel strictement positif.

Pour des valeurs de  $u$ ,  $\theta$  et  $\tau$  données,  $\tilde{y}$  suit une  $\mathcal{N}(\theta_0 + \theta_1 u, \tau)'$ .

Supposons que  $\mathcal{U} = [-1, 1]$  et considérons un plan  $\mathbf{u} = (u_{(1)}, \dots, u_{(N)})$  avec donc  $-1 \leq u_{(i)} \leq 1$ .



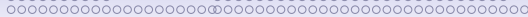
## Exemple 1

### Régression linéaire simple, gaussienne et contrôlée

Un calcul direct permet de montrer que

$$V^{-1}(\mathbf{u}; \theta, \tau) = \frac{1}{\tau} \begin{pmatrix} N & \sum_{i=1}^N u_{(i)} \\ \sum_{i=1}^N u_{(i)} & \sum_{i=1}^N u_{(i)}^2 \end{pmatrix}.$$

Voici un cas où  $V(\mathbf{u}; \theta, \tau)$  se simplifie comme indiqué ci-dessus. Il sera alors possible de trouver un plan  $\mathbf{u}_*$  qui sera le « meilleur » pour toutes les valeurs de  $\theta$  et  $\tau$  possibles.



## Exemple 1

### Régression linéaire simple, gaussienne et contrôlée

Si nous sommes intéressés par l'estimation de  $\theta$ , le problème se transforme ainsi en celui de rendre la matrice

$$M(\mathbf{u}) = \frac{1}{\tau} \begin{pmatrix} N & \sum_{i=1}^N u(i) \\ \sum_{i=1}^N u(i) & \sum_{i=1}^N u(i)^2 \end{pmatrix}.$$

« aussi grande que possible ».

1. Il est possible de montrer qu'il n'existe pas, même dans ce cas simple, de plan  $\mathbf{u}_*$  qui soit le « meilleur » au sens matriciel fort, ordre de Loewner, suivant :  $M(\mathbf{u}_*) - M(\mathbf{u})$  est semi-définie positive pour tout plan  $\mathbf{u}$ .

Par conséquent, et pour remédier à cette difficulté, nous essayerons de **rendre maximale une fonction** de  $V^{-1}$ .



## Exemple 1

### Régression linéaire simple, gaussienne et contrôlée

2. Dans le cas de ce modèle, même si nous avons travaillé en l'absence de l'hypothèse de normalité mais en conservant les

$$\begin{aligned}\mathbb{E}[\tilde{y}|u, \theta, \tau] &= \theta_0 + \theta_1 u \\ \text{Var}[\tilde{y}|u, \theta, \tau] &= \tau\end{aligned}$$

$V(\mathbf{u}; \theta, \tau)$  resterait la matrice de variance-covariance de l'estimateur de Gauss-Markov de  $\theta$ , ce qui rejoint la remarque que nous avons faite précédemment sur la recherche de plans dans le contexte du modèle linéaire homoscédastique.







## Exemple 2

### Plans factoriel complet $2^2$

Supposons que nous soyons intéressés par l'estimation des effets séparés (principaux) et combinés (interactions) de deux traitements  $T_1$  et  $T_2$  sur une réponse expérimentale  $y$ .

Nous supposons, comme généralement pour ce type de modèle, qu'une combinaison quelconque des modalités des traitements ne peut modifier que la valeur moyenne de  $\tilde{y}$ .

Notons :

- $\theta_1$  l'effet de  $T_1$  seul ;
- $\theta_2$  l'effet de  $T_2$  seul ;
- $\theta_1 + \theta_2 + \theta_3$  l'effet de  $T_1$  et  $T_2$  lorsqu'ils sont combinés.



## Exemple 2

### Plans factoriel complet $2^2$

Il est possible de mettre chacune des unités expérimentales dans l'**une** des quatre conditions suivantes :

1. aucun traitement ;
2. seulement  $T_1$  ;
3. seulement  $T_2$  ;
4.  $T_1$  et  $T_2$ .





## Exemple 2

### Plans factoriel complet $2^2$

Par conséquent, l'ensemble  $\mathcal{U}$  est réduit à quatre points.

En formulant les hypothèses usuelles pour nos modèles, nous obtenons :

$$\begin{aligned}\mathbb{E}[\tilde{y} | \mathbf{u}, \theta, \tau] &= \theta_0 + \theta_1 u_1 + \theta_2 u_2 + \theta_3 u_1 u_2 \\ \text{Var}[\tilde{y} | \mathbf{u}, \theta, \tau] &= \tau,\end{aligned}$$

ce qui correspond exactement à un modèle de régression linéaire tel que nous l'avons vu dans l'exemple 1.





## Exemple 3

### Régression quadratique avec une variable contrôlée, fonction non-linéaire du paramètre

Un calcul direct permet de montrer que, pour un  $N$ -plan,

$$V^{-1}(\mathbf{u}; \theta, \tau) = \frac{1}{\tau} \begin{pmatrix} N & \sum_{i=1}^N u_{(i)} & \sum_{i=1}^N u_{(i)}^2 \\ \sum_{i=1}^N u_{(i)} & \sum_{i=1}^N u_{(i)}^2 & \sum_{i=1}^N u_{(i)}^3 \\ \sum_{i=1}^N u_{(i)}^2 & \sum_{i=1}^N u_{(i)}^3 & \sum_{i=1}^N u_{(i)}^4 \end{pmatrix}.$$

$V(\mathbf{u}; \theta, \tau)$  se simplifie de la même manière que pour l'exemple 1. Néanmoins, il y a une différence fondamentale liée à la présence de la fonction non-linéaire  $g$ .



## Exemple 3

### Régression quadratique avec une variable contrôlée, fonction non-linéaire du paramètre

Rappelons que dans le cas d'un ESB de  $g(\theta)$ , la borne inférieure de l'inégalité de Fréchet-Darmonis-Cramer-Rao est :

$$\{Dg(\theta)\}' V(\mathbf{u}; \theta, \tau) \{Dg(\theta)\},$$

avec ici  $\{Dg(\theta)\}' = \frac{1}{2\theta_2^2} (-\theta_2 \quad \theta_1)$ .

La quantité précédente dépend de  $\theta$  et rien ne permet d'assurer qu'il existe un plan  $\mathbf{u}_*$  qui la rende minimale pour toutes les valeurs de  $\theta$  possible.

À ce titre, cette situation se distingue des deux précédentes.





## Exemple 4

### Analyse probit : un modèle non-gaussien et non linéaire

Des sujets peuvent être soumis à un stimulus avec différents niveaux d'intensité  $u$  pouvant appartenir à un intervalle  $\mathcal{U}$ .

Un sujet **peut ou non** réagir à un stimulus de niveau  $u$ , et la probabilité pour qu'il réagisse est donnée avec une bonne approximation par :

$$\Phi\left(\frac{u - \theta_1}{\theta_2}\right)$$

où  $\Phi$  est la fonction de répartition de la loi normale centrée et réduite.



## Exemple 4

### Analyse probit : un modèle non-gaussien et non linéaire

Supposons que nous disposions de  $N$  sujets pour réaliser l'expérience.

À quels niveaux du stimulus  $u$  devons-nous les soumettre pour estimer  $\theta_1$  et  $\theta_2$  le plus précisément possible ?



## Exemple 4

### Analyse probit : un modèle non-gaussien et non linéaire

Introduisons la variable aléatoire binaire  $\tilde{y}$  qui prend :

- la valeur 1 si le sujet réagit ;
- la valeur 0 si le sujet n'as pas de réaction.

La densité de  $\tilde{y}$  est alors donnée par l'expression

$$p(y|u, \theta) = \left[ \Phi \left( \frac{u - \theta_1}{\theta_2} \right) \right]^y \left[ 1 - \Phi \left( \frac{u - \theta_1}{\theta_2} \right) \right]^{1-y} .$$



## Exemple 4

### Analyse probit : un modèle non-gaussien et non linéaire

Un calcul direct permet de montrer que pour un plan  $\mathbf{u}$ , la matrice d'information au sens de Fisher pour les paramètres  $\theta_1$  et  $\theta_2$  est

$$\begin{pmatrix} \sum_{i=1}^N \frac{1}{\Phi_i(1-\Phi_i)} \left( \frac{\partial \Phi_i}{\partial \theta_1} \right)^2 & \sum_{i=1}^N \frac{1}{\Phi_i(1-\Phi_i)} \frac{\partial \Phi_i}{\partial \theta_1} \frac{\partial \Phi_i}{\partial \theta_2} \\ \sum_{i=1}^N \frac{1}{\Phi_i(1-\Phi_i)} \frac{\partial \Phi_i}{\partial \theta_1} \frac{\partial \Phi_i}{\partial \theta_2} & \sum_{i=1}^N \frac{1}{\Phi_i(1-\Phi_i)} \left( \frac{\partial \Phi_i}{\partial \theta_2} \right)^2 \end{pmatrix}$$

où  $\Phi_i$  désigne  $\Phi \left( \frac{u_{(i)} - \theta_1}{\theta_2} \right)$ .



## Exemple 4

### Analyse probit : un modèle non-gaussien et non linéaire

En vertu du principe que nous avons énoncé plus haut, nous souhaitons choisir un  $\mathbf{u}$  qui rende cette matrice la plus grande possible en un sens que nous préciserons dans la suite.

À nouveau cette matrice dépend de  $\theta$ , ce qui empêche a priori l'existence d'un plan  $\mathbf{u}_*$  qui rende l'information maximale pour tout  $\theta$ .



## Exemple 4

### Analyse probit : un modèle non-gaussien et non linéaire

Cet exemple permet d'avoir une idée concrète du type problème rencontré dans le cas non-linéaire.

Nous ne voulons pas en effet réaliser d'expériences pour des valeurs de  $u$  pour lesquelles  $\Phi\left(\frac{u-\theta_1}{\theta_2}\right)$  est soit proche de 0, soit proche de 1, parce que l'observation de telles valeurs sera peu informatif sur  $\theta_1$  et  $\theta_2$ .

**Or les valeurs de  $u$  pour lesquelles ceci se produit dépendent de  $\theta_1$  et  $\theta_2$  !**



## Exemple 4

### Analyse probit : un modèle non-gaussien et non linéaire

Puisque nous ignorons les valeurs réelles des paramètres  $\theta_1$  et  $\theta_2$ , comment pouvons-nous savoir si nous sommes en train de réaliser des expériences pour des valeurs non-informatives de  $u$  ?

La seule solution envisageable à ce problème est de mettre en place une **expérimentation séquentielle** où nous choisirons les niveaux à utiliser pour les nouveaux sujets à partir des réponses obtenues pour les niveaux que nous avons déjà utilisés pour les sujets pour lesquels nous avons déjà fait les expériences.



## Exemple 4

### Analyse probit : un modèle non-gaussien et non linéaire

Une telle **approche séquentielle** sera généralement nécessaire pour tous les problèmes de **planification non-linéaire** pour lesquels nous ne disposons **pas de connaissance préalable** sur les valeurs plausibles des **paramètres inconnus**.









## Modèles de régression linéaire soumis aux hypothèses statistiques usuelles :

- une réponse  $Y$  à valeurs réelles
- $\mathcal{U}$  une partie d'un espace vectoriel de dimension finie  $\mathbb{R}^V$  muni de sa structure euclidienne canonique ;
- une variable indépendante  $u \in \mathcal{U}$  ;
- $\mathbf{f} = (f_1, \dots, f_k)'$  est une fonction connue définie sur  $\mathcal{U}$  et à valeurs dans  $\mathbb{R}^k$ , posons  $\mathcal{X} = f(\mathcal{U})$  ;
- $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)' \in \mathbb{R}^k$  est un vecteur de paramètres inconnus.

$$y(u) = \sum_{j=1}^k \theta_j f_j(u) = \boldsymbol{\theta}' \mathbf{f}(u), \quad (2)$$



Valeurs de la réponse  $y$  aux points  $u_1, \dots, u_n \in \mathcal{U}$  représentées par des variables aléatoires à valeurs réelles  $\tilde{y}_1, \dots, \tilde{y}_n$  décorrélées telles que

$$\mathbb{E}[\tilde{y}_i | u, \theta, \tau] = y(u_i) \quad i = 1, \dots, n, \quad (3)$$

$$\text{Var}[\tilde{y}_i | u, \theta, \tau] = \tau \quad i = 1, \dots, n. \quad (4)$$

$\tau \in ]0, +\infty[$  est inconnue.



Il est possible de se placer dans un cadre légèrement plus général si nous supposons que

$$\text{Var}[\tilde{y}_i | u, \theta, \tau] = \tau v(u), \quad i = 1, \dots, n,$$

où  $v$  est une fonction de  $u$  connue.

En effet, il suffit alors de ramener au cas précédent en considérant :

- $y' = \frac{y}{v(u)^{1/2}}$  à la place de  $y$  ;
- $f_i(u)' = \frac{f_i(u)}{v(u)^{1/2}}$  à la place de  $f_i$ .



Les valeurs  $u_1, \dots, u_n$  pour lesquelles les observations sont effectuées sont :

1. connues exactement
2. contrôlées par l'expérimentateur.

## Définition

Un plan approché  $\xi$  est un couple  $(\mathbf{U}, \mathbf{w})$  où :

1.  $\mathbf{U}$  est un ensemble fini de points  $\{u_1, \dots, u_r\} \in \mathcal{U}$ , le **support du plan**,
2.  $\mathbf{w} = (w_1, \dots, w_r)$  sont les poids des points support du plan.



Remarquons que, sous les hypothèses (3) et (4), le choix d'un élément  $u \in \mathcal{U}$  est équivalent au choix d'un vecteur à  $k$  composante dans l'espace induit par le plan  $\mathcal{X} = f(\mathcal{U})$ .

Choisir un  $N$ -plan revient donc à choisir  $N$  vecteurs dans l'espace induit par le plan  $\mathcal{X}$ .

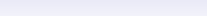
Par conséquent, nous pouvons, sans perte de généralité, adopter la notation suivante, en remplaçant (3) par (3'),

$$\mathbb{E}[\tilde{y} | \mathbf{x}, \theta, \tau] = \theta' \mathbf{x}.$$

et représenter un plan  $\xi$  par un couple  $(\mathbf{X}, \mathbf{w})$ , où  $\mathbf{X}$  est un ensemble fini de points  $\{\mathbf{x}_1, \dots, \mathbf{x}_r\} \in \mathcal{X}$ , l'espace induit par le plan.







## Définition

*La matrice des moments d'un plan est la matrice  $M(\xi)$  définie ainsi :*

$$\begin{aligned} M(\xi) &= \sum_{i=1}^r \xi(u_i) \mathbf{f}(u_i) \mathbf{f}(u_i)' \\ &= \sum_{i=1}^r w_i x_i x_i'. \end{aligned}$$

*Pour un plan exact, nous avons :*

$$\begin{aligned} M(\xi_n) &= \sum \mathbf{f}(u_i) \mathbf{f}(u_i)' \\ &= \sum x_i x_i'. \end{aligned}$$









Cet ordre peut servir pour comparer les plans même si la matrice des moments n'est pas inversible.

## Définition

*Un plan  $\xi_0$  est dit admissible s'il n'existe pas de plan  $\xi$  tel que :*

$$M(\xi_0) \leq M(\xi) \quad \text{et} \quad M(\xi_0) \neq M(\xi), \quad (6)$$

*où  $M(\xi_0)$  et  $M(\xi)$  sont les matrices des moments respectivement du plan  $\xi_0$  et du plan  $\xi$ .*

Propriété d'admissibilité **indépendante des poids** du plan.



Ordre **partiel** de Loewner : **défaut majeur**.

Ne permet pas de sélectionner un optimum unique : si  $k \neq 1$ .

Chacune des matrices des moments qui ne sont pas comparables associées à plusieurs plans.

De nombreux dispositifs expérimentaux non **comparables**.

Il est d'usage de spécifier un **critère d'optimalité** réel : une fonction de la matrice des moments à **valeurs réelles**.

# Optimalité

## Définition (Critère d'optimalité)

Une fonction  $\Phi : \mathcal{A} \rightarrow \mathbb{R}$  est un critère d'optimalité si :

1.  $\mathcal{A}$  est un cône convexe inclus dans  $S_k$  tel que  $S_k^{++} \subset \mathcal{A} \subset S_k^+$ , où  $S_k^{++}$  est le cône convexe étoilé en 0 des matrices réelles symétriques définies positives d'ordre  $k$ .
2.  $\Phi$  est décroissante pour l'ordre partiel de Loewner :

$$A \leq B \iff \Phi(A) \geq \Phi(B), \quad \forall (A, B) \in \mathcal{A}^2. \quad (7)$$

3.  $\Phi$  est convexe.



Recherche d'un plan  $\Phi$ -optimal  $\xi$ .

Chercher une solution au problème de minimisation suivant :

$$\min_{\xi | M(\xi) \in \mathcal{A}} \Phi(M(\xi)). \quad (8)$$

Posons  $\mathcal{M}$  l'ensemble des matrices des moments  $M(\xi)$

lorsque  $\xi$  parcourt l'ensemble de tous les plans.

Le problème de minimisation donné par l'équation 8 devient alors :

$$\min_{M \in \mathcal{M} \cap \mathcal{A}} \Phi(M). \quad (9)$$



# Optimalité

## Définition alternative d'un critère d'optimalité

Chez certains auteurs la fonction  $\Phi$  est

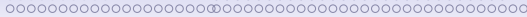
1. *Condition inchangée.*
2.  $\Phi$  est **croissante** pour l'ordre partiel de Loewner :

$$A \leq B \iff \Phi(A) \leq \Phi(B), \quad \forall(A, B) \in \mathcal{A}^2. \quad (10)$$

3.  $\Phi$  est **concave**.

La recherche d'un plan optimal est alors un problème de maximisation défini sur les mêmes ensembles.

Pour passer de l'une à l'autre des conventions, remplacer  $\phi$  par  $-\phi$ .



## Définition (Concavité et concavité stricte)

*Une application  $\Phi$  à valeur dans  $\mathbb{R} \cup \{-\infty\}$  est concave sur  $\mathcal{A}$  si*

$$\phi(\lambda M_1 + (1 - \lambda)M_2) \geq \lambda\phi(M_1) + (1 - \lambda)\phi(M_2),$$

*pour tout  $\lambda \in [0; 1]$ , pour tout  $(M_1, M_2) \in \mathcal{A}^2$ .*

*Si l'inégalité précédente est stricte pour  $0 < \lambda < 1$  sur  $\mathcal{A}^+$ , l'ensemble des points  $A$  de  $\mathcal{A}$  pour lesquels  $\Phi(A)$  est fini, alors  $\phi$  est strictement concave.*

## Proposition

*Un critère strictement concave admet sur  $\mathcal{A}$  un unique maximum  $A_*$ .*



Pour résoudre le problème d'optimalité, il suffit alors de :

1. Déterminer une matrice des moments optimale  $M^*$ .  
Problème de minimisation convexe.
2. Construire un plan  $\xi$  de matrice des moments  $M(\xi) = M^*$ .

Plusieurs **difficultés** :

1. Dimension du problème rapidement très élevée.
2. Résolution numérique du problème  $\Rightarrow$  connaissance approchée de la matrice des moments optimales.
3. Exhiber un plan ayant la bonne matrice des moments.  
Coordonnées des points support du plan connues de manière aproximative.



Remarquons que solution du problème d'optimalité tel que nous l'avons posé n'a aucune raison pour être un plan exact.

C'est même volontairement que nous avons décidé de recherché les plans optimaux dans un ensemble, les plans approchés, qui contient **strictement** les plans exacts.

En effet, l'ensemble des  $N$ -plans, ou plans exacts à  $N$  essais, est un ensemble discret et minimiser le critère  $\phi$  sur celui-ci revient à chercher le minimum d'une fonction à valeurs réelles définie uniquement sur les nombres entiers !





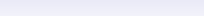
Classe importante de critères d'optimalité : critères orthogonalement invariants sur  $\mathcal{S}_k^{++}$ .

## Définition

*Un critère d'optimalité sur  $\mathcal{S}_k^{++}$  est orthogonalement invariant s'il est invariant pour l'action du groupe orthogonal de  $\mathbb{R}^k$  par congruence sur  $\mathcal{S}_k^{++}$ .*

**Objectif** : Utiliser les symétries du problème de minimisation pour :

1. Réduire la dimension.
2. Obtenir une résolution exacte.



## Exemples

Pour  $-\infty \leq p \leq 1$ , les fonctions  $\Phi_p$  suivantes, définies sur  $\mathcal{S}_k^{++}$ , sont des critères d'optimalité orthogonalement invariants. Ce sont les critères d'optimalité  $\Phi_p$  de Kiefer (1974).

$$\left\{ \begin{array}{l} \Phi_p(\mathbf{A}) = \left( \frac{1}{k} \sum_{j=1}^k \lambda_j(\mathbf{A})^p \right)^{\frac{-1}{p}}, \quad \text{si } p \notin \{-\infty, 0\}, \\ \Phi_0(\mathbf{A}) = \left( \prod_{j=1}^k \lambda_j(\mathbf{A}) \right)^{\frac{-1}{k}} = (\det(\mathbf{A}))^{\frac{-1}{k}}, \\ \Phi_{-\infty}(\mathbf{A}) = \left( \min_{j=1, \dots, k} \lambda_j(\mathbf{A}) \right)^{-1}. \end{array} \right. \quad (11)$$

Les critères obtenus pour  $p = 0$ ,  $p = -1$  et  $p = -\infty$  sont appelés respectivement critères de  $D$ -optimalité,  $A$ -optimalité et  $E$ -optimalité.



## Remarque

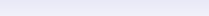
La  $D$ -optimalité trouve son interprétation naturelle dans le contexte du modèle linéaire gaussien.

Un ellipsoïde de confiance pour  $\theta$ , pour un niveau de confiance fixé et une somme des carrés résiduelles donnée, provenant de l'utilisation du plan  $\xi$  est de la forme

$$\mathcal{E} = \left\{ \theta \in \mathbb{R}^k : (\theta - \hat{\theta})' M(\xi) (\theta - \hat{\theta}) \leq \text{constante} \right\}$$

où  $\hat{\theta}$  est l'estimateur de Gauss-Markov de  $\theta$ .





## Remarque

Le volume de cet ellipsoïde  $\mathcal{E}$ ,  $\text{vol}(\mathcal{E})$ , est proportionnel à  $(\det(M(\xi)))^{-\frac{1}{2}}$ .

Le critère de  $D$ -optimalité vise donc à rendre cette ellipsoïde le plus petit possible en minimisant le déterminant de  $M(\xi)^{-1}$  ou, en d'autres termes, en maximisant celui de  $M(\xi)$ .



## Définition (Optimalité minimax)

*Pour toute partie  $\mathcal{C} \subset \mathbb{R}^k$  telle que  $f$  soit prolongeable naturellement à  $\mathcal{C}$ , un critère d'optimalité est défini par la formule :*

$$\Phi_{\mathcal{C}}(A) = \max_{u \in \mathcal{C}} \mathbf{f}(u)' A^{-1} \mathbf{f}(u). \quad (12)$$

*En utilisant ce critère, pour choisir un plan nous cherchons à rendre minimale la valeur maximale de la variance de l'estimateur de  $\theta' \mathbf{f}(u)$  lorsque  $u$  décrit  $\mathcal{C}$ .*



## Définition ( $G$ -optimalité)

*Le critère de  $G$ -optimalité, qui est orthogonalement invariant, est défini en prenant  $\mathcal{C} = \mathcal{X}$  :*

$$\Phi_G(A) = \max_{u \in \mathcal{U}} \mathbf{f}(u)' A^{-1} \mathbf{f}(u). \quad (13)$$

*En utilisant ce critère, pour choisir un plan nous cherchons à rendre minimale la valeur maximale de la variance de l'estimateur de  $\theta' \mathbf{f}(u)$  lorsque  $u$  décrit  $\mathcal{U}$ .*

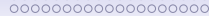


## Définition ( $E$ -optimalité)

*Le critère de  $E$ -optimalité est défini en prenant  $\mathcal{C}$  égal à la sphère unité :*

$$\Phi_E(A) = \max_{x \in \mathbb{R}^k : x'x=1} x' A^{-1} x. \quad (14)$$

*En utilisant ce critère, pour choisir un plan nous cherchons à rendre minimale la valeur maximale de la variance de l'estimateur de  $\theta'x$  lorsque  $x$  décrit la sphère unité, ce qui revient bien à chercher à rendre maximale la plus petite valeur propre de  $M(\xi)$ .*



## Définition ( $I$ -optimalité)

*Le critère de  $I$ -optimalité, qui n'est pas orthogonalement invariant, est défini pour tout domaine compact  $\mathcal{U}$  d'intérieur non vide par la formule :*

$$\Phi_I(\mathbf{A}) = \frac{1}{\text{vol}(\mathcal{U})} \int_{\mathcal{U}} \mathbf{f}(u)' \mathbf{A}^{-1} \mathbf{f}(u) du, \quad (15)$$

*où  $\text{vol}(\mathcal{U})$  désigne le volume du domaine expérimental compact  $\mathcal{U}$ .*

# Sommaire

## Introduction

Plans pour pesée

Définition générale du problème

Exemples

## Planification, modèle linéaire et symétries

Définitions, admissibilité, optimalité

**Invariances**

Cas de l'isovariance

## Caractérisation différentielle de l'optimalité

Définitions, première propriétés

Théorèmes de caractérisation



# Invariances

Différentes notions d'invariance pour l'action d'un groupe  $\mathcal{G}$  existent (Pukelsheim 1993, Gaffke, Heiligers 1996).

1. La  $(\mathcal{G}, \mathcal{Q})$ -équivalence = propriétés de symétrie héritées de la forme du domaine expérimental  $\mathcal{U}$  et du modèle de régression.
2. La  $\mathcal{G}$ -invariance = équivalence et invariance de l'ensemble des points support du plan pour l'action de  $\mathcal{G}$ .
3. La  $\mathcal{G}$ -invariance faible = équivalence et invariance de la matrice des moments du plan pour l'action de  $\mathcal{G}$  sur les points support du plan.



## Définition (Équivariance)

Considérons un modèle de régression linéaire,  $y(u) = \theta' \mathbf{f}(u)$ ,  $u \in \mathcal{U}$ ,  $\theta \in \mathbb{R}^k$ . Soit  $\mathcal{G}$  un groupe de transformations bijectives de  $\mathcal{U}$  sur  $\mathcal{U}$  et  $\mathcal{Q}$  un groupe de matrices réelles carrées d'ordre  $k$  pour la multiplication matricielle usuelle. Le modèle de régression est **équivariant** pour  $\mathcal{G}$  et  $\mathcal{Q}$ , ce que nous abrégeons également en  $(\mathcal{G}, \mathcal{Q})$ -**équivariant**, s'il existe une application surjective de  $\mathcal{G}$  sur  $\mathcal{Q}$  qui à un élément  $g$  de  $\mathcal{G}$  associe un élément  $Q_g$  de  $\mathcal{Q}$  tel que :

$$\mathbf{f}(g(u)) = Q_g \mathbf{f}(u), \quad \forall (u, g) \in \mathcal{U} \times \mathcal{G}. \quad (16)$$

## Exemple

Tout modèle de régression est  $(\{\text{id}_{\mathcal{U}}\}, \{I_k\})$ -équivariant.





## Définition

*Une application  $g$  de  $\mathcal{U}$  dans  $\mathcal{U}$  induit une opération sur l'ensemble des plans dont le support est inclus dans  $\mathcal{U}$  de la manière suivante :*

$$\xi^g = \begin{pmatrix} g(u_1) & \dots & g(u_r) \\ \xi(u_1) & \dots & \xi(u_r) \end{pmatrix}, \quad (17)$$

où  $\xi^g$  est l'image par  $g$  du plan

$$\xi = \begin{pmatrix} u_1 & \dots & u_r \\ \xi(u_1) & \dots & \xi(u_r) \end{pmatrix}. \quad (18)$$

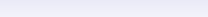


## Proposition

*Si un modèle de régression est  $(\mathcal{G}, \mathcal{Q})$ -équivariant alors le groupe  $\mathcal{G}$  induit une opération sur l'ensemble des matrices des moments des plans dont le support est inclus dans  $\mathcal{U}$  de la manière suivante :*

$$M(\xi^g) = Q_g M(\xi) Q_g', \quad \forall \xi \subset \mathcal{U}, \forall g \in \mathcal{G}. \quad (19)$$

*Il s'agit donc d'une action par congruence du groupe  $\mathcal{Q}$ .*



## Définition (Invariance, invariance faible)

Soit un modèle de régression  $(\mathcal{G}, \mathcal{Q})$ –équivariant. Un plan  $\xi$  est :

1.  $\mathcal{G}$ –invariant si pour tout  $g \in \mathcal{G}$ ,  $\xi^g = \xi$ ,
2.  $\mathcal{G}$ –faiblement invariant si pour tout  $g \in \mathcal{G}$ ,  $M(\xi^g) = M(\xi)$ .

## Lemme

Soit un modèle de régression  $(\mathcal{G}, \mathcal{Q})$ –équivariant avec  $\mathcal{Q}$  un groupe compact. Pour tout plan  $\xi$  il existe un plan faiblement invariant  $\tilde{\xi}$  tel que

$$M(\tilde{\xi}) \in \text{Conv} \{M(\xi^g), g \in \mathcal{G}\}. \quad (20)$$

Si  $\mathcal{G}$  est fini alors, pour tout plan faiblement invariant  $\tilde{\xi}$ , il existe un plan invariant  $\bar{\xi}$  dont le support,  $\text{Supp} \bar{\xi}$ , est inclus dans celui de  $\tilde{\xi}$  et pour lequel  $M(\bar{\xi}) = M(\tilde{\xi})$ .



## Définition (Modèle de régression polynomiale)

Soient  $d$  un entier naturel et  $A$  un sous-ensemble fini de  $\mathbb{N}^V$ .

Les éléments de  $A$  notés  $\alpha = (\alpha_1, \dots, \alpha_V)$  sont tels que :

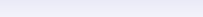
1.  $|\alpha| = \sum_{i=1}^V \alpha_i \leq d$  pour tout  $\alpha \in A$
2. il existe  $\alpha_0 \in A$  tel que  $|\alpha_0| = d$ .

Un modèle de régression multiple polynomiale  $A$  de degré  $d$  sur un domaine expérimental  $\mathcal{U} \in \mathbb{R}^V$  est alors défini par :

$$y(u) = \sum_{\alpha \in A} \theta_{\alpha} u^{\alpha}, \quad u = (u_1, \dots, u_V)' \in \mathcal{U}. \quad (21)$$

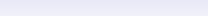
Le modèle de régression multiple polynomiale complet  $A_d$  de degré  $d$  sur un domaine expérimental  $\mathcal{U} \in \mathbb{R}^V$  est défini par :

$$A_d = \{\alpha \in \mathbb{N}^V, |\alpha| \leq d\}. \quad (22)$$



## Exemples

1. Soit  $\mathcal{G}_s$  le groupe des changements de signe.  
Si le domaine expérimental  $\mathcal{U}$  est  $\mathcal{G}_s$ -invariant alors tout modèle de régression multiple polynomiale  $A$  est  $(\mathcal{G}_s, \mathcal{Q}_s)$ -équivariant.
2. Soit  $\mathcal{G}_p$  le groupe des permutations des coordonnées.  
Si le domaine expérimental  $\mathcal{U}$  est  $\mathcal{G}_p$ -invariant et l'ensemble des degrés  $A$  est invariant par permutation alors le modèle de régression multiple polynomiale  $A$  est  $(\mathcal{G}_p, \mathcal{Q}_p)$ -équivariant.
3. Soit  $\mathcal{G}_{sp}$  le groupe engendré par  $\mathcal{G}_s$  et  $\mathcal{G}_p$ .  
Si le domaine expérimental  $\mathcal{U}$  est  $\mathcal{G}_{sp}$ -invariant et l'ensemble des degrés  $A$  est invariant par permutation alors le modèle de régression multiple polynomiale  $A$  est  $(\mathcal{G}_{sp}, \mathcal{Q}_{sp})$ -équivariant.



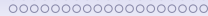
## Exemple

Le domaine expérimental  $\mathcal{U}$  la boule de  $\mathbb{R}^V$  centrée en  $\mathbf{0}$ . Soit  $\mathcal{G} = \mathcal{G}_{orth}$  le groupe des transformations orthogonales de  $\mathbb{R}^V$ . Le modèle de régression multiple polynomiale complet  $A_d$  est  $(\mathcal{G}_{orth}, \mathcal{Q}_{orth})$ -équivariant.

## Définition (Isovariance)

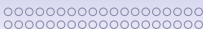
Soient un domaine expérimental  $\mathcal{U}$  et un le modèle de régression multiple polynomiale complet  $A_d$   $(\mathcal{G}_{orth}, \mathcal{Q}_{orth})$ -équivariant.

Un plan  $\xi$  est dit **isovariant** s'il est  $\mathcal{G}_{orth}$ -**faiblement invariant**.



## Remarques

1. Seuls les trois groupes  $\mathcal{Q}_s$ ,  $\mathcal{Q}_p$  et  $\mathcal{Q}_{sp}$  sont des sous-groupes du groupe orthogonal.
2. Les plans **invariants** pour l'action du groupe  $\mathcal{G}_{orth}$  sont réduits au plan formé par le point  $\mathbf{0}$ .
3. Nous verrons dans la suite que les plans **faiblement invariants** pour l'action du groupe  $\mathcal{G}_{orth}$  ne se réduisent pas à celui formé par le point  $\mathbf{0}$ .



## Invariance, admissibilité et optimalité

Intérêt marqué pour la combinaison invariance, admissibilité et optimalité : Gaffke et Heiligers 96, Schwabe 96 et Liski, Mandal, Shah et Sinha 02.

### Définition (Critère invariant)

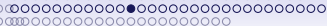
Soit  $\mathcal{Q}$  un groupe de matrices réelles carrées d'ordre  $k$ . Un critère d'optimalité  $\Phi$  défini sur  $\mathcal{A}$  est  $\mathcal{Q}$ -invariant si :

$$QAQ' \in \mathcal{A} \text{ et } \Phi(QAQ') = \Phi(A), \quad \forall (A, Q) \in \mathcal{A} \times \mathcal{Q}. \quad (23)$$

### Exemples

Le critère  $\Phi_0$  de  $D$ -optimalité est  $\mathcal{Q}_{orth}$ -invariant.





## Lemme

*Soit un modèle de régression multiple polynomiale  $A$   $(\mathcal{G}, \mathcal{Q})$ –équivariant, avec  $\mathcal{Q}$  est un groupe compact, et  $\Phi$  un critère d'optimalité défini sur  $\mathcal{A}$  et  $\mathcal{Q}$ –invariant. Pour tout plan  $\xi$  tel que  $M(\xi) \in \mathcal{A}$  et tout plan  $\tilde{\xi}$   $\mathcal{G}$ –faiblement invariant avec  $M(\tilde{\xi}) \in \text{Conv} \{M(\xi^g), g \in \mathcal{G}\}$  nous avons :*

$$M(\tilde{\xi}) \in \mathcal{A}, \quad \Phi \left( M \left( \tilde{\xi} \right) \right) \leq \Phi \left( M \left( \xi \right) \right). \quad (24)$$



Importance des plans isovariants parmi tous les plans dont le support est inclus dans une boule de  $\mathbb{R}^V$  centrée en  $\mathbf{0}$  pour les critères de  $D$ -optimalité.

## Corollaire

*Considérons un domaine expérimental  $\mathcal{U}$  égal à une boule de  $\mathbb{R}^V$  centrée en  $\mathbf{0}$ . Pour un modèle polynomial de degré  $d$  isovariant et le critère d'optimalité  $\Phi_0$  et  $\Phi_I$ , nous pouvons restreindre la recherche d'un plan optimal pour la classe des plans dont le support est dans  $\mathcal{U}$  à la sous-classe des plans isovariants.*



## Définition

Un plan  $\xi$  est **admissible** pour un modèle de régression polynomial  $A$  de degré  $d$  est dit  $A$ -**admissible**.

## Lemme (Gaffke, Heiligers 96)

Soit  $\mathcal{G}$  un groupe compact de transformations agissant linéairement sur  $\mathcal{U}$ . Supposons que le modèle de régression polynomial  $A$  de degré  $d$  est  $(\mathcal{G}, \mathcal{Q})$ -équivariant. Alors, pour tout plan faiblement invariant  $\xi$  il existe un plan faiblement invariant  $\tilde{\xi}$  qui soit  $A$ -admissible et  $A_d$ -admissible et tel que :

$$M(\tilde{\xi}) \geq M(\xi). \quad (25)$$



## Corollaire

*Lors de la recherche d'une solution optimale pour un critère d'optimalité  $\mathcal{Q}$ -invariant dans la classe des plans dont le support est dans  $\chi$ , nous pouvons nous restreindre à chercher un plan optimal dans la sous-classe des plans faiblement invariants et  $A_d$ -admissibles.*

Nous pouvons désormais caractériser les plans  $A_d$ -admissibles pour un domaine  $\chi$  compact. Le cas unidimensionnel a été traité en 1959 par Kiefer.



## Théorème

Soit  $v = 1$ , et  $\chi = [a, b]$  avec  $(a, b) \in \mathbb{R}^2$  tels que  $a < b$ . Un plan  $\xi$  est  $A_d$ -admissible si et seulement si :

$$\text{Card}(\text{supp}(\xi) \cap ]a, b[) \leq d - 1. \quad (26)$$



Nous introduisons désormais la notion de segment et d'intérieur relatif d'un segment pour appliquer le théorème 2.1 au cas d'un modèle polynomial  $A_d$  avec  $v \geq 2$ .

### Définition (Segment)

*Un segment  $S$  d'extrémités  $\mathbf{x}^{(0)}$  et  $\mathbf{x}^{(1)}$ , deux points de  $\mathbb{R}^v$ , est l'enveloppe convexe de ces points :*

$$S = \left\{ \lambda \mathbf{x}^{(0)} + (1 - \lambda) \mathbf{x}^{(1)}, \lambda \in [0, 1] \right\}. \quad (27)$$

*Le segment  $S$  n'est pas dégénéré si ces extrémités,  $\mathbf{x}^{(0)}$  et  $\mathbf{x}^{(1)}$ , ne sont pas égales.*



## Définition (Intérieur relatif)

L'intérieur relatif d'un segment  $S$ , noté  $\text{ri}(S)$ , d'extrémités  $\mathbf{x}^{(0)}$  et  $\mathbf{x}^{(1)}$ , deux points de  $\mathbb{R}^V$ , est l'ensemble des barycentres à coefficients strictement positifs de ces points :

$$S = \left\{ \lambda \mathbf{x}^{(0)} + (1 - \lambda) \mathbf{x}^{(1)}, \lambda \in ]0, 1[ \right\}. \quad (28)$$

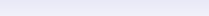


## Théorème

*Soit  $v \geq 1$ , et  $\mathcal{S} \subset \chi$  un segment non dégénéré inclus dans le domaine expérimental. Si un plan  $\xi$  est  $A_d$ -admissible :*

$$\text{Card}(\text{supp}(\xi) \cap \text{ri}(\mathcal{S})) \leq d - 1. \quad (29)$$





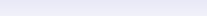
## Définition (Face)

Une face  $\mathcal{F}$  d'un ensemble  $\chi$  est un sous-ensemble convexe de  $\chi$  tel que :

$$S \subset \chi \text{ est un segment et } \mathcal{F} \cap \text{ri}(S) \neq \emptyset, \implies S \subset \mathcal{F}. \quad (30)$$

## Remarque

L'ensemble  $\chi$  est une face de  $\chi$  et chaque point extrémal de  $\chi$  est une face de  $\chi$ .



Il existe plusieurs liens entre un compact convexe  $\chi$  et ses faces. La proposition suivante explicite l'un d'entre eux.

### Proposition

*Si  $\chi$  est un compact convexe de  $\mathbb{R}^V$ , alors :*

$$\chi = \bigcup_{\mathcal{F} \text{ face de } \chi} \text{ri}(\mathcal{F}). \quad (31)$$



## Corollaire

*Nous reprenons le contexte du théorème 2.2 et nous supposons de surcroît que le domaine expérimental  $\chi$  est compact et convexe.*

- 1. Si  $d = 1$ , alors les points support d'un plan  $A_1$ –admissible sont les points extrêmes de  $\chi$ .*
- 2. Si  $d = 2$ , alors pour toute face  $\mathcal{F}$  de  $\chi$  un plan  $A_2$ –admissible a au plus un point support dans  $\text{ri}(\mathcal{F})$ .*

Pour obtenir des résultats plus précis nous étudions des domaines expérimentaux particuliers.



## Définition (Domaine cubique)

Le domaine expérimental  $\mathcal{C}_b$  est un  $v$ -hypercube symétrique centré en  $\mathbf{0}$  :

$$\mathcal{C}_b = \left\{ \mathbf{x} = (x_1, \dots, x_v)' \in \mathbb{R}^v, |x_i| \leq b_i, 1 \leq i \leq v \right\}, \quad (32)$$

où  $b_i \in ]0, +\infty[$ ,  $1 \leq i \leq v$ .

L'ensemble des sommets de  $\mathcal{C}_b$  est  $\mathcal{V}_b$  avec

$$\mathcal{V}_b = \left\{ \mathbf{x} = (x_1, \dots, x_v)' \in \mathbb{R}^v, |x_i| = \pm b_i, 1 \leq i \leq v \right\}. \quad (33)$$



## Définition (Domaine elliptique)

Le domaine expérimental  $\mathcal{E}_H$  est un  $v$ -ellipsoïde centré en  $\mathbf{0}$  :

$$\mathcal{E}_H = \left\{ \mathbf{x} = (x_1, \dots, x_v)' \in \mathbb{R}^v, \mathbf{x}' H \mathbf{x} \leq 1 \right\}, \quad (34)$$

où  $H \in \mathcal{S}_k^{++}$ .

La frontière de  $\mathcal{E}_H$  est  $\partial \mathcal{E}_H$  :

$$\partial \mathcal{E}_H = \left\{ \mathbf{x} = (x_1, \dots, x_v)' \in \mathbb{R}^v, \mathbf{x}' H \mathbf{x} = 1 \right\}. \quad (35)$$



## Corollaire

Soit un modèle de régression multiple polynomiale de degré  $d = 1$  sur  $\chi = \mathcal{C}_b$  ou  $\chi = \mathcal{E}_H$ . Un plan  $\xi$  est  $A_1$ -admissible si et seulement si :

$$\text{supp}(\xi) \subset \begin{cases} \mathcal{V}_b, & \text{si } \chi = \mathcal{C}_b, \\ \partial\mathcal{E}_H, & \text{si } \chi = \mathcal{E}_H. \end{cases} \quad (36)$$



## Remarque

Dans le cas d'un domaine  $\mathcal{E}_H$  avec  $H = (1/r^2)I_V$ , c'est-à-dire si  $\chi$  est la boule de rayon  $r$  centrée en  $\mathbf{0}$ , la matrice des moments d'un plan faiblement invariant pour  $\mathcal{G}_{sp}$  et  $A_1$ -admissible est

$M^* = \text{Diag} \left( 1, \frac{r^2}{V}, \dots, \frac{r^2}{V} \right)$ . Il s'agit de la même matrice des moments que celle d'un plan isovariant et  $A_1$ -admissible.

Ainsi tout plan sphérique est un plan  $A_1$ -admissible si le domaine expérimental est si le support du plan est inclus dans la frontière de cette boule.

Dans la suite nous nous concentrons sur les résultats d'admissibilité concernant les domaines de type  $\mathcal{E}_H$ .



## Corollaire

*Soit un modèle de régression multiple polynomiale de degré  $d = 2$ , un domaine expérimental  $\chi$  égal à  $\mathcal{E}_H$  et le groupe de transformations constitué de deux éléments, l'identité et la symétrie centrale par rapport à l'origine du repère.*

- 1. Si un plan  $\xi$  est faiblement invariant et  $A_2$ -admissible alors  $\text{supp}(\xi) \subset \mathcal{E}_H \cup \{\mathbf{0}\}$ .*
- 2. Réciproquement, si un plan, non nécessairement faiblement invariant,  $\xi$  est tel que  $\text{supp}(\xi) \subset \mathcal{E}_H \cup \{\mathbf{0}\}$  alors il est  $A_2$ -admissible.*





Le corollaire 2.5 justifie pleinement l'utilisation de plans comportant des points répartis sur une même sphère, éventuellement complétés par des points au centre pour un modèle de régression polynomiale multiple de degré 2 et un domaine expérimental de type boule ; le support du plan étant inclus dans la frontière de cette boule.

## Corollaire

*Tout plan dont le support est contenu dans la sphère unité de  $\mathbb{R}^V$  est  $A_2$ -admissible pour un domaine  $\chi$  égal à la boule unité de  $\mathbb{R}^V$ .*



Dans le cas d'un domaine expérimental de type cube centré en  $\mathbf{0}$  et symétrique ou de la boule de rayon  $r$  centrée en  $\mathbf{0}$ , quelques résultats généraux concernant les modèles de degré supérieurs ou égaux 3 existent, voir Gaffke et Heiligers (1996). Toutefois si nous faisons l'hypothèse supplémentaire que le plan est isovariant il est possible de démontrer le résultat que nous allons exposer dans la section suivante.



## Classe essentiellement complète

Lors de la recherche d'une solution optimale pour un critère d'optimalité  $Q$ -invariant dans la classe des plans dont le support est dans  $\mathcal{U}$ , nous pouvons nous restreindre à chercher un plan optimal dans la sous-classe des plans faiblement invariants et  $A_d$ -admissibles.

## Réduction de dimension

Les **classes essentiellement complètes** formées de plans faiblement invariants, de plans admissibles, ou de plans faiblement invariants et admissibles donnent un cadre formel pour la **réduction** de la dimension du problème de minimisation.



# Sommaire

## Introduction

- Plans pour pesée
- Définition générale du problème
- Exemples

## Planification, modèle linéaire et symétries

- Définitions, admissibilité, optimalité
- Invariances
- Cas de l'isovariance

## Caractérisation différentielle de l'optimalité

- Définitions, première propriétés
- Théorèmes de caractérisation



Il a été étudié par de nombreux auteurs depuis l'article fondateur de Box et Hunter (1957) :

- Draper, Gaffke et Pukelsheim (1993) ;
- Gaffke et Heiligers (1996) ;
- Myers et Montgomery (2002).

Lorsqu'un plan possède la propriété d'isovariance, la variance de l'estimateur de Gauss-Markov en un point du domaine expérimental ne dépend que de la distance de ce point au centre du domaine.

Ainsi, à une distance fixe de l'origine, la précision de l'estimation est constante.



Soit  $\chi$  un domaine expérimental égal à la boule centrée en  $\mathbf{0}$  et de rayon  $r > 0$  et un modèle de régression multiple isovariant de degré  $d \geq 1$ .

Nous supposons que, pour le problème d'optimalité considéré, la restriction à l'ensemble des plans isovariants est justifiée comme c'est le cas pour les critères de  $D$ -optimalité et de  $I$ -optimalité.



## Proposition

*Soit un modèle polynomial complet de degré  $d \geq 1$  sur la boule  $\mathcal{B}_r$  de rayon  $r$  centrée en  $\mathbf{0}$  et le groupe de transformations  $\mathcal{G}_{orth}$ . Pour tout  $\rho \in [0, r]$ , l'ensemble des plans isovariants dont le support est inclus dans la sphère  $S_\rho$  est non-vidé et tous ces plans ont la même matrice des moments, notée  $\overline{M}_{d,\rho}$ . Nous avons de plus, pour tout plan  $\xi$  isovariant sur la boule  $\mathcal{B}_r$  la décomposition suivante pour la matrice de ses moments :*

$$M_d(\xi) = \sum_{\rho \in R(\xi)} \xi(S_\rho) \overline{M}_{d,\rho}, \quad (37)$$

*où  $\xi(S_\rho)$  est le poids total de tous les points du plan  $\xi$  appartenant à la sphère  $S_\rho$  et  $R(\xi)$  l'ensemble, fini, des rayons  $\rho \in [0, r]$  tels que  $\xi(S_\rho) > 0$ .*



## Définition (Multiplicité d'un rayon)

Soit  $\xi$  un plan isovariant sur la boule  $\mathcal{B}_r$ . La multiplicité  $m_\xi(\rho)$  d'un rayon  $\rho \in R(\xi)$  est égale à :

$$m_\xi(\rho) = \begin{cases} 0, & \text{si } \rho = r \\ 1, & \text{si } 0 < \rho < r \\ \frac{1}{2}, & \text{si } \rho = 0 \end{cases} . \quad (38)$$





## Définition (Cardinal effectif)

Soit  $\xi$  un plan isovariant sur la boule  $\mathcal{B}_r$ . Le cardinal effectif  $\gamma_\xi$  de  $R(\xi)$  est égal à :

$$\gamma_\xi = \sum_{\rho \in R(\xi)} m_\xi(\rho). \quad (39)$$



## Théorème

Soit un modèle polynomial complet d'ordre  $d \geq 1$  sur la boule  $\mathcal{B}_r$ . Si le plan  $\xi$  est isovariant et  $A_d$ -admissible alors :

$$\gamma_\xi \leq \frac{d-1}{2}. \quad (40)$$

Réciproquement, si  $\xi$  est un plan, non nécessairement isovariant, tel que :

$$\gamma_\xi \leq \frac{d-1}{2}, \quad (41)$$

alors il est  $A_d$ -admissible.



## Corollaire

1. *Si  $d = 1$ ,  $\gamma_\xi = 0$ . Ainsi le support d'un plan isovariant et  $A_1$ -admissible est inclus dans la frontière de la boule  $\mathcal{B}_r$ . Réciproquement tout plan dont le support est inclus dans la sphère de rayon  $r$  est  $A_1$ -admissible.*
2. *Si  $d = 2$ ,  $\gamma_\xi \leq 1/2$ . Ainsi le support d'un plan isovariant et  $A_2$ -admissible est :*
  - *soit inclus dans la frontière de la boule  $\mathcal{B}_r$ ,*
  - *soit dans l'ensemble formé par l'union de cette frontière et de l'origine  $\mathbf{0}$ .*

*Réciproquement si un plan vérifie l'une des conditions ci-dessus, il est  $A_2$ -admissible.*



## Corollaire

1. *Si  $d = 3$ ,  $\gamma_\xi \leq 1$ . Ainsi le support d'un plan isovariant et  $A_3$ -admissible est :*
  - *soit inclus dans la frontière de la boule  $\mathcal{B}_r$ ,*
  - *soit dans l'ensemble formé par l'union de cette frontière et de l'origine  $\mathbf{0}$ ,*
  - *soit dans l'ensemble formé par l'union de cette frontière et d'une sphère de rayon  $0 < \rho < R$  centrée en  $\mathbf{0}$ .*

*Réciproquement si un plan vérifie l'une des conditions ci-dessus, il est  $A_3$ -admissible.*



## Corollaire

1. Si  $d = 4$ ,  $\gamma_\xi \leq 3/2$ . Ainsi le support d'un plan isovariant et  $A_4$ -admissible est :
  - soit inclus dans la frontière de la boule  $\mathcal{B}_r$ ,
  - soit dans l'ensemble formé par l'union de cette frontière et de l'origine  $\mathbf{0}$ ,
  - soit dans l'ensemble formé par l'union de cette frontière et d'une sphère de rayon  $0 < \rho < R$  centrée en  $\mathbf{0}$ ,
  - soit dans l'ensemble formé par l'union de cette frontière, d'une sphère de rayon  $0 < \rho < R$  centrée en  $\mathbf{0}$  et d'une sphère de rayon  $0 < \rho < R$  centrée en  $\mathbf{0}$ .

Réciproquement si un plan vérifie l'une des conditions ci-dessus, il est  $A_4$ -admissible.



## Conclusion

Ce résultat permet de justifier la forme de la plupart des plans qui sont communément utilisés pour l'étude des modèles de régression.

Nous rencontrerons plusieurs plans de ce type au cours des exercices que nous traiterons.



La proposition 2.5 indique que la matrice des moments d'un plan isovariant se décompose de la manière suivante :

$$M_d(\xi) = \sum_{\rho \in R(\xi)} \xi(S_\rho) \bar{M}_{d,\rho}, \quad (42)$$

Les matrices  $\bar{M}_{d,\rho}$  vérifient la relation :

$$\bar{M}_{d,\rho} = D_\rho \bar{M}_{d,1} D_\rho \quad (43)$$

où  $D_\rho$  est la matrice diagonale d'ordre  $\begin{pmatrix} v+d \\ d \end{pmatrix}$  avec

$$D_{\rho\alpha,\alpha} = \rho^{|\alpha|}, |\alpha| \leq d.$$



Les matrices des moments des plans isovariants  $\xi$  sont donc paramétrées de manière linéaire par le vecteur :

$$m(\xi) = \sum_{\rho \in R(\xi)} \xi(S_\rho) m(\rho), \quad (44)$$

avec  $m(\rho) = (\rho^2, \dots, \rho^{2d})'$ , pour  $0 \leq \rho \leq r$ .

Lorsque  $\xi$  décrit l'ensemble des plans isovariants,  $m(\xi)$  décrit  $\mathcal{M}$  où :

$$\mathcal{M} = \text{Conv} \{ m(\rho), 0 \leq \rho \leq r \}. \quad (45)$$





Une résolution explicite d'un problème d'optimisation portant sur les plans isovariants nécessite donc le calcul du terme général de la matrice  $\overline{M}_{d,\rho}$ . Comme nous l'avons rappelé ci-dessus, il suffit de déterminer le terme général de la matrice  $\overline{M}_{d,1}$ .



## Fonction génératrice des moments

Mélange entre approche décrite dans Gaffke et Heiligers (1996) et Myers et Montgomery (2002).

### Définition (Fonction génératrice des moments)

*La fonction génératrice des moments d'un plan  $\xi$ , pour un modèle polynomial complet  $A_d$  de degré  $d$ , notée  $MGF^{A_d}(\xi)$ , est définie par :*

$$MGF^{A_d}(\xi)(\mathbf{t}) = \mathbb{E}_{\xi} \left[ (1 + \mathbf{t}'\mathbf{X})^{2d} \right], \forall \mathbf{t} \in \mathbb{R}^v. \quad (46)$$

*Ainsi dans le cas où les poids sont tous égaux :*

$$MGF^{A_d}(\xi)(\mathbf{t}) = \frac{1}{r} \sum_{i=1}^r (1 + \mathbf{t}'u_i)^{2d}, \forall \mathbf{t} \in \mathbb{R}^v. \quad (47)$$



Transmission de propriétés d'invariance de la matrice des moments à la fonction génératrice des moments du plan.

## Proposition

*Lorsque le plan  $\xi$  est isovariant nous avons l'identité :*

$$\text{MGF}^{A_d}(\xi)(\mathbf{t}) = \sum_{j=1}^d a_{2j} \left( \sum_{i=1}^k t_i^2 \right)^j, \quad \forall \mathbf{t} \in \mathbb{R}^V. \quad (48)$$

$\text{MGF}^{A_d}(\xi)$  est une fonction **radiale**.



## Proposition

Soit  $\xi$  un plan expérimental inclus dans le domaine expérimental  $\chi$ . Le coefficient d'un monôme  $\mathbf{t}^\gamma = t_1^{\gamma_1} \times \dots \times t_v^{\gamma_v}$  de  $\text{MGF}^{A_d}(\xi)$ , la fonction génératrice des moments du plan  $\xi$  pour un modèle polynomial complet  $A_d$  de degré  $d$ , est égal à :

$$\frac{(2d)!}{\prod_{i=1}^v (\gamma_i)! (2d - |\gamma|)!} \mu_\gamma(\xi), \quad (49)$$

où  $|\gamma| = \sum_{i=1}^v \gamma_i \leq 2d$ .





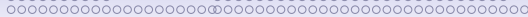
## Proposition (suite)

$$\lambda_{|\gamma|} = \frac{a_{|\gamma|} 2^{|\gamma|/2} (|\gamma|/2)! (2d - |\gamma|)!}{(2d)!}, \quad (51)$$

où  $|\gamma| = \sum_{i=1}^v \gamma_i$  et les  $a_{|\gamma|}$ , définis pour  $0 \leq |\gamma| \leq 2d$  et  $|\gamma|$  pair, sont des nombres réels.

Nous avons alors l'identité :

$$\text{MGF}^{A_d}(\boldsymbol{\xi})(\mathbf{t}) = \sum_{j=1}^d a_{2j} \left( \sum_{i=1}^k t_i^2 \right)^j, \quad \forall \mathbf{t} \in \mathbb{R}^v. \quad (52)$$



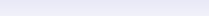
## Remarque

Soient  $\mathcal{U}$  un domaine expérimental de  $\mathbb{R}^V$  d'intérieur non vide, un plan  $\xi$  analysé à l'aide d'un modèle de régression polynomiale complet de degré  $d$   $(\mathcal{G}, \mathcal{Q})$ —équivariant pour l'action linéaire d'un sous-groupe compact de  $\mathcal{GL}_V(\mathbb{R})$ .

**Méthode de construction** d'un plan expérimental

**$\mathcal{G}$ -faiblement invariant :**

1. Déterminer  $A \in \mathcal{GL}_V(\mathbb{R})$  et  $\mathcal{K}$  un sous-groupe de  $\mathcal{O}_V(\mathbb{R})$  tels que  $\mathcal{G} = A\mathcal{K}A^{-1}$ .
2. Choisir un plan  $\eta$ , tel que
  - le support de  $\eta^A$  est inclus dans  $\mathcal{U}$
  - la fonction génératrice des moments du plan expérimental  $\eta$ ,  $\text{MGF}^{A_d}(\eta)$ , est  $\mathcal{K}$ -invariante.
3. Le plan expérimental  $\xi = \eta^A$  est alors un plan  $\mathcal{G}$ -faiblement invariant.



# Applications naturelles des plans isovariants

Enfin les plans isovariants ont été utilisés pour l'étude des phénomènes suivants :

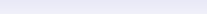
1. Étude des propriétés de certains tenseurs physiques (Hext 1963)
2. Régression sphérique (Kupper 1970, 1973)
3. Données directionnelles (Mardia 1975, Prentice 86, Prentice 89, Mardia et Jupp 2000)
4. Reconnaissance de formes (Dette, Melas et Pepelyshev 2005)











## Rappel du problème

### Mesure de planification, matrice d'information

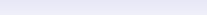
Posons  $H$  la classe des mesures de probabilité sur les boréliens de  $\mathcal{X}$ .

Chaque élément  $\eta \in H$  est appelé mesure (de planification).  
 Pour chaque  $\eta \in H$ , la matrice d'information, ou matrice des moments, de la mesure  $\eta$  est :

$$M(\eta) = \mathbb{E} [\tilde{x}\tilde{x}']$$

où  $\tilde{x}$  est un vecteur aléatoire distribué suivant la loi  $\eta$ .

L'existence de  $M(\eta)$ , pour tout  $\eta \in H$ , est assurée le fait que  $\mathcal{X}$  est compact.



## Rappel du problème

### Critère d'optimalité

Nous avons introduit précédemment l'ensemble  $\mathcal{M}$  des matrices des moments  $M(\eta)$  lorsque  $\eta$  parcourt l'ensemble de tous les plans :

$$\mathcal{M} = \{M(\eta) : \eta \in H\}.$$

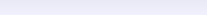
Supposons qu'un critère d'optimalité  $\phi$  soit défini sur  $\mathcal{S}_k$ , majoré sur  $\mathcal{M}$  sans être nécessairement minoré.

Par exemple, si  $\phi = \log \det$  et si  $M(\eta)$  n'est pas inversible, nous poserions  $\phi(M(\eta)) = -\infty$ .



## Kiefer 1974

Comme nous l'avons déjà indiqué dans la section précédente, la plupart des critères d'optimalité  $\Phi$  sont des fonctions concaves ou peuvent être remplacés par des fonctions qui sont concaves et qui sont associés à des critères équivalents.



# Rappel du problème

## Objectif

Notre objectif est de trouver un plan approché  $\eta_*$   $\phi$ -optimal, c'est-à-dire tel que :

$$\eta_* = \operatorname{argmax}_{\eta \in H} (\phi(M(\eta)))$$



## Structure de $\mathcal{M}$

### Proposition

1. *Chaque élément de  $\mathcal{M}$  est une matrice symétrique semi-définie positive et peut donc être représentée par un point de  $\mathbb{R}^{k(k+1)/2}$ , le point de coordonnées  $(m_{ij}; 1 \leq i \leq j \leq n)$  lorsque  $M = ((m_{ij}))$ .*
2.  *$\mathcal{M}$  est un ensemble convexe puisque  $\mathcal{M}$  est l'adhérence de l'enveloppe convexe de  $\{xx', x \in \mathcal{X}\}$ .*

Remarquons que si  $\eta_x$  est la masse de Dirac en  $x$ , alors  $M(\eta_x) = xx'$ .





## Structure de $\mathcal{M}$

Une application du théorème de Carathéodory permet de montrer la proposition suivante.

### Proposition

*Chaque élément  $M$  de  $\mathcal{M}$  peut s'écrire comme la combinaison convexe d'au plus  $k(k+1)/2 + 1$  éléments de  $\mathcal{M}$ .*

*Si  $M$  appartient à la frontière de  $\mathcal{M}$  alors il n'est nécessaire d'utiliser qu'au plus  $k(k+1)/2$  termes dans la combinaison convexe.*

Cette proposition est extrêmement importante puisqu'elle nous assure que, si une matrice d'information  $M_\star \in H$  maximise notre critère, il sera toujours possible de trouver d'un **plan approché** porté par au plus  $k(k+1)/2 + 1$  points et tel que  $M_\star = M(\eta_\star)$ .

## Structure de $\mathcal{M}$

En d'autres termes, il existe toujours un **plan approché**, donc **discret**, qui sera solution du problème d'optimisation, ce qui est de bon augure puisque notre objectif final est d'approcher la mesure  $\eta_*$  par un plan exact, qui est nécessairement discret.

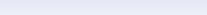


## Définition (Croissance stricte)

*Un critère d'optimalité est strictement croissant si l'implication suivante est vérifiée :*

$$\phi(M_1) \in \mathbb{R} \text{ et } M_1 - M_2 > 0 \quad \Rightarrow \quad \phi(M_1) > \phi(M_2).$$

*Par exemple, le critère  $\phi(M) = \det M$  est strictement croissant.*



## Proposition

*Si le critère  $\phi$  est strictement croissant, il existe un plan optimal dont le support est formé d'au plus  $k(k+1)/2$  points.*

## Démonstration

Si la dimension de  $\mathcal{M}$  est  $k(k+1)/2$  et si  $M$  est un point intérieur de  $\mathcal{M}$ , il existe  $a > 1$  tel que  $aM \in \mathcal{M}$  et ainsi

$$\phi(aM) = \phi(M + (a-1)M) > \phi(M).$$

Sinon la dimension de  $\mathcal{M}$  est  $< k(k+1)/2$  et tout point  $M$  de  $\mathcal{M}$  est la combinaison d'au plus  $k(k+1)/2$  points.



## Proposition

- *L'ensemble  $H$  est convexe : si  $\eta_1$  et  $\eta_2$  sont deux mesures de  $H$  et  $0 \leq \lambda \leq 1$ , il en va de même de  $\lambda\eta_1 + (1 - \lambda)\eta_2$ .*
- *De plus,  $M(\lambda\eta_1 + (1 - \lambda)\eta_2) = \lambda M(\eta_1) + (1 - \lambda)M(\eta_2)$ .*

## Proposition

*Si  $\phi$  est concave sur  $\mathcal{M}$ , alors  $f(\eta)$  est concave sur  $H$  et l'ensemble  $H_\star$  des mesures  $\phi$ -optimale est une partie convexe de  $H$ .*



## Remarque

L'application  $\eta \in H \mapsto M(\eta) \in \mathcal{M}$  n'est généralement pas injective et ainsi la concavité stricte de  $\phi$  n'implique généralement pas celle de  $f$  ni l'unicité d'une mesure  $\eta_\star$   $\phi$ -optimale.

Néanmoins dans la plupart des applications pratiques, la mesure  $\eta_\star$  sera unique.



## Définition (Dérivée directionnelle de Gâteaux)

*La dérivée directionnelle, ou dérivée de Gâteaux, de  $\phi$  en  $M_1$  dans la direction de  $M_2$  est :*

$$G_{\phi}(M_1, M_2) = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} [\phi(M_1 + \varepsilon M_2) - \phi(M_1)].$$

Deux propriétés nous utiles par la suite.



## Proposition

*Si  $\phi$  est différentiable en  $x$ , alors  $G_\phi(x, y)$  existe et*

$$G_\phi = \langle \nabla \phi(x), y \rangle = \sum y_i \frac{\partial \phi(x)}{\partial x_i}.$$

*Si  $\phi$  est une fonction concave et  $x$  un point où  $\phi$  est finie, alors  $\phi$  est différentiable en  $x$  si, et seulement si,  $G_\phi = (x, y)$  est linéaire par rapport à la seconde variable :*

$$G_\phi(x, \sum a_i y_i) = \sum a_i G_\phi(x, y_i),$$

*pour tout réel  $a_i$  et  $y_i \in \mathbb{R}^n$ .*





## Définition (Dérivée de Fréchet)

La dérivée de Fréchet de  $\phi$  en  $M_1$  dans la direction de  $M_2$  est :

$$F_{\phi}(M_1, M_2) = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} [\phi((1 - \varepsilon)M_1 + \varepsilon M_2) - \phi(M_1)].$$

C'est cette dérivée dont nous nous servons par la suite.

## Remarques

1.  $M_1, M_2 \in \mathcal{M}$  implique  $(1 - \varepsilon)M_1 + \varepsilon M_2 \in \mathcal{M}$  et ainsi  $\phi((1 - \varepsilon)M_1 + \varepsilon M_2)$  est toujours défini.



## Remarques (suite)

2. La concavité de  $\phi$  implique que

$$\frac{1}{\varepsilon} [\phi((1 - \varepsilon)M_1 + \varepsilon M_2) - \phi(M_1)]$$

est une fonction décroissante de  $\varepsilon$ , pour  $0 < \varepsilon \leq 1$ .

Par conséquent, comme  $\phi$  est concave,  $F_{\phi(M_1, M_2)}$  existe toujours si nous permettons à  $F_{\phi}$  de prendre la valeur  $+\infty$ .

3. Pour  $\varepsilon = 1$ , nous obtenons

$$F_{\phi}(M_1, M_2) \geq \phi(M_2) - \phi(M_1).$$



## Remarques (suite)

4. Par définition,  $F_\phi(M_1, M_2) = G_\phi(M_1, M_2 - M_1)$ . Par conséquent, si  $\phi$  est différentiable et si  $\sum a_i = 1$ ,

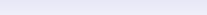
$$F_\phi(M_1, \sum a_i M_i) = \sum F_\phi(M_1, M_i).$$

Si  $\tilde{M}$  est une matrice aléatoire,  $\phi$  est différentiable, cette propriété de linéarité implique

$$\mathbb{E} \left[ F_\phi(M_1, \tilde{M}) \right] = F_\phi(M_1, \mathbb{E} \left[ \tilde{M} \right]).$$

5. Par définition, pour tout  $M$  fixé

$$F_\phi(M, M) = 0.$$



## Remarques (suite)

6. Supposons que  $\tilde{x}$  est un vecteur aléatoire à  $k$  composantes de loi  $\eta$ , et comme d'habitude  $M(\eta) = \mathbb{E}[\tilde{x}\tilde{x}']$ . Ainsi, en combinant (4) et (5), nous obtenons le résultat que si  $\phi$  est différentiable en  $M(\eta)$  alors

$$\mathbb{E}[F_\phi(M(\eta, \tilde{x}\tilde{x}'))] = F_\phi(M(\eta, \mathbb{E}[\tilde{x}\tilde{x}'])) = F_\phi(M(\eta, \eta)) = 0$$

## Transition

Ces résultats préliminaires vont nous permettre d'obtenir les principaux résultats théoriques pour construire des mesures  $\phi$ -optimales.



# Sommaire

## Introduction

Plans pour pesée

Définition générale du problème

Exemples

## Planification, modèle linéaire et symétries

Définitions, admissibilité, optimalité

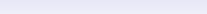
Invariances

Cas de l'isovariance

## Caractérisation différentielle de l'optimalité

Définitions, première propriétés

Théorèmes de caractérisation



# Théorèmes de caractérisation

## Théorème 1

Si  $\phi$  est concave sur  $\mathcal{M}$ , la mesure  $\eta_*$  est  $\phi$ -optimale si et seulement si  $F_\phi(M(\eta_*), M(\eta)) \leq 0$  pour tout  $\eta \in H$ .

## Démonstration

- **Condition suffisante.**

Si  $F_\phi(M(\eta_*), M(\eta)) \leq 0, \forall \eta \in H$ , la propriété (3) implique que  $\phi(M(\eta)) - \phi(M(\eta_*)) \leq 0$ , c'est-à-dire que la mesure  $\eta_*$  est  $\phi$ -optimale.



# Théorèmes de caractérisation

## Démonstration

- **Condition nécessaire.**

Si  $\phi$  est maximal en  $M(\eta_*)$  alors

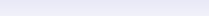
$$\phi((1 - \varepsilon)M(\eta_*) + \varepsilon M(\eta)) - \phi(M(\eta_*)) \leq 0$$

pour tout  $\varepsilon$  dans  $[0; 1]$  et tout  $\eta \in H$ , puisque

$$(1 - \varepsilon)M(\eta_*) + \varepsilon M(\eta) = M(1 - \varepsilon)\eta_* + \varepsilon\eta$$

ce qui implique, d'après la définition de  $F_\phi$ , que

$$F_\phi(M(\eta_*), M(\eta)) \leq 0, \quad \forall \eta \in H.$$



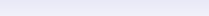
## Théorèmes de caractérisation

### Remarque

La preuve de ce théorème est réellement simple, ce qui n'est pas surprenant puisqu'il exprime l'information suivante : nous savons que nous sommes sur le sommet d'une « montagne » concave et il n'est pas possible à partir de cet endroit de regarder en l'aire vers un autre point de cette « montagne ».

Ce théorème n'a malheureusement que peu d'intérêt pratique car  $\mathcal{M}$  est une partie de  $\mathbb{R}^{k(k+1)/2}$  et ainsi il existe, même si  $k$  est modérément élevé, rapidement un très grand nombre de directions dans lesquelles nous devons regarder pour savoir si la condition est vérifiée.





## Théorèmes de caractérisation

Si la « montagne » est différentiable en son sommet, il n'est en fait pas nécessaire de regarder dans toutes les directions mais seulement dans la direction des points extrémaux de sa base convexe.

### Théorème 2

Si  $\phi$  est concave sur  $\mathcal{M}$  et différentiable en  $M(\eta_*)$  alors  $\eta_*$  est  $\phi$ -optimal si, et seulement si,

$$F_{\phi}(M(\eta_*), xx') \leq 0, \quad x \in \mathcal{X}.$$



## Théorèmes de caractérisation

### Démonstration

- **Condition nécessaire.**

Conséquence immédiate du résultat précédent.

- **Condition suffisante.**

Chaque  $M(\eta)$  peut se mettre sous la forme :

$$M(\eta) = \sum_{i=1}^I \lambda_i x_i x_i'$$

où  $\sum \lambda_i = 1$  et chaque  $\lambda_i > 0$ . Alors si  $\phi$  est différentiable en  $M(\eta_*)$

$$F_\phi(M(\eta_*), M(\eta)) = \sum_i \lambda_i F_\phi(M(\eta_*), x_i x_i'),$$

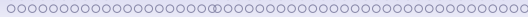


## Théorèmes de caractérisation

d'après la propriété (4). Par conséquent,  
 $F_\phi(M(\eta_\star), xx') \leq 0$  pour tout  $x \in \mathcal{X}$  implique  
 $F_\phi(M(\eta_\star), M(\eta)) \leq 0$  pour tout  $\eta \in H$  et la condition  
 suffisante découle du théorème précédent.

### Remarque

Ce théorème est le théorème clef pour caractériser les mesures  $\phi$ -optimale pour les modèles de régression linéaire parce qu'il fournit une condition qui est être vérifiée facilement en pratique.



## Théorèmes de caractérisation

### Régression quadratique avec une variable contrôlée

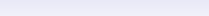
Supposons que nous avons une seule variable contrôlée  $u$ ,  
 $U = [-1; 1]$  et

$$\mathbb{E}[\tilde{y}|u, \theta] = \theta_0 + \theta_1 u + \theta_2 u^2.$$

Par conséquent l'espace induit par le plan  $\mathcal{X}$  est une partie de  $\mathbb{R}^3$  :

$$\mathcal{X} = \left\{ x = \begin{pmatrix} 1 \\ u \\ u^2 \end{pmatrix}; -1 \leq u \leq 1 \right\},$$

qui est un arc de parabole dans le plan  $x_1 = 1$ .



## Théorèmes de caractérisation

### Régression quadratique avec une variable contrôlée

Supposons que le critère qui nous intéresse soit celui de  $D$ -optimalité :

$$\begin{aligned} \phi(M) &= \log \det M, & \text{si } M \text{ est inversible,} \\ &= -\infty, & \text{si } M \text{ n'est pas inversible.} \end{aligned}$$

### Remarque

Nous utilisons  $\phi = \log \det$  à la place de  $\det$  non seulement pour que notre critère soit concave mais aussi pour ne retenir des mesures  $\eta$  que pour lesquelles  $M(\eta)$  est inversible.



## Théorèmes de caractérisation

### Régression quadratique avec une variable contrôlée

Commençons par calculer  $F_\phi(M_1, M_2)$  pour  $M_1$  inversible. Il est légèrement plus simple, et c'est généralement le cas, de commencer par calculer  $G_\phi(M_1, M_2)$ .

$$\begin{aligned} \log \det M_1 + \varepsilon M_2 - \log \det M_1 &= \log \det(I + \varepsilon M_2 M_1^{-1}) \\ &= \log(1 + \varepsilon \operatorname{tr}(M_2 M_1^{-1})) + O(\varepsilon^2) \\ &= \varepsilon \operatorname{tr}(M_2 M_1^{-1}) + O(\varepsilon^2). \end{aligned}$$

Par conséquent,  $G_\phi(M_1, M_2) = \operatorname{tr}(M_2 M_1^{-1})$ .



## Théorèmes de caractérisation

### Régression quadratique avec une variable contrôlée

Cette application est linéaire par rapport à la seconde variable, le critère est différentiable pour chaque mesure  $\eta$  pour lequel il est fini et le deuxième théorème s'applique puisque une mesure  $\eta$  ne peut être  $D$ -optimale que si  $M(\eta)$  est inversible.

Par conséquent,

$$F_{\phi}(M_1, M_2) = G_{\phi}(M_1, M_2 - M_1) = \text{tr}(M_2 M_1^{-1}) - 3,$$

et

$$F_{\phi}(M_1, xx') = x' M_1^{-1} x - 3.$$



## Théorèmes de caractérisation

### Régression quadratique avec une variable contrôlée

Par conséquent, le second théorème nous indique que  $\eta_*$  est  $D$ -optimal si, et seulement si,

$$x' M(\eta_*)^{-1} x \leq 3, \quad x \in \mathcal{X}.$$

Si  $N = 3m$ , soit le plan constitué de  $m$  observations en  $-1$ ,  $m$  en  $0$  et  $m$  en  $1$  et  $\eta_0$  la mesure discrète associée.

Nous vérifierons en exercice que le plan  $\eta_0$  est  $D$ -optimal.





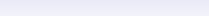
## Théorèmes de caractérisation

### Théorème 3

Si  $\phi$  est différentiable en tous les points de  $\mathcal{M}^+$ , le sous-ensemble de  $\mathcal{M}$  où  $\phi(M) > -\infty$ , et si une mesure  $\phi$ -optimale existe, alors  $\eta_\star$  est  $\phi$ -optimale si, et seulement si,

$$\max_{x \in \mathcal{X}} F_\phi(M(\eta_\star), xx') = \min_{\eta} \max_{x \in \mathcal{X}} F_\phi(M(\eta), xx').$$

Le minimum par rapport à  $\eta$  est pris sur l'ensemble  $\{\eta; M(\eta) \in \mathcal{M}^+\}$ .



## Théorèmes de caractérisation

### Démonstration

Nous avons vu à la proposition (6) que si  $\tilde{x}$  est un vecteur aléatoire de loi  $\eta$  et  $\phi$  différentiable en  $M(\eta)$  alors :

$$\mathbb{E}[F_\phi(M(\eta), \tilde{x}\tilde{x}')] = 0.$$

Par conséquent, pour tout  $\eta$  de  $\{\eta : M(\eta) \in \mathcal{M}^+\}$ , nous avons

$$\max_{x \in \mathcal{X}} F_\phi(M(\eta), xx') \geq 0.$$

Or, d'après le théorème 2,  $\eta_\star$  est  $\phi$ -optimal si, et seulement si,

$$\max_{x \in \mathcal{X}} F_\phi(M(\eta), xx') \leq 0,$$



## Théorèmes de caractérisation

### Démonstration

et ainsi

$$\min_{\eta} \max_{x \in \mathcal{X}} F_{\phi}(M(\eta), xx') = 0,$$

le minimum étant atteint pour un plan  $\eta_{\star}$   $\phi$ -optimal lorsqu'un tel plan existe. Par conséquent, la condition indiquée est nécessaire.

D'autre part, si  $\eta_{\dagger}$  satisfait la condition et si une mesure  $\phi$ -optimale  $\eta$  existe, alors

$$\max_{x \in \mathcal{X}} F_{\phi}(M(\eta_{\dagger}), xx') = 0,$$

et en appliquant le second théorème,  $\eta_{\dagger}$  est  $\phi$ -optimal. Ainsi, la condition est également suffisante.



## Théorèmes de caractérisation

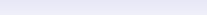
### Corollaire

*Si  $\eta_*$  est  $\phi$ -optimale et si  $\phi$  est différentiable en  $M(\eta_*)$  alors nous avons à la fois :*

$$\max_{x \in \mathcal{X}} F_\phi(M(\eta_*), xx') = 0$$

$$\mathbb{E}[F_\phi(M(\eta_*), \tilde{x}\tilde{x}')] = 0,$$

*où  $\tilde{x}$  est un vecteur aléatoire de loi  $\eta_*$ .*



## Théorèmes de caractérisation

### Corollaire (suite)

Ceci ne peut se produire que si

$$F_{\phi}(M(\eta_{\star}), xx') = 0$$

$\eta_{\star}$  presque sûrement. Si  $\eta_{\star}$  est discret, de support fini  $x_1, \dots, x_n$ , alors

$$F_{\phi}(M(\eta_{\star}), x_i x_i') = 0, \quad i = 1, 2, \dots, l.$$

### Remarque

L'intérêt pratique du théorème 3, n'est pas aussi important que celui du théorème 2 car le critère de  $\phi$ -optimalité obtenu n'est pas facilement vérifiable en pratique.



## Théorèmes de caractérisation

### Théorème 3 et $D$ -optimalité

L'intérêt du théorème 3 est principalement confiné au cas du critère  $\phi = \log \det$ . En effet, nous avons vu dans l'exemple que

$$F_{\phi}(M, xx') = x'M^{-1}x - k.$$

Par définition, une mesure  $\eta_{\star}$  est  $G$ -optimale si

$$\max_{x \in \mathcal{X}} F_{\phi}(M(\eta_{\star}), xx') = \min_{\eta} \max_{x \in \mathcal{X}} F_{\phi}(M(\eta), xx').$$

Ainsi en prenant le critère  $\phi = \log \det$  et en appliquant le théorème 3, nous obtenons l'équivalence pour les mesures des critères de  $D$ -optimalité et de  $G$ -optimalité.

# Théorèmes de caractérisation

## Théorème

*Théorème d'équivalence de Kiefer et Wolfowitz (1960) Pour les mesures  $\eta$ , les critères de  $D$ -optimalité et de  $G$ -optimalité coïncident.*

## Remarque

Le résultat d'équivalence précédent n'est démontré que pour les **mesures** et n'est pas nécessairement vrai pour les  $N$ -plans.