Planification optimale des expériences

Introduction et cas du modèle linéaire

Frédéric Bertrand[1]

[1]IRMA, Université de Strasbourg
Strasbourg, France

ENSAI 3e Année
2014-2015

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**, 7th Edition, aux éditions Wiley, 2009.
5. le livre de Samuel D. Silvey, **Optimal Design**, aux éditions Chapman and Hall, 1980.

Sommaire

Sommaire

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.

Sommaire

Plans pour pesée

Contexte

Un exemple simple pour comprendre pourquoi il est important de réfléchir à la planification les expériences.

Supposons que nous devions **peser quatre objets** avec une balance ayant **deux plateaux** en réalisant au maximum **quatre pesées**.

À chacune des pesées, nous pouvons répartir tous ou une partie des objets sur **l'un ou l'autre des deux plateaux** de la balance.

Plans pour pesée

Définition

La répartition des objets lors des quatre pesées est appelée un *plan pour pesée*.

Mise en équation

Soient β_1, β_2, ..., et β_n les poids, inconnus, des objets.

Nous posons $u_{ij} = 1$ si le $j-$ème objet est dans le plateau de droite au cours de la $i-$ème pesée, $u_{ij} = -1$ si le $j-$ème objet est dans le plateau de gauche au cours de la $i-$ème pesée et $u_{ij} = 0$ si le $j-$ème objet n'est pas utilisé lors de la $i-$ème pesée.

Plans pour pesée

Mise en équation (suite)

Afin d'équilibrer les deux plateaux de la balance lors de la $i-$ème pesée il est possible de **rajouter une masse** à l'un des deux plateaux de la balance.

Le signe de cette masse dépend du plateau sur lequel elle a été posée : positif si elle se trouve sur le plateau de gauche et négatif si elle se trouve sur le plateau de droite.

Nous désignons par y_i cette masse signée utilisée lors de la $i-$ème pesée.

Plans pour pesée

Mise en équation (suite)

Supposons que les erreurs de mesure ne sont pas corrélées, de dispersion stable et sans biais systématique, ce qui revient à supposer que les y_i sont des variables aléatoires **non corrélées**, de **même variance**, notée σ_e^2, et d'**espérance** égale à la valeur de la masse qu'il faut ajouter lors de la $i-$ème pesée afin d'équilibrer les deux plateaux de la balance.

Plans pour pesée

Comparaisons de deux plans

Le premier plan pour pesée.

1. Pesée 1 : objet 1 à droite ;
2. Pesée 2 : objet 2 à droite ;
3. Pesée 3 : objet 3 à droite ;
4. Pesée 4 : objet 4 à droite.

La matrice $\boldsymbol{X}$ associée est

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Plans pour pesée

Comparaisons de deux plans

Le second plan pour pesée.

1. Pesée 1′ : objets 1, 2, 3 et 4 à droite ;
2. Pesée 2′ : objets 3 et 4 à gauche, 1 et 2 à droite ;
3. Pesée 3′ : objets 2 et 4 à gauche, 1 et 3 à droite ;
4. Pesée 4′ : objets 2 et 3 à gauche, 1 et 4 à droite.

La matrice $\boldsymbol{X}$ associée est

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Plans pour pesée

Comparaisons de deux plans

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

1. $\widehat{\beta_1} = y_1$, $\mathrm{Var}[\widehat{\beta_1}] = \sigma_e^2$;
2. $\widehat{\beta_2} = y_2$, $\mathrm{Var}[\widehat{\beta_2}] = \sigma_e^2$;
3. $\widehat{\beta_3} = y_3$, $\mathrm{Var}[\widehat{\beta_3}] = \sigma_e^2$;
4. $\widehat{\beta_4} = y_4$, $\mathrm{Var}[\widehat{\beta_4}] = \sigma_e^2$.

Plans pour pesée

Comparaisons de deux plans

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

1. Pesée 1 : $\widehat{\beta_1} = \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$, $\mathrm{Var}[\widehat{\beta_1}] = \frac{\sigma_e^2}{4}$;
2. Pesée 2 : $\widehat{\beta_2} = \frac{1}{4}(y_1 + y_2 - y_3 - y_4)$, $\mathrm{Var}[\widehat{\beta_2}] = \frac{\sigma_e^2}{4}$;
3. Pesée 3 : $\widehat{\beta_3} = \frac{1}{4}(y_1 - y_2 - y_3 + y_4)$, $\mathrm{Var}[\widehat{\beta_3}] = \frac{\sigma_e^2}{4}$;
4. Pesée 4 : $\widehat{\beta_4} = \frac{1}{4}(y_1 - y_2 + y_3 - y_4)$, $\mathrm{Var}[\widehat{\beta_4}] = \frac{\sigma_e^2}{4}$.

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 sur la balance pour les peser tous quatre fois, plutôt que de les peser chacun directement. En effet, dans le premier cas, l'estimateur est une moyenne de quatre termes non corrélés, ce qui réduit sa variance.

Sommaire

Définition générale du problème

Modélisation de la réponse $\tilde{y}$

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

1. un vecteur colonne $u = (u_1, u_2, \ldots, u_r)'$ 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 $\tilde{y}$

2. un vecteur colonne $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$ 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 $\tau = (\tau_1, \tau_2, \ldots, \tau_r)'$ 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 θ strictement à inférieur au nombre k de paramètres.

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

Définition générale du problème

Domaines d'appartenance des vecteurs

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

Définition générale du problème

Loi de $\tilde{y}$

Nous supposons que, pour des valeurs données de u, θ et τ, 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 $u_{(1)}, u_{(2)}, \ldots, u_{(N)}$ 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 θ et τ 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 souhaitions estimer θ.

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 (θ, τ) 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 $u_{(1)}$, $u_{(2)}$, ..., $u_{(N)}$ est

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

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

Définition générale du problème

Information de Fisher

Supposons maintenant que $\boldsymbol{u}$ est tel que la matrice $L(\boldsymbol{u}; \theta, \tau)$ est inversible [1].

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

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

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

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 $\boldsymbol{u}$ 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, $V(\boldsymbol{u}; \theta, \tau)$ est exactement la variance de l'estimateur de Gauss-Markov de θ ;
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 θ est approximativement $V(\boldsymbol{u}; \theta, \tau)$ puisque ces estimateurs sont asymptotiquement efficaces.

Définition générale du problème

Estimation d'une fonction du paramètre

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

Posons g la fonction $(g_1, \ldots, g_s)'$ à valeurs dans $\mathbb{R}^s$ et notons $Dg(\theta) = \left(\left(\dfrac{\partial g_j(\theta)}{\partial \theta_i} \right) \right)_{1 \leqslant i \leqslant k, 1 \leqslant j \leqslant s}$ sa différentielle au point θ lorsqu'elle existe.

Définition générale du problème

Estimation d'une fonction du paramètre

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

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

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

Définition générale du problème

Estimation d'un fonction du paramètre

Le cas de l'estimation du paramètre θ correspond au cas où $g_i(\theta) = \theta$ pour tout $i = 1, \ldots, k$ 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 θ.

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 $\boldsymbol{u}$ 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 $\boldsymbol{u}$ 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 pourrons 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 $V_g(\boldsymbol{u};\theta,\tau)$ « petite » ?
2. est-il possible de trouver un plan $\boldsymbol{u}$, noté $\boldsymbol{u}_\star$, indépendant de θ et τ, qui rende $V_g(\boldsymbol{u};\theta,\tau)$ « petite », au sens où nous le souhaitons, pour toutes les valeurs de θ et τ ?

Définition générale du problème

Objectif de la planification

Pour toute une classe de problèmes, le modèle linéaire par exemple, l'expression de $V_g(\boldsymbol{u};\theta,\tau)$ 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 $\boldsymbol{u}_\star$ n'existe pas, nous devrons 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

Exemple 1

Régression linéaire simple, gaussienne et contrôlée

Dans cet exemple :

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

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

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

Exemple 1

Régression linéaire simple, gaussienne et contrôlée

Un calcul direct permet de montrer que

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

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

Exemple 1

Régression linéaire simple, gaussienne et contrôlée

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

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

« aussi grande que possible ».

1. Il est possible de montrer qu'il n'existe pas, même dans ce cas simple, de plan $\boldsymbol{u}_\star$ qui soit le « meilleur » au sens matriciel fort, ordre de Loewner, suivant : $M(\boldsymbol{u}_\star) - M(\boldsymbol{u})$ est semi-définie positive pour tout plan $\boldsymbol{u}$.
   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 avions travaillé en l'absence de l'hypothèse de normalité mais en conservant les

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

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

Exemple 2

Plans factoriel complet 2^2

Une partie de la planification d'expériences traite de la manière de répartir les traitements de plusieurs facteurs à des unités expérimentales. La différence apparente avec notre approche est qu'il s'agit de planifier des expériences avec des variables qualitatives et non quantitatives.

Montrons, à l'aide de l'exemple simple suivant, comment inclure également ce domaine dans notre cadre général.

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 :

- θ_1 l'effet de T_1 seul ;
- θ_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

Codons ces situations, avec $u = (u_1, u_2)$ de la manière suivante :

1. $(0, 0)$;
2. $(1, 0)$;
3. $(0, 1)$;
4. $(1, 1)$.

Le rôle du planificateur est de choisir, pour chaque unité expérimentale, les valeurs de u_1 et u_2 dans $\{0; 1\}$.

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 :

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

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

Dans cet exemple :

- u est un nombre réel ;
- $\theta = (\theta_0,\theta_1,\theta_2)'$, avec $\theta_2 < 0$;
- τ est un nombre réel strictement positif.

Pour des valeurs de u, θ et τ données, $\tilde{y}$ suit une loi $\mathcal{N}(\theta_0+\theta_1 u+\theta_2 u^2,\tau)$.

Supposons que nous souhaitions trouver la valeur de u pour laquelle $\mathbb{E}(\tilde{y})$ est maximale, c'est-à-dire estimer

$$g(\theta) = \frac{-\theta_1}{2\theta_2}.$$

Exemple 3

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

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

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

$V(\boldsymbol{u};\theta,\tau)$ 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-Darmois-Cramer-Rao est :

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

avec ici $\{Dg(\theta)\}' = \dfrac{1}{2\theta_2^2}\begin{pmatrix}0 & -\theta_2 & \theta_1\end{pmatrix}$.

La quantité précédente dépend de θ et rien ne permet d'assurer qu'il existe un plan $\boldsymbol{u}_\star$ qui la rende minimale pour toutes les valeurs de θ 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 Φ 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 θ_1 et θ_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'a 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 $\boldsymbol{u}$, la matrice d'information au sens de Fisher pour les paramètres θ_1 et θ_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 Φ_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 $\boldsymbol{u}$ qui rende cette matrice la plus grande possible en un sens que nous préciserons dans la suite.

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

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 θ_1 et θ_2.

Or les valeurs de u pour lesquelles ceci se produit dépendent de θ_1 et θ_2 !

Exemple 4

Analyse probit : un modèle non-gaussien et non linéaire

Puisque nous ignorons les valeurs réelles des paramètres θ_1 et θ_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**.

Sommaire

Sommaire

Introduction
- Plans pour pesée
- Définition générale du problème
- Exemples

Planification, modèle linéaire et symétries
- Définitions, admissibilité, optimalité
- Caractérisation différentielle de l'optimalité

Théorèmes de caractérisation
- Matrices d'information inversibles
- Matrices d'information singulières

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 $u \in \mathcal{U}$;
- $\boldsymbol{f} = (f_1, \ldots, f_k)'$ est une fonction connue définie sur $\mathcal{U}$ et à valeurs dans $\mathbb{R}^k$, posons $\mathcal{X} = f(\mathcal{U})$;
- $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)' \in \mathbb{R}^k$ est un vecteur de paramètres inconnus.

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

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

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

$$\mathrm{Var}[\tilde{y}_i | u, \theta, \tau] = \tau \quad i = 1, \ldots, n. \tag{4}$$

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

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

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

où v est une fonction de u connue.

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

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

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

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

Définition

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

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

r est le nombre de points distincts utilisés pour construire le plan.

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

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

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

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

et représenter un plan ξ par un couple $(\boldsymbol{X},\boldsymbol{w})$, où $\boldsymbol{X}$ est un ensemble fini de points $\{x_1,\ldots,x_r\} \in \boldsymbol{\mathcal{X}}$, l'espace induit par le plan.

Plan approché = mesure de probabilité de support fini sur $\boldsymbol{\mathcal{X}}$.

Plan exact de taille n = plan avec des poids $\boldsymbol{\xi_n}(x_i) = \frac{n_i}{n}$, n_i entiers de somme n.
Dans la réalité, il n'est possible de n'organiser des expériences que selon ces plans d'où leur qualificatif.

La qualité d'un plan expérimental est traduite par sa **matrice des moments** $M(\boldsymbol{\xi})$.

Définition

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

$$M(\xi) = \sum_{i=1}^{r} \xi(u_i)\boldsymbol{f}(u_i)\boldsymbol{f}(u_i)' = \sum_{i=1}^{r} w_i x_i x_i'.$$

Pour un plan exact, nous avons :

$$M(\xi_{\boldsymbol{n}}) = \sum_{i=1}^{n} \boldsymbol{f}(u_i)\boldsymbol{f}(u_i)' = \sum_{i=1}^{n} x_i x_i'.$$

Proposition

1. *$M(\xi)$ matrice carrée d'ordre k, matrice polaire d'une forme quadratique positive.*
2. *Si $\xi_{\boldsymbol{n}}$ plan exact et $M(\xi_{\boldsymbol{n}})$ inversible alors $\frac{\tau}{n}M(\xi_{\boldsymbol{n}})^{-1}$ matrice de variance-covariance de l'estimateur des moindres carrés ordinaires de $\boldsymbol{\theta}$.*
3. *Si modèle linéaire gaussien, alors $\frac{n}{\tau}M(\xi_{\boldsymbol{n}})$ **matrice d'information de Fisher**.*

$\Rightarrow$ Naturel de comparer les plans au travers des propriétés de leur matrice des moments.

Notions abordées dans la suite pour comparer les plans.

1. **admissibilité**
2. **optimalité alphabétique**
3. diverses notions d'invariance associées à un plan comme : l'**équivariance**, l'**invariance** et l'**invariance faible**.

Admissibilité

Il existe un ordre partiel sur les formes quadratiques positives : l'ordre de Loewner.

Définition

$$A \leqslant B \Longleftrightarrow B - A \in \mathcal{S}_k^+ \quad \left(\forall (A,B) \in \mathcal{S}_k^2\right) \tag{5}$$

où $\mathcal{S}_k^+$ est le cône convexe des matrices symétriques réelles positives et $\mathcal{S}_k$ l'espace vectoriel réel des matrices symétriques réelles.

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

Définition

Un plan ξ_0 est dit admissible s'il n'existe pas de plan ξ tel que :

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

où $M(\xi_0)$ et $M(\xi)$ sont les matrices des moments respectivement du plan ξ_0 et du plan ξ.

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

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

$$A \leqslant B \Longleftrightarrow \Phi(A) \geqslant \Phi(B), \quad \forall (A,B) \in \mathcal{A}^2. \tag{7}$$

3. *Φ est convexe.*

Recherche d'un plan $\Phi-$optimal ξ.

Chercher une solution au problème de minimisation suivant :

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

Posons $\mathcal{M}$ l'ensemble des matrices des moments $M(\xi)$ lorsque ξ parcourt l'ensemble de tous les plans.
Le problème de minimisation donné par l'équation 8 devient alors :

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

Optimalité

Définition alternative d'un critère d'optimalité

Chez certains auteurs la fonction Φ est

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

$$A \leqslant B \Longleftrightarrow \Phi(A) \leqslant \Phi(B), \quad \forall (A,B) \in \mathcal{A}^2. \tag{10}$$

3. Φ 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 ϕ par $-\phi$.

Définition (Concavité et concavité stricte)

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

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

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 ϕ est strictement concave.

Proposition

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

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

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

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 approximative.

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 ϕ sur celui-ci revient à chercher le minimum d'une fonction à valeurs réelles définie uniquement sur les nombres entiers !

Dans le cas d'une fonction à valeurs réelles définie uniquement sur les nombres entiers mais naturellement prolongeable à $\mathbb{R}_+$, nous chercherions ses variations et son minimum $t_\star$ sur $\mathbb{R}_+$ à l'aide des outils du calcul différentiel, puis nous chercherions à montrer que le minimum sur les entiers est en fait atteint pour un entier proche de $t_\star$.

C'est cette idée qui a amené Kiefer à introduire les plans approchés.

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

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

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 θ, pour un niveau de confiance fixé et une somme des carrés résiduelles donnée, provenant de l'utilisation du plan $\boldsymbol{\xi}$ est de la forme

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

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

Remarque

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

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

Remarque

Supposons que nous nous intéressions prioritairement à certaines combinaisons linéaires fixées des paramètres $\theta_1,\ldots,\theta_n$, ces s combinaisons linéaires sont les éléments du vecteur

$$A'\theta$$

où A' une matrice $s \times k$ de rang $s < k$.

Remarque

Si $\boldsymbol{x}$ est un N-plan pour lequel $M(\boldsymbol{x})$ est inversible, la variance de l'estimateur des moindres carrées de $A'\theta$ est proportionnelle à

$$A'M(\boldsymbol{x})^{-1}A$$

et pour les mêmes raisons qu'au paragraphe précédent, un critère naturel d'optimalité est de minimiser

$$\Phi_{D_A}(M(\boldsymbol{x})) = \det(A'M(\boldsymbol{x})^{-1}A)$$

Il s'agit du critère de D_A-optimalité (Sibson 1974).

Remarque

La situation canonique d'utilisation de ce critère se présente dans les conditions suivantes :

$$A' = \begin{pmatrix} I_s & 0 \end{pmatrix},$$

c'est-à-dire lorsque nous sommes intéressés dans les estimations des s premiers paramètres $\theta_1, \ldots, \theta_s$ avec $s < k$.

Nous partitionnons alors $M(\boldsymbol{x})$ en conséquence

$$M(\boldsymbol{x}) = \begin{pmatrix} M_{11}(\boldsymbol{x}) & M_{12}(\boldsymbol{x}) \\ M'_{12}(\boldsymbol{x}) & M_{22}(\boldsymbol{x}) \end{pmatrix}.$$

Remarque

Un calcul rapide d'algèbre linéaire (complément de Schur) permet de montrer que

$$A'M(\boldsymbol{x})^{-1}A = M_{11}(\boldsymbol{x}) - M_{12}(\boldsymbol{x})M_{22}(\boldsymbol{x})^{-1}M'_{12}(\boldsymbol{x}).$$

Notre critère revient donc à trouver un N-plan $\boldsymbol{x}$ qui minimise le déterminant de cette matrice.

$$\Phi_{D_A}(M(\boldsymbol{x})) = \det\left(M_{11}(\boldsymbol{x}) - M_{12}(\boldsymbol{x})M_{22}(\boldsymbol{x})^{-1}M'_{12}(\boldsymbol{x})\right) = \Phi_{D_s}(M(\boldsymbol{x}))$$

Ce critère Φ_{D_s} est appelé D_s-optimalité (Karlin & Studden 1976).

Remarque

Il existe une difficulté cachée qui amène à des complications théoriques importantes.

Si nous souhaitons estimer $A'\theta$, il est possible qu'un plan $\boldsymbol{x}$ permette d'estimer $A'\theta$ de manière optimale sans qu'il ne soit possible d'estimer θ !

Exemple

Considérons :

- deux variables contrôlées x_1 et x_2 ;
- $\mathbb{E}[\tilde{y}|x_1, x_2, \theta] = \theta_1 x_1 + \theta_2 x_2$;
- $\mathcal{X}$ le triangle de sommets $(0;0)$, $(1;0)$ et $(0;1)$.

Si nous nous intéressions à l'estimation de θ_1, alors il est intuitif que le N-plan qui place N observations en $(1,0)$ sera au minimum un bon candidat pour finir le meilleur estimateur de θ_1.

Exemple

Pour ce plan, nous avons

$$M(\boldsymbol{x}) = \begin{pmatrix} N & 0 \\ 0 & 0 \end{pmatrix},$$

qui, bien entendu, n'est pas inversible.

Remarque

Par conséquent, dans le développement de la théorie de D_A-optimalité, nous devons tenir compte des plans $\boldsymbol{x}$ tels que $A'\theta$ est estimable alors que $M(\boldsymbol{x})$ n'est pas inversible.

Définition (Optimalité minimax)

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

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

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

Définition ($G-$optimalité)

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

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

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

Définition ($E-$optimalité)

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

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

En utilisant ce critère, pour choisir un plan nous cherchons à rendre minimale la valeur maximale de la variance de l'estimateur de $\boldsymbol{\theta}' x$ 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 :

$$\Phi_I(A) = \frac{1}{\mathrm{vol}(\mathcal{U})} \int_{\mathcal{U}} \boldsymbol{f}(u)' A^{-1} \boldsymbol{f}(u) \mathrm{d}u, \tag{15}$$

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

Remarque ($I-$optimalité généralisée)

Considérons la moyenne par rapport à une mesure de probabilité μ définie sur une région $\mathcal{C}$. L'analogue du critère de I-optimalité est alors donné par la formule :

$$\Phi_{I_\mu}(A) = \int_{\mathcal{C}} \boldsymbol{c}' A^{-1} \boldsymbol{c} \mu(\mathrm{d}\boldsymbol{c}) = \mathrm{tr}(A^{-1}B), \qquad (16)$$

où tr désigne la trace et $B = \int_{\mathcal{C}} \boldsymbol{c}\boldsymbol{c}' \mu(\mathrm{d}c)$ qui est encore une matrice semi-définie positive.

Remarque ($I-$optimalité généralisée)

Si B est de rang s, alors $B = RR'$ où R est une matrice $k \times s$ de rang s. Ce critère se met alors sous la forme :

$$\Phi_{I_\mu}(A) = \mathrm{tr}(R'A^{-1}R),$$

ce qui met en évidence son lien avec la D_A-optimalité. À nouveau si $s < k$, nous devons faire face à la même difficulté qu'avec la D_A-optimalité et devons considérer la possibilité que des plans avec une matrice M non-inversible soient optimaux.

Définition ($c-$optimalité)

C'est un cas particulier du critère précédent de I-optimalité généralisée lorsque μ attribue une probabilité de 1 à une valeur c donnée.

$$\Phi_c(A) = \boldsymbol{c}' A^{-1} \boldsymbol{c}. \qquad (17)$$

Sommaire

Rappel du problème

Espace induit

Considérons une partie $\mathcal{X}$ compacte d'un espace euclidien de dimension k.

Nous verrons cette partie $\mathcal{X}$ comme l'**espace induit par le plan**.

Il est naturel de supposer que la partie $\mathcal{X}$ est compacte car cela sera presque systématiquement le cas lors des applications réelles.

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 η est :

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

où $\tilde{x}$ est un vecteur aléatoire distribué suivant la loi η. 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 η parcourt l'ensemble de tous les plans :

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

Supposons qu'un critère d'optimalité ϕ 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 Φ 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 une mesure de planification $\eta_\star$ $\phi-$optimale, c'est-à-dire tel que :

$$\eta_\star = \text{argmax}_{\eta \in H} (\phi(M(\eta)))$$

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 $(m_{ij}; 1 \leqslant i \leqslant j \leqslant n)$ 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 η_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 $\{xx',\ x \in \mathcal{X}\}$.

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 $M_\star \in \mathcal{M}$ maximise notre critère, il sera toujours possible de trouver 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 $\eta_\star$ par un plan exact, qui est nécessairement discret.

Définition (Croissance stricte)

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

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

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

Introduction Planification, modèle linéaire et symétries Théorèmes de caractérisation

Proposition

Si le critère ϕ 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.

Introduction Planification, modèle linéaire et symétries Théorèmes de caractérisation

Proposition

- *L'ensemble H est convexe : si η_1 et η_2 sont deux mesures de H et $0 \leqslant \lambda \leqslant 1$, 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 ϕ est concave sur $\mathcal{M}$, alors $f = \phi \circ M$ est concave sur H et l'ensemble $H_\star$ des mesures $\phi-$optimales est une partie convexe de H.

Introduction Planification, modèle linéaire et symétries Théorèmes de caractérisation

Remarque

L'application $\eta \in H \mapsto M(\eta) \in \mathcal{M}$ n'est généralement pas injective et ainsi la concavité stricte de ϕ 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 ϕ en M_1 dans la direction de M_2 est :

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

Deux propriétés nous utiles par la suite.

Proposition

Si ϕ 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 ϕ est une fonction concave et x un point où ϕ est finie, alors ϕ 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 ϕ en M_1 dans la direction de M_2 est :

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

C'est cette dérivée dont nous nous servirons 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 ϕ implique que

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

est une fonction décroissante de ε, pour $0 < \varepsilon \leqslant 1$. Par conséquent, comme ϕ est concave, $F_\phi(M_1, M_2)$ existe toujours si nous permettons à F_ϕ de prendre la valeur $+\infty$.

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

$$F_\phi(M_1, M_2) \geqslant \phi(M_2) - \phi(M_1).$$

Remarques (suite)

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

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

Si $\tilde{M}$ est une matrice aléatoire, ϕ 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 η, et comme d'habitude $M(\eta) = \mathbb{E}[\tilde{x}\tilde{x}']$. Ainsi, en combinant (4) et (5), nous obtenons le résultat que si ϕ est différentiable en $M(\eta)$ alors

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

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

Sommaire

Sommaire

Introduction
Plans pour pesée
Définition générale du problème
Exemples

Planification, modèle linéaire et symétries
Définitions, admissibilité, optimalité
Caractérisation différentielle de l'optimalité

Théorèmes de caractérisation
Matrices d'information inversibles
Matrices d'information singulières

Théorèmes de caractérisation

Matrices d'information inversibles

Théorème 1

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

Démonstration

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

Théorèmes de caractérisation

Matrices d'information inversibles

Démonstration

- **Condition nécessaire.**
  Si ϕ est maximal en $M(\eta_\star)$ alors

$$\phi((1-\varepsilon)M(\eta_\star) + \varepsilon M(\eta)) - \phi(M(\eta_\star)) \leqslant 0$$

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

$$(1-\varepsilon)M(\eta_\star) + \varepsilon M(\eta) = M\left((1-\varepsilon)\eta_\star + \varepsilon\eta\right)$$

ce qui implique, d'après la définition de F_ϕ, que

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

Théorèmes de caractérisation

Matrices d'information inversibles

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

Matrices d'information inversibles

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

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

Théorèmes de caractérisation

Matrices d'information inversibles

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 ϕ est différentiable en $M(\eta_\star)$

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

Théorèmes de caractérisation

Matrices d'information inversibles

d'après la propriété (4). Par conséquent,
$F_\phi(M(\eta_\star), xx') \leqslant 0$ pour tout $x \in \mathcal{X}$ implique
$F_\phi(M(\eta_\star), M(\eta)) \leqslant 0$ 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

Matrices d'information inversibles

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 $\mathbb{R}^3$:

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

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

Théorèmes de caractérisation

Matrices d'information inversibles

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 η que pour lesquelles $M(\eta)$ est inversible.

Théorèmes de caractérisation

Matrices d'information inversibles

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 $G_\phi(M_1, M_2)$.

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

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

Théorèmes de caractérisation

Matrices d'information inversibles

Régression quadratique avec une variable contrôlée

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

Par conséquent,

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

et

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

Théorèmes de caractérisation

Matrices d'information inversibles

Régression quadratique avec une variable contrôlée

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

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

Si $N = 3m$, soit le plan constitué de m observations en -1, m en 0 et m en 1 et η_0 la mesure discrète associée.

Nous vérifierons en exercice que le plan η_0 est $D-$optimal.

Théorèmes de caractérisation

Matrices d'information inversibles

Théorème 3

Si ϕ 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 η est pris sur l'ensemble $\{\eta; M(\eta) \in \mathcal{M}^+\}$.

Théorèmes de caractérisation
Matrices d'information inversibles

Démonstration

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

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

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

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

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

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

Théorèmes de caractérisation
Matrices d'information inversibles

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 η 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
Matrices d'information inversibles

Corollaire

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

$$\max_{x \in \mathcal{X}} F_\phi(M(\eta_\star), xx') = 0$$
$$\mathbb{E}[F_\phi(M(\eta_\star), \tilde{x}\tilde{x}')] = 0,$$

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

Théorèmes de caractérisation
Matrices d'information inversibles

Corollaire (suite)

Ceci ne peut se produire que si

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

$\eta_\star$ presque surement. Si $\eta_\star$ est discret, de support fini $x_1, \ldots, x_n$, alors

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

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
Matrices d'information inversibles

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
Matrices d'information inversibles

Théorème (Théorème d'équivalence de Kiefer et Wolfowitz (1960))

Pour les mesures η, 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.

Sommaire

Théorèmes de caractérisation
Matrices d'information singulières

Remarque

Le théorème 2 est un outil théorique puissant. Devons-nous réellement nous inquiéter de pouvoir rencontrer un plan η pour lequel $G_\phi(M(\eta), M(\eta'))$ est non-linéaire par rapport à la deuxième variable de telle sorte que la condition suffisante du théorème 2 ne soit plus vérifiée ?

Malheureusement oui !

Voici un exemple de ce type de situation

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

Considérons

- $x' = (x_1, x_2)$ et $\theta' = (\theta_1, \theta_2)$;
- $\mathcal{X}$ le quadrilatère de sommets $(0;0), (1;0), (4;1), (4;2)$;
- $\mathbb{E}[\tilde{y}|x, \theta] = \theta_1 x_1 + \theta_2 x_2$.

Cherchons à minimiser la variance de l'estimateur par moindres carrés de θ_1 ; il en résulte que θ_2 est vu comme un paramètre de nuisance.

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

Il va de soi que tous les N-plans pour lesquels la matrice d'information n'est pas singulière permettent d'estimer θ_1.

Il en va de même de ceux pour lesquels les points appartiennent à l'intervalle $(0;1]$ sur l'axe (Ox_1).

Par contre aucun autre N-plan ne le permet.

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

Si un N-plan $\boldsymbol{x}$ a une matrice d'information :

- inversible $M(\boldsymbol{x}) = (m_{ij}(\boldsymbol{x}))$, alors la variance de l'estimateur des moindres carrés de θ_1 obtenu à partir de celui-ci est proportionnelle à

$$\frac{1}{m_{11}(\boldsymbol{x}) - \frac{m_{12}^2(\boldsymbol{x})}{m_{22}(\boldsymbol{x})}};$$

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

- singulière avec $m_{22}(\boldsymbol{x}) = m_{12}(\boldsymbol{x}) = 0$ et $m_{11}(\boldsymbol{x}) \neq 0$, alors la variance de l'estimateur des moindres carrés de θ_1 obtenu à partir de celui-ci est proportionnelle à

$$\frac{1}{m_{11}(\boldsymbol{x})};$$

- dans les autres cas, θ_1 n'est pas estimable.

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

Pour les mesures η, une définition appropriée du critère d'optimalité ϕ à maximiser est ainsi :

$$\begin{aligned} \phi(M(\eta)) &= \log\left(m_{11}(\eta) - \frac{m_{12}^2(\eta)}{m_{22}(\eta)}\right), \text{ si } M(\eta) \text{ est inversible ;} \\ &= \log\left(m_{11}(\eta)\right), \text{ si } m_{11}(\eta) \neq 0 \text{ et } m_{22}(\eta) = m_{12}(\eta) = 0\,; \\ &= -\infty, \text{ sinon.} \end{aligned}$$

Nous avons introduit la fonction log pour simplifier les calculs à venir.

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

Considérons la mesure η_0 qui attribue la probabilité 1 à $(1; 0)$.

C'est un candidat sérieux pour pour être la meilleure mesure pour estimer θ_1, c'est-à-dire un plan ϕ-optimal. Nous avons :

$$M(\eta_0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

et un calcul rapide permet de montrer que :

$$\begin{aligned} F_\phi(M(\eta_0), M(\eta)) &= m_{11}(\eta) - \frac{m_{12}^2(\eta)}{m_{22}(\eta)} - 1, \text{ si } m_{22}(\eta) \neq 0\,; \\ &= m_{11}(\eta) - 1, \text{ si } m_{22}(\eta) = m_{12}(\eta) = 0. \end{aligned}$$

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

Ainsi la fonction $M(\eta) \mapsto F_\phi(M(\eta_0), M(\eta))$ n'est pas linéaire. Voyons maintenant comment le théorème 2 est mis en échec.

Nous avons l'égalité suivante :

$$\begin{aligned} F_\phi(M(\eta_0), xx') &= -1, \text{ si } x_2 \neq 0\,; \\ &= x_1^2 - 1, \text{ si } x_2 = 0. \end{aligned}$$

Par conséquent, pour tout $x \in \boldsymbol{\mathcal{X}}$, $F_\phi(M(\eta_0), xx') \leqslant 0$. Néanmoins, il ne faudrait pas conclure que la mesure η_0 est ϕ-optimale puisque nous ne nous pouvons pas appliquer le théorème 2.

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

En fait cette mesure n'est pas ϕ-optimale puisque pour la mesure η_1, qui affecte la probabilité $\frac{1}{2}$ à $(4; 1)$ et $(4; 2)$, nous obtenons :

$$F_\phi(M(\eta_0), M(\eta_1)) = \frac{3}{5} > 0.$$

Par conséquent, η_0 ne remplit pas la condition nécessaire et suffisante établie au théorème 1 qui, qant à lui, s'applique que ϕ soit différentiable, ou non, en $M(\eta_0)$.

Théorèmes de caractérisation

Matrices d'information singulières

Non-linéarité

Remarquons au passage que ϕ est différentiable en tout M inversible et que, pour une telle M,

$$F_\phi(M(\eta_0), M(\eta)) = m_{11}(\eta) - \frac{m_{12}^2(\eta)}{m_{22}(\eta)} - 1.$$

Par conséquent, la mesure attribuant une probabilité de $\frac{2}{3}$ à $(4,1)$ et $\frac{1}{3}$ à $(4,2)$ est ϕ-optimale.

Il s'agit d'un cas particulier de la D_s-optimalité que nous étudierons plus en détails dans la suite.

Théorèmes de caractérisation

Matrices d'information singulières

Mise en équation du problème

Supposons que nous souhaitions estimer certaines combinaisons linéaires des paramètres inconnus regroupées dans le vecteur

$$A'\theta,$$

où A' est une matrice de taille $s \times k$ de rang $s < k$. Comme $s < k$, certains N-plans, bien que possédant des matrices d'information singulière, permettront d'estimer le vecteur $A'\theta$.

Théorèmes de caractérisation

Matrices d'information singulières

Mise en équation du problème

Par conséquent, nous souhaitons que cette possibilité existe également pour les mesures de planification. Par analogie avec la théorie des moindres carrés, nous considérons qu'une matrice d'information M permet d'estimer $A'\theta$ si l'implication suivante est vraie

$$Mz = 0 \Rightarrow A'z = 0$$

ou de manière équivalente

$$\exists Y \in \mathcal{M}_s(\mathbb{R}) \text{ telle que } A = MY.$$

Théorèmes de caractérisation

Matrices d'information singulières

Mise en équation du problème

Posons $\mathcal{M}_A$ le sous-ensemble des matrices de $\mathcal{M}$ possédant cette propriété.

$\mathcal{M}_A$ contient :

- $\mathcal{M} \cap \mathcal{GL}_k(\mathbb{R})$;
- des matrices de $\mathcal{M}$ non-inversibles.

Théorèmes de caractérisation

Matrices d'information singulières

Mise en équation du problème

En poursuivant l'analogie avec l'estimation par moindres carrés, la matrice de variance-covariance de l'estimateur par moindres carrés de $A'\theta$ associée à une mesure de matrice d'information $M \in \mathcal{M}_A$ est proportionnelle à

$$A'M^{-}A,$$

où M^{-} est un g-inverse **quelconque** de M, c'est-à-dire une matrice $M^{-} \in \mathcal{M}_k(\mathbb{R})$ telle que $MM^{-}M = M$.

1. $A'M^{-}A$ ne dépend pas du g-inverse utilisé pour calculer cette matrice ;
2. $A'M^{-}A$ est matrice carrée d'ordre s symétrique définie positive.

Théorèmes de caractérisation

Matrices d'information singulières

Mise en équation du problème

Notre intérêt pour l'estimation de $A'\theta$, nous amène donc à vouloir « minimiser » $A'M^{-}A$:

- minimiser $\log\det(A'M^{-}A)$ ou $\mathrm{tr}(A'M^{-}A)$ ou, de manière équivalente,
- maximiser $-\log\det(A'M^{-}A)$ ou $-\mathrm{tr}(A'M^{-}A)$.

Le premier de ces critères est simplement celui de D_A optimalité et le second l'extension, aux matrices singulières $M \in \mathcal{M}_A$, de la reformulation du critère de I optimalité généralisée.

Théorèmes de caractérisation

Matrices d'information singulières

Formulation du problème d'optimalité

Plus généralement, supposons que nous souhaitions maximiser une fonction $\phi(M)$ définie par

$$\begin{aligned}\phi(M) &= \psi(A'M^{-}A), \text{ si } M \in \mathcal{M}_A\,; \\ &= -\infty, \text{ sinon,}\end{aligned}$$

où ψ est une fonction définie et prenant des valeurs finies sur $\mathcal{S}_s^{++}(\mathbb{R})$, concave sur $\mathcal{M}_A$, différentiable en tout M inversible et non nécessairement différentiable en M singulière.

Ces conditions sont vérifiées par $\psi = -\log\det$ et $\psi = -\mathrm{tr}$.

Théorèmes de caractérisation

Matrices d'information singulières

Formulation du problème d'optimalité

Si M_* est une matrice inversible appartenant à $\mathcal{M}_A$, alors le théorème 2 permet d'affirmer que M_* est la matrice d'information d'une mesure ϕ-optimale si et seulement si :

$$F_\phi(M_*, xx') \leqslant 0, \text{ pour tout } x \in \mathcal{X}.$$

Théorèmes de caractérisation

Matrices d'information singulières

Vers une condition suffisante

Au contraire, supposons que M_* est une matrice de rang $r < k$ et notons-là M_r pour souligner cette propriété.

Nous ne pouvons plus utiliser le théorème 2 et devons retourner vers le théorème 1 dont les hypothèses sont beaucoup plus fastidieuses à remplir.

Théorèmes de caractérisation

Matrices d'information singulières

Vers une condition suffisante

Comme ψ est définie sur $\mathcal{S}_s^{++}$, ϕ est défini sur la classe $\mathcal{L}_A$ des matrices $\mathcal{S}_k^+$ Y ainsi que sur $\mathcal{M}_A$.

Pour une matrice M_r de rang $r < k$, il existe des matrices H de taille $k \times (n-k)$ et de rang $k-r$, telles que $M_r + HH'$ est inversible ; n'importe quelle H dont les colonnes et celles de M_r forment une base de R^k convient.

Soit $\mathcal{H}(M_r)$ la classe formée de ces matrices H.

Théorèmes de caractérisation

Matrices d'information singulières

Théorème : Condition suffisante

Soit ϕ, M_r et $\mathcal{H}(M_r)$ définis comme précédemment. Une condition suffisante pour que ϕ soit maximal sur $\mathcal{M}$ en M_r est qu'il existe $H \in \mathcal{H}(M_r)$ tel que

$$F_\phi(M_r + HH', xx') \leqslant 0$$

pour tout $x \in \boldsymbol{\mathcal{X}}$.

Théorèmes de caractérisation

Matrices d'information singulières

Démonstration

Puisque $M_r + HH'$ est inversible, ϕ est différentiable en ce point. La condition énoncée implique, comme pour la preuve du théorème 2, que $\phi(M_r + HH') \geqslant \phi(M)$ pour tout $M \in \mathcal{M}$.

Puis $(M_r + HH')^{-1}$ est un g-inverse de M_r, par conséquent, comme la valeur de $A'M_r^- A$ ne dépend pas du g-inverse choisi et que l'inverse est un g-inverse :

$$\begin{aligned}\phi(M_r) &= \psi(A'M_r^- A) = \psi(A'(M_r + HH')^{-1}A) \\ &= \phi(M_r + HH').\end{aligned}$$

En conclusion,

$$\phi(M_r) \geqslant \phi(M), \text{ pour tout } M \in \mathcal{M}.$$

Théorèmes de caractérisation

Matrices d'information singulières

Remarque

Il est beaucoup plus complexe de déterminer si cette condition est également nécessaire dans le cas général.

Une preuve de ce point a été apportée pour les deux critères

- $\psi = -\log\det$;
- $\psi = -\mathrm{tr}$.