Procédures adaptatives d'estimation et de tests : sélection de modèles et tests multiples

Béatrice LAURENT

INSA de Toulouse
Institut de Mathématiques de Toulouse

GdR Mascot-Num, Avignon, 17 Mars 2010

Plan de l'exposé

1. Introduction

Plan de l'exposé

1. Introduction
2. Estimation adaptative par sélection de modèles via un critère pénalisé
   - Heuristique de Mallows
   - Sélection de modèles en régression gaussienne (L. Birgé et P. Massart)

Plan de l'exposé

1. Introduction
2. Estimation adaptative par sélection de modèles via un critère pénalisé
    - Heuristique de Mallows
    - Sélection de modèles en régression gaussienne (L. Birgé et P. Massart)
3. Tests adaptatifs via des procédures de tests multiples
    - Tests d'hypothèses linéaires en régression gaussienne (Y. Baraud, S. Huet, B. L.)

Plan de l'exposé

1. Introduction
2. Estimation adaptative par sélection de modèles via un critère pénalisé
    - Heuristique de Mallows
    - Sélection de modèles en régression gaussienne (L. Birgé et P. Massart)
3. Tests adaptatifs via des procédures de tests multiples
    - Tests d'hypothèses linéaires en régression gaussienne (Y. Baraud, S. Huet, B. L.)
4. Applications en génomique
    - Applications à l'inférence de réseaux de gènes (N. Verzelen et F. Villers)

INTRODUCTION

- On considère le modèle

$$Y = \mu + \sigma\epsilon$$

Y, μ et ϵ sont des vecteurs de $\mathbb{R}^n$,
$\epsilon_1, \ldots, \epsilon_n$ sont i.i.d. de loi $\mathcal{N}(0, 1)$.

INTRODUCTION

- On considère le modèle

$$Y = \mu + \sigma\epsilon$$

Y, μ et ϵ sont des vecteurs de $\mathbb{R}^n$,
$\epsilon_1, \ldots, \epsilon_n$ sont i.i.d. de loi $\mathcal{N}(0, 1)$.

- A partir de l'estimation de Y, on souhaite
  - Estimer μ
  - Réaliser des tests d'hypothèses sur μ.

INTRODUCTION

- On considère le modèle

$$Y = \mu + \sigma\epsilon$$

Y, μ et ϵ sont des vecteurs de $\mathbb{R}^n$,
$\epsilon_1, \ldots, \epsilon_n$ sont i.i.d. de loi $\mathcal{N}(0,1)$.

- A partir de l'estimation de Y, on souhaite
  - Estimer μ
  - Réaliser des tests d'hypothèses sur μ.
- μ est de même dimension que le vecteur des observations Y, il s'agit d'un problème **non paramétrique.**

Exemples d'applications :

- **Estimation d'une fonction de régression**

$$Y_i = f(x_i) + \sigma\epsilon_i, \quad 1 \leq i \leq n, \quad 0 \leq x_1 < \ldots < x_n \leq 1.$$

Dans le cas où f constante par morceaux, l'estimation de μ permet d'estimer les points de ruptures de f.

Exemples d'applications :

- **Estimation d'une fonction de régression**

$$Y_i = f(x_i) + \sigma\epsilon_i, \quad 1 \leq i \leq n, \quad 0 \leq x_1 < \ldots < x_n \leq 1.$$

Dans le cas où f constante par morceaux, l'estimation de μ permet d'estimer les points de ruptures de f.

- **Sélection de variables**

$$Y_i = \theta_1 X_i^{(1)} + \ldots + \theta_p X_i^{(p)} + \sigma\epsilon_i, \quad 1 \leq i \leq n,$$

où $X^{(1)}, \ldots, X^{(p)}$ sont p vecteurs de $\mathbb{R}^n$.

Estimation de μ :

- Pour un estimateur $\hat{\mu}$ de μ, on définit le risque quadratique

$$R(\hat{\mu}, \mu) = E(\|\hat{\mu} - \mu\|_n^2)$$

où $\|u\|_n^2 = \sum_{i=1}^n u_i^2 / n$.

Estimation de μ :

- Pour un estimateur $\hat{\mu}$ de μ, on définit le risque quadratique

$$R(\hat{\mu}, \mu) = E(\|\hat{\mu} - \mu\|_n^2)$$

où $\|u\|_n^2 = \sum_{i=1}^n u_i^2 / n$.

- L'objectif de la sélection de modèle est de sélectionner parmi une collection d'estimateurs $(\hat{\mu}_m, m \in \mathcal{M})$ de μ un estimateur $\hat{\mu}_{\hat{m}}$ qui minimise le risque quadratique.

Estimation de μ :

- Pour un estimateur $\hat{\mu}$ de μ, on définit le risque quadratique

$$R(\hat{\mu}, \mu) = E(\|\hat{\mu} - \mu\|_n^2)$$

où $\|u\|_n^2 = \sum_{i=1}^n u_i^2/n$.

- L'objectif de la sélection de modèle est de sélectionner parmi une collection d'estimateurs $(\hat{\mu}_m, m \in \mathcal{M})$ de μ un estimateur $\hat{\mu}_{\hat{m}}$ qui minimise le risque quadratique.
- Inégalité "oracle" :

$$\forall \mu \in \mathbb{R}^n,\ R(\hat{\mu}_{\hat{m}}, \mu) \leq \quad \inf_{m \in \mathcal{M}} R(\hat{\mu}_m, \mu)$$

Estimation de μ :

- Pour un estimateur $\hat{\mu}$ de μ, on définit le risque quadratique

$$R(\hat{\mu}, \mu) = E(\|\hat{\mu} - \mu\|_n^2)$$

où $\|u\|_n^2 = \sum_{i=1}^n u_i^2 / n$.

- L'objectif de la sélection de modèle est de sélectionner parmi une collection d'estimateurs $(\hat{\mu}_m, m \in \mathcal{M})$ de μ un estimateur $\hat{\mu}_{\hat{m}}$ qui minimise le risque quadratique.
- Inégalité "oracle" :

$$\forall \mu \in \mathbb{R}^n,\ R(\hat{\mu}_{\hat{m}}, \mu) \leq C \inf_{m \in \mathcal{M}} R(\hat{\mu}_m, \mu) + \gamma_n$$

Tests d'hypothèses linéaires :

- On souhaite tester l'hypothèse

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \notin V$$

où V est un sous-espace vectoriel de $\mathbb{R}^n$.

Tests d'hypothèses linéaires :

- On souhaite tester l'hypothèse

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \notin V$$

où V est un sous-espace vectoriel de $\mathbb{R}^n$.

- Pour tester

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \in W$$

où W est un s.e.v. de $\mathbb{R}^n$ de petite dimension devant n tel que $V \subset W$, on réalise un test de Fisher.

Tests multiples :

- Dans un contexte non paramétrique, on se donne une collection $(W_m, m \in \mathcal{M})$ de sous-espaces vectoriels de $\mathbb{R}^n$, contenant V. Pour tout $m \in \mathcal{M}$, on considère le test de Fisher d'hypothèses :

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \in W_m.$$

Tests multiples :

- Dans un contexte non paramétrique, on se donne une collection $(W_m, m \in \mathcal{M})$ de sous-espaces vectoriels de $\mathbb{R}^n$, contenant V. Pour tout $m \in \mathcal{M}$, on considère le test de Fisher d'hypothèses :

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \in W_m.$$

- A partir de cette collection de tests, nous allons définir une procédure de tests multiples.

Tests multiples :

- Dans un contexte non paramétrique, on se donne une collection $(W_m, m \in \mathcal{M})$ de sous-espaces vectoriels de $\mathbb{R}^n$, contenant V. Pour tout $m \in \mathcal{M}$, on considère le test de Fisher d'hypothèses :

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \in W_m.$$

- A partir de cette collection de tests, nous allons définir une procédure de tests multiples.
- L'objectif sera de prouver une inégalité de type "oracle" pour les tests : la puissance du test multiple en tout vecteur $\mu \notin V$ est "comparable" à la meilleure puissance des tests de la collection en μ.

Application en génomique :

- On considère des réseaux d'interactions entre gènes.
- N. Verzelen et F. Villers ont proposé des tests d'adéquation qui s'appuient sur les procédures de tests multiples.

ESTIMATION ADAPTATIVE

$$Y = \mu + \sigma\epsilon$$

Heuristique de Mallows

- Soit V un s.e.v. de $\mathbb{R}^n$ de dimension D, et $\hat{\mu} = \Pi_V(Y)$.

$$\begin{aligned}
R(\Pi_V(Y), \mu) &= E(\|\Pi_V(Y) - \mu\|_n^2) \\
&= \|\Pi_V(\mu) - \mu\|_n^2 + E(\|\Pi_V(\sigma\epsilon)\|_n^2) \\
&= \|\Pi_V(\mu) - \mu\|_n^2 + \sigma^2 \frac{D}{n}.
\end{aligned}$$

ESTIMATION ADAPTATIVE

$$Y = \mu + \sigma\epsilon$$

Heuristique de Mallows

- Soit V un s.e.v. de $\mathbb{R}^n$ de dimension D, et $\hat{\mu} = \Pi_V(Y)$.

$$\begin{aligned} R(\Pi_V(Y), \mu) &= E(\|\Pi_V(Y) - \mu\|_n^2) \\ &= \|\Pi_V(\mu) - \mu\|_n^2 + E(\|\Pi_V(\sigma\epsilon)\|_n^2) \\ &= \|\Pi_V(\mu) - \mu\|_n^2 + \sigma^2 \frac{D}{n}. \end{aligned}$$

- L' heuristique de Mallows consiste à
  - Estimer sans biais le risque $R(\Pi_V(Y), \mu)$
  - Sélectionner parmi une collection d'estimateurs un estimateur qui minimise le risque estimé.

Estimation sans biais du risque

$$R(\Pi_V(Y), \mu) = \|\mu\|_n^2 - \|\Pi_V(\mu)\|_n^2 + \sigma^2 \frac{D}{n}$$

$$R(\Pi_V(Y), \mu) - \|\mu\|_n^2 = -\|\Pi_V(\mu)\|_n^2 + \sigma^2 \frac{D}{n}$$

$$E(\|\Pi_V(Y)\|_n^2) = \|\Pi_V(\mu)\|_n^2 + \sigma^2 \frac{D}{n}$$

$$-\|\Pi_V(Y)\|_n^2 + 2\sigma^2 \frac{D}{n}$$

est un estimateur sans biais de

$$R(\Pi_V(Y), \mu) - \|\mu\|_n^2.$$

Critère Cp de Mallows (1973)

- On se donne une collection de s.e.v. de $\mathbb{R}^n$: $(S_m, m \in \mathcal{M})$, avec $\dim(S_m) = D_m$.
- On considère

$$\hat{m} = \mathrm{Argmin}_{m \in \mathcal{M}}(-\|\Pi_{S_m}(Y)\|_n^2 + 2\sigma^2 \frac{D_m}{n})$$

- On estime μ par $\Pi_{S_{\hat{m}}}(Y)$.

Critère Cp de Mallows (1973)

- On se donne une collection de s.e.v. de $\mathbb{R}^n$: $(S_m, m \in \mathcal{M})$, avec $\dim(S_m) = D_m$.
- On considère

$$\hat{m} = \mathrm{Argmin}_{m \in \mathcal{M}}\left(-\|\Pi_{S_m}(Y)\|_n^2 + 2\sigma^2 \frac{D_m}{n}\right)$$

- On estime μ par $\Pi_{S_{\hat{m}}}(Y)$.

- Quel choix pour la collection de "modèles" $(S_m, m \in \mathcal{M})$?
- Que peut-on dire du risque de cet estimateur ? Est-il possible de prouver une inégalité oracle ?

Choix de la collection de modèles

$$Y_i = f(x_i) + \sigma\epsilon_i, \quad 1 \leq i \leq n, \quad 0 \leq x_1 < \ldots < x_n \leq 1.$$

Choix de la collection de modèles

$$Y_i = f(x_i) + \sigma\epsilon_i, \quad 1 \leq i \leq n, \quad 0 \leq x_1 < \ldots < x_n \leq 1.$$

- On note $(\phi_j)_{j\geq 1}$ la base Fourier de $\mathbb{L}^2([0,1])$.
  Pour tout $m = 1, \ldots, n$, on considère l'espace S_m engendré par les m vecteurs de $\mathbb{R}^n$ $(v_1, \ldots, v_m)$ où

$$v_j = (\phi_j(x_1), \ldots, \phi_j(x_n)), \quad 1 \leq j \leq m.$$

Choix de la collection de modèles

$$Y_i = f(x_i) + \sigma\epsilon_i, \quad 1 \leq i \leq n, \quad 0 \leq x_1 < \ldots < x_n \leq 1.$$

- On note $(\phi_j)_{j\geq 1}$ la base Fourier de $\mathbb{L}^2([0,1])$.
  Pour tout $m = 1, \ldots, n$, on considère l'espace S_m engendré par les m vecteurs de $\mathbb{R}^n$ $(v_1, \ldots, v_m)$ où

$$v_j = (\phi_j(x_1), \ldots, \phi_j(x_n)), \quad 1 \leq j \leq m.$$

- Fonctions constantes par morceaux sur une partition régulière de $[0,1]$ en m intervalles pour $m = 1, \ldots, n$.

Choix de la collection de modèles

$$Y_i = f(x_i) + \sigma\epsilon_i, \quad 1 \leq i \leq n, \quad 0 \leq x_1 < \ldots < x_n \leq 1.$$

- On note $(\phi_j)_{j\geq 1}$ la base Fourier de $\mathbb{L}^2([0,1])$.
  Pour tout $m = 1, \ldots, n$, on considère l'espace S_m engendré par les m vecteurs de $\mathbb{R}^n$ $(v_1, \ldots, v_m)$ où

$$v_j = (\phi_j(x_1), \ldots, \phi_j(x_n)), \quad 1 \leq j \leq m.$$

- Fonctions constantes par morceaux sur une partition régulière de $[0,1]$ en m intervalles pour $m = 1, \ldots, n$.
- Fonctions constantes par morceaux sur une partition irrégulière de $[0,1]$ en m intervalles pour $m = 1, \ldots, n$, obtenue à partir d'une grille fine de pas $1/n$.

Choix de la collection de modèles

$$Y_i = \theta_1 X_i^{(1)} + \ldots + \theta_p X_i^{(p)} + \sigma\epsilon_i, \quad 1 \leq i \leq n,$$

Choix de la collection de modèles

$$Y_i = \theta_1 X_i^{(1)} + \ldots + \theta_p X_i^{(p)} + \sigma\epsilon_i, \quad 1 \leq i \leq n,$$

- **Sélection de variables ordonnée**
  Pour tout $m = 1, \ldots, p$, on considère l'espace S_m engendré par les vecteurs $(X^{(1)}, X^{(2)}, \ldots, X^{(m)})$.

Choix de la collection de modèles

$$Y_i = \theta_1 X_i^{(1)} + \ldots + \theta_p X_i^{(p)} + \sigma\epsilon_i, \quad 1 \leq i \leq n,$$

- **Sélection de variables ordonnée**
  Pour tout $m = 1, \ldots, p$, on considère l'espace S_m engendré par les vecteurs $(X^{(1)}, X^{(2)}, \ldots, X^{(m)})$.
- **Sélection de variables non ordonnée**
  Pour tout $m \subset \{1, 2, \ldots, p\}$, on considère l'espace S_m engendré par les vecteurs $(X^{(j)}, j \in m)$.

Sélection de modèles en régression gaussienne

L. Birgé et P. Massart *Minimal penalties for Gaussian model selection. Probab. Theory Related Fields (2007).*

Théorème

$$Y = \mu + \sigma\epsilon$$

On se donne une collection $(S_m, m \in \mathcal{M})$ *de s.e.v. de* $\mathbb{R}^n$ *tels que* $\dim(S_m) = D_m$.

$$\hat{m} = \ Argmin\ _{m \in \mathcal{M}}(-\|\Pi_{S_m}(Y)\|_n^2 + pen(m)).$$

Sélection de modèles en régression gaussienne

L. Birgé et P. Massart *Minimal penalties for Gaussian model selection. Probab. Theory Related Fields (2007).*

Théorème

$$Y = \mu + \sigma\epsilon$$

On se donne une collection $(S_m, m \in \mathcal{M})$ *de s.e.v. de* $\mathbb{R}^n$ *tels que* $dim(S_m) = D_m$.

$$\hat{m} = \ Argmin\ _{m\in\mathcal{M}}(-\|\Pi_{S_m}(Y)\|_n^2 + pen(m)).$$

Soit $(L_m, m \in \mathcal{M})$ *une collection de nombres* ≥ 0 *tels que*

$$\Sigma = \sum_{m\in\mathcal{M}} \exp(-L_m D_m) < +\infty.$$

$$pen(m) \geq \frac{D_m}{n}\sigma^2(\kappa + 2(2-\theta)\sqrt{L_m} + \frac{2}{\theta}L_m),\ \ avec\ \theta \in]0,1[, \kappa > 2-\theta,$$

Sélection de modèles en régression gaussienne

L. Birgé et P. Massart *Minimal penalties for Gaussian model selection. Probab. Theory Related Fields (2007).*

Théorème

$$Y = \mu + \sigma\epsilon$$

On se donne une collection $(S_m, m \in \mathcal{M})$ *de s.e.v. de* $\mathbb{R}^n$ *tels que* $\dim(S_m) = D_m$.

$$\hat{m} = \ Argmin \ _{m\in\mathcal{M}}(-\|\Pi_{S_m}(Y)\|_n^2 + pen(m)).$$

Soit $(L_m, m \in \mathcal{M})$ *une collection de nombres* ≥ 0 *tels que*

$$\Sigma = \sum_{m\in\mathcal{M}} \exp(-L_m D_m) < +\infty.$$

$$pen(m) \geq \frac{D_m}{n}\sigma^2(\kappa + 2(2-\theta)\sqrt{L_m} + \frac{2}{\theta}L_m), \ avec \ \theta \in]0,1[, \kappa > 2-\theta,$$

$$E(\|\Pi_{S_{\hat{m}}}(Y) - \mu\|_n^2) \leq \frac{1}{1-\theta} \inf_{m\in\mathcal{M}} \left(\|\Pi_{S_m}(\mu) - \mu\|_n^2 + pen(m) - \sigma^2 \frac{D_m}{n} \right) + C_2(\theta)\frac{\Sigma}{n}.$$

Applications

On se donne une collection $(S_m, m \in \mathcal{M})$ de s.e.v. de $\mathbb{R}^n$.
Pour tout $D \geq 1$, on note $N(D)$ le nombre d'élements de la collection de dimension D.

Applications

On se donne une collection $(S_m, m \in \mathcal{M})$ de s.e.v. de $\mathbb{R}^n$.
Pour tout $D \geq 1$, on note $N(D)$ le nombre d'élements de la collection de dimension D.

- Si $N(D) \leq 1$ (partition régulière, sélection de variable ordonnée), on peut prendre $L_m = \delta > 0$

$$\Sigma = \sum_{m \in \mathcal{M}} \exp(-L_m D_m) \leq \sum_{D \geq 1} \exp(-\delta D) < +\infty.$$

Applications

On se donne une collection $(S_m, m \in \mathcal{M})$ de s.e.v. de $\mathbb{R}^n$. Pour tout $D \geq 1$, on note $N(D)$ le nombre d'élements de la collection de dimension D.

- Si $N(D) \leq 1$ (partition régulière, sélection de variable ordonnée), on peut prendre $L_m = \delta > 0$

$$\Sigma = \sum_{m \in \mathcal{M}} \exp(-L_m D_m) \leq \sum_{D \geq 1} \exp(-\delta D) < +\infty.$$

- Le théorème s'applique pour

$$\mathrm{pen}(m) \geq \frac{D_m}{n}\sigma^2(\kappa + 2(2-\theta)\sqrt{\delta} + \frac{2}{\theta}\delta), \text{ avec } \theta \in]0,1[, \kappa > 2-\theta.$$

- L'heuristique de Mallows est justifiée dans ce cas.

Applications

- $N(D) = C_n^D$ (partition irrégulière, sélection de variable non ordonnée avec $p = n$).
- En posant $L_m = L(D_m)$,

$$\sum_{m \in \mathcal{M}} \exp(-L_m D_m) \leq \sum_{D=1}^{n} C_n^D \exp(-DL(D)) \leq \sum_{D=1}^{n} \exp(D + D\ln(\frac{n}{D}) - DL(D))$$

Applications

- $N(D) = C_n^D$ (partition irrégulière, sélection de variable non ordonnée avec $p = n$).
- En posant $L_m = L(D_m)$,

$$\sum_{m \in \mathcal{M}} \exp(-L_m D_m) \leq \sum_{D=1}^{n} C_n^D \exp(-DL(D)) \leq \sum_{D=1}^{n} \exp(D + D\ln(\frac{n}{D}) - DL(D))$$

- On peut prendre $L(D) = L + \ln(\frac{n}{D}), L > 1$

$$\mathrm{pen}(m) \geq \frac{D_m}{n}\sigma^2 \left(\kappa + C(\theta)\ln\left(\frac{n}{D_m}\right)\right).$$

Applications

- $N(D) = C_n^D$ (partition irrégulière, sélection de variable non ordonnée avec $p = n$).
- En posant $L_m = L(D_m)$,

$$\sum_{m \in \mathcal{M}} \exp(-L_m D_m) \leq \sum_{D=1}^{n} C_n^D \exp(-DL(D)) \leq \sum_{D=1}^{n} \exp(D + D\ln(\frac{n}{D}) - DL(D))$$

- On peut prendre $L(D) = L + \ln(\frac{n}{D}), L > 1$

$$\mathrm{pen}(m) \geq \frac{D_m}{n}\sigma^2 \left(\kappa + C(\theta)\ln\left(\frac{n}{D_m}\right)\right).$$

- Le théorème ne permet pas de justifier l'heuristique de Mallows dans ce cas.

Applications

- $N(D) = C_n^D$ (partition irrégulière, sélection de variable non ordonnée avec $p = n$).
- En posant $L_m = L(D_m)$,

$$\sum_{m \in \mathcal{M}} \exp(-L_m D_m) \leq \sum_{D=1}^{n} C_n^D \exp(-DL(D)) \leq \sum_{D=1}^{n} \exp(D + D\ln(\frac{n}{D}) - DL(D))$$

- On peut prendre $L(D) = L + \ln(\frac{n}{D}), L > 1$

$$\mathrm{pen}(m) \geq \frac{D_m}{n}\sigma^2 \left(\kappa + C(\theta)\ln\left(\frac{n}{D_m}\right)\right).$$

- Le théorème ne permet pas de justifier l'heuristique de Mallows dans ce cas.
- L. Birgé et P. Massart montrent que, pour des collection de modèles complexes, le C_p de Mallows ne fonctionne pas.

Références Sélection de modèles

- Y. Baraud, C. Giraud, S. Huet (2009) *Gaussian model selection with an unknown variance*, Ann. Statist.
- L. Birgé, P. Massart (2007) *Minimal penalties for Gaussian model selection*, Probab. Theory Related Fields.
- X. Gendre (2008) *Simultaneous estimation of the mean and the variance in heteroscedastic Gaussian regression*, EJS.
- B. L., C. Ludena, C. Prieur (2008) *Adaptive estimation of a linear functionals by model selection*, EJS.
- P. Massart (2007) *Concentration inequalities and model selection*. Cours de l'école d'été de St Flour 2003, Springer.

- S. Arlot, P. Massart (2009). *Data-driven calibration of penalties for least squares regression*, Journal of Machine Learning Research.
- M. Lerasle (2009) *Optimal model selection in density estimation*, Preprint. Hal.

TESTS ADAPTATIFS

Tests d'hypothèses linéaires

Y. Baraud , S. Huet, B.L. (2003), Ann. Statist.

$$Y = \mu + \sigma\epsilon$$

avec $\epsilon_1, \ldots, \epsilon_n$ i.i.d. de la $\mathcal{N}(0, 1)$.

- On souhaite tester l'hypothèse $\mathrm{H}_0 : \mu \in V$ contre $\mathrm{H}_1 : \mu \notin V$ où V est un sous-espace vectoriel de $\mathbb{R}^n$.

TESTS ADAPTATIFS

Tests d'hypothèses linéaires

Y. Baraud , S. Huet, B.L. (2003), Ann. Statist.

$$Y = \mu + \sigma\epsilon$$

avec $\epsilon_1, \ldots, \epsilon_n$ i.i.d. de la $\mathcal{N}(0,1)$.

- On souhaite tester l'hypothèse $\mathrm{H}_0 : \mu \in V$ contre $\mathrm{H}_1 : \mu \notin V$ où V est un sous-espace vectoriel de $\mathbb{R}^n$.
- On se donne un niveau $\alpha \in]0,1[$ et une fonction de test $\Phi_\alpha(Y) \in \{0,1\}$.
- On accepte H_0 si $\Phi_\alpha(Y) = 0$, on rejette H_0 si $\Phi_\alpha(Y) = 1$.

TESTS ADAPTATIFS

Tests d'hypothèses linéaires

Y. Baraud , S. Huet, B.L. (2003), Ann. Statist.

$$Y = \mu + \sigma\epsilon$$

avec $\epsilon_1, \ldots, \epsilon_n$ i.i.d. de la $\mathcal{N}(0,1)$.

- On souhaite tester l'hypothèse $\mathrm{H}_0 : \mu \in V$ contre $\mathrm{H}_1 : \mu \notin V$ où V est un sous-espace vectoriel de $\mathbb{R}^n$.
- On se donne un niveau $\alpha \in]0,1[$ et une fonction de test $\Phi_\alpha(Y) \in \{0,1\}$.
- On accepte H_0 si $\Phi_\alpha(Y) = 0$, on rejette H_0 si $\Phi_\alpha(Y) = 1$.
- **Test de niveau α :**

$$\mathbb{P}_{\mathrm{H}_0}(\Phi_\alpha(Y) = 1) \leq \alpha.$$

- **Puissance du test :**

$$\mu \mapsto \mathbb{P}_\mu(\Phi_\alpha(Y) = 1).$$

Test de Fisher

Pour tester

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \in W$$

où W est un s.e.v. strict de $\mathbb{R}^n$ tel que $V \subset W$, on réalise un test de Fisher. $dim(V) = q$, $dim(W) = p$.

$$F(Y) = \frac{\|\Pi_W(Y) - \Pi_V(Y)\|^2/(p-q)}{\|Y - \Pi_W(Y)\|^2/(n-p)}$$

On rejette H_0 si

$$F(Y) > \bar{F}^{-1}_{p-q,n-p}(\alpha)$$

où $\bar{F}^{-1}_{p-q,n-p}(\alpha)$ désigne le $1-\alpha$ quantile de la loi de Fisher de paramètres $(p-q, n-p)$.

$$\Phi_\alpha(Y) = \mathbb{1}_{F(Y) > \bar{F}^{-1}_{p-q,n-p}(\alpha)}.$$

Puissance du test

- A quelle distance de l'hypothèse H_0, peut-on garantir que le test est puissant ?

Puissance du test

- A quelle distance de l'hypothèse H_0, peut-on garantir que le test est puissant ?
- Soit $\beta \in]0,1[$, on peut montrer que $\mathbb{P}_\mu(\Phi_\alpha(Y)=1) \geq 1-\beta$ si

$$d_n^2(\mu, V) = \|\mu - \Pi_V(\mu)\|_n^2 \geq C_1 \|\mu - \Pi_W(\mu)\|_n^2 + C_2 \frac{\sigma^2}{n} \sqrt{(p-q)\ln\left(\frac{2}{\alpha\beta}\right)}.$$

Puissance du test

- A quelle distance de l'hypothèse H_0, peut-on garantir que le test est puissant ?
- Soit $\beta \in]0,1[$, on peut montrer que $\mathbb{P}_\mu(\Phi_\alpha(Y)=1) \geq 1-\beta$ si

$$d_n^2(\mu, V) = \|\mu - \Pi_V(\mu)\|_n^2 \geq C_1 \|\mu - \Pi_W(\mu)\|_n^2 + C_2 \frac{\sigma^2}{n} \sqrt{(p-q)\ln\left(\frac{2}{\alpha\beta}\right)}.$$

- En pratique, comment choisir W ?
- Si on choisit W, le test sera peu puissant si $d_n(\mu, W)$ est grande.

Procédure de tests multiples

On se donne une collection $(S_m, m \in \mathcal{M})$ de s.e.v. de $V^\perp$, tels que $\dim(S_m) = D_m$.

- On pose pour tout $m \in \mathcal{M}$, $W_m = V + S_m$,
  on note $N_m = n - dim(W_m)$.

Procédure de tests multiples

On se donne une collection $(S_m, m \in \mathcal{M})$ de s.e.v. de $V^{\perp}$, tels que $\dim(S_m) = D_m$.

- On pose pour tout $m \in \mathcal{M}$, $W_m = V + S_m$,
  on note $N_m = n - dim(W_m)$.
- On considère la statistique de test de Fisher pour tester

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \in W_m$$

$$F_m(Y) = \frac{\|\Pi_{S_m}(Y)\|^2 / D_m}{\|Y - \Pi_{W_m}(Y)\|^2 / N_m}.$$

Procédure de tests multiples

On se donne une collection $(S_m, m \in \mathcal{M})$ de s.e.v. de $V^\perp$, tels que $\dim(S_m) = D_m$.

- On pose pour tout $m \in \mathcal{M}$, $W_m = V + S_m$, on note $N_m = n - dim(W_m)$.
- On considère la statistique de test de Fisher pour tester

$$\mathrm{H}_0 : \mu \in V \text{ contre } \mathrm{H}_1 : \mu \in W_m$$

$$F_m(Y) = \frac{\|\Pi_{S_m}(Y)\|^2 / D_m}{\|Y - \Pi_{W_m}(Y)\|^2 / N_m}.$$

- On rejette H_0 si $\exists\, m \in \mathcal{M}, F_m(Y) > \bar{F}^{-1}_{D_m,N_m}(\alpha_n)$ où α_n vérifie :

$$\mathbb{P}\left(\sup_{m\in\mathcal{M}} \left\{F_m(\epsilon) - \bar{F}^{-1}_{D_m,N_m}(\alpha_n)\right\} > 0\right) = \alpha.$$

- $\alpha/|\mathcal{M}| \leq \alpha_n \leq \alpha$.

Puissance de la procédure de tests multiples

Théorème

$$Y = \mu + \sigma \epsilon$$

$$\mathrm{H}_0 : \mu \in V \ \textit{contre}\ \mathrm{H}_1 : \mu \notin V$$

$$\begin{aligned} \Phi_\alpha(Y) &= 1 \textit{ si } \exists m \in \mathcal{M}, F_m > \bar{F}^{-1}_{D_m,N_m}(\alpha_n) \\ &= 0 \textit{ sinon} \end{aligned}$$

Soit $\beta \in]0,1[$, $\mathbb{P}_\mu(\Phi_\alpha(Y) = 1) \geq 1 - \beta$ *si*

$$d_n^2(\mu, V) \geq \inf_{m \in \mathcal{M}} \left\{ C_1 \|\mu - \Pi_{W_m}(\mu)\|_n^2 + C_2 \frac{\sigma^2}{n} \sqrt{D_m \ln(\frac{2}{\alpha_n \beta})}. \right\}$$

Vitesse de détection de signal

Corollaire

$$Y_i = f(x_i) + \sigma\epsilon_i, \ldots, 1 \leq i \leq n.$$

$$\mu = E(Y) = (f(x_1), \ldots, f(x_n))'.$$

$$\mathrm{H}_0 : \mu = 0 \textit{ contre } \mathrm{H}_1 : \mu \neq 0$$

$$V = \{0\} \quad , W_m = S_m.$$

Vitesse de détection de signal

Corollaire

$$Y_i = f(x_i) + \sigma\epsilon_i, \dots, 1 \le i \le n.$$

$$\mu = E(Y) = (f(x_1), \dots, f(x_n))'.$$

$$\mathrm{H}_0 : \mu = 0 \textit{ contre } \mathrm{H}_1 : \mu \neq 0$$

$$V = \{0\} \quad , W_m = S_m.$$

Pour tout $m \in \mathcal{M} = \{2^j,\ j \geq 0\} \cap \{1, ..., [n/2]\}$

$$S_m = Vect\left\{ \left(\mathbb{1}_{]\frac{j-1}{m}, \frac{j}{m}]}(x_1), ..., \ \mathbb{1}_{]\frac{j-1}{m}, \frac{j}{m}]}(x_n) \right)', \ \ j = 1, ..., m \right\}.$$

Vitesse de détection de signal

Corollaire

$$Y_i = f(x_i) + \sigma\epsilon_i, \dots, 1 \leq i \leq n.$$

$$\mu = E(Y) = (f(x_1), \dots, f(x_n))'.$$

$$\mathrm{H}_0 : \mu = 0 \textit{ contre } \mathrm{H}_1 : \mu \neq 0$$

$$V = \{0\} \quad , W_m = S_m.$$

Pour tout $m \in \mathcal{M} = \{2^j,\ j \geq 0\} \cap \{1, ..., [n/2]\}$

$$S_m = Vect\left\{\left(\mathbb{1}_{]\frac{j-1}{m}, \frac{j}{m}]}(x_1), ..., \ \mathbb{1}_{]\frac{j-1}{m}, \frac{j}{m}]}(x_n)\right)', \ \ j = 1, ..., m\right\}.$$

$$\forall R > 0, s \in]0,1], \quad \mathcal{H}_s(R) = \left\{f, \forall(x,y) \in [0,1]^2, |f(x) - f(y)| \leq R|x-y|^s\right\}.$$

Vitesse de détection de signal

Corollaire

$$Y_i = f(x_i) + \sigma\epsilon_i, \ldots, 1 \leq i \leq n.$$

$$\mu = E(Y) = (f(x_1), \ldots, f(x_n))'.$$

$$\mathrm{H}_0 : \mu = 0 \textit{ contre } \mathrm{H}_1 : \mu \neq 0$$

$$V = \{0\} \quad , W_m = S_m.$$

Pour tout $m \in \mathcal{M} = \{2^j,\ j \geq 0\} \cap \{1, ..., [n/2]\}$

$$S_m = Vect\left\{\left(\ \mathbb{1}_{]\frac{j-1}{m},\frac{j}{m}]}(x_1), ...,\ \mathbb{1}_{]\frac{j-1}{m},\frac{j}{m}]}(x_n)\right)',\ \ j = 1, ..., m\right\}.$$

$$\forall R > 0, s \in]0,1], \quad \mathcal{H}_s(R) = \left\{f, \forall (x,y) \in [0,1]^2, |f(x) - f(y)| \leq R|x-y|^s\right\}.$$

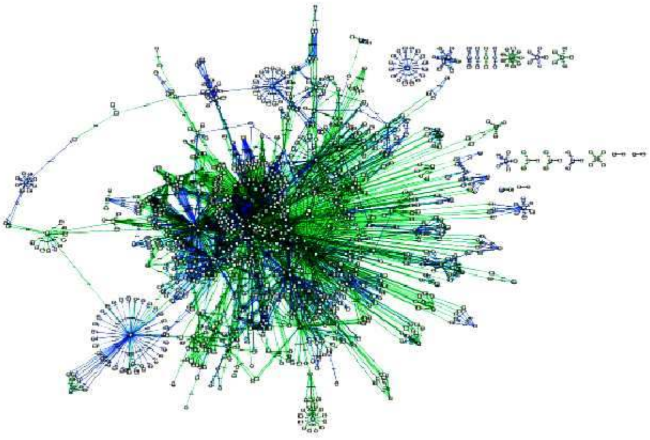
$$\rho_n^2 = R^{\frac{2}{1+4s}} \left(\frac{\sqrt{\ln\ln(n)}}{n}\sigma^2\right)^{\frac{4s}{1+4s}} + R^2 n^{-2s} + \frac{\ln\ln(n)}{n}\sigma^2$$

$\forall\ R > 0, s \in]0,1]$, *si* $f \in \mathcal{H}_s(R)$ *et* $\|\mu\|_n^2 \geq C(\alpha,\beta)\rho_n^2$, $\mathbb{P}_\mu(\Phi_\alpha(Y) = 1) \geq 1 - \beta.$

APPLICATIONS EN GENOMIQUE

- **Tests pour les modèles graphiques gaussiens**
  Réf. N. Verzelen et F. Villers
  - *Tests for Gaussian graphical models.* CSDA, 2009
  - *Goodness-of-fit Tests for high-dimensional Gaussian linear models.* à paraitre Ann. Statist.

Réseau de régulation génique d'E. Coli



Modèle graphique gaussien

On considère $X = (X_1, \ldots, X_p) \sim \mathcal{N}_p(0, \Sigma)$

Σ inversible
$\Gamma := \{1, \ldots, p\}$
$G = (\Gamma, E)$ graphe non orienté fini
$\mathrm{ne}_G(a)$: voisins de a dans G.

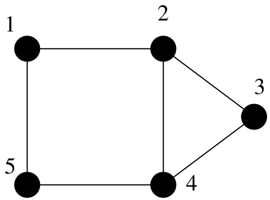
Modèle graphique gaussien

On considère $X = (X_1, \ldots, X_p) \sim \mathcal{N}_p(0, \Sigma)$

Σ inversible
$\Gamma := \{1, \ldots, p\}$
$G = (\Gamma, E)$ graphe non orienté fini
$\mathrm{ne}_G(a)$: voisins de a dans G.

X vérifie la propriété de Markov local en a par rapport à G si

$$(X_a \perp\!\!\!\perp X_{-\{a,\mathrm{ne}_G(a)\}})|X_{\mathrm{ne}_G(a)}$$



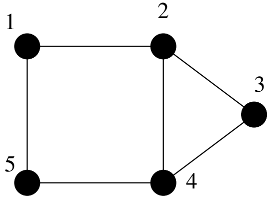
Modèle graphique gaussien

On considère $X = (X_1, \ldots, X_p) \sim \mathcal{N}_p(0, \Sigma)$

Σ inversible
$\Gamma := \{1, \ldots, p\}$
$G = (\Gamma, E)$ graphe non orienté fini
$\mathrm{ne}_G(a)$: voisins de a dans G.

X vérifie la propriété de Markov local en a par rapport à G si

$$(X_a \perp\!\!\!\perp X_{-\{a, \mathrm{ne}_G(a)\}}) | X_{\mathrm{ne}_G(a)}$$



X est un modèle graphique gaussien par rapport à G

$\Longleftrightarrow$

X vérifie la propriété de Markov local pour tout sommet $a \in \Gamma$.

Tests pour les modèles graphiques gaussiens

- n observations de $X = (X_1, \ldots, X_p)$, avec $X \sim \mathcal{N}_p(0, \Sigma)$.
  n Microarrays sur lesquels on mesure l'expression de p gènes ($n << p$)

Tests pour les modèles graphiques gaussiens

- n observations de $X = (X_1, \ldots, X_p)$, avec $X \sim \mathcal{N}_p(0, \Sigma)$.
  n Microarrays sur lesquels on mesure l'expression de p gènes ($n << p$)
- **Test de voisinage :** On se donne un noeud a.

$H_{0,a}$: X satisfait la propriété de Markov locale au noeud a relativement au graphe G.

$H_{0,a}$: $X_a \perp\!\!\!\perp X_{\Gamma \setminus \{a, \mathrm{ne}_G(a)\}} | X_{\mathrm{ne}_G(a)}$.

Tests pour les modèles graphiques gaussiens

- n observations de $X = (X_1, \ldots, X_p)$, avec $X \sim \mathcal{N}_p(0, \Sigma)$.
  n Microarrays sur lesquels on mesure l'expression de p gènes ($n << p$)
- **Test de voisinage :** On se donne un noeud a.

$H_{0,a}$: X satisfait la propriété de Markov locale au noeud a relativement au graphe G.

$H_{0,a}$: $X_a \perp\!\!\!\perp X_{\Gamma \setminus \{a, \mathrm{ne}_G(a)\}} | X_{\mathrm{ne}_G(a)}$.

- Propriété des vecteurs gaussiens :

$$X_a = \sum_{b \in \Gamma \setminus \{a\}} \theta_b^a X_b + \epsilon_a$$

  - $\sum_{b \in \Gamma \setminus \{a\}} \theta_b^a X_b = E(X_a / X_{-a})$
  - $\epsilon_a \perp\!\!\!\perp X_{-a}$ de loi $\mathcal{N}(0, \sigma_a^2)$.

Tests pour les modèles graphiques gaussiens

- n observations de $X = (X_1, \ldots, X_p)$, avec $X \sim \mathcal{N}_p(0, \Sigma)$.
  n Microarrays sur lesquels on mesure l'expression de p gènes ($n << p$)
- **Test de voisinage :** On se donne un noeud a.

$H_{0,a}$: X satisfait la propriété de Markov locale au noeud a relativement au graphe G.

$H_{0,a}$: $X_a \perp\!\!\!\perp X_{\Gamma \setminus \{a, \mathrm{ne}_G(a)\}} | X_{\mathrm{ne}_G(a)}$.

- Propriété des vecteurs gaussiens :

$$X_a = \sum_{b \in \Gamma \setminus \{a\}} \theta_b^a X_b + \epsilon_a$$

  - $\sum_{b \in \Gamma \setminus \{a\}} \theta_b^a X_b = E(X_a / X_{-a})$
  - $\epsilon_a \perp\!\!\!\perp X_{-a}$ de loi $\mathcal{N}(0, \sigma_a^2)$.

$$H_{0,a} : \theta_b^a = 0 \ \forall b \notin \mathrm{ne}_G(a).$$

Procédure de tests multiples

- On note $\mathbf{X}_j$ le vecteur de $\mathbb{R}^n$ correspondant aux n observations de X_j pour tout $j \in \Gamma$, $\mathbf{X}_{-a} = (\mathbf{X}_b, b \neq a)$

$$H_{0,a} : \theta_b^a = 0 \; \forall b \notin \mathrm{ne}_G(a).$$

Procédure de tests multiples

- On note $\mathbf{X}_j$ le vecteur de $\mathbb{R}^n$ correspondant aux n observations de X_j pour tout $j \in \Gamma$, $\mathbf{X}_{-a} = (\mathbf{X}_b, b \neq a)$
$$H_{0,a} : \theta_b^a = 0 \; \forall b \notin \mathrm{ne}_G(a).$$
- On se donne une collection $\mathcal{M}$ de sous-ensembles de $\Gamma \backslash \{\mathrm{ne}_G(a), a\}$ Pour tout $m \in \mathcal{M}$,
$$H_{1,a,m} : \theta_b^a = 0 \; \forall b \notin \mathrm{ne}_G(a) \cup m.$$

Procédure de tests multiples

- On note $\mathbf{X}_j$ le vecteur de $\mathbb{R}^n$ correspondant aux n observations de X_j pour tout $j \in \Gamma$, $\mathbf{X}_{-a} = (\mathbf{X}_b, b \neq a)$

$$H_{0,a} : \theta_b^a = 0 \ \forall b \notin \mathrm{ne}_G(a).$$

- On se donne une collection $\mathcal{M}$ de sous-ensembles de $\Gamma \backslash \{\mathrm{ne}_G(a), a\}$ Pour tout $m \in \mathcal{M}$,

$$H_{1,a,m} : \theta_b^a = 0 \ \forall b \notin \mathrm{ne}_G(a) \cup m.$$

$$V = Vect(\mathbf{X}_b, b \in \mathrm{ne}_G(a)), \quad S_m = Vect(\mathbf{X}_b, b \in m).$$

$$\phi_m(\mathbf{X}_a, \mathbf{X}_{-a}) = \frac{\|\Pi_{V+S_m}(\mathbf{X}_a) - \Pi_V(\mathbf{X}_a)\|^2 / D_m}{\|\mathbf{X}_a - \Pi_{V+S_m}(\mathbf{X}_a)\|^2 / N_m}.$$

Procédure de tests multiples

- L'hypothèse $H_{0,a}$ est rejetée si

$$\exists m \in \mathcal{M},\ \phi_m(\mathbf{X}_a, \mathbf{X}_{-a}) > \bar{F}^{-1}_{D_m,N_m}(\alpha_n),$$

$$\alpha_n = \sup\left\{u, P\left(\sup_{m \in \mathcal{M}} (\phi_m(\frac{\epsilon_a}{\sigma_a}, \mathbf{X}_{-a}) - \bar{F}^{-1}_{D_m,N_m}(u)) > 0 / \mathbf{X}_{-a}\right) \leq \alpha\right\}$$

Puissance du test

- Collection de modèles : $\mathcal{M} = \{\{b\}\,, b \notin \{\mathrm{ne}(a) \cup a\}\}\,.$

Puissance du test

- Collection de modèles : $\mathcal{M} = \{\{b\}, b \notin \{\mathrm{ne}(a) \cup a\}\}$.

Théorème

(N.Verzelen, F. Villers)
Soit $\alpha, \beta \in]0,1[$. On note $d_a = |\mathrm{ne}(a)|$, $p = |\Gamma|$.

$$\rho^2 = \frac{C_1}{n - d_a} \ln\left(\frac{p - d_a - 1}{\alpha\beta}\right).$$

$P(T_\alpha > 0) \geq 1 - \beta$ dès que :

$$\exists b \in \Gamma \backslash \{\mathrm{ne}(a) \cup a\}, \ \frac{Var(X_a/X_{\mathrm{ne}(a)}) - Var(X_a/X_{\mathrm{ne}(a)\cup\{b\}})}{Var(X_a/X_{\mathrm{ne}(a)\cup\{b\}})} \geq \rho^2$$

Puissance du test

- Collection de modèles : $\mathcal{M} = \{\{b\}, b \notin \{\mathrm{ne}(a) \cup a\}\}$.

Théorème

(N.Verzelen, F. Villers)
Soit $\alpha, \beta \in]0,1[$. *On note* $d_a = |\mathrm{ne}(a)|$, $p = |\Gamma|$.

$$\rho^2 = \frac{C_1}{n - d_a} \ln\left(\frac{p - d_a - 1}{\alpha\beta}\right).$$

$P(T_\alpha > 0) \geq 1 - \beta$ *dès que :*

$$\exists b \in \Gamma \backslash \{\mathrm{ne}(a) \cup a\}, \ \frac{Var(X_a/X_{\mathrm{ne}(a)}) - Var(X_a/X_{\mathrm{ne}(a)\cup\{b\}})}{Var(X_a/X_{\mathrm{ne}(a)\cup\{b\}})} \geq \rho^2$$

- Propriétés d'optimalités de la procédure sous certaines hypothèses.
- Application sur des simulations et des données réelles.

Références Tests adaptatifs

- Y. Baraud, S. Huet, B. L. (2003) *Adaptive tests of linear hypotheses by model selection*, Ann. Statist.
- Y. Baraud, S. Huet, B. L. (2005) *Adaptive tests of qualitative hypotheses* , Ann. Statist.
- M. Fromont, B. L. (2006) *Adaptive goodness-of-fit tests in a density model* , Ann. Statist.
- N. Verzelen, F. Villers (2009) *Tests for Gaussian graphical models.* CSDA .
- N. Verzelen, F. Villers (2010) *Goodness-of-fit Tests for high-dimensional Gaussian linear models.* à paraitre Ann. Statist.
- M. Fromont, B. L., P. Reynaud (2010) *Adaptive tests of homogeneity for Poisson processes* Preprint.