

# Traitement des paramètres de nuisance : orthogonalité et insensibilité

Michèle Basseville

► **To cite this version:**

Michèle Basseville. Traitement des paramètres de nuisance : orthogonalité et insensibilité. [Rapport de recherche] PI 1860, 2007, pp.25. inria-00165485

**HAL Id: inria-00165485**

**<https://hal.inria.fr/inria-00165485>**

Submitted on 26 Jul 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PUBLICATION  
INTERNE  
N° 1860



TRAITEMENT DES PARAMÈTRES DE NUISANCE :  
ORTHOGONALITÉ ET INSENSIBILITÉ

MICHÈLE BASSEVILLE



## Traitement des paramètres de nuisance : orthogonalité et insensibilité

Michèle Basseville \*

Systèmes numériques  
Projet Sisthem

Publication interne n1860 — Juillet 2007 — 25 pages

**Résumé :** On s'intéresse au traitement des paramètres de nuisance à des fins d'isolation de pannes pour le diagnostic. On se place d'abord dans le cas d'un vecteur Gaussien dont on surveille la moyenne. On joue avec différents points de vue : transformation, information, projection, réjection, avec le souci de préservation de l'information ou celui du caractère bloc-diagonal de certaines matrices. Ensuite, on passe brièvement en revue des concepts et méthodes pour l'élimination de paramètres de nuisance, en particulier la notion de paramètres orthogonaux. Enfin on examine une notion d'insensibilité au paramètre de nuisance, qui généralise cette propriété d'orthogonalité au cas de fonctions d'estimation autres que celle du maximum de vraisemblance.

**Mots clés :** Diagnostic de pannes, paramètres de nuisance, paramètres orthogonaux, sensibilité, minmax, information de Fisher, information de Godambe.

*(Abstract: pto)*

\* CNRS-IRISA. [Michele.Basseville@irisa.fr](mailto:Michele.Basseville@irisa.fr)

## Handling nuisance parameters : orthogonality and insensitivity

**Abstract:** The handling of nuisance parameters is discussed with the objective of fault isolation for diagnosis. First, the monitoring of the mean of a Gaussian vector is investigated based on different points of view: transformation, information, projection, rejection. The main concerns are preserving information or the block-diagonal structure of some matrices. Then some key concepts and methods for nuisance parameters elimination are briefly overviewed, among which the notion of orthogonal parameters. Finally a property of insensitivity to the nuisance parameter, which generalizes that orthogonality to other estimating functions than the likelihood.

**Key-words:** Fault diagnosis, nuisance parameters, orthogonal parameters, sensitivity, minmax, Fisher information, Godambe information.

## Introduction

On s'intéresse au traitement des paramètres de nuisance à des fins d'isolation de pannes pour le diagnostic. On se place d'abord dans le cas d'un vecteur Gaussien dont on surveille la moyenne. On rappelle les outils "maison" (sensibilité, minmax) pour l'isolation de pannes. On exhibe le piètre caractère isolant de celui-là, et des formulations (très) antérieures de celui-ci. On décrit deux autres approches, qui, dans le présent cas Gaussien, se réduisent au test minmax. Ce faisant, on joue avec différents points de vue : transformation, information, projection, réjection, et on (re)donne le bréviaire des formules correspondantes.

Ces petits jeux sont guidés par le souci de préservation de l'information relative au paramètre lors des transformations du vecteur, et aussi le souci du caractère bloc-diagonal de la matrice d'incidence et/ou de la matrice de covariance et/ou de la matrice d'information de Fisher. On rappelle aussi que la mise en oeuvre numérique de ces outils peut nécessiter des calculs sensiblement plus élaborés.

Ensuite, dans une deuxième partie, on passe brièvement en revue des méthodes d'élimination de paramètres de nuisance. En particulier, une des notions-clefs pour examiner le découplage (l'orthogonalité) entre deux sous-ensembles de composantes d'un vecteur de paramètres, paramètres d'intérêt et paramètres de nuisance, est le caractère bloc-diagonal de la matrice d'information de Fisher.

Une notion d'insensibilité au paramètre de nuisance, qui généralise cette propriété d'orthogonalité de paramètres au cas de fonctions d'estimation autres que celle du maximum de vraisemblance, a été récemment proposée. On la résume dans une troisième partie.

Un point important est que les matrices de sensibilité et de covariance d'une fonction d'estimation sont en général distinctes, alors qu'elles sont toutes deux égales à la matrice d'information de Fisher dans le cas de la vraisemblance. Et que c'est la matrice de sensibilité qui doit intervenir dans la réjection du paramètre de nuisance, et non pas la matrice d'information.

## 1 Inspection de la moyenne d'un vecteur Gaussien

On considère le problème de l'isolation des pannes sous l'angle de l'inspection des composantes de la moyenne d'un vecteur Gaussien de matrice de covariance connue. Grâce à la méthode générale de conception d'algorithmes de surveillance décrite dans [2], ce problème particulier est de portée très large : on sait en effet réduire une large classe de problèmes de surveillance de processus complexes à ce problème de surveillance de la moyenne d'un vecteur Gaussien. Voir aussi [3, 4].

On ne traite pas le cas où le nombre de pannes est inconnu : une procédure heuristique incrémentale est proposée en [26]. À nombre de pannes simples connu, il s'agit de prendre une décision en présence d'un paramètre de nuisance. Sur ce sujet, voir par exemple [14, 23, 29, 33, 35] et les références qui s'y trouvent.

L'essentiel de cette section est un condensé de [1], à l'exception des parties 1.4.4, 1.4.5 et 1.6.

### 1.1 Préambule

On considère donc un vecteur aléatoire Gaussien de vecteur moyenne  $M\eta$  et de matrice de covariance  $\Sigma$ , ce que l'on note :

$$Z \sim \mathcal{N}(M\eta, \Sigma), \quad \Sigma \triangleq \bar{\Sigma} \bar{\Sigma}^T, \quad M : \text{incidence}$$

Dans ce qui suit, pour chaque test  $t_\alpha$ , on introduit sa définition – le plus souvent *via* un rapport de vraisemblance, et on en donne différentes expressions basées sur :

- une projection, telle que celle définie par  $P_M$  en (5),
- une transformation du vecteur  $Z$  en un vecteur tel le score  $\xi$  en (3) ou un autre "résidu"  $\zeta$  en (10).

On explicite aussi la transformation  $\mathcal{F}$  et les matrices d'incidence  $L$  et de covariance  $R$  définies en (10), et on calcule l'information de Fisher  $\mathbf{F}_\zeta$  donnée par (11).

Le caractère bloc-diagonal de la matrice d'incidence  $L$  et/ou de covariance  $R$  et/ou d'information de Fisher  $\mathbf{F}_\zeta$  peut alors être examiné.

### 1.2 Notations et formules de base

On donne les éléments relatifs à l'approche par maximum de vraisemblance, puis les formules liées à la recherche d'un vecteur transformé.

$$\text{log-vraisemblance} \quad l_\eta(Z) \triangleq -2 \ln p_\eta(Z) \triangleq (Z - M\eta)^T \Sigma^{-1} (Z - M\eta) + C \quad (1)$$

$$\text{information Fisher} \quad \mathbf{F}_Z = M^T \Sigma^{-1} M \triangleq \mathbf{F} \quad (2)$$

$$\text{score} \quad \xi \triangleq \nabla_{\eta} l_{\eta}(Z) = M^T \Sigma^{-1} Z - M^T \Sigma^{-1} M \eta \sim \mathcal{N}(0, \mathbf{F}) \quad (3)$$

$$\text{MLE} \quad \hat{\eta} \triangleq \arg \min l_{\eta}(Z) = (M^T \Sigma^{-1} M)^{-1} M^T \Sigma^{-1} Z \quad (4)$$

### Projection et réjection

$$\text{projection} \quad P_M = I - \bar{\Sigma}^{-1} M (M^T \Sigma^{-1} M)^{-1} M^T \bar{\Sigma}^{-T} \quad (5)$$

$$\text{réjection} \quad \tilde{P}_M M = 0 \quad (6)$$

$$\text{où } \tilde{P}_M \triangleq \bar{\Sigma} P_M \bar{\Sigma}^{-1} \quad (7)$$

$$\text{résidu LS} \quad Z - M \hat{\eta} = \tilde{P}_M Z \quad (8)$$

$$\text{MLE} \quad \hat{\eta} = \mathbf{F}^{-1} M^T \Sigma^{-1} Z \sim \mathcal{N}(\eta, \mathbf{F}^{-1}) \quad (9)$$

### Transformation et information

$$\text{transformation} \quad \zeta \triangleq \mathcal{F} Z \sim \mathcal{N}(L \eta, R), \quad L \triangleq \mathcal{F} M, \quad R \triangleq \mathcal{F} \Sigma \mathcal{F}^T \quad (10)$$

$$\text{information} \quad \mathbf{F}_{\zeta} = L^T R^{-1} L = M^T \mathcal{F}^T (\mathcal{F} \Sigma \mathcal{F}^T)^{-1} \mathcal{F} M \quad (11)$$

## 1.3 Détection

$$\text{Tester } \eta = 0 \text{ contre } \eta \neq 0 \quad (12)$$

### Rapport de vraisemblance

$$t_1 \triangleq 2 \ln \frac{\max_{\eta} p_{\eta}(Z)}{p_0(Z)} = l_0(Z) - \min_{\eta} l_{\eta}(Z) \quad (13)$$

### Projection

$$t_1 = Z^T \bar{\Sigma}^{-T} (I - P_M) \bar{\Sigma}^{-1} Z \quad (14)$$

### Transformation

$$\text{score efficace} \quad \xi \triangleq \nabla_{\eta} l_{\eta}(Z)|_{\eta=0} = \mathcal{F} Z \triangleq M^T \Sigma^{-1} Z \quad (15)$$

$$\mathcal{F} M = \mathbf{F} = \mathcal{F} \Sigma \mathcal{F}^T, \quad \mathbf{F}_{\xi} = \mathbf{F} \quad (16)$$



$$\xi \sim \mathcal{N}(\mathbf{F} \eta, \mathbf{F}) \quad (17)$$

$$\text{test} \quad t_1 = \xi^T \mathbf{F}^{-1} \xi \quad (18)$$

#### Paramètre de non-centralité

$$\tau = \mathbf{K}(\eta) = \eta^T \mathbf{F} \eta \quad \text{sous } \eta \neq 0 \quad (19)$$

#### 1.4 Isolation

On partitionne le vecteur moyenne  $\eta$  en deux sous-ensembles de composantes, les paramètres d'intérêt  $\varphi$  et les paramètres de nuisance  $\psi$  :

$$\eta = \begin{pmatrix} \varphi \\ \psi \end{pmatrix} \quad (20)$$

On note  $p_{\varphi, \psi}(Z)$  la distribution du vecteur Gaussien  $Z \sim \mathcal{N}(M \eta, \Sigma)$ , et on partitionne les matrices correspondantes, soit :

$$M = \begin{pmatrix} M_\varphi & M_\psi \end{pmatrix} \quad (21)$$

$$\mathbf{F} = M^T \Sigma^{-1} M = \begin{pmatrix} M_\varphi^T \Sigma^{-1} M_\varphi & M_\varphi^T \Sigma^{-1} M_\psi \\ M_\psi^T \Sigma^{-1} M_\varphi & M_\psi^T \Sigma^{-1} M_\psi \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{F}_{\varphi\varphi} & \mathbf{F}_{\varphi\psi} \\ \mathbf{F}_{\psi\varphi} & \mathbf{F}_{\psi\psi} \end{pmatrix} \quad (22)$$

$$\mathbf{F}^{-1} = \begin{pmatrix} \mathbf{F}_\varphi^{*-1} & -\mathbf{F}_{\varphi\varphi}^{-1} \mathbf{F}_{\varphi\psi} \mathbf{F}_\psi^{*-1} \\ -\mathbf{F}_{\psi\psi}^{-1} \mathbf{F}_{\psi\varphi} \mathbf{F}_\varphi^{*-1} & \mathbf{F}_\psi^{*-1} \end{pmatrix} \quad (23)$$

où :

$$\mathbf{F}_i^* \triangleq \mathbf{F}_{ii} - \mathbf{F}_{ij} \mathbf{F}_{jj}^{-1} \mathbf{F}_{ji} \quad (24)$$

À nombre de pannes  $n_\varphi$  supposé connu, il s'agit de :

$$\text{Tester } \varphi = 0 \text{ contre } \varphi \neq 0 \text{ avec } \psi \text{ inconnu} \quad (25)$$

Noter néanmoins que  $\psi$  est alors de dimension  $n_\psi$  connue ( $n_\psi \triangleq n - n_\varphi$ ).

On traite d'abord le cas où on suppose connaître les  $n_\varphi$  composantes susceptibles d'être affectées par des pannes et où on s'est arrangé pour qu'elles figurent justement dans  $\varphi$ . Le cas où on ne connaît que le nombre  $n_\varphi$ , mais pas les composantes elles-mêmes, est traité en 1.5.

#### 1.4.1 Remarque : décomposition de Karhunen-Loeve

Cette décomposition permet de transformer  $Z$  en un vecteur de composantes décorréelées (ce qui n'est ni nécessaire ni suffisant pour l'isolation), au moyen de la transformation orthogonale définie par la matrice  $\check{\mathcal{F}}$  telle que :

$$\Sigma = \check{\mathcal{F}} \Delta \check{\mathcal{F}}^T$$

Le vecteur transformé qui en résulte admet les propriétés (incidence, covariance, information) suivantes.

#### Découplage, décorrélation et information

$$\check{\xi} \triangleq \check{\mathcal{F}}^T Z \sim \mathcal{N}(\check{L} \eta, \check{\Sigma}), \quad \check{L} = \check{\mathcal{F}}^T M, \quad \check{\Sigma} = \Delta, \quad \check{\mathbf{F}} = \mathbf{F}$$

#### 1.4.2 Méthode de sensibilité

Il s'agit du problème (25) projeté sur  $\psi = 0$ .

#### Rapport de vraisemblance

$$\tilde{t}_\varphi \triangleq 2 \ln \frac{\max_\varphi p_{\varphi,0}(Z)}{p_{0,0}(Z)} = l_{0,0}(Z) - \min_\varphi l_{\varphi,0}(Z) \quad (26)$$

#### Projection

$$\tilde{t}_\varphi = Z^T \bar{\Sigma}^{-T} (I - P_{M_\varphi}) \bar{\Sigma}^{-1} Z \quad (27)$$

#### Transformation

$$\text{score partiel} \quad \tilde{\xi}_\varphi \triangleq \nabla_\varphi l_{\varphi,0}(Z) = \tilde{\mathcal{F}}_\varphi Z \triangleq M_\varphi^T \Sigma^{-1} Z \quad (28)$$

$$\tilde{\mathcal{F}}_\varphi M_\varphi = \mathbf{F}_{\varphi\varphi} = \tilde{\mathcal{F}}_\varphi \Sigma \tilde{\mathcal{F}}_\varphi^T, \quad \mathbf{F}_{\tilde{\xi}_\varphi} = \mathbf{F}_{\varphi\varphi} \quad (29)$$

$$\tilde{\xi}_\varphi \sim \mathcal{N}(\mathbf{F}_{\varphi\varphi} \varphi, \mathbf{F}_{\varphi\varphi}) \quad (30)$$

$$\text{test} \quad \tilde{t}_\varphi = \tilde{\xi}_\varphi^T \mathbf{F}_{\varphi\varphi}^{-1} \tilde{\xi}_\varphi \quad (31)$$

### Paramètre de non-centralité

$$\tilde{\mathbf{K}}(\eta) = (\varphi + \mathbf{F}_{\varphi\varphi}^{-1} \mathbf{F}_{\varphi\psi} \psi)^T \mathbf{F}_{\varphi\varphi} (\varphi + \mathbf{F}_{\varphi\varphi}^{-1} \mathbf{F}_{\varphi\psi} \psi) \quad (32)$$

Noter que, sous  $\varphi \neq 0$  ET  $\psi = 0$ , le paramètre de non-centralité vaut  $\tilde{\mathbf{K}}(\varphi) = \varphi^T \mathbf{F}_{\varphi\varphi} \varphi$ . Mais qu'il peut être non nul sous  $\varphi = 0$  et  $\psi \neq 0$ . D'où le piètre caractère isolant de ce test.

Le vecteur constitué des deux scores partiels admet les propriétés suivantes.

### Découplage, décorrélation et information

$$\tilde{\xi} \triangleq \begin{pmatrix} \tilde{\xi}_\varphi \\ \tilde{\xi}_\psi \end{pmatrix} \sim \mathcal{N}(\tilde{L} \eta, \tilde{\Sigma}), \quad \tilde{L} = \tilde{\Sigma} = \tilde{\mathbf{F}} = \mathbf{F} \quad (33)$$

#### 1.4.3 Méthode de réjection (minmax)

Il s'agit de résoudre le problème (25) pour la nuisance  $\psi$  la moins favorable au sens de la vraisemblance.

### Rapport de vraisemblance

$$t_\varphi^* \triangleq 2 \ln \frac{\max_{\varphi, \psi} p_{\varphi, \psi}(Z)}{\max_\psi p_{0, \psi}(Z)} = \min_\psi l_{0, \psi}(Z) - \min_{\varphi, \psi} l_{\varphi, \psi}(Z) \quad (34)$$

### Projection

$$t_\varphi^* = Z^T \bar{\Sigma}^{-T} (P_{M_\psi} - P_M) \bar{\Sigma}^{-1} Z \quad (35)$$

### Transformation

$$\xi_\varphi^* \triangleq \mathcal{F}_\varphi^* Z \triangleq M_\varphi^T \Sigma^{-1} \tilde{P}_{M_\psi} Z \quad (36)$$

$$\text{réjection} \quad \mathcal{F}_\varphi^* M_\psi = 0, \quad \text{i.e.} \quad \mathcal{F}_\varphi^* (\nabla_\psi \mathbf{E} l_{\varphi, \psi}(Z)) = 0 \quad (37)$$

$$\mathcal{F}_\varphi^* M_\varphi = \mathbf{F}_\varphi^* = \mathcal{F}_\varphi^* \Sigma \mathcal{F}_\varphi^{*T}, \quad \mathbf{F}_{\xi_\varphi^*} = \mathbf{F}_\varphi^* \quad (38)$$

$$\text{score effectif} \quad \xi_\varphi^* = \tilde{\xi}_\varphi - \mathbf{F}_{\varphi\psi} \mathbf{F}_{\psi\psi}^{-1} \tilde{\xi}_\psi \quad (39)$$

régression du score partiel sur le score de nuisance [29, 14]

$$\xi_\varphi^* \sim \mathcal{N}(\mathbf{F}_\varphi^* \varphi, \mathbf{F}_\varphi^*) \quad (40)$$

$$\text{test} \quad t_\varphi^* = \xi_\varphi^{*T} \mathbf{F}_\varphi^{*-1} \xi_\varphi^* \quad (41)$$

### Paramètre de non-centralité

$$\mathbf{K}^*(\varphi) = \varphi^T \mathbf{F}_\varphi^* \varphi \quad \text{sous } \varphi \neq 0 \text{ et } \forall \psi \quad (42)$$

D'où le caractère parfaitement isolant de ce test.

Le vecteur constitué des deux scores effectifs admet les propriétés suivantes.

### Découplage, décorrélation et information

$$\xi^* \triangleq \begin{pmatrix} \xi_\varphi^* \\ \xi_\psi^* \end{pmatrix} \sim \mathcal{N}(L^* \eta, \Sigma^*)$$

$$L^* = \begin{pmatrix} \mathbf{F}_\varphi^* & 0 \\ 0 & \mathbf{F}_\psi^* \end{pmatrix}, \quad \Sigma^* = \begin{pmatrix} \mathbf{F}_\varphi^* & -\mathbf{F}_\varphi^* \mathbf{F}_{\varphi\psi}^{-1} \mathbf{F}_{\psi\psi} \\ -\mathbf{F}_\psi^* \mathbf{F}_{\psi\psi}^{-1} \mathbf{F}_{\psi\varphi} & \mathbf{F}_\psi^* \end{pmatrix}, \quad \mathbf{F}^* = \mathbf{F}$$

#### 1.4.4 Méthode de Wald

L'idée, pour tester si  $f(\eta) \neq 0$ , est d'examiner – à l'aide sa normalité asymptotique – si  $f(\hat{\eta})$  est proche de zéro ou non [35, 27].

Pour  $f(\eta) = F\eta$ , où  $F = ( I \ 0 )$ , et dans le présent cas particulier de l'inférence sur la moyenne d'un vecteur Gaussien – voir (9), ceci donne :

$$\xi_\varphi^\sharp \triangleq F \hat{\eta} = F \mathbf{F}^{-1} M^T \Sigma^{-1} Z \triangleq \mathcal{F}_\varphi^\sharp Z \quad (43)$$

$$\text{réjection} \quad \mathcal{F}_\varphi^\sharp M_\psi = 0 \quad (44)$$

$$\mathcal{F}_\varphi^\sharp M_\varphi = I, \quad \mathcal{F}_\varphi^\sharp \Sigma \mathcal{F}_\varphi^{\sharp T} = \mathbf{F}_\varphi^{*-1} \quad (45)$$

$$\xi_\varphi^\# \sim \mathcal{N}(\varphi, \mathbf{F}_\varphi^{*-1}) \quad (46)$$

$$\text{test} \quad t_\varphi^\# \triangleq \xi_\varphi^{\#T} \mathbf{F}_\varphi^* \xi_\varphi^\# \quad (47)$$

$$t_\varphi^\# = t_\varphi^* \quad (48)$$

Autrement dit, dans ce cas particulier, le test de Wald est identique au test de réjection minmax  $t_\varphi^*$ .

#### 1.4.5 Méthode du multiplicateur de Lagrange

L'idée est de se servir du multiplicateur de Lagrange  $\lambda$  correspondant au Lagrangien

$$l_\eta - f^T(\eta) \lambda$$

de l'estimation de  $\eta$  par maximum de vraisemblance sous la contrainte vectorielle  $f(\eta) = 0$ , et de la normalité asymptotique de  $\hat{\lambda}$ .

Pour  $f(\eta) = F\eta$ , où  $F = \begin{pmatrix} \mathbf{0} & \mathbf{I} \end{pmatrix}$ , et dans le présent cas particulier de l'inférence sur la moyenne d'un vecteur Gaussien, on obtient [27] :

$$\text{test} \quad t_\lambda \triangleq \hat{\lambda}^T \Sigma_\lambda^{-1} \hat{\lambda} \quad (49)$$

$$t_\lambda = t_\varphi^* \quad (50)$$

La conclusion est donc la même que pour le test de Wald.

D'autres tests asymptotiques sont décrits en [27]. Dans le présent cas particulier de l'inférence sur la moyenne d'un vecteur Gaussien, ils sont tous identiques au test minmax  $t_\varphi^*$  défini en (41).

#### 1.4.6 Mixer réjection et sensibilité

L'idée est de surveiller les composantes d'intérêt  $\varphi$  avec le test minmax (41) et les composantes de nuisance  $\psi$  avec le test de sensibilité (31).

Le vecteur constitué des deux scores (respectivement, effectif et partiel) correspondants admet les propriétés suivantes.

#### Découplage, décorrélation et information

$$\check{\xi} \triangleq \begin{pmatrix} \xi_\varphi^* \\ \xi_\psi \end{pmatrix} \sim \mathcal{N}(\check{L} \eta, \check{\Sigma}), \quad \check{L} = \begin{pmatrix} \mathbf{F}_\varphi^* & \mathbf{0} \\ \mathbf{F}_{\psi\varphi} & \mathbf{F}_{\psi\psi} \end{pmatrix}, \quad \check{\Sigma} = \begin{pmatrix} \mathbf{F}_\varphi^* & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{\psi\psi} \end{pmatrix}, \quad \check{\mathbf{F}} = \mathbf{F}$$

Cette idée de mélange des deux méthodes est issue de la remarque suivante.

### Décomposition du test global

$$\begin{aligned}
 t_1 &= -2 \ln \frac{\max_{\varphi, \psi} p_{\varphi, \psi}}{p_{0,0}} \\
 &= -2 \ln \frac{\max_{\varphi, \psi} p_{\varphi, \psi}}{\max_{\psi} p_{0, \psi}} - 2 \ln \frac{\max_{\psi} p_{0, \psi}}{p_{0,0}} \\
 &= t_{\varphi}^* + \tilde{t}_{\psi}
 \end{aligned} \tag{51}$$

Le caractère additif de cette décomposition est vrai pour une distribution générale (pas seulement Gaussienne), et dans le cas de plus de deux pannes.

Le caractère orthogonal – au sens :  $\text{cov}(t_{\varphi}^*, \tilde{t}_{\psi}) = 0$  – de cette décomposition est vrai seulement pour une distribution Gaussien, et seulement dans le cas de deux pannes [1].

## 1.5 Procédures d'isolation des $n_{\varphi}$ pannes

### 1.5.1 Par sensibilité

- Calculer tous les tests  $\tilde{t}_{\varphi}$  correspondants à toutes les partitions possibles de  $\eta$  en 2 ensembles de composantes de dimensions  $(n_{\varphi}, n - n_{\varphi})$ .
- Pannes isolées : celles qui correspondent au test  $\tilde{t}_{\varphi}$  le plus grand.

Cette méthode correspond à une décision par maximum *a posteriori* (sous l'hypothèse d'un *a priori* uniforme sur toutes les pannes simples).

### 1.5.2 Par réjection (minmax)

- Calculer tous les tests  $t_{\varphi}^*$  correspondants à toutes les partitions possibles de  $\eta$  en 2 ensembles de composantes de dimensions  $(n_{\varphi}, n - n_{\varphi})$ .
- Pannes isolées : celles qui correspondent au test  $t_{\varphi}^*$  le plus grand.

Cette procédure possède un caractère d'optimalité. On sait que la collection des tests minmax optimise deux critères de puissance (soit le minimum, soit la 'somme' des puissances des tests minmax individuels) [26, 34, 1].

## 1.6 Calculs numériques

Les tests ci-dessus font intervenir des inversions de matrices et leur mise en oeuvre requiert un minimum de précautions numériques, lorsque les matrices à inverser ont un mauvais conditionnement – ce qui n'est pas rare dans le contexte général [2, 3].

Les recommandations principales [37] sont :

- éviter au maximum les inversions de matrices,
- calculer les formes quadratiques  $\tilde{t}_\varphi$ ,  $t_\varphi^*$  comme des carrés.

Ceci est réalisé dans le logiciel [36].

## 2 Traitement des paramètres de nuisance

Les éléments de cette section sont extraits de [6]. On considère le cas général d'observations suivant une distribution  $p_\theta$ , dont le vecteur de paramètres  $\theta$  est partitionné en deux sous-ensembles de composantes, les paramètres d'intérêt  $\varphi$  et les paramètres de nuisance  $\psi$  :

$$\theta = \begin{pmatrix} \varphi \\ \psi \end{pmatrix}$$

Réduire l'influence ou éliminer des paramètres de nuisance est une question ancienne, reconnue comme un problème majeur de l'inférence statistique. Une étude parue il y a une trentaine d'années [7] identifiait déjà dix catégories de méthodes, certaines d'entre elles remontant aux travaux de R.A. Fisher. On trouvera en [25] une revue plus récente. On esquisse ici quelques idées pour réduire ou éliminer l'effet de paramètres de nuisance.

### 2.1 Approche Bayésienne

D'un point de vue Bayésien, l'élimination d'un paramètre de nuisance peut sembler ne pas poser de problème, en ce sens que le calcul de la densité marginale *a posteriori* du paramètre d'intérêt fournit la solution [8]. Cependant, en pratique la situation peut être beaucoup moins simple ; voir [25], qui contient une revue de l'approche Bayésienne *integrated likelihood*.

### 2.2 Vraisemblances marginale et conditionnelle

Dans l'approche par vraisemblance, le problème est de trouver une fonction de vraisemblance pour le seul paramètre d'intérêt [31]. Dans certains cas, il peut arriver que les distributions marginales de certaines composantes de l'observation ne dépendent pas du paramètre de nuisance  $\psi$ . Une autre idée consiste à baser l'inférence au sujet de  $\varphi$  sur une distribution conditionnelle des observations, étant donnée une statistique exhaustive pour  $\psi$  à  $\varphi$  fixé [24]. De telles distributions, appelées respectivement *vraisemblance marginale* et *vraisemblance conditionnelle*, peuvent alors être

utilisées comme pseudo-vraisemblances pour  $\varphi$ . Quelques contre-exemples – e.g. une distribution marginale et une distribution conditionnelle fournissant des résultats différents [7] – suggèrent que de telles approches doivent être utilisées avec précaution [25]. Cependant, l'utilisation de fonctions scores conditionnelles s'appuie sur des résultats d'optimalité, tandis que celle des fonctions scores marginales ne l'est pas [24]. On trouvera en [32][Chap.8] une discussion détaillée de développements modernes, y compris des approximations pour ces pseudo-vraisemblances. La manière dont l'information de nuisance est paramétrée peut être d'importance cruciale dans ce contexte.

### 2.3 Orthogonalité des paramètres

Une propriété particulièrement intéressante est celle de l'*orthogonalité des paramètres* : on dit que le paramètre d'intérêt  $\varphi$  est orthogonal (par rapport à l'information de Fisher) au paramètre de nuisance  $\psi$  si les fonctions scores correspondantes sont décorréélées, ou de manière équivalente si le bloc  $\mathbf{F}_{\varphi\psi}$  hors de la diagonale de la matrice d'information de Fisher, et donc le bloc  $\mathbf{F}_{\psi\varphi}$ , sont nuls [20, 18, 19, 10]. On parle d'orthogonalité locale (resp. globale) lorsque ces conditions sont satisfaites pour une valeur du paramètre (resp. toutes). En cas d'orthogonalité, les estimateurs du maximum de vraisemblance (MLE) des deux paramètres  $\varphi$  et  $\psi$  sont asymptotiquement indépendants.

Quand  $\varphi$  est scalaire, on peut toujours trouver une transformation en une paramétrisation orthogonale globale qui préserve  $\varphi$  [10].

L'orthogonalité des paramètres est un cas particulier de l'orthogonalité des fonctions d'estimation discutée dans [24].

L'orthogonalité des fonctions d'estimation projetées explique un paradoxe associé à l'estimation de paramètres en présence de paramètres de nuisance : l'efficacité de certains estimateurs peut décroître lorsque le paramètre de nuisance est connu [15].

### 2.4 Exploitation de propriétés d'invariance

L'exploitation de l'invariance de la distribution de probabilité sous certaines transformations est connue sous le nom d'*approche invariante* de l'inférence [23], et étudiée plus avant pour le traitement des paramètres de nuisance dans [5].

### 2.5 Approche minmax ou vraisemblance profil

L'approche à base de vraisemblance la plus générale consiste à maximiser la vraisemblance par rapport au paramètre de nuisance. La fonction de *vraisemblance*



*profil*, également appelée *vraisemblance concentrée* ou *maximale*, est :

$$\mathcal{L}_p(\varphi) \triangleq \mathcal{L}(\varphi, \hat{\psi}_\varphi) \quad (52)$$

où  $\hat{\psi}_\varphi$  est le MLE de  $\psi$  pour  $\varphi$  fixé [28].

Des approximations jusqu'au troisième ordre ont été étudiées pour l'évaluation de la vraisemblance profil [30].

La fonction score correspondant à la vraisemblance profil n'est plus centrée. Quand le nombre de paramètres de nuisance est petit devant la taille de l'échantillon, ce biais est souvent négligeable devant l'erreur d'estimation. Ceci n'est plus vrai dans le cas de paramètres de nuisance multiples, où le profil peut induire en erreur. Des corrections de la vraisemblance profil ont été proposées [10].

Le principe minmax d'élimination de nuisance (par maximisation par rapport à  $\psi$ ) a aussi été appliqué à la fonction de risque de procédures de décision [23].

### 3 Insensibilité au paramètre de nuisance

On suppose toujours observer un processus  $Z$  de distribution  $p_\theta$ , dont le paramètre  $\theta$  est partitionné en deux sous-ensembles de composantes  $\varphi$  et  $\psi$ .

Pour le traitement du paramètre de nuisance  $\psi$ , on rappelle la notion – symétrique en  $\varphi$ ,  $\psi$  – de découplage des scores partiels ou **F**-orthogonalité, et on en décrit une extension *asymétrique* dans le cas de fonctions d'estimation autres que le score de vraisemblance.

#### 3.1 Score efficace et information de Fisher

On note  $\xi$  le score efficace associé :

$$\xi(\theta, Z) \triangleq \nabla_\theta l_\theta(Z) \quad (53)$$

où :

$$l_\theta(Z) \triangleq -\ln p_\theta(Z)$$

Le score est sans biais, i.e.

$$\mathbf{E}_\theta (\xi(\theta, Z)) = 0$$

où  $\mathbf{E}_\theta$  est l'espérance sous la loi  $p_\theta$ . Dans ce qui suit, on omet la dépendance en  $Z$ , voire en  $\theta$ , quand il n'y a pas d'ambiguïté.

Le score se décompose en :

$$\xi = \begin{pmatrix} \xi_\varphi \\ \xi_\psi \end{pmatrix}$$

où les deux scores partiels  $\xi_\varphi$  et  $\xi_\psi$  associés à  $l_\theta \triangleq l_{\varphi,\psi}$  sont définis par :

$$\begin{aligned} \xi_\varphi &\triangleq \nabla_\varphi l_{\varphi,\psi}(Z) \\ \xi_\psi &\triangleq \nabla_\psi l_{\varphi,\psi}(Z) \end{aligned} \quad (54)$$

La matrice de covariance du score  $\xi$  est la matrice d'information de Fisher :

$$\mathbf{F} \triangleq \mathbf{E} \left( \xi \xi^T \right) \quad (55)$$

qui s'écrit aussi :

$$\mathbf{F} = - \mathbf{E} (\nabla_\theta \xi) \quad (56)$$

Elle se partitionne en :

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{\varphi\varphi} & \mathbf{F}_{\varphi\psi} \\ \mathbf{F}_{\psi\varphi} & \mathbf{F}_{\psi\psi} \end{pmatrix} \quad (57)$$

où :

$$\begin{aligned} \mathbf{F}_{\varphi\varphi} &\triangleq \mathbf{E} \left( \xi_\varphi \xi_\varphi^T \right) = - \mathbf{E} (\nabla_\varphi \xi_\varphi) \\ \mathbf{F}_{\psi\psi} &\triangleq \mathbf{E} \left( \xi_\psi \xi_\psi^T \right) = - \mathbf{E} (\nabla_\psi \xi_\psi) \\ \mathbf{F}_{\varphi\psi} &\triangleq \mathbf{E} \left( \xi_\varphi \xi_\psi^T \right) = - \mathbf{E} (\nabla_\varphi \xi_\psi) \end{aligned} \quad (58)$$

Son inverse est donnée par (23)-(24).

Le *score effectif* s'obtient par projection :

$$\xi_\varphi^* \triangleq \mathcal{F}_\varphi^* \xi \quad (59)$$

où  $\mathcal{F}_\varphi^*$  est telle que :

$$\mathcal{F}_\varphi^* (\nabla_\psi \mathbf{E} l_{\varphi,\psi}(Z)) = 0 \quad (60)$$

Le score effectif est aussi le résidu de la régression du score partiel d'intérêt sur le score partiel de nuisance [29, 14] :

$$\xi_\varphi^* = \xi_\varphi - \mathbf{F}_{\varphi\psi} \mathbf{F}_{\psi\psi}^{-1} \xi_\psi \quad (61)$$

Dans le cas d'un vecteur Gaussien, (59)-(61) correspond bien à la méthode min-max de réjection (37)-(39).

### 3.1.1 Découplage (symétrique) par $\mathbf{F}$ -orthogonalité

Pour l'estimation par maximum de vraisemblance, la notion-clef pour examiner le découplage (l'orthogonalité) entre les deux sous-ensembles de paramètres d'intérêt  $\varphi$  et de nuisance  $\psi$  est le caractère bloc-diagonal de la matrice d'information de Fisher (57). En effet, si  $\mathbf{F}$  est bloc-diagonale,

$$\mathbf{F}_{\varphi\psi} = \mathbf{F}_{\psi\varphi}^T = 0 \quad (62)$$

et, en vertu de la première égalité de (58), les deux scores partiels  $\xi_\varphi$  et  $\xi_\psi$  (54) sont décorrélés<sup>1</sup>. Les paramètres  $\varphi$  et  $\psi$  sont alors dits orthogonaux [18, 10]. Ce que l'on désigne par  $\mathbf{F}$ -orthogonalité (au sens de l'information de Fisher) entre les paramètres  $\varphi$  et  $\psi$ . De la dernière égalité de (58), on peut dire que  $\xi_\varphi$  est  $\psi$ -insensible.

Cette propriété est symétrique en les deux composantes  $\varphi$  et  $\psi$  de  $\theta$ .

Elle implique quatre propriétés fondamentales du MLE pour un échantillon  $Z_1, \dots, Z_n$  [10, 21, 22].

1. Indépendance asymptotique des estimateurs MLE  $\hat{\varphi}$  et  $\hat{\psi}$ ;
2. Estimation efficace stable de  $\varphi$ , i.e. la variance asymptotique de  $\hat{\varphi}$  est la même, que  $\psi$  soit traité comme connu ou inconnu ;
3. Simplification de l'algorithme d'estimation ;
4. L'estimateur  $\hat{\varphi}_\psi$  de  $\varphi$  à  $\psi$  fixé varie lentement avec  $\psi$ .

La préservation de la propriété 2, non garantie dans le cas de fonctions d'estimation autres que de vraisemblance, a motivé l'introduction d'une notion d'insensibilité étendant la propriété de  $\mathbf{F}$ -orthogonalité.

### 3.2 Fonction d'estimation, sensibilité et information de Godambe

Soit  $\zeta$ , de dimension au moins égale à celle de  $\theta$ , une fonction d'estimation de  $\theta$ , i.e. une fonction telle que, sur la base de l'échantillon  $Z_1, \dots, Z_n$ , la solution en  $\theta$  de l'équation :

$$\sum_{i=1}^n \zeta(\theta, Z_i) = 0$$

fournisse un estimateur  $\hat{\theta}$  de  $\theta$ . Dans le cas de l'estimation par maximum de vraisemblance, la fonction  $\zeta$  n'est autre que le score efficace :

$$\zeta(\theta, Z_i) = \nabla_\theta l_\theta(Z_i) = \xi(\theta, Z_i)$$

---

<sup>1</sup>Voir (33) dans le cas Gaussien.

On suppose  $\zeta$  sans biais, i.e. :

$$\mathbf{E}_\theta(\zeta(\theta, Z)) = 0$$

ce qui est effectivement le cas du score efficace.

Sous des conditions de régularité [11],  $\hat{\theta}$  est consistant et asymptotiquement Gaussien de matrice de covariance :

$$\text{var}(\hat{\theta}) = \mathbf{G}^{-1}$$

où :

$$\mathbf{G} \triangleq \mathcal{J}^T \Sigma^{-1} \mathcal{J} \quad (63)$$

$$\Sigma \triangleq \mathbf{E} \begin{pmatrix} \zeta \\ \zeta^T \end{pmatrix} \quad (64)$$

$$\mathcal{J} \triangleq - \mathbf{E} \begin{pmatrix} \zeta \\ \xi^T \end{pmatrix} \quad (65)$$

$$= + \mathbf{E}(\nabla_\theta \zeta) \quad (66)$$

La matrice  $\mathbf{G}$  est appelée *matrice d'information de Godambe* [13, 17]. Noter la définition inhabituelle de  $\mathcal{J}$  en (65), introduite en [16, 17, 21].

Dans le cas du score efficace :  $\zeta = \xi$  en (53), toutes ces matrices se réduisent à l'information de Fisher :  $-\mathcal{J} = \Sigma = \mathbf{G} = \mathbf{F}$ .

Partitionnons la fonction d'estimation  $\zeta$  en :

$$\zeta = \begin{pmatrix} \zeta_\varphi \\ \zeta_\psi \end{pmatrix}$$

et la sensibilité  $\mathcal{J}$  en<sup>2</sup> :

$$\mathcal{J} = \begin{pmatrix} \mathcal{J}_{\varphi\varphi} & \mathcal{J}_{\varphi\psi} \\ \mathcal{J}_{\psi\varphi} & \mathcal{J}_{\psi\psi} \end{pmatrix} \quad (67)$$

où :

$$\begin{aligned} \mathcal{J}_{\varphi\varphi} &\triangleq - \mathbf{E} \begin{pmatrix} \zeta_\varphi \\ \xi_\varphi^T \end{pmatrix} = \mathbf{E}(\nabla_\varphi \zeta_\varphi) \\ \mathcal{J}_{\psi\psi} &\triangleq - \mathbf{E} \begin{pmatrix} \zeta_\psi \\ \xi_\psi^T \end{pmatrix} = \mathbf{E}(\nabla_\psi \zeta_\psi) \\ \mathcal{J}_{\varphi\psi} &\triangleq - \mathbf{E} \begin{pmatrix} \zeta_\varphi \\ \xi_\psi^T \end{pmatrix} = \mathbf{E}(\nabla_\psi \zeta_\varphi) \end{aligned} \quad (68)$$

---

<sup>2</sup>Noter que le partitionnement de  $\mathcal{J}$  en (67) n'est pas le même que celui de  $M$  en (21).

Les fonctions  $\zeta_\varphi$  et  $\zeta_\psi$  sont appelées *fonctions d'estimation marginales* :  $\zeta_\varphi$  est une fonction d'estimation de  $\varphi$  quand  $\psi$  est connu, et  $\zeta_\psi$  est une fonction d'estimation de  $\psi$  quand  $\varphi$  est connu. Les matrices  $\mathcal{J}_{\varphi\varphi}$  et  $\mathcal{J}_{\psi\psi}$  sont appelées *sensibilités marginales*, et les matrices  $\mathcal{J}_{\varphi\psi}$  et  $\mathcal{J}_{\psi\varphi}$  les  $\psi$ - et  $\varphi$ -sensibilités de  $\zeta_\varphi$  et  $\zeta_\psi$ , respectivement [21].

Enfin partitionnons l'information de Godambe  $\mathbf{G}$  (63) et son inverse comme il est fait pour l'information de Fisher  $\mathbf{F}$  en (57) et (23), soit :

$$\mathbf{G} = \mathcal{J}^T \Sigma^{-1} \mathcal{J} \triangleq \begin{pmatrix} \mathbf{G}_{\varphi\varphi} & \mathbf{G}_{\varphi\psi} \\ \mathbf{G}_{\psi\varphi} & \mathbf{G}_{\psi\psi} \end{pmatrix} \quad (69)$$

et :

$$\mathbf{G}^{-1} = \begin{pmatrix} \mathbf{G}_\varphi^{*-1} & -\mathbf{G}_{\varphi\varphi}^{-1} \mathbf{G}_{\varphi\psi} \mathbf{G}_\psi^{*-1} \\ -\mathbf{G}_{\psi\psi}^{-1} \mathbf{G}_{\psi\varphi} \mathbf{G}_\varphi^{*-1} & \mathbf{G}_\psi^{*-1} \end{pmatrix} \quad (70)$$

où :

$$\mathbf{G}_i^* = \mathbf{G}_{ii} - \mathbf{G}_{ij} \mathbf{G}_{jj}^{-1} \mathbf{G}_{ji} \quad (71)$$

Quand on utilise une fonction d'estimation autre que celle du maximum de vraisemblance, i.e. une fonction  $\zeta$  autre que le score  $\xi$ , la propriété 2 n'est pas toujours satisfaite. Il existe même des situations paradoxales où la connaissance du paramètre de nuisance fait augmenter la variance (diminuer l'efficacité) de l'estimation du paramètre d'intérêt [15].

**Fonction d'estimation projetée ou standardisée.** Ce paradoxe se produit lorsque les deux composantes de la projection de la fonction d'estimation considérée sur le score efficace sont ortho-gonales. On appelle fonction d'estimation projetée [15], ou parfois standardisée [17], la fonction :

$$\bar{\zeta} = -\mathcal{J}^T \Sigma^{-1} \zeta \quad (72)$$

Cette fonction est sans biais d'information, parce que :  $-\mathcal{J}_{\bar{\zeta}} = \Sigma_{\bar{\zeta}} = \mathbf{G}_{\bar{\zeta}}$ .

Elle est utile pour les reparamétrisations  $\mathbf{G}$ -orthogonales [21].

### 3.2.1 Insensibilité (asymétrique)

La propriété 2 étant considérée comme cruciale, une notion étendue d'orthogonalité pour fonctions d'estimation, dite d'insensibilité au paramètre de nuisance, est introduite en [21, 22], qui satisfait les propriétés 2 à 4.

On dit qu'une fonction d'estimation marginale  $\zeta_\varphi$  est *insensible au paramètre de nuisance*  $\psi$ , ou en abrégé  *$\psi$ -insensible*, lorsque :

$$\mathcal{J}_{\varphi\psi} = - \mathbf{E} \left( \zeta_\varphi \xi_\psi^T \right) = 0 \quad (73)$$

Dans le cas du score efficace :  $\zeta = \xi$  en (53), et lorsque  $\xi_\varphi$  est  $\psi$ -insensible, on retrouve la  $\mathbf{F}$ -orthogonalité comme cas particulier.

Si  $\zeta$  satisfait (73), alors à cause de (66),  $\zeta$  est insensible à  $\psi$ , au sens où :

$$\mathbf{E}(\nabla_\psi \zeta_\varphi) = 0$$

Cette propriété est non-symétrique en les deux composantes  $\varphi$  et  $\psi$  de  $\theta$ .

Le théorème d'insensibilité dit qu'une fonction d'estimation marginale  $\zeta_\varphi$  est  $\psi$ -insensible si et seulement si :

$$\mathbf{G}_\varphi^* = \mathcal{J}_{\varphi\varphi}^T \Sigma_{\varphi\varphi}^{-1} \mathcal{J}_{\varphi\varphi} \quad (74)$$

Montrer (74) pour une  $\zeta_\varphi$   $\psi$ -insensible est immédiat. La réciproque l'est moins [21, 22].

**Fonction d'estimation marginale propre.** On dit  $\zeta_\varphi$  est une fonction d'estimation marginale propre lorsqu'elle dépend de  $(\varphi, \psi)$  seulement *via*  $\varphi$ . Dans ce cas il est facile de montrer que (74) est vérifiée et donc que  $\zeta_\varphi$  est  $\psi$ -insensible.

C'est le cas d'une fonction sous forme produit :

$$\zeta_\varphi(\varphi) A(\psi) \quad (75)$$

où  $A$  est une matrice non aléatoire et non singulière. La même conclusion vaut pour des fonctions d'estimation qui sont des sommes de termes de la forme (75) [21].

Les équations de Yule-Walker, pour un modèle AR( $p$ ) de vecteur de paramètres autorégressifs  $\varphi$ , constituent un exemple de fonction d'estimation marginale propre [22]. En effet, ces équations s'écrivent :

$$\Gamma \varphi - \gamma = 0 \quad (76)$$

où  $\gamma$  est le vecteur des covariances empiriques jusqu'à l'ordre  $p$  et  $\Gamma$  est la matrice  $p \times p$  des covariances empiriques correspondante. Écrire l'équation (76) sous la forme autocovariance, plutôt que sous la forme autocorrélation, permet de conclure que le membre de gauche de (76) est une fonction d'estimation marginale sans biais pour  $\varphi$ ,

et évidemment propre par rapport à la variance du processus. (D'où l'efficacité stable de cet estimateur par rapport à cette variance).

Les fonctions d'estimation martingales [9] sont un autre exemple de fonctions d'estimation marginales propres [21].

### 3.2.2 **G-orthogonalité**

Les paramètres  $\varphi$  et  $\psi$  sont **G**-orthogonaux par rapport à  $\zeta$  si :

$$\mathbf{G}_{\varphi\psi} = \mathbf{G}_{\psi\varphi}^T = 0 \quad (77)$$

i.e. si **G** est bloc-diagonale.

Bien qu'apparemment similaire à la **F**-orthogonalité, la **G**-orthogonalité est plus compliquée, à cause de la partition de  $\mathcal{J}$  en (67).

Si  $\zeta_\varphi$  est  $\psi$ -insensible, alors  $\varphi$  et  $\psi$  sont **G**-orthogonaux si et seulement si [21, 22] :

$$\mathbf{E} \begin{pmatrix} \zeta_\psi & \bar{\zeta}_\varphi^T \end{pmatrix} = \mathbf{E} \begin{pmatrix} \zeta_\psi & \xi_\varphi^T \end{pmatrix} = - \mathcal{J}_{\psi\varphi}$$

où  $\bar{\zeta}_\varphi$  est la fonction d'estimation partielle standardisée introduite en (72) :

$$\bar{\zeta}_\varphi = - \mathcal{J}_{\varphi\varphi}^T \Sigma_{\varphi\varphi}^{-1} \zeta_\varphi$$

et  $\xi_\varphi$  est le score partiel.

Il y a donc une différence fondamentale entre **G**- et **F**-orthogonalité, en ce sens que, sous la **G**-orthogonalité,  $\zeta_\psi$  n'a pas besoin d'être  $\varphi$ -insensible (puisque  $\mathcal{J}_{\psi\varphi}$  n'est pas forcément nulle) bien que  $\zeta_\varphi$  soit  $\psi$ -insensible.

De nombreux aspects de la **G**-orthogonalité sont étudiés en [13, 12].

### 3.2.3 **Insensibilisation par projection**

Comme dans [33], on peut insensibiliser par projection. Étant données des fonctions d'estimation marginales régulières  $\zeta_\varphi$  et  $\zeta_\psi$ , on peut construire une autre fonction d'estimation marginale régulière pour  $\varphi$  *via* :

$$\zeta_\varphi^* = \zeta_\varphi - \mathcal{J}_{\varphi\psi} \mathcal{J}_{\psi\psi}^{-1} \zeta_\psi \quad (78)$$

i.e. la projection de  $\zeta_\varphi$  sur le complémentaire orthogonal de  $\xi_\psi$ . La nouvelle paire de fonctions d'estimation  $(\zeta_\varphi^*, \zeta_\psi)$  est équivalente à  $(\zeta_\varphi, \zeta_\psi)$ . On vérifie que  $\zeta_\varphi^*$  est  $\psi$ -insensible.

L'application de (78) au score partiel  $\xi_\varphi$  permet de retrouver le score effectif (61) :

$$\xi_\varphi^* = \xi_\varphi - \mathbf{F}_{\varphi\psi} \mathbf{F}_{\psi\psi}^{-1} \xi_\psi \quad (79)$$

Noter que ce qui joue, dans (78), le rôle joué par l'information de Fisher  $\mathbf{F}$  dans (79), ce n'est pas l'information de Godambe  $\mathbf{G}$  (63), mais c'est la sensibilité  $\mathcal{J}$  (65). Comparer en effet (55)-(56) à (65)-(66).

Par analogie avec (59)-(61), la fonction d'estimation marginale  $\psi$ -insensible  $\zeta_\varphi^*$  définie en (78) s'écrit aussi :

$$\zeta_\varphi^* \triangleq \mathcal{G}_\varphi^* \zeta \quad (80)$$

où  $\mathcal{G}_\varphi^*$  est telle que :

$$\mathcal{G}_\varphi^* (\nabla_\psi \mathbf{E} \zeta) = 0 \quad (81)$$

### 3.3 Rôle des reparamétrisations

Le rôle joué par les reparamétrisations est complètement différent pour la notion d'insensibilité au paramètre de nuisance (73), par rapport au rôle joué pour les  $\mathbf{F}$ -et  $\mathbf{G}$ -orthogonalités (62) et (77).

Si  $\varphi$  est scalaire, on peut étendre à la  $\mathbf{G}$ -orthogonalité le résultat de [10] pour la  $\mathbf{F}$ -orthogonalité, i.e. on peut trouver une transformation du vecteur de paramètres  $\theta$  en une paramétrisation  $\mathbf{G}$ -orthogonale [21].



## Références

- [1] M. Basseville. Information criteria for residual generation and fault detection and isolation. *Automatica*, 33(5) :783–803, 1997.
- [2] M. Basseville. On-board component fault detection and isolation using the statistical local approach. *Automatica*, 34(11) :1391–1416, 1998.
- [3] M. Basseville, M. Abdelghani, and A. Benveniste. Subspace-based fault detection algorithms for vibration monitoring. *Automatica*, 36(1) :101–109, January 2000.
- [4] M. Basseville, L. Mevel, and M. Goursat. Statistical model-based damage detection and localization : subspace-based residuals and damage-to-noise sensitivity ratios. *Jal Sound and Vibration*, 275(3-5) :769–794, 2004.
- [5] M. Basseville and I. Nikiforov. Fault isolation for diagnosis : Nuisance rejection and multiple hypothesis testing. *Annual Reviews in Control*, 26(2) :189–202, 2002.
- [6] M. Basseville and I. Nikiforov. Handling nuisance parameters in systems monitoring. In *CDC-ECC'05, Sevilla, S.*, 2005.
- [7] D. Basu. On the elimination of nuisance parameters. *Jal American Statistical Association*, 72(358) :355–366, 1977.
- [8] J.O. Berger, B. Liseo, and R.L. Wolpert. Integrated likelihood methods for eliminating nuisance parameters (with discussion). *Statistical Science*, 14(1) :1–28, 1999.
- [9] B.M. Bibby and M. Sorensen. Martingales estimating functions for discretely observed diffusion processes. *Bernoulli*, 1 :17–39, 1995.
- [10] D.R. Cox and N. Reid. Parameter orthogonality and approximate conditional inference (with discussion). *Jal Royal Statistical Society*, 49(1) :1–39, 1987.
- [11] B. Delyon, A. Juditsky, and A. Benveniste. On the relationship between identification and local tests. Research Report 1104, IRISA, May 1997.
- [12] V. P. Godambe. Orthogonality of estimating functions and nuisance parameters. *Biometrika*, 78(1) :143–151, 1991.
- [13] V. P. Godambe and M. E. Thompson. An extension of quasi-likelihood estimation. *Jal Statistical Planning and Inference*, 22(2) :137–172, 1989.
- [14] W.J. Hall and D.J. Mathiason. On large sample estimation and testing in parametric models. *International Statistical Review*, 58(1) :77–97, 1990.

- [15] M. Henmi and S. Eguchi. A paradox concerning nuisance parameters and projected estimating functions. *Biometrika*, 91(4) :929–942, 2004.
- [16] C.C. Heyde. Quasi-likelihood and optimality for estimating functions : some current unifying themes (Fisher lecture). In *47th session of the ISI (Int. Stat. Inst.)*, pages 19–29, Paris, F., September 1989.
- [17] C.C. Heyde. *Quasi-Likelihood and its Applications*. Series in Statistics. Springer Verlag, Berlin, 1997.
- [18] V.S. Huzurbazar. Probability distributions and orthogonal parameters. *Proc. Cambridge Philosophical Society*, 46 :281–284, 1950.
- [19] V.S. Huzurbazar. Sufficient statistics and orthogonal parameters. *Sankhya : The Indian Journal of Statistics*, 17 :217–220, 1956.
- [20] H. Jeffreys. *Theory of Probability (2nd ed.)*. Oxford Classic Texts in the Physical Sciences, 1948.
- [21] B. Jørgensen and S.J. Knudsen. Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal Statistics*, 31(1) :93–114, 2004.
- [22] S.J. Knudsen. *Estimating Functions and Separate Inference*. Cand. Scient. Thesis, Århus University, 1999. <http://www.stat.sdu.dk/oldpublications/monographs.html>.
- [23] E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses (3rd ed.)*. Springer, 2005.
- [24] K.-Y. Liang and S.L. Zeger. Inference based on estimating functions in the presence of nuisance parameters. *Statistical Science*, 10(2) :158–172, 1995.
- [25] B. Liseo. The elimination of nuisance parameters. In D. Dey and C.R. Rao, editors, *Handbook of Statistics*, volume 25, pages 811–825. Elsevier, 2005.
- [26] G. Mathis. *Outils de Détection de Ruptures et de Diagnostic : Application à la Surveillance de Turbines à Gaz*. Ph.D. thesis, Rennes I University, 1994. <http://www.irisa.fr/bibli/publi/theses/1994/mathis/mathis.html>.
- [27] A. Monfort. Statistique. Cours, École Polytechnique, 2000. In French.
- [28] S.A. Murphy and A.W. van der Vaart. On profile likelihood (with discussion). *Journal American Statistical Association*, 95(450) :449–485, 2000.
- [29] J. Neyman. Optimal asymptotic tests of composite statistical hypotheses. In *The Harald Cramer Volume*, pages 213–234. Wiley, 1959.
- [30] N. Reid and D.A.S. Fraser. Likelihood inference in the presence of nuisance parameters. In *Proc. PhyStat 2003, SLAC, Stanford, CA*, pages 265–271, September 2003.

- 
- [31] T.A. Severini. Likelihood functions for inference in the presence of a nuisance parameter. *Biometrika*, 85 :507–522, 1998.
  - [32] T.A. Severini. *Likelihood Methods in Statistics*. Oxford Univ. Press, 2000.
  - [33] C.G. Small and D.L. McLeish. Projection as a method for increasing sensitivity and eliminating nuisance parameters. *Biometrika*, 76(4) :693–703, 1989.
  - [34] E. Spjøtvoll. On the optimality of some multiple comparison procedures. *Annals Statistics*, 21(3) :1486–1521, 1972.
  - [35] A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. American Mathematical Society*, 54 :426–482, 1943.
  - [36] Q. Zhang. Gaussian vector computations. Technical report, IRISA. Private Matlab toolbox.
  - [37] Q. Zhang and M. Basseville. Advanced numerical computation of  $\chi^2$ -tests for fault detection and isolation. In *5th Symp. Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, Washington, pages 211–216, June 2003.

## Table des matières

<b>Introduction</b>	<b>3</b>
<b>1 Inspection de la moyenne d'un vecteur Gaussien</b>	<b>4</b>
1.1 Préambule . . . . .	4
1.2 Notations et formules de base . . . . .	4
1.3 Détection . . . . .	5
1.4 Isolation . . . . .	6
1.4.1 Remarque : décomposition de Karhunen-Loeve . . . . .	7
1.4.2 Méthode de sensibilité . . . . .	7
1.4.3 Méthode de réjection (minmax) . . . . .	8
1.4.4 Méthode de Wald . . . . .	9
1.4.5 Méthode du multiplicateur de Lagrange . . . . .	10
1.4.6 Mixer réjection et sensibilité . . . . .	10
1.5 Procédures d'isolation des $n_\varphi$ pannes . . . . .	11
1.5.1 Par sensibilité . . . . .	11
1.5.2 Par réjection (minmax) . . . . .	11
1.6 Calculs numériques . . . . .	11
<b>2 Traitement des paramètres de nuisance</b>	<b>12</b>
2.1 Approche Bayésienne . . . . .	12
2.2 Vraisemblances marginale et conditionnelle . . . . .	12
2.3 Orthogonalité des paramètres . . . . .	13
2.4 Exploitation de propriétés d'invariance . . . . .	13
2.5 Approche minmax ou vraisemblance profil . . . . .	13
<b>3 Insensibilité au paramètre de nuisance</b>	<b>14</b>
3.1 Score efficace et information de Fisher . . . . .	14
3.1.1 Découplage (symétrique) par $\mathbf{F}$ -orthogonalité . . . . .	16
3.2 Fonction d'estimation, sensibilité et information de Godambe . . . . .	16
3.2.1 Insensibilité (asymétrique) . . . . .	18
3.2.2 $\mathbf{G}$ -orthogonalité . . . . .	20
3.2.3 Insensibilisation par projection . . . . .	20
3.3 Rôle des reparamétrisations . . . . .	21
<b>Références</b>	<b>22</b>