



**HAL**  
open science

## Contrôle non destructif, détection, courbes ROC

Michel Broniatowski, Hebert Gayard

► **To cite this version:**

Michel Broniatowski, Hebert Gayard. Contrôle non destructif, détection, courbes ROC. 2014. hal-01058750

**HAL Id: hal-01058750**

**<https://hal.science/hal-01058750>**

Preprint submitted on 9 Sep 2014

**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.

Université Pierre et Marie Curie

ISUP-LSTA

Contrôle non destructif, détection, courbes ROC

Michel Broniatowski, Hebert Gayraud

**2014**

## Résumé

---

Le contrôle non destructif (CND) est un ensemble de méthodes permettant de caractériser l'état d'intégrité de structures ou de matériaux sans avoir recours à leur dégradation. Ces vérifications sont généralement effectuées par un opérateur à l'aide de divers instruments dont font partie les transducteurs ultrasons et électromagnétiques. Ceux-ci se caractérisent par l'émission d'un signal interagissant avec la pièce à inspecter. Le signal réfracté vers le transducteur porte la signature des spécificités de la structure et des défauts y étant éventuellement présents. Ce dernier signal doit alors faire l'objet d'une interprétation permettant de prendre une décision quant à la présence ou non d'une anomalie. Cette procédure consiste généralement à comparer l'amplitude maximale du signal à la valeur d'un seuil fixé au préalable. Ainsi, on ne conclura à la présence d'un défaut que si l'amplitude dépasse ce seuil. L'objectif principal consiste à évaluer la performance de détection d'un capteur CND tandis que celui du stage entre dans le cadre de l'analyse des méthodes d'évaluation.

Le signal, résultant de la combinaison de phénomènes déterministes et aléatoires, rend l'approche probabiliste nécessaire à la réalisation de ce projet. Elle conduit à la définition et au calcul d'une probabilité de détection (POD) fonction de la grandeur caractéristique du défaut présent dans la pièce (probabilité conditionnelle). La première partie de cette étude est ainsi dédiée à l'analyse des modèles proposés par Alan Berens pour l'estimation de la POD. Si le modèle « Hit/Miss » permet de traiter les données d'inspections de types binaires « détections/non détections », le modèle « Signal/Réponse » s'applique quant à lui aux données composées directement de retours sur le signal. Dans chacun des cas, la procédure d'estimation se décompose en deux étapes : modélisation appropriée aux données puis dérivation de la POD. Nous examinerons les pièges à éviter dans le cadre de l'utilisation de ces modèles. N'ayant à disposition qu'une estimation de la POD malheureusement soumise aux aléas d'échantillonnage, les experts en CND sont attachés à la donnée du niveau de confiance associé. Cela se traduit par la détermination d'une bande de confiance dont nous verrons le principe général de calcul. Les industriels s'intéressent d'autre part à la taille de défaut  $a_{0,9}$  qui engendre une probabilité de détection de 90%. Cette taille constitue en effet un bon indicateur de la performance générale d'un détecteur.

Il arrive que le signal se caractérise par un bruit important. Cette situation peut avoir pour origine un manque d'expérience de l'opérateur exécutant le contrôle ou bien encore, être engendrée par les caractéristiques même de la pièce inspectée (bruit de pièce). Un niveau de bruit trop élevé est gênant car il peut générer des détections alors qu'aucun défaut n'est présent dans la structure (fausses alarmes). Or ces erreurs viennent contrebalancer le niveau de POD, qui n'est alors plus suffisant dans l'évaluation de la performance. Nous justifierons par conséquent la nécessité d'introduire la notion de probabilité de fausse alarme (PFA). Les courbes ROC se présenteront dès lors comme un outil de visualisation de la performance intrinsèque du capteur à dissocier bruit et signal. Nous en étudierons les principales caractéristiques ainsi que les diverses méthodes d'estimation. Concernant ce dernier point, on s'attardera particulièrement sur les procédures autorisant la prise en compte d'une ou plusieurs covariable(s). Ces méthodes permettent en effet par exemple de mesurer l'impact de la grandeur caractéristique du défaut sur la performance et de se replacer ainsi dans un cadre plus proche de la problématique originelle. Nous étudierons enfin les concepts de droites d'isoperformance et d'enveloppe convexe des courbes ROC (ROCCH) et leurs applications. Ces notions ont été introduites de façon à prendre en compte des facteurs venant moduler l'appréciation de la performance globale d'un capteur. Les coûts associés aux erreurs de détection et la fréquence des défauts dans la population générale des pièces (prévalence) en sont de bons exemples.

Non-destructive evaluation (NDE) is a set of methods aimed at characterizing the integrity state of structures or materials without resorting to their destruction. These checkouts are generally performed by an operator thanks to various devices including ultrasonic or electromagnetic transducers. They are characterized by a signal emission interacting with the inspected piece. The refracted signal towards the transducer is modulated by structure specificities and especially by defects potentially included. This latter signal has to be interpreted in order to make a decision about the presence or absence of an anomaly. This procedure consists in comparing the signal maximal amplitude with a threshold set beforehand. Thereby, one will conclude to the presence of a defect only if the maximal amplitude exceeds the threshold value. The main goal consists in the overall performance assessment of an NDE detector whereas those of this internship fall within the framework of the captor evaluation methods analysis.

Signal, arising from the combination of deterministic and random phenomena, makes the probabilistic approach necessary for the execution of such a project. It leads to the definition and calculation of a probability of detection (POD) which is function of defect's characteristic size (conditional probability). The first part of this study is dedicated to the analysis of models suggested by Alan Berens for the POD estimation. If the Hit/Miss model enables to deal with binary inspections data "detection/no detection", the "Signal/Response" model applies to data composed of experience feedback about the signal itself. In each case, the estimation procedure is broke down into two steps: appropriate data modeling and POD derivation. We will examine pitfalls which have to be avoided when using these types of models. However, since NDE experts only have an access to an estimation of the POD unfortunately suffering from sampling hazard, they are wedded to its associated level of confidence. This request results in the determination of a confidence band whose general principle of calculation will be presented. Moreover, for practical purposes, industrialists are interested in the defect size  $a_{0,9}$  involving a probability of detection equal to 90%. This size is a good indicator of the detector overall performance.

It happens to retrieve a signal characterized by an important noise. Such an event can be explained by the operator lack of experience or by some characteristics of the inspected piece. A too high level of noise is annoying because it can be at the origin of detections while no defect is present in the structure (false alarms). These errors counterbalance the POD level which is no longer sufficient in describing the captor performance. We therefore try to give proof of the need to introduce the notion of probability of false alarm (PFA). In this context, ROC curves turn out to be a complete visualization tool of the intrinsic captor performance in discriminating noise and signal. We will study their main characteristics as well as their estimation. We will particularly dwell upon procedures enabling to take into account one or several covariables since they give the possibility to assess the impact of the defect characteristic size on the overall performance for instance. This way, we will set the analysis back to the original context (Berens). We will finally study concepts of isoperformance lines and of convex ROC hull (ROCCH), and their associated applications. These notions have been introduced to consider some factors effect in the modulation of the performance assessment. Defects frequency (prevalence) and costs linked to detection errors are both good example of such factors.

# Sommaire

---

I. Le contrôle non destructif-CND	
I.1. Description générale	6
I.2. Les différentes méthodes utilisées en CND	6
II. Evaluation des méthodes d'inspection	8
II.1. Probabilité de détection (POD)	8
II.1.A. Une approche probabiliste	8
II.1.B. Facteurs dans le cadre CND	8
II.2. Le modèle POD de Berens	8
II.2.A. Le modèle succès/échec	9
a. Estimation empirique de la POD	9
b. Estimation paramétrique de la POD	9
i. Nécessité d'un modèle paramétrique	9
ii. Un modèle logistique	10
iii. Reparamétrage du modèle logistique	10
iv. Estimation des paramètres – Maximum de vraisemblance	11
II.2.B. Le modèle signal/réponse	12
a. Description générale – Un modèle Tobit à double censure	12
b. Estimations des paramètres – Maximum de vraisemblance	13
c. Expression de la POD	13
II.2.C. Quelques remarques concernant à le modèle de Berens	15
a. Erreurs multiplicatives et limitation du modèle	15
b. L'hypothèse de linéarité dans le modèle signal/réponse – Plan d'expérience	15
II.2.D. Bande de confiance pour la POD	17
a. Problématique industrielle	17
b. Bandes de confiance non-paramétriques - limitations	17
c. Bande de confiance pour les fonctions de répartition de variables continues	18
i. Principe général pour une bande de confiance bilatérale	18
ii. Cas des modèles de translation - changement d'échelle	19
• Région de confiance pour les paramètres (R)	19
• Bornes de confiance pour la fonction de répartition – Détermination géométrique	20
• Bornes de confiance pour la fonction de répartition – Détermination à l'aide d'un Lagrangien	21
iii. Bande de confiance unilatérale	21
• Intérêt d'une bande de confiance unilatérale	21
• Méthode de calcul – bande de confiance unilatérale supérieure	21
II.3. Courbes ROC	22
II.3.A. Contexte	22
II.3.B. Probabilité de détection et probabilité de fausse alarme	22
II.3.C. Les courbes ROC	23
a. Définition et premières propriétés	23
b. Estimations	24
i. Cadre commun	25
• Estimation non-paramétrique – Principe général	25
• Estimation paramétrique – Remarques générales	25

• Estimation semi-paramétrique – Remarques.....	27
ii. Cadre de la Yes-No method - Données binaires .....	27
• Estimation non-paramétrique.....	28
• Estimation semi-paramétrique.....	28
iii. Cadre de la Rating Method - Données ordinales .....	29
• Estimation non-paramétrique.....	30
• Estimation semi-paramétrique.....	30
iv. Cadre de la Scoring Method - Données continues :.....	32
• Estimation non-paramétrique.....	32
• Estimation semi-paramétrique.....	32
II.3.D. Classification robuste dans le cadre des environnements imprécis – ROCCH .....	38
a. Courbes ROC – isolement de la performance pure .....	38
b. Projection de l’information analytique de décision sur l’espace ROC.....	39
c. Identification des traducteurs potentiellement optimaux – ROCCH .....	39
d. Sensibilité des résultats aux imprécisions sur l’environnement.....	40
e. Robustesse – Hybrides de classification basés sur la ROCCH .....	41
II.3.E. Etiquetage incertain ou impossible .....	42
a. Estimation semi-paramétrique bayésienne.....	42
i. Extension de la méthode d’Erkanli & Al (2006) .....	42
ii. Méthode de Choi & Al 2006 .....	42
b. Estimation non-paramétrique bayésienne.....	43
.....	<b>Error! Bookmark not defined.</b>
III. Conclusion .....	45
Annexes .....	47
Principe du maximum de vraisemblance pour les lois continues .....	47
Algorithme de Newton-Raphson et algorithme du scoring – équations de vraisemblance.....	47
Rappels sur la loi normale, logistique, log-normale et log-logistique .....	48
Propriété de la matrice de Fisher d’un modèle de translation –changement d’échelle .....	49
Démonstrations des propriétés de la courbes ROC .....	50
Bibliographie.....	52

## I.1. Le contrôle non destructif – CND

### I.1.A. Description générale

Le contrôle non destructif est [1,2,3] un ensemble de méthodes permettant de caractériser l'état d'intégrité de structures ou de matériaux, sans les dégrader, lorsqu'ils sont :

- **Soit en cours de production :**  
Les pièces sortant des fonderies contiennent toujours quelques défauts.
- **Soit en cours d'utilisation :**  
Les conditions de travail des pièces durant cette période favorisent l'apparition de défauts de fatigue.
- **Soit en cours de maintenance :**  
Contrôles a posteriori généralement nécessaires avant la remise en service de la pièce ou du système dans lequel elle est présente.

Le diagnostic d'intégrité de la pièce passe par l'estimation d'un ensemble de paramètres la caractérisant. Ce procédé d'estimation correspond à ce que l'on appelle une étape d'« inversion » dans la démarche CND, qui dans son ensemble peut être décrite par le schéma de la Figure 1.

Le système CND émet d'abord des signaux en direction de la cible à inspecter. Ces signaux se voient modifiés lorsqu'ils interagissent avec la cible et plus particulièrement avec le défaut présent. Ce dernier laisse une empreinte, « sa signature », dans le signal renvoyé au système CND. Il est donc fonction des paramètres du défaut. L'inversion consiste à retrouver la valeur de ces paramètres à partir du signal, ou tout du moins, à en donner une estimation suffisamment précise.

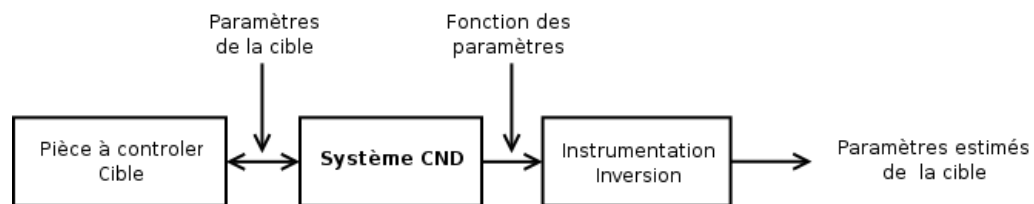


Figure 1 : La démarche CND

Avant leur mise en service, les méthodes de CND doivent être calibrées. Des essais empiriques, souvent onéreux et pendant lesquels les pièces doivent fréquemment être détruites sont généralement menés dans le but de :

- déterminer ou de vérifier expérimentalement les valeurs acceptables (respect des normes imposées par exemple) pouvant être prises par le ou les paramètre(s) caractérisant les défauts : taille, profondeur, orientation, etc.
- régler la sensibilité des outils de CND de façon à s'assurer qu'ils détectent *efficacement* un défaut dont les grandeurs caractéristiques vérifient certaines propriétés.

Les méthodes CND sont classiquement utilisées dans les secteurs industriels produisant des pièces à coûts de production élevés et en faible quantité, ou fabriquant des pièces dont la fiabilité de fonctionnement est un élément critique.

### I.1.B. Les différentes méthodes utilisées en CND

Il existe de nombreuses méthodes de contrôle non destructif. Chacune possède bien évidemment ses avantages et ses inconvénients fonctions de l'efficacité du diagnostique, du coût, des types de défauts à détecter, de la géométrie de la pièce à inspecter, de l'accessibilité de la pièce, etc.

Parmi ces techniques, on peut citer l'examen visuel, le ressuage, les rayons X, les ultrasons, l'émission acoustique, la thermographie, la magnétoscopie, l'étanchéité, les courants de Foucault. Plus récemment, des méthodes couplées sont aussi apparues.

Le LIST a axé ses recherches sur les dispositifs CND basés sur les ultrasons, les courants de Foucault et les rayons X.

- **Les ultrasons :**

Des trains d'onde ultrasonore sont émis par un transducteur, et se propagent dans la pièce à contrôler. Ces ondes se reflètent sur le fond de la pièce et sur les défaut(s) puis retournent sur le transducteur qui les réceptionne. Les signaux, portant l'emprunte des défauts rencontrés, sont interprétés. Cette interprétation donne la possibilité de localiser les défauts dont une représentation tridimensionnelle est possible à conditions de contrôler différentes surfaces de la pièce.

Cette méthode est relativement lente car nécessite un balayage mécanique exhaustif de la pièce.

- **Les courants de Foucault :**

Ils correspondent aux courants électriques créés dans un milieu conducteur, soit par la variation au cours du temps d'un champ magnétique traversant ce milieu, soit par le déplacement de ce milieu dans un champ magnétique constant. En CND, un champ magnétique variable est généré à l'aide d'une bobine dans laquelle circule un courant électrique alternatif à fréquence variable. Cette bobine, placée à proximité de la pièce à inspecter, y produit des courants induits : les courants de Foucault. Ces courants induits génèrent à leur tour un champ magnétique.

La présence d'un défaut engendre des variations de conductivité électrique à l'intérieur de la pièce, perturbant les courants de Foucault. Le champ magnétique qu'ils génèrent se voit lui aussi perturbé, et ses variations peuvent être mesurées par l'intermédiaire de la bobine émettrice. Celle-ci fait donc aussi office de récepteur.

Ce dispositif peut être automatisé ce qui a l'avantage d'être rapide et fonctionnel même en environnement dangereux.

- **Les rayons X et gamma :**

Dans le cadre du CND, les rayons X et gamma sont utilisés dans le but d'obtenir des images directement exploitables de la pièce à inspecter. Une interprétation correcte des radiographies requière cependant un haut niveau d'expertise de l'opérateur.



## II. Evaluation des méthodes d'inspection

Encore aujourd'hui, l'appréciation de la fiabilité des techniques de contrôle non destructif repose en grande majorité sur des retours d'expériences réalisées dans des conditions fixées, et dont les données résultantes sont analysées par l'intermédiaire des méthodes classiques de statistiques descriptives et par l'établissement d'une **probabilité de détection (POD)**.

Cependant, on l'a vu, ces situations expérimentales ont généralement le désavantage d'être coûteuses et de réclamer des délais non négligeables dans l'obtention des résultats. Les progrès et les innovations dans la conception des pièces à inspecter multiplient les configurations à étudier et nécessitent par la même occasion, une ré-estimation systématique de la POD. Dans ce contexte, mener à bien l'évaluation de la fiabilité des techniques CND tout en respectant l'ensemble des contraintes est complexe. Mais l'évolution des moyens informatiques de simulation laisse entrevoir des solutions.

Un traitement statistique correct nécessite toujours à la fois une quantité de données suffisante, et une bonne répartition de celles-ci. Pour satisfaire ces deux critères tout en conservant un coût raisonnable, une idée consiste à introduire des données simulées afin de « compléter » les informations expérimentales recueillies. La répartition de l'ensemble de ces données devra se baser sur des plans expérimentaux optimaux préalablement établis.

### II.1. Probabilité de détection (POD)

Par probabilité de détection ou POD (Probability Of Detection), on entend de façon générale la probabilité que le système d'inspection détecte un défaut de taille donnée. Mais pourquoi doit-on associer une probabilité à un tel évènement ?

#### II.1.A. Une approche probabiliste

Les instruments de CND basés sur les ultrasons ou les courants de Foucault produisent un signal qu'il faut interpréter. Une amplitude importante de ce signal suggère la présence d'une anomalie dans la structure inspectée. L'amplitude est considérée comme « importante » à partir du moment où elle dépasse un seuil défini au préalable. La mise en place d'une telle précaution repose sur le fait que le signal est toujours *bruité*. Le seuil doit être positionné de façon à discerner le mieux possible les situations où l'on est en présence de bruit seul, de celles où le bruit s'accompagne d'un signal.

La majorité de la variabilité inhérente au signal peut être expliquée par l'effet de divers facteurs que l'on classe en deux grandes catégories : **les facteurs fixes et les facteurs aléatoires**. Les facteurs fixes peuvent être contrôlés pendant le processus d'inspection, tandis que les facteurs aléatoires correspondent aux facteurs qui ne sont pas bien maîtrisés, pour lesquels on ne possède que peu d'information ou qui repose sur des phénomènes physiques aléatoires par nature. Ils peuvent donc éventuellement varier d'une expérimentation à une autre. Dans le cas où l'on souhaite mesurer l'influence d'un facteur sur le signal, il est nécessaire que celui-ci soit un facteur fixe. Il faut cependant noter qu'il restera toujours une part de variabilité « incompressible » que l'on ne pourra expliquer, attribuable à des facteurs non identifiés. Ce que l'on a désigné par « bruit » dans le paragraphe précédent peut correspondre soit à de la variabilité incompressible, soit à de la variabilité incompressible associée à de la variance provenant de facteurs aléatoires.

La présence et la structure du bruit montre que l'on travaillera toujours dans un contexte incertain. L'approche probabiliste se révèle donc indispensable si l'on souhaite tirer des conclusions de portée générale.

#### II.1.B. Facteurs dans le cadre CND

Le signal traité par les instruments de CND étant le résultat de phénomènes complexes, la recherche des facteurs influençant le signal nous conduit à une liste assez large. De façon non exhaustive, on distinguera les facteurs liés [4] :

- aux caractéristiques du détecteur (composants, électronique,...)
- aux caractéristiques de la pièce inspectée (géométrie, matériaux, rugosité,...)
- aux caractéristiques du défaut (taille, position, densité, forme, composition,...)

On pourrait aussi ajouter les facteurs « extérieurs » associés à l'environnement hors détecteur-pièce-défaut, comme ceux liés à l'opérateur effectuant l'inspection.

Le classement entre aléatoire et fixe peut varier selon les circonstances. Par exemple, la morphologie du défaut peut être considérée comme un facteur fixe si les expériences sont menées sur le même défaut.

Parmi les facteurs aléatoires on trouvera classiquement tous les détails de microstructure de la pièce inspectée (bruit de structure), la position du défaut par rapport à la sonde du détecteur, certaines composantes de la morphologie du défaut, etc.

## II.2. Le modèle POD de Berens

Dans le problème d'évaluation de la fiabilité des détecteurs en CND, la « taille du défaut » est un facteur que l'on souhaite relier au signal.

A taille de défaut fixée, de nombreux autres paramètres relatifs soit au défaut lui-même (forme, orientation, propriété élastique, conductivité,...), soit au reste de l'environnement (détecteur, pièce inspectée) peuvent varier lors des expérimentations. Cela conduit à

observer des résultats de détection différents à taille de défaut fixé, voir même lors de mesures répétées sur un même défaut. Tous ces facteurs conduisent ainsi à une caractérisation probabiliste de la fiabilité des inspections.

Alan P. Berens, chercheur à l'université de Dayton aux Etats-Unis a proposé deux modèles [5] permettant d'estimer une probabilité de détection selon le genre de données recueillies. Celles-ci peuvent en effet être de **type succès/échec** (« hit/miss ») ou de **type signal/réponse** (« signal/response »).

A l'origine, les données d'inspections étaient enregistrées seulement sous forme binaire, selon la présence ou non d'un défaut. Elles ont donc été qualifiées comme étant des *données succès/échec* et leur analyse s'est développée à partir de la caractérisation binomiale. Plus tard, il a été mis en évidence que le signal du système CND à partir duquel étaient prises les décisions succès/échec contenait d'avantage d'information. Les données de signal brut sont donc dorénavant conservées et classées parmi les données de type signal/réponse.

## II.2.A. Le modèle succès/échec

Dans le cadre succès/échec, on observe les réalisations d'une variable binaire  $\delta$  qui indique la conclusion de l'expérience :

- Si  $\delta = 1$ , la détection du défaut est signalée.
- Si  $\delta = 0$ , aucune détection de défaut n'est signalée.

Cette variable  $\delta$  est implicitement liée à la valeur du signal  $S$  ainsi qu'au niveau du seuil de détection imposé  $s^{dec}$  par le détecteur. Elle peut être vue comme une fonction indiquant si le seuil a été dépassé ou non :

$$\delta = 1_{\{S(a) \geq s^{th}\}} = \begin{cases} 1 & \text{si } S(a) \geq s^{dec} \\ 0 & \text{sinon} \end{cases} \quad (1)$$

On s'appuie alors sur cette variable pour proposer une définition de la POD :

<b>Définition</b>	La POD dans le cas des données <i>hit-miss</i> se définit comme étant la probabilité de détection d'un défaut sachant que sa taille est $a$ ,
	$POD(a) = P(\delta = 1/a)$

### a. Estimation empirique de la POD

La variable  $\delta/a$  («  $\delta$  sachant  $a$  ») décrit l'issue d'une expérience sachant que la taille du défaut est  $a$ . Elle est distribuée selon une loi binomiale de paramètre  $p_a = POD(a)$ . L'espérance d'une telle loi étant égale à son paramètre, on a :  $E_a(\delta) = p_a = POD(a)$ . Adoptons la notation  $\delta/a = \delta^a$ . La méthode des moments permet de donner une estimation de la POD :

$$\widehat{POD}(a) = \hat{E}(\delta / a) = \frac{1}{n_a} \sum_{i=1}^{n_a} \delta_i(a) = \frac{\text{Nombre de détections}(a)}{\text{Nombre de tentatives}(a)}$$

- où :
- $\delta_i(a)$  sont les réalisations de  $\delta$  pour les défauts de taille  $a$ .
  - $n_a$  est le nombre total de défaut de taille  $a$ .

### b. Estimation paramétrique de la POD

#### i. Nécessité d'un modèle paramétrique

Si l'on effectue une série de  $N_1$  inspections sur une pièce contenant un défaut de taille  $a$ , on obtient un ensemble de réalisations  $(\delta_i^a)_{i=1, \dots, N_1}$  indépendantes de la variable  $\delta^a$ . Supposons que ces résultats d'expériences comportent  $D_1$  détections. La méthode des moments permet de donner une première estimation empirique de la POD qui correspond au rapport entre le nombre de détections et le nombre total d'inspections :

$$\widehat{POD}_{emp}^1(a) = \hat{E}(\delta^a) = \frac{D_1}{N_1} = \frac{1}{N_1} \sum_{i=1}^{N_1} \delta_i^a$$

Il est ensuite possible de réitérer ce processus sur une deuxième pièce contenant toujours un défaut de taille  $a$ . On aura dans ce cas  $N_2$  résultats d'inspections  $(\delta_i^a)_{i=1, \dots, N_2}$  dont  $D_2$  se concluront par une détection. On estime alors la POD de la même façon :

$$\widehat{POD}_{emp}^2(a) = \frac{D_2}{N_2} = \frac{1}{N_2} \sum_{i=1}^{N_2} \delta_i^a$$

Cette procédure répétée sur  $k$  pièces fournira ainsi  $k$  estimations  $\widehat{POD}_{emp}^1(a), \dots, \widehat{POD}_{emp}^k(a)$  de la POD.

La variabilité du signal pour une taille de défaut donnée, s'expliquant par l'effet des facteurs aléatoires non contrôlés (le défaut, bien que de taille identique dans tous les cas, peut être orienté de façon légèrement différente dans chacune des pièces par exemple), engendre la plupart du temps des valeurs différentes pour chacune de ces estimations. Cette remarque reste vraie quelle que soit la taille du défaut considérée (voir Figure 4).

Dans un tel contexte, on obtient plusieurs estimations de la POD pour une même taille de défaut, ce qui n'est pas acceptable d'un point de vue moral. L'objectif consiste alors à ajuster une courbe de régression, reliant la valeur moyenne de la POD à la taille du défaut. La fonction recherchée devra vérifier les propriétés suivantes :

- Croissance avec la taille du défaut ou son logarithme
- A valeurs dans l'intervalle [0; 1] puisqu'on modélise une probabilité

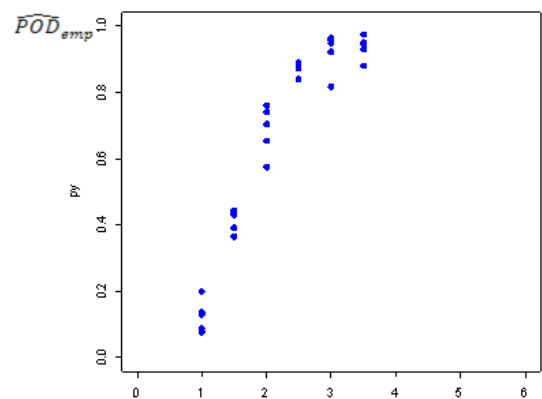


Figure 2 : POD estimées empiriquement pour diverses tailles de défaut

### ii. Un modèle logistique

Un ensemble de fonctions vérifiant les propriétés demandées ci-dessus est constitué des fonctions de répartition des lois de probabilité. On utilisera en particulier celles associées à la loi logistique ou log-logistique. Dans ce cas, on parle en général de **modèles logit**. Si  $a$  est la taille du défaut, alors un tel modèle conduit à une expression de la POD de la forme :

$$\text{Modèle logistique : } POD(a) = \frac{e^{\alpha+\beta \cdot a}}{1 + e^{\alpha+\beta \cdot a}} = \frac{1}{1 + e^{-(\alpha+\beta \cdot a)}} \quad (2)$$

$$\text{Modèle log-logistique : } POD(a) = \frac{e^{\alpha+\beta \ln(a)}}{1 + e^{\alpha+\beta \ln(a)}} = \frac{1}{1 + e^{-(\alpha+\beta \ln(a))}} \quad (3)$$

On peut remarquer qu'en utilisant la fonction logit, le modèle devient linéaire en la taille du défaut ou son logarithme :

$$\text{logit}(POD(a)) = \log\left(\frac{POD(a)}{1 - POD(a)}\right) = \alpha + \beta \cdot a \quad \text{ou} \quad \alpha + \beta \ln(a)$$

La quantité  $\frac{POD(a)}{1 - POD(a)} = \frac{P(\delta=1/a)}{P(\delta=0/a)}$  est appelée l'*odds-ratio* ou le *rapport des chances*.

### iii. Reparamétrage du modèle logistique

On remarquera qu'étant donné la proximité de la loi logistique et de la loi normale, un choix alternatif pour la POD aurait pu s'appuyer sur la fonction de répartition d'une loi Normale. On aurait alors obtenu un **modèle probit** :

$$POD(a) = \Phi\left(\frac{\ln(a) - \theta}{\zeta}\right) = \text{f.r. d'une log-normale}(\theta, \zeta^2)$$

La fonction  $\Phi$  n'admettant pas de forme analytique explicite, il reste plus simple de travailler avec la fonction de répartition de la loi logistique. Cependant, en s'appuyant sur l'approximation,

$$\Phi\left(\frac{x - \mu}{\sigma}\right) \approx \frac{1}{1 + e^{-\frac{\pi}{\sqrt{3}} \frac{(x - \mu)}{\sigma}}} \quad \forall x \in \mathbb{R} \quad (4)$$

il est possible de proposer le nouveau paramétrage suivant :

$$POD(a) = \frac{1}{1 + e^{-\frac{\pi}{\sqrt{3}} \left(\frac{\ln(a) - \mu}{\sigma}\right)}} \quad \text{où} \quad \begin{cases} \mu = -\frac{\alpha}{\beta} \\ \sigma = \frac{\pi}{\beta \sqrt{3}} \end{cases} \quad (5)$$

L'approximation (4) provient de la proximité entre les lois logistique et normale de même moyenne et variance. On peut en voir une illustration sur la Figure 5.

Afin retrouver le paramétrage (5), adoptons le style de notation suivant :

$$\text{Nom\_de\_la\_loi}_{\text{paramètres}}(\text{espérance, variance})$$

Considérons dès lors deux variables telles que (on se réfèrera aux Annexes pour des rappels concernant les lois logistique, log-logistique, normale et log-normale) :

- $X_1$  de loi logistique de paramètre  $\mu$  et  $s$  :

$$X_1 \sim \text{logistique}_{(\mu,s)} \left( \mu, \frac{s^2 \pi^2}{3} \right)$$

- $X_2$  de loi normale de paramètre  $\mu$  et  $\sigma^2$  :

$$X_2 \sim \text{normale}_{(\mu,\sigma)}(\mu, \sigma^2)$$

Supposons qu' $X_1$  et  $X_2$  ont la même espérance  $\mu$ . Si l'on souhaite de plus qu' $X_1$  et  $X_2$  aient les mêmes variances, il est nécessaire que :

$$V(X_1) = V(X_2) \Leftrightarrow \frac{s^2 \pi^2}{3} = \sigma^2 \Leftrightarrow s = \frac{\sqrt{3}}{\pi} \sigma$$

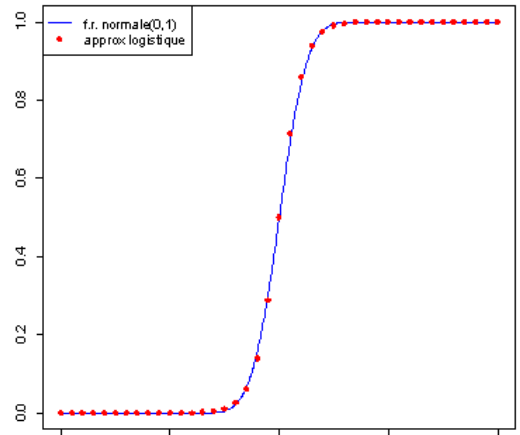


Figure 3 : Approximation (4) pour  $\mu = 0$  et  $\sigma = 1$

Ainsi, la fonction de répartition d'une variable  $X_3 \sim \text{Logistique}_{(\mu,\sigma)}(\mu, \sigma^2)$  de même moyenne et variance qu'une loi Normale $_{(\mu,\sigma)}(\mu, \sigma^2)$  vérifie :

$$F_{X_3}(x) = \frac{1}{1 + e^{-\frac{\pi}{\sqrt{3}} \frac{(x-\mu)}{\sigma}}}$$

et l'on retrouve bien l'expression (5).

#### iv. Estimation des paramètres – Maximum de vraisemblance

Même s'il est possible d'estimer les paramètres par régression linéaire, l'estimation par maximum de vraisemblance se révèle plus intéressante dans le sens où les résultats seront meilleurs en cas d'échantillon relativement restreint. Les résultats dans le cadre hit-miss sont constitués de l'observation d'une suite de 0 et de 1 correspondant à des réalisations de  $\delta$ . Supposons que l'on dispose de  $n$  pièces contenant chacune un défaut de taille donnée et que l'on effectue  $k$  inspections sur chacune d'entre elles. Notons  $p$  la probabilité d'observer  $\delta = 1$ . Alors  $(1 - p)$  est la probabilité d'observer  $\delta = 0$  et la vraisemblance associée à ces données est celle d'une loi de Bernouilli :

$$L(\mu, \sigma) = \prod_{i=1}^n \prod_{j=1}^k p_i^{\delta_{ij}} (1 - p_i)^{1 - \delta_{ij}}$$

où : 

- Les paramètres  $\mu$  et  $\sigma$  apparaissent à travers l'expression de  $p_i$ , la probabilité de détecter un défaut de taille  $a_i$ , soit :

$$p_i = \text{POD}(a_i) = \frac{1}{1 + e^{-\frac{\pi}{\sqrt{3}} \left( \frac{\ln(a_i) - \mu}{\sigma} \right)}}$$

- $\delta_{ij} = 1$  si la  $j^{\text{ième}}$  inspection du défaut  $i$  a produit une détection, et  $\delta_{ij} = 0$  dans le cas contraire.

Les paramètres sélectionnés maximiseront la vraisemblance  $L$ . La fonction logarithme étant monotone croissante, il est équivalent de chercher à maximiser  $\ln(L)$ . Les solutions seront donc obtenues en résolvant les équations de vraisemblance ci-après à l'aide de l'algorithme de Newton-Raphson :

$$\frac{\partial \ln(L)}{\partial \mu} = 0 \quad \text{et} \quad \frac{\partial \ln(L)}{\partial \sigma} = 0$$

Le choix des valeurs initiales est crucial pour obtenir la convergence. Berens propose à cet effet des estimateurs de la moyenne  $\mu$  et de la variance  $\sigma$  basés sur des statistiques d'ordre (échantillon ordonné) :

$$\left\{ \begin{array}{l} \mu_{init} = \ln(a_{(n)}) - \frac{\ln(a_{(1)}) \hat{p}_1^{emp}}{2} - \frac{1}{2} \sum_{i=1}^n [\ln(a_{(i)}) - \ln(a_{(i-1)})] [\hat{p}_i^{emp} - \hat{p}_{i-1}^{emp}] \\ \sigma_{init} = \left\{ \ln(a_{(n)})^2 - \frac{\hat{p}_1^{emp}}{2} \ln(a_{(1)})^2 - \mu_{init}^2 - \frac{1}{2} \sum_{i=2}^n [\ln(a_{(i)})^2 - \ln(a_{(i-1)})^2] [\hat{p}_i^{emp} - \hat{p}_{i-1}^{emp}] \right\}^{\frac{1}{2}} \end{array} \right.$$

où : 

- $\ln(a_{(1)}) < \dots < \ln(a_{(n)})$  sont les statistiques d'ordres associées aux  $\ln(a_i)$  pour  $i=1, \dots, n$

- $\hat{p}_i^{emp}$  = pourcentage de détection associée à la taille de défaut  $a_{(i)}$

Ces différentes étapes terminées, on obtient une estimation des paramètres du modèle et la POD modélisée se caractérise par une forme sigmoïdale :

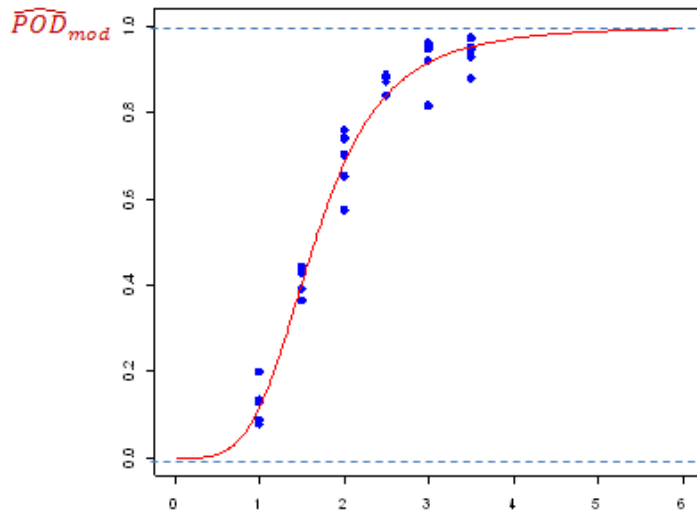


Figure 4 : En rouge, la POD estimée à l'aide du modèle Hit/Miss - En bleu, estimations empiriques de la POD

## II.2.B. Le modèle signal/réponse

### a. Description générale – Un modèle Tobit à double censure

Pour les systèmes CND reposant sur les ultrasons et les courants de Foucault, la réponse mesurée peut être un pic de voltage. On suppose que cette réponse est enregistrée sous la forme d'un paramètre  $\hat{a}$  corrélé à la taille du défaut (*attention* :  $\hat{a}$  ne correspond pas à une estimation).

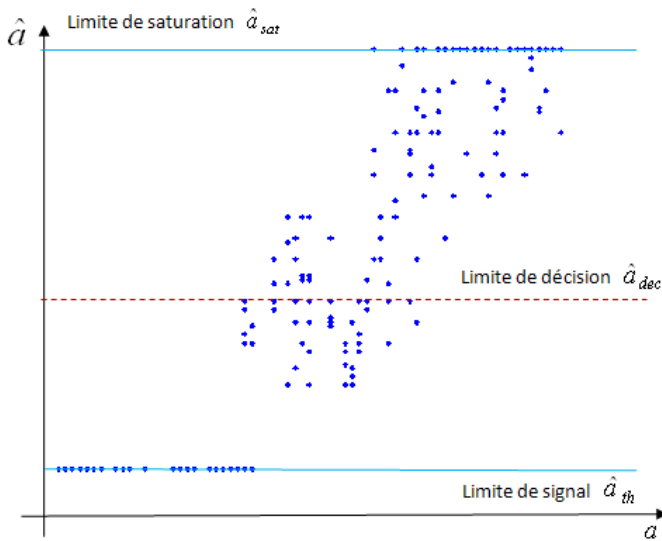


Figure 5 : Observations de signal - Courant de Foucault

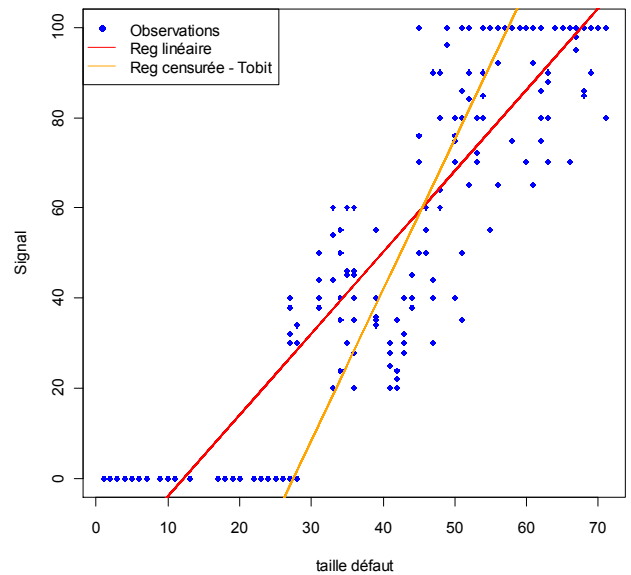


Figure 6 : Régression linéaire contre régression censurée

La Figure 5 contient un exemple de valeurs réelles relevées lors d'un contrôle non destructif en courant de Foucault. Certaines amplitudes, trop faibles, ont été ramenées à la valeur  $\hat{a}_{th}$  tandis que les mesures dépassant les possibilités d'affichage de l'écran de l'instrument d'inspection, trop élevées, ont été ramenées à  $\hat{a}_{sat}$ . Ces derniers relevés correspondent à des observations censurées. On déclare le défaut détecté uniquement lorsque le signal dépasse la limite (seuil) de décision  $\hat{a}_{dec}$ .

Sur ce graphique, on remarque aussi que dans l'intervalle  $]\hat{a}_{th}; \hat{a}_{dec}[$ , une relation linéaire entre signal et dimension du défaut semble acceptable. En première intention, on serait tenté d'effectuer une régression linéaire classique. Toutefois, les points ramenés aux limites de saturation la rendraient caduque. Pour comprendre pourquoi, il faut rappeler les hypothèses sur lesquelles elle est basée. Pour rester dans le cadre qui nous intéresse ici, on supposerait que pour *toutes les valeurs de a* :

$$\hat{a} = \beta_0 + \beta_1 a + \delta, \text{ ou bien, } \ln(\hat{a}) = \beta_0 + \beta_1 \ln(a) + \delta$$

avec  $\delta$  correspondant aux fluctuations distribuées comme une loi Normale de moyenne nulle et de variance  $\sigma_\delta^2$  indépendante de  $a$ . Cette modélisation reviendrait à dire que,

$$\hat{a} \sim \mathcal{N}(\beta_0 + \beta_1 a, \sigma_\delta^2) \quad \text{ou bien que} \quad \ln(\hat{a}) \sim \mathcal{N}(\beta_0 + \beta_1 \ln(a), \sigma_\delta^2), \quad \forall a \in [0; +\infty[$$

La distribution du signal se caractérise ainsi par une densité continue. Or sous cette hypothèse, la probabilité de l'évènement consistant à observer plusieurs fois la même valeur de signal est nulle. Il y a donc a priori une incompatibilité entre le modèle classique de régression et les observations répétées que l'on obtient en pratique.

Afin de contourner ce problème, Berens a proposé un modèle de type **Tobit à double censure**, aussi connu sous le nom de *modèle de Rosett* ou encore de *régression censurée*. (Illustration : Figure 6).

Il présente l'avantage de déplacer la continuité de la loi sur une variable latente (cachée)  $\hat{a}^*$  et d'être ainsi en accord avec les multiples observations saturées de  $\hat{a}$ . Si l'hypothèse de linéarité est toujours présente, elle se limite dorénavant à l'intervalle  $]\hat{a}_{th}; \hat{a}_{dec}[$  et concerne  $\hat{a}^*$ . La variance  $\sigma_\delta^2$  des erreurs sera encore supposée indépendante de  $a$  :

$$\hat{a} = \begin{cases} \hat{a}_{th} & \text{si } \hat{a}^* \leq \hat{a}_{th} \\ \hat{a}^* & \text{si } \hat{a}_{th} \leq \hat{a}^* \leq \hat{a}_{sat} \\ \hat{a}_{sat} & \text{si } \hat{a}^* \geq \hat{a}_{sat} \end{cases} \quad \text{où } \hat{a}^* = \beta_0 + \beta_1 a + \delta \text{ avec } \delta \sim \mathcal{N}(0, \sigma_\delta^2) \quad (6)$$

Les paramètres du modèle sont  $\beta_0$ ,  $\beta_1$  et  $\sigma_\delta$ . Ils doivent être estimés à l'aide des données recueillies.

De la même façon, on proposera le modèle ci-dessous concernant le logarithme du signal. Celui-ci présente un intérêt lorsque les données présentent une non-linéarité. Une transformation logarithmique leur étant alors en effet appliquée pour tenter de les redresser.

$$\ln(\hat{a}) = \begin{cases} \ln(\hat{a}_{th}) & \text{si } \ln(\hat{a}^*) \leq \ln(\hat{a}_{th}) \\ \ln(\hat{a}^*) & \text{si } \ln(\hat{a}_{th}) \leq \ln(\hat{a}^*) \leq \ln(\hat{a}_{sat}) \\ \ln(\hat{a}_{sat}) & \text{si } \ln(\hat{a}^*) \geq \ln(\hat{a}_{sat}) \end{cases} \quad \text{où } \ln(\hat{a}^*) = \beta_0 + \beta_1 \ln(a) + \delta \text{ avec } \delta \sim \mathcal{N}(0, \sigma_\delta^2) \quad (7)$$

## b. Estimations des paramètres – Maximum de vraisemblance

S'il existe une méthode en deux étapes dites « de Heckman » pour estimer les paramètres  $\beta_0$ ,  $\beta_1$  et  $\sigma_\delta$ , les moyens de calculs autorisent aujourd'hui l'estimation directe par maximum de vraisemblance. Calculons les contributions de chaque observation de  $\hat{a}$  à la vraisemblance. Pour le modèle (6), on a d'une part,

$$P(\hat{a} = \hat{a}_{th}) = P(\hat{a}^* \leq \hat{a}_{th}) = P\left(\frac{\delta}{\sigma_\delta} \leq \frac{\hat{a}_{th} - (\beta_0 + \beta_1 a)}{\sigma_\delta}\right) = \Phi\left(\frac{\hat{a}_{th} - (\beta_0 + \beta_1 a)}{\sigma_\delta}\right) \quad (8)$$

$$P(\hat{a} = \hat{a}_{sat}) = P(\hat{a}^* \geq \hat{a}_{sat}) = 1 - \Phi\left(\frac{\hat{a}_{sat} - (\beta_0 + \beta_1 a)}{\sigma_\delta}\right) = \Phi\left(\frac{(\beta_0 + \beta_1 a) - \hat{a}_{sat}}{\sigma_\delta}\right) \quad (9)$$

et d'autre part, si  $\hat{a}^* \in [\hat{a}_{th}; \hat{a}_{sat}]$ , alors  $\hat{a}$  est distribuée comme  $\hat{a}^*$ . Les contributions pour de telles observations sont données par la densité d'une loi normale de moyenne  $\beta_0 + \beta_1 a$  et de variance  $\sigma_\delta^2$ , soit :

$$f(\hat{a}) = \frac{1}{\sigma_\delta} \phi\left(\frac{\hat{a} - (\beta_0 + \beta_1 a)}{\sigma_\delta}\right) \quad (10)$$

où  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$  est la densité d'une loi normale centrée réduite.

La vraisemblance associée à l'échantillon  $(\hat{a}_i)_{i=1\dots n}$  s'exprime à partir des quantités (8), (9) et (10) comme le produit suivant :

$$L = \prod_{\hat{a}_i = \hat{a}_{th,i}} \Phi\left(\frac{\hat{a}_{th,i} - (\beta_0 + \beta_1 a_i)}{\sigma_\delta}\right) \prod_{\hat{a}_i = \hat{a}^*} \frac{1}{\sigma_\delta} \phi\left(\frac{\hat{a}_i - (\beta_0 + \beta_1 a_i)}{\sigma_\delta}\right) \prod_{\hat{a}_i = \hat{a}_{sat,i}} \Phi\left(\frac{(\beta_0 + \beta_1 a_i) - \hat{a}_{sat,i}}{\sigma_\delta}\right)$$

On en déduit la log-vraisemblance, que l'on cherchera à maximiser en  $\beta_0$ ,  $\beta_1$  et  $\sigma_\delta$  en résolvant les équations (11), éventuellement à l'aide de l'algorithme de Newton-Raphson, ou de sa forme modifiée connue sous le nom de *méthode du scoring* (voir Annexes) :

$$\frac{\partial \ln(L)}{\partial \beta_0} = 0 \quad \frac{\partial \ln(L)}{\partial \beta_1} = 0 \quad \frac{\partial \ln(L)}{\partial \sigma_\delta} = 0 \quad (11)$$

Là encore, le choix des valeurs initiales pour chaque paramètre conditionne en grande partie la convergence de l'algorithme. Berens propose d'utiliser les estimations issues d'une régression linéaire classique.

## c. Expression de la POD

Une fois les paramètres  $\beta_0$ ,  $\beta_1$  et  $\sigma_\delta$  estimés, nous pouvons nous intéresser à l'expression de la POD en elle-même. Avant d'en proposer une définition dans le cadre Signal/Réponse, il convient de faire une remarque importante sur la distinction entre les variables  $\hat{a}$  et  $\hat{a}^*$ . Subissant un traitement de la part de l'opérateur, les données observées ne correspondent pas exactement à des réalisations du phénomène aléatoire d'intérêt. Elles ne portent qu'une information partielle le concernant et la variable pour laquelle il est intéressant

d'isoler la partie déterministe n'est pas  $\hat{a}$ , mais bien la variable latente  $\hat{a}^*$ . La définition de la POD doit par conséquent s'articuler autour de cette dernière :

<b>Définition</b>	La POD dans le cas des données signal-réponse se définit comme la probabilité que le signal, modélisé par $\hat{a}^*$ (et non $\hat{a}$ , cf. remarque ci-dessus) dépasse le seuil de décision sachant que le défaut est de taille $a$ , $POD(a) = P(\hat{a}^* > \hat{a}_{dec}/a)$
-------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Comme nous l'avons vu précédemment, la modélisation de  $\hat{a}^*$  s'appuie sur trois hypothèses :

**Hypothèse 1 :** Le signal  $\hat{a}^*$ , ou son logarithme, est une fonction linéaire de la taille du défaut  $a$  :

$$\hat{a}^* = \beta_0 + \beta_1 a + \delta \quad \text{ou} \quad \ln(\hat{a}^*) = \beta_0 + \beta_1 \ln(a) + \delta$$

**Hypothèse 2 :** Les fluctuations aléatoires  $\delta$  sont distribuées selon une loi normale de moyenne nulle et de variance  $\sigma_\delta^2$  :

$$\delta \sim \mathcal{N}(0, \sigma_\delta^2)$$

**Hypothèse 3 :** La variance des fluctuations est indépendante de  $a$  (pas d'hétéroscédasticité).

Notons alors  $g_a(\hat{a}^*)$  la densité de  $\hat{a}^*$  sachant  $a$ , la POD s'écrit :

$$POD(a) = P(\hat{a}^* > \hat{a}_{dec}/a) = \int_{\hat{a}_{dec}}^{+\infty} g_a(\hat{a}^*) d\hat{a}^*$$

Examinons les deux possibilités :

▪ Le modèle  $\hat{a}^*$  vs  $a$  implique :

$$POD(a) = \int_{\hat{a}_{dec}}^{+\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(\hat{a}^* - (\beta_0 + \beta_1 a))^2}{2\sigma_\delta^2}} d\hat{a}^* = 1 - \Phi\left(\frac{\hat{a}_{dec} - (\beta_0 + \beta_1 a)}{\sigma_\delta}\right) = \Phi\left(\frac{\beta_0 + \beta_1 a - \hat{a}_{dec}}{\sigma_\delta}\right)$$

où  $\Phi$  est la fonction de répartition d'une loi  $\mathcal{N}(0,1)$ . La POD correspond à la *fonction de répartition d'une loi normale* :

$$POD(a) = \Phi\left(\frac{a - \mu}{\sigma}\right) \quad \text{avec} \quad \begin{cases} \mu = \frac{\beta_0 - \hat{a}_{dec}}{\beta_1} \\ \sigma = \frac{\sigma_\delta}{\beta_1} \end{cases} \quad (12)$$

L'estimation de la POD est alors obtenue en utilisant les valeurs des paramètres déterminées à l'aide du maximum de vraisemblance, soit :

$$\widehat{POD}(a) = \Phi\left(\frac{a - \hat{\mu}}{\hat{\sigma}}\right)$$

▪ Le modèle  $\ln(\hat{a}^*)$  vs  $\ln(a)$  (ou log-log) implique :

$$POD(a) = \Phi\left(\frac{a - \ln(\mu)}{\sigma}\right) \quad \text{avec} \quad \begin{cases} \mu = \frac{\beta_0 - \ln(\hat{a}_{dec})}{\beta_1} \\ \sigma = \frac{\sigma_\delta}{\beta_1} \end{cases} \quad (13)$$

La POD s'exprime donc comme la *fonction de répartition d'une loi log-normale*. De la même façon, la POD estimée s'obtient en remplaçant les paramètres par leur estimation au maximum de vraisemblance :

$$\widehat{POD}(a) = \Phi\left(\frac{a - \ln(\hat{\mu})}{\hat{\sigma}}\right)$$

**Remarque :** Dans les deux cas, la définition de l'écart-type implique  $\sigma > 0$  et  $\sigma_\delta > 0$ , et impose par conséquent  $\beta_1 \geq 0$ .

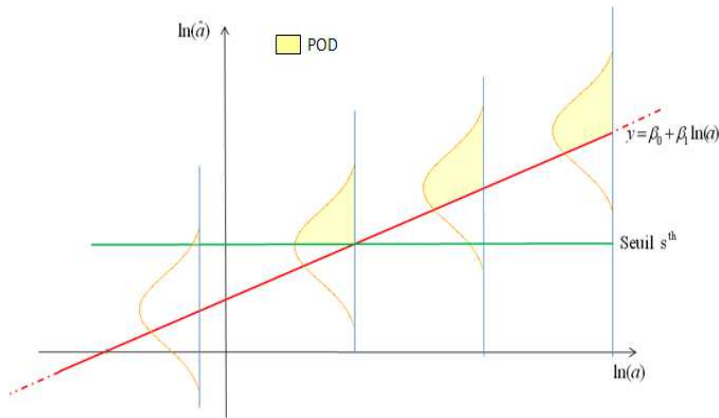


Figure 7 : Modèle de Berens – données signal/réponse

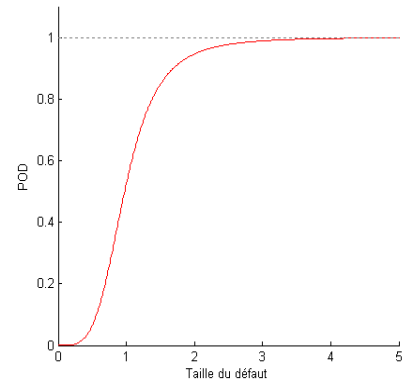


Figure 8 : Exemple de POD – modèle signal réponse de Berens

La Figure 7 illustre graphiquement la modélisation de la relation entre  $\ln(\hat{a}^*)$  et  $\ln(a)$ , tandis que la Figure 8 donne un exemple de courbe POD obtenue.

## II.2.C. Quelques remarques concernant à le modèle de Berens

### a. Erreurs multiplicatives et limitation du modèle

Le modèle signal-réponse de Berens impose un modèle statistique de la forme :

1.  $\ln(\hat{a}^*) = \beta_0 + \beta_1 \ln(a) + \delta \Leftrightarrow \hat{a}^* = e^{\beta_0} a^{\beta_1} e^{\delta}$ , en échelle log-log.
2.  $\hat{a}^* = \beta_0 + \beta_1 a + \delta$ , en échelle classique.

Ces expressions engendrent quelques remarques :

- Le terme multiplicatif en  $e^{\delta}$  du modèle log-log revient à dire que le rapport entre le signal émis et la taille du défaut est proportionnel aux fluctuations. Cette propriété est difficilement interprétable physiquement :
- Plus généralement, on s'attend à ce que la réponse d'un défaut dépende :
  - Du type de matériau constituant le défaut. Ce paramètre entre en jeu dans les modèles physique sous la forme d'un contraste.
  - De la géométrie du défaut. Elle est introduite dans les modèles à travers une surface ou un volume.
  - De la position du défaut par rapport au traducteur.

Que ce soit en acoustique ou en électromagnétique, la réponse du défaut est intégrée sur son volume, et on a :

$$\text{Amplitude Signal Réponse} \propto \text{Contraste} \cdot \text{Volume défaut} \cdot \frac{\text{Champ incident}}{\text{distance au traducteur}}$$

Ainsi, si le contraste entre la pièce et le défaut n'est pas suffisant, l'amplitude du signal tendra vers zéro à volume et distance au traducteur fixés et il ne pourra donc y avoir détection. De la même façon, si le défaut est trop loin du traducteur, ce dernier ne pourra le percevoir. Il serait par conséquent intéressant de pouvoir intégrer ces informations dans le modèle de POD.

### b. L'hypothèse de linéarité dans le modèle signal/réponse – Plan d'expérience

Le modèle signal-réponse repose sur l'hypothèse d'une relation linéaire entre l'amplitude maximale du signal et la taille du défaut dans un repère éventuellement munit d'une échelle log-log. Si dans la pratique, les données ne semblent a priori pas la contredire, elles ne permettent cependant pas non plus d'en assurer la validité sur toute l'étendue des tailles de défaut : la saturation cache en effet le comportement du signal en dehors de l'intervalle  $[\hat{a}_{th}; \hat{a}_{sat}]$ , zone dans laquelle rien ne l'empêche d'adopter un profil non-linéaire.



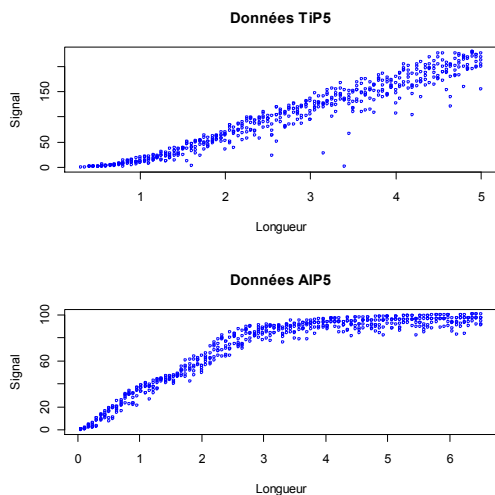


Figure 9 : Données simulées - signal vs longueur

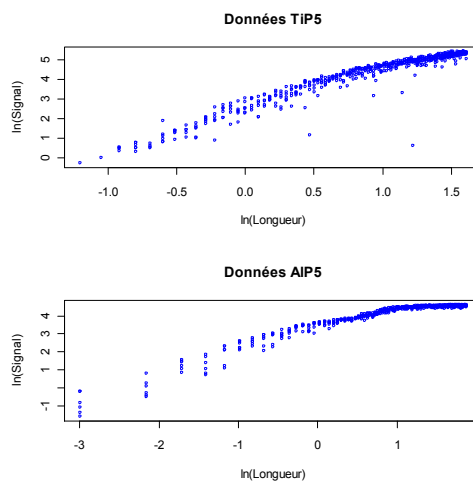


Figure 10 : Données simulées - ln(signal) vs ln(longueur)

Les simulations issues de CIVA semblent par ailleurs confirmer cette tendance. La Figure 9 et la Figure 10 présentent des échantillons simulés. Les configurations à partir desquelles ont été générées sont basées sur des pièces en titanes ou en aluminium.

Globalement, la relation entre l'amplitude du signal et la taille du défaut (en l'occurrence, sa longueur) s'avère assez fortement non-linéaire, même si le passage à l'échelle log-log « redresse » les observations. On remarque aussi la présence d'une hétéroscédasticité relativement prononcée sur les données signal vs longueur : la variance des observations croît avec l'amplitude du signal (forme en entonnoir du nuage de points).

La Figure 11 et la Figure 12 présentent une régression linéaire et une régression censurée sur les données de simulation (triangles bleus) artificiellement saturées : autrement dit, ces régressions ont été ajustées sur les données saturées (cercles verts). On remarque dans les deux cas que la régression linéaire est fortement influencée par les observations aux limites de saturation avec une pente trop faible pour s'ajuster correctement aux données. La régression censurée donne à l'inverse des résultats satisfaisants. La Figure 12 présente un cas pathologique. Les observations (cercles verts) suggèrent une relation linéaire entre le signal et la longueur du défaut mais il s'avère cependant que la partie inobservable des données (triangles bleus) se caractérise par une importante non linéarité. Ce cas illustre par la même occasion l'intérêt que peut revêtir le redressement des données par leur transformation logarithmique.

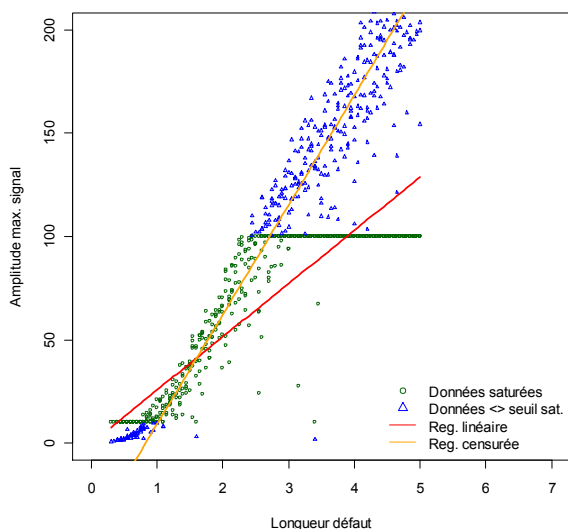


Figure 11 : Régression linéaire et censurée ajustées sur données simulées - TiP5

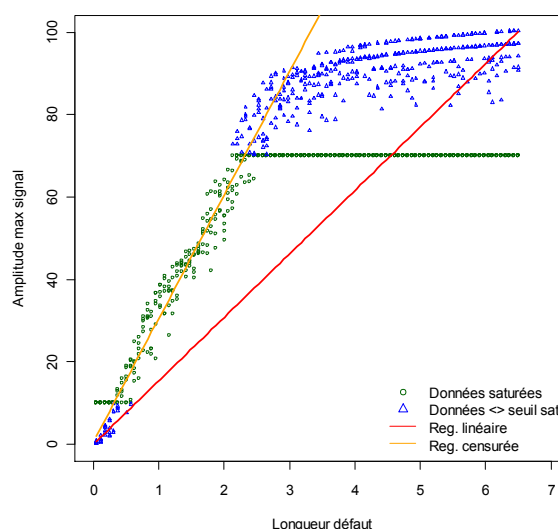


Figure 12 : Régression linéaire et censurée ajustées sur données simulées - AIP5

Les interpolations effectuées à l'aide du modèle de Berens pour les valeurs du signal en dehors de l'intervalle non censuré devront donc être considérées avec précaution. Il reste en pratique impossible de savoir si les hypothèses du modèle sont bien vérifiées dans cette zone. On remarquera cependant que cette censure n'est pas nécessairement très handicapante car les régions inobservables seront a priori associées à de très faibles ou très fortes probabilités de détection.

On signale d'autre part que l'analyse de données réelles non simulées se caractérise souvent par une quantité de mesures relativement restreinte. Optimiser l'estimation des paramètres du modèle sur la base d'échantillons de taille moindre suggère la mise en place d'un plan

d'expérience adapté. L'objectif est alors de déterminer le nombre mesures du signal à effectuer pour des valeurs de  $a$  réparties de façon spécifique sur un intervalle  $[a_{min}; a_{max}]$  de façon à minimiser un critère tel que la variance des estimations.

## II.2.D. Bande de confiance pour la POD

### a. Problématique industrielle

L'estimation de la POD étant soumise aux aléas d'échantillonnage, la donnée de cette estimation n'est pas suffisante dans le cadre de la pratique industrielle. Il est en effet nécessaire de quantifier l'incertitude l'entourant. Cela se traduit par la détermination d'une bande de confiance pour la POD.

D'autre part, les industriels s'intéressent particulièrement à la taille de défaut  $a_{0,9}$  pour laquelle la probabilité de détection est égale à 0.9 :

$$a_{0,9} = a \text{ tq } \{POD(a) = 0.9\}$$

Le modèle de Berens décrit la POD par l'intermédiaire d'une fonction monotone croissante (propriété des fonctions de répartition). Cette fonction est donc inversible et il est donc aisé de déterminer une estimation de  $a_{0,9}$  :

$$\hat{a}_{0,9} = \overline{POD}^{-1}(0.9)$$

Pourquoi  $a_{0,9}$  ? Pour comprendre, intéressons-nous à un exemple de problématique industrielle pour un constructeur en aéronautique. Imaginons qu'un tel constructeur ne puisse garantir la fiabilité d'une pièce d'avion que si celle-ci contient un défaut de taille supérieure à 1mm. Celui-ci sera rassuré s'il sait que  $a_{0,9} = 0.5mm$ . En effet, il aura dès lors plus de 90% de chances de détecter un défaut de taille supérieure ou égale à 1mm. Ainsi, en répétant les mesures, les chances de laisser une pièce contenant un défaut critique en fonctionnement seront limitées. S'intéresser à une POD de 0.9 permet d'autre part d'abaisser la taille de défaut maximale acceptable tout en garantissant un niveau de détection élevée.

La détermination de la bande de confiance comme sur la Figure 13 donne une borne inférieure de la POD avec un niveau de confiance de 95%. Cette borne fournit une autre estimation de  $a_{0,9}$  que l'on note  $\hat{a}_{0,9;0,95}$ .

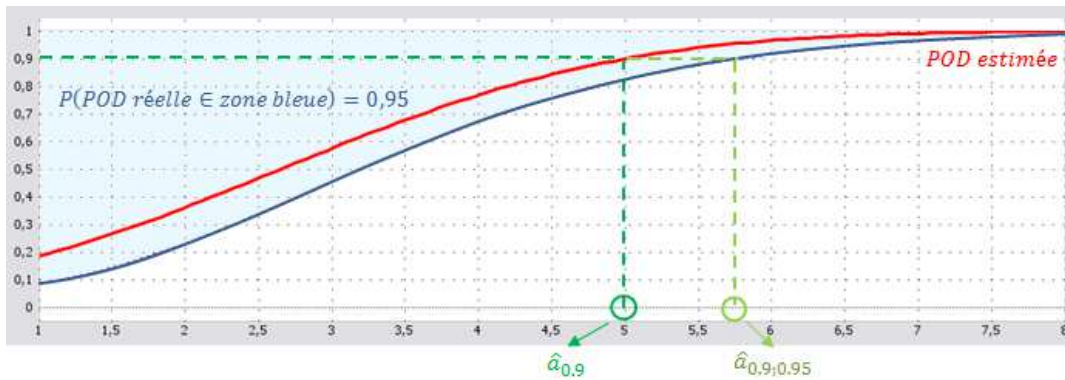


Figure 13 : POD estimée - Bande de confiance (niveau 95%) – Estimations de  $a_{0,9}$

Nous abordons dans les paragraphes suivants, la construction de la bande de confiance associée à la POD.

### b. Bandes de confiance non-paramétriques - limitations

Classiquement, on construit des bandes de confiance pour les fonctions de répartition empiriques en suivant un principe proche des tests de Kolmogorov-Smirnov. L'inégalité de Massart permet d'affirmer que si l'on dispose de  $n$  variables aléatoires réelles i.i.d :  $X_1, \dots, X_n$  de fonction de répartition  $F$  et de fonction de répartition empirique  $F_n$ , alors :

$$P\left(\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \geq \frac{z}{\sqrt{n}}\right) \leq 2e^{-2nz^2} \text{ pour tout } z > 0$$

Si l'on souhaite obtenir des bandes de confiance de niveau  $1 - \alpha$ , il suffit de déterminer  $z_\alpha$  tel que :

$$2e^{-2nz_\alpha^2} = \alpha \text{ soit } z_\alpha = \sqrt{\frac{-\log\left(\frac{\alpha}{2}\right)}{2n}}$$

Et on a alors :

$$P\left(\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{z_\alpha}{\sqrt{n}}\right) \geq 1 - \alpha$$

En fixant  $\alpha = 0.05$ , on obtient une bande de confiance de niveau 95% pour la fonction de répartition.

Cette procédure est cependant uniquement valable dans le cas où l'on donne une estimation non paramétrique de la fonction de répartition. D'autre part, les bandes de confiance ainsi obtenues sont de largeur verticale constante et peuvent se révéler trop large au

niveau des queues de distributions, avec un passage possible sous zéro ou au-dessus de l'une de ses bornes. Une telle situation n'est pas acceptable dans la pratique.

Le modèle de Berens conduit, selon les cas, à une POD s'exprimant comme la fonction de répartition d'une loi logistique, log-logistique, normale ou encore log-normale dont on cherche à estimer les paramètres. Une méthode de construction de bande de confiance adaptée doit par conséquent être utilisée. Il est souhaitable d'éviter l'écueil des bandes de confiance dont la largeur dépasse le domaine de validité de la POD.

### C. Bande de confiance pour les fonctions de répartition de variables continues

#### i. Principe général pour une bande de confiance bilatérale

La POD s'exprimant comme une fonction de répartition, les méthodes permettant d'obtenir une bande de confiance pour de telles fonctions vont pouvoir être appliquées.

La procédure proposée par Cheng et Iles [6] repose sur l'estimation par maximum de vraisemblance des paramètres d'une distribution continue. Ces derniers font par exemple référence à la moyenne  $\mu$  et à la variance  $\sigma$  d'une loi normale  $\mathcal{N}(\mu, \sigma^2)$ . Comme nous l'avons vu dans la description du modèle de Berens, la fonction de répartition de cette famille de lois est donnée par :

$$F_{\mu,\sigma}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Considérons ce dernier exemple. La procédure de Cheng et Iles se décompose en deux étapes :

1. On détermine une région de confiance  $R$  pour les paramètres inconnus ( $\mu$  et  $\sigma$  par exemple) :

$$R = [\mu_{min}; \mu_{max}] \times [\sigma_{min}; \sigma_{max}] \subset \mathbb{R}^2 \quad \text{où :} \quad P(\mu \in [\mu_{min}; \mu_{max}]) = 0.95 \quad \text{et} \quad P(\sigma \in [\sigma_{min}; \sigma_{max}]) = 0.95$$

2. On rappelle que le  $p^{ième}$  quantile d'une variable aléatoire  $X$  est la valeur  $X_{(p)}$  pour laquelle  $p\%$  des réalisations de  $X$  sont inférieures à  $X_{(p)}$ . Posons  $p = \Phi\left(\frac{x_p - \mu}{\sigma}\right)$ . Alors  $x_p = \sigma \Phi^{-1}(p) + \mu$  correspond au  $p^{ième}$  quantile de la variable dont la fonction de répartition est donnée par  $F_{\mu,\sigma}$ .

La bande de confiance est obtenue en déterminant, pour chaque valeur de  $p$ , les bornes de variation du quantile  $x_p$  lorsque les paramètres  $\mu$  et  $\sigma$  fluctuent dans la région de confiance  $R$ .

L'idée est donc de transformer la région de confiance sur l'espace des paramètres en une bande de confiance pour la fonction de répartition (la POD) en utilisant la relation entre le  $p^{ième}$  quantile et les paramètres de la variable aléatoire.

De façon plus formelle, considérons maintenant une variable  $\mathbf{X}$  de fonction de répartition  $F(\mathbf{x}, \boldsymbol{\theta})$  paramétrée par un vecteur  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$ . Dans notre exemple,  $\boldsymbol{\theta} = (\mu, \sigma)'$  et  $F(\mathbf{x}, \boldsymbol{\theta}) = F_{\mu,\sigma}(x)$ . Soient  $R$  la région de confiance pour  $\boldsymbol{\theta}$  dans l'espace des paramètres et  $B$  la bande de confiance pour la fonction de répartition.

On considère un  $n$ -échantillon  $X_1, \dots, X_n$  de cette variable ( $n$  observations indépendantes de  $\mathbf{X}$ ). Le but est d'étudier les variations du graphe  $y = F(\mathbf{x}, \boldsymbol{\theta})$  lorsque  $\boldsymbol{\theta}$  varie dans  $R$ . Détaillons la marche à suivre pour chacune des deux étapes précédemment présentées :

- Si  $R$  est telle que :

$$P(\boldsymbol{\theta} \in R) = 1 - \alpha \tag{14}$$

Cela signifie que le vrai  $\boldsymbol{\theta}$  a une probabilité  $1 - \alpha$  de se trouver dans  $R$ , et que par conséquent, la vraie fonction de répartition a une probabilité d'au moins  $1 - \alpha$  d'appartenir à  $B = \{F(\mathbf{x}, R), \mathbf{x} \in \mathbb{R}\}$ , soit :

$$P(F \in B) \geq 1 - \alpha \tag{15}$$

Le premier objectif est donc de déterminer  $R$  de façon à ce qu'elle vérifie la propriété (14).

On se base sur l'estimation de  $\boldsymbol{\theta}$  par maximum de vraisemblance  $\hat{\boldsymbol{\theta}}_{ML}$  et sur les propriétés asymptotiques suivantes :

$$\hat{\boldsymbol{\theta}}_{ML} \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta}))$$

où  $I(\boldsymbol{\theta})$  est la matrice d'information de Fisher composée des termes  $I_{ij}(\boldsymbol{\theta}) = -E\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln(L)\right)$  avec  $L$  la vraisemblance des observations sous le modèle considéré. Cette propriété implique :

$$Q(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta})' I(\boldsymbol{\theta}) (\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \chi_k^2$$

La quantité  $Q(\boldsymbol{\theta})$  étant une forme quadratique en des variables de loi asymptotique  $\mathcal{N}(0, I^{-1}(\boldsymbol{\theta}))$ , elle suit asymptotiquement un  $\chi^2$  à  $k$  degrés de liberté,  $k$  étant le nombre de paramètres du modèle. Cette quantité peut être interprétée comme une distance de  $\hat{\boldsymbol{\theta}}_{ML}$  à  $\boldsymbol{\theta}$  vis-à-vis de la norme induite par la matrice  $I(\boldsymbol{\theta})$ . Ainsi, la distribution de  $Q(\boldsymbol{\theta})$  nous renseigne sur la distribution de  $\hat{\boldsymbol{\theta}}_{ML}$  autour de  $\boldsymbol{\theta}$ . Définissons

R' comme étant la zone autour de  $\theta$  (inconnu) dans laquelle  $\hat{\theta}$  se trouve avec une probabilité  $1 - \alpha$ . Déterminer R' revient à chercher  $\gamma_\alpha$  vérifiant :

$$P(\hat{\theta} \in R') = P(Q(\theta) \leq \gamma_\alpha) \underset{as.}{\approx} P(\chi_k^2 \leq \gamma_\alpha) = 1 - \alpha \quad (16)$$

Cependant, la problématique originelle consistait plutôt à déterminer la zone R autour de  $\hat{\theta}$  dans laquelle se trouve  $\theta$  avec une probabilité  $1 - \alpha$  (équation (14)). Dans la partie suivante consacrée au cas des modèles de translation – changement d'échelle, nous verrons que (16) peut être inversée pour donner R.

▪ Une fois la région R identifiée, l'objectif suivant consiste à évaluer l'enveloppe inférieure et supérieure de B. On va pour cela s'intéresser aux quantiles de X. Le p<sup>ième</sup> quantile de X, noté  $x_p$ , est défini par :

$$P(X \leq x_p) = F(x_p, \theta) = p$$

A p fixé, si  $\theta$  varie dans R alors  $x_p$  varie dans  $[\hat{x}_p(\min); \hat{x}_p(\max)]$ , où :

$$\hat{x}_p(\min) = \min_{\theta \in R} x_p \quad \text{et} \quad \hat{x}_p(\max) = \max_{\theta \in R} x_p$$

Chaque point de l'une des bornes de la bande de confiance est donné par l'intersection entre la droite  $y = p$  et de la droite  $x = \hat{x}_p(\min)$  ou  $x = \hat{x}_p(\max)$  (voir Figure 14) :

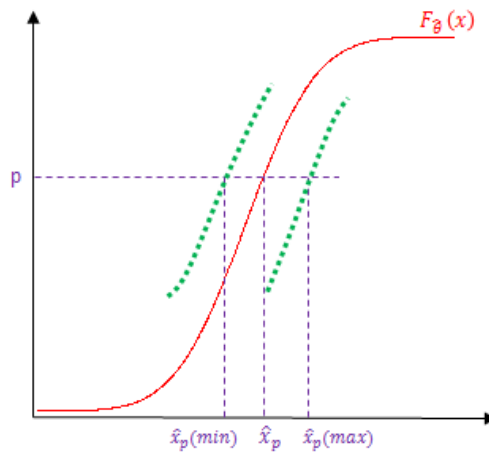


Figure 14 : Bande de confiance et quantiles - Méthode de Cheng et Iles

Nous allons voir dans la suite, comment, dans le cas des modèles de translation - changement d'échelle, l'estimation des quantiles  $\hat{x}_p(\min)$  et  $\hat{x}_p(\max)$  est obtenue.

## ii. Cas des modèles de translation - changement d'échelle

- Région de confiance pour les paramètres (R)

Un modèle est dit de translation – changement d'échelle lorsque la fonction de répartition G associée peut s'écrire en fonction d'une fonction de répartition canonique F indépendante des paramètres du modèle et vérifiant :

$$G(x) = F\left(\frac{x - \mu}{\sigma}\right) \quad (17)$$

où  $\mu$  et  $\sigma$  sont les paramètres du modèle.

Un exemple classique est celui de la loi normale  $\mathcal{N}(\mu, \sigma^2)$  de moyenne  $\mu$  et de variance  $\sigma^2$  dont la fonction de répartition est donnée par :

$$F_{\mathcal{N}(\mu, \sigma^2)}(x) = F_{\mathcal{N}(0,1)}\left(\frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

où  $\Phi$  est la fonction de répartition d'une  $\mathcal{N}(0,1)$  et correspond à la fonction de répartition canonique.

Les estimations par maximum de vraisemblance des paramètres d'un modèle de translation – changement d'échelle possèdent des propriétés particulières. En dérivant l'équation (17) par rapport à x, on trouve la densité g du modèle en fonction de la densité f de la distribution canonique :

$$g(x) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

A partir de cette relation, il est possible d'écrire la log-vraisemblance pour un n-échantillon  $(x_1, \dots, x_n)$  constitué de n réalisations indépendantes de la variable considérée. On en dérive la matrice de Fisher :

$$I(\mu, \sigma) = -\frac{n}{\sigma^2} \begin{bmatrix} k_0 & -k_1 \\ -k_1 & k_2 \end{bmatrix} \quad (18)$$

Dans un tel modèle, il est possible de démontrer que **les différents composants  $k_0, k_1, k_2$  de la matrice de Fisher sont indépendants de  $\mu$  et  $\sigma$**  (voir annexes)

L'expression de la matrice de Fisher donnée dans l'équation (45) (voir Annexes) nous permet d'écrire la quantité  $Q(\theta) = Q((\mu, \sigma)')$  sous une nouvelle forme :

$$\begin{aligned} Q(\theta) &= \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \end{pmatrix}' I(\mu, \sigma) \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \end{pmatrix} \\ &= nk_0 \frac{(\hat{\mu} - \mu)^2}{\sigma^2} - 2nk_1 \frac{(\hat{\mu} - \mu)(\hat{\sigma} - \sigma)}{\sigma^2} + nk_2 \frac{(\hat{\sigma} - \sigma)^2}{\sigma^2} \end{aligned}$$

Ainsi l'équation en posant  $x = \hat{\mu} - \mu$  et  $y = \hat{\sigma} - \sigma$ , et en modifiant l'expression ci-dessus, on fait clairement apparaître l'équation d'une conique :

$$Q(\theta) = \gamma_\alpha \Leftrightarrow nk_0 x^2 - 2nk_1 xy + (nk_2 - \gamma_\alpha)y^2 + 2\hat{\sigma}\gamma_\alpha y = \hat{\sigma}^2\gamma_\alpha \quad (19)$$

Par la caractérisation cartésienne d'une conique, on sait que l'équation (19) conduit à une conique du genre ellipse si le discriminant  $n^2k_1^2 - nk_0(nk_2 - \gamma_\alpha)$  est strictement négatif, soit si :

$$\frac{k_0k_2 - k_1^2}{k_0} > \frac{\gamma_\alpha}{n}$$

Par conséquent, pour un échantillon de taille assez grande, cette relation sera nécessairement vérifiée et **la région R sera délimitée par une ellipse**.

- Bornes de confiance pour la fonction de répartition - Détermination géométrique

Il reste à déterminer les variations des quantiles lorsque les paramètres évoluent dans R. Dans le cas du modèle de translation - changement d'échelle défini par la fonction de répartition G, nous avons vu que le  $p^{\text{ième}}$  quantile  $x_p$  vérifie :

$$G_{(\mu, \sigma)}(x_p) = F\left(\frac{x_p - \mu}{\sigma}\right) = p$$

L'inversibilité de F permet d'en donner une expression explicite :

$$x_p = \mu + \sigma F^{-1}(p) = \mu + \sigma x_p^{can}, \quad \text{où } p \in [0; 1]$$

Les quantiles  $x_p$  sont donc identiques à ceux  $x_p^{can}$  de la variable canonique à une translation et un changement d'échelle près. Cette relation permet d'autre part, à **p fixé**, d'écrire  $\sigma$  comme une fonction de  $\mu$  :

$$\sigma = -\frac{1}{F^{-1}(p)}\mu + \frac{x_p}{F^{-1}(p)}$$

A p fixé, le point  $(\mu, \sigma)$  appartient à la droite d'équation  $y = -\frac{1}{F^{-1}(p)}x + \frac{x_p}{F^{-1}(p)}$ . Si l'on souhaite que  $(\mu, \sigma)$  puisse parcourir l'ensemble R entier tout en respectant cette contrainte, il est nécessaire d'accorder un degré de liberté en laissant varier le quantile  $x_p$ . La région R coïncide alors avec l'union des droites de pente  $-\frac{1}{F^{-1}(p)}$  interceptées avec la conique donnée par (19). Les quantiles estimés  $\hat{x}_p(\min)$  et  $\hat{x}_p(\max)$  correspondent aux intersections des tangentes à la conique (ellipse) de pente  $-\frac{1}{F^{-1}(p)}$  avec l'axe des abscisses. La Figure 15 illustre ces propriétés. On réitère ce procédé pour chaque valeur de p.

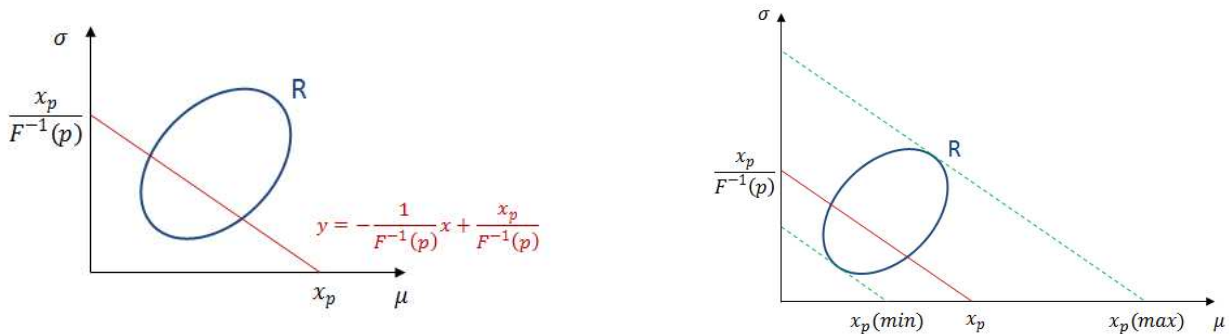


Figure 15 : Détermination géométrique de la variation des quantiles engendrée par la variation des paramètres

- Bornes de confiance pour la fonction de répartition – Détermination à l'aide d'un Lagrangien

Le principe décrit ci-dessus permet de déterminer de façon géométrique l'intervalle de variation des quantiles. Il est cependant aussi possible de le retrouver par le calcul. Ce qui nous donnera une forme explicite pour :

$$\hat{x}_p(\min) = \min_{(\mu, \sigma) \in R} x_p(\mu, \sigma) \quad \text{et} \quad \hat{x}_p(\max) = \max_{(\mu, \sigma) \in R} x_p(\mu, \sigma)$$

On introduit un Lagrangien  $\mathcal{L} = x_p + \eta Q(\mu, \sigma)$  puis on cherche à résoudre les équations :  $\frac{\partial \mathcal{L}}{\partial \mu}(\mu, \sigma) = 0$  et  $\frac{\partial \mathcal{L}}{\partial \sigma}(\mu, \sigma) = 0$ . Le multiplicateur  $\eta$  peut alors être supprimé par division :

$$\frac{\partial x_p}{\partial \mu} / \frac{\partial x_p}{\partial \sigma} = \frac{\partial Q}{\partial \mu} / \frac{\partial Q}{\partial \sigma} \quad (20)$$

La résolution de l'équation (20) conduit à deux solutions :

$$\begin{cases} \hat{x}_p(\min) = \hat{x}_p + \hat{\sigma} b(N - k_1 - ak_0)^{-1} \\ \hat{x}_p(\max) = \hat{x}_p - \hat{\sigma} b(N + k_1 + ak_0)^{-1} \end{cases} \quad (21)$$

où  $a = F^{-1}(p)$ ,  $b = k_2 + 2k_1a + k_0a$ ,  $\hat{x}_p = \hat{\mu} + a\hat{\sigma}$  et  $N = \sqrt{(k_0k_2 - k_1^2) \left(\frac{nb}{\gamma} - 1\right)}$ .

En inversant ces solutions, il est possible de déterminer l'expression des bornes inférieure et supérieure de la bande de confiance :

$$\begin{cases} \hat{F}_{\min}(x) = F(\hat{\zeta} - h) \\ \hat{F}_{\max}(x) = F(\hat{\zeta} + h) \end{cases} \quad (22)$$

avec  $\hat{\zeta} = \frac{x - \hat{\mu}}{\hat{\sigma}}$  et  $h = \sqrt{\gamma n^{-1} k_0^{-1} \left[1 + (k_0 \hat{\zeta} + k_1)^2 (k_0 k_2 - k_1^2)^{-1}\right]}$

### iii. Bande de confiance unilatérale

- Intérêt d'une bande de confiance unilatérale

La démarche présentée précédemment permet de calculer une bande de confiance bilatérale autour de la POD estimée. Or en pratique, on préférera s'assurer d'une valeur minimale de la POD et on sera donc plus intéressé par une bande de confiance unilatérale de la forme  $[borne_{inf}; 1]$  pour chaque taille de défaut. Cette approche correspond au calcul d'une « bande de confiance unilatérale supérieure ».

- Méthode de calcul – bande de confiance unilatérale supérieure

Le principe de calcul de la bande de confiance unilatérale supérieure (ou inférieure) [7] repose sur les mêmes bases que celles utilisées pour la détermination d'une bande de confiance bilatérale. Dans le cas « supérieure », on recherche les fonctions de répartition (correspondant à la donnée de  $\mu$  et  $\sigma$ ) dont les quantiles sont toujours (c'est-à-dire pour tout  $p$  dans  $[0; 1]$ ) inférieurs à  $\hat{x}_p(\max)$ . Notons  $B_U$  la bande de confiance unilatérale supérieure :

$$B_U = \{ \text{fonctions de répartition telles que } x_p \leq \hat{x}_p(\max) \}$$

et  $R_U$  l'ensemble suivant :

$$R_U = \{ (\mu, \sigma), x_p(\mu, \sigma) \leq \hat{x}_p(\max) \forall p \in [0; 1] \}$$

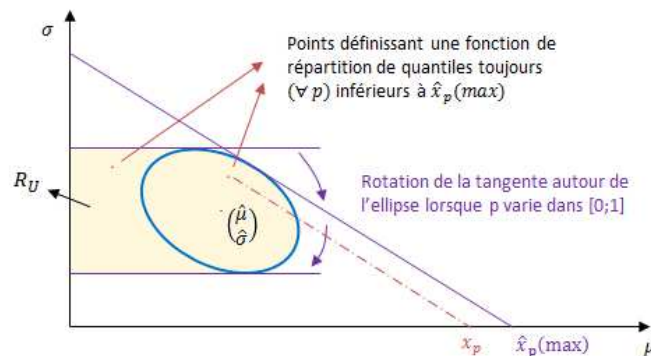


Figure 16 : Détermination de  $R_U$

A l'aide du mécanisme de la tangente à R (voir Figure 15), il est possible de déterminer  $R_U$  (voir Figure 16). Lorsque  $p$  varie entre 0 et 1, la pente de la tangente permettant de déterminer  $\hat{x}_p(max)$  tourne autour de R en passant de la tangente horizontale au-dessus de R à la tangente horizontale au-dessous de R en longeant le côté droit de l'ellipse. La zone délimitée par l'ensemble de ces tangentes correspond à  $R_U$ .

A partir de  $R_U$ , on peut définir, comme dans le cas de la bande de confiance bilatérale, une région  $R'_U$  dans le plan  $(\hat{\mu}, \hat{\sigma})$  correspondante. A la fois  $R_U$  et  $R'_U$  dépendent de  $\gamma$  et  $\gamma$  doit être calculée de façon à ce que :

$$P((\hat{\mu}, \hat{\sigma}) \in R'_U) = 1 - \alpha$$

La valeur de  $\gamma$  est alors déterminée à l'aide des propriétés de la distribution du vecteur :

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \sigma^{-1} n^{\frac{1}{2}} k_0^{-\frac{1}{2}} \begin{pmatrix} k_0 & -k_1 \\ 0 & d \end{pmatrix} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \end{pmatrix}$$

$\hat{\beta}_1$  et  $\hat{\beta}_2$  sont asymptotiquement distribués selon une loi normale et indépendants. Cette propriété conduit à l'expression du niveau de confiance asymptotique associé à  $R_U$  :

$$P((\mu, \sigma) \in R_U) = \frac{1}{2} [P(\chi^2_2 \leq \gamma) + 2\Phi(\sqrt{\gamma}) - 1]$$

Il faut alors déterminer  $\gamma$  de façon à ce que :

$$\frac{1}{2} [P(\chi^2_2 \leq \gamma) + 2\Phi(\sqrt{\gamma}) - 1] = 1 - \alpha$$

Un exemple de bande de confiance unilatérale supérieure peut être observé sur la figure Figure 13.

## II.3. Courbes ROC

### II.3.A. Contexte

Jusqu'ici, nous nous sommes intéressés à la construction de méthodes permettant d'estimer une probabilité de détection. Comme nous l'avons vu dans la partie II.1, cette dernière correspond à la probabilité conditionnelle de détection ( $\hat{a}$  supérieur à un seuil) sachant que la taille du défaut vaut  $a$ . Ainsi, afin de pouvoir en donner une estimation, les données recueillies doivent provenir de signaux issus d'une expérimentation avec défauts.

Mais se reposer uniquement sur la probabilité de détection comme indicateur de performance d'un capteur CND n'est pas suffisant dans le contexte général des applications industrielles. En effet, plusieurs facteurs peuvent venir moduler l'appréciation d'un capteur.

Dans un premier temps, nous verrons qu'il est possible qu'un capteur avec une probabilité de détection élevée peut en même temps se caractériser par une **probabilité de fausse alarme** (probabilité de conclure à la présence d'un défaut alors qu'il n'y en a pas) importante, et inversement. Dans le premier cas, le capteur sera qualifié de « conservateur » tandis qu'il sera jugé « libéral » dans le second. L'objectif consistera à sélectionner les caractéristiques du capteur de façon à obtenir un compromis optimal.

Un critère d'évaluation construit sur les bases de la POD et de la PFA et éventuellement de facteurs comme les coûts associés aux erreurs du capteur ou la prévalence des défauts dans la population des pièces inspectées, précisera la notion d'optimalité. Nous verrons qu'une façon d'obtenir les meilleures caractéristiques vis-à-vis de ce critère consistera à résoudre un problème d'optimisation sous contraintes.

Les courbes ROC pourront à la fois se présenter comme un outil de visualisation des différents compromis possibles et d'interprétation graphique du problème d'optimisation.

### II.3.B. Probabilité de détection et probabilité de fausse alarme

La probabilité de fausse alarme (PFA) est indissociable de la probabilité de détection (POD). Pour justifier cette assertion, plaçons-nous dans le contexte simplifié décrit ci-après.

Supposons que l'on effectue une série d'expérimentations sur des pièces pouvant contenir un défaut ou non, et supposons d'autre part que chaque défaut est identique à un défaut générique. Nous avons ainsi deux réalités possibles : présence ou non d'un défaut dans la pièce, et deux décisions possibles correspondant de la même façon à la présence ou non d'un défaut. On ne cherchera pas ici à relier la taille du défaut à sa probabilité de détection, mais à introduire des pièces sans défauts dans les expérimentations afin de faire apparaître le concept de probabilité de fausse alarme et en justifier son utilité.

On dispose donc de deux populations **disjointes** de pièces (avec et sans défaut) sur lesquelles deux décisions peuvent être prises. Quatre événements sont ainsi possibles. Une fois les expériences réalisées, le Tableau 1 des effectifs croisés, aussi appelé **table des confusions**, permet de calculer une estimation des probabilités (Tableau 2) de détection et de fausse alarme.

Effectifs		Réalité	
		Présence défaut r+	Absence défaut r-
Décision	Présence défaut d+	vp	fp
	Absence défaut d-	fn	vn
		p = vp+fn	n = fp+vn

Tableau 1 : Table des confusions

Probabilités		Réalité	
		Présence défaut r+	Absence défaut r-
Décision	Présence défaut d+	$P(d+ / r+) \approx \frac{vp}{p}$ POD estimée	$P(d- / r-) \approx \frac{fp}{p}$ PFA estimée
	Absence défaut d-	$1 - P(d- / r+)$	$1 - P(r- / d-)$
		1	1

Tableau 2 : Probabilités estimées

En raisonnant en termes de probabilités conditionnelles (relativement aux deux populations de pièce), on fait apparaître la POD et la PFA (Tableau 2).

Les deux populations étant disjointes et les relations suivantes vérifiées :

$$P(d+ / r-) = 1 - P(d+ / r+) = 1 - POD$$

$$P(d- / r-) = 1 - P(d- / r+) = 1 - PFA$$

la connaissance de la POD et de la PFA est nécessaire et suffisante pour reconstituer le tableau des probabilités. Par conséquent, **POD et PFA sont deux indicateurs indissociables dans l'évaluation du capteur.**

### II.3.C. Les courbes ROC

#### a. Définition et premières propriétés

Le contexte du paragraphe II.3.B correspond en fait à une problématique abordée dès les années 50 dans le domaine de la théorie de la détection du signal. Celle-ci est posée comme un problème de décision : lors d'une expérience, deux événements sont possibles mais un seul est effectivement réalisé à chaque fois. Dans tous les cas, les événements sont inaccessibles à l'opérateur. Il souhaite cependant prendre une décision en faveur de la réalisation de l'un ou l'autre de ces événements. On suppose qu'il a tout de même la possibilité de relever indirectement des informations par l'intermédiaire d'une variable S (souvent appelé **score**), à valeurs dans un espace continu et unidimensionnel, dans l'une des conditions suivantes :

- soit l'opérateur a un accès total à S et les données qu'il a recueillies contiennent ainsi toute l'information sur S.
- soit l'opérateur n'a qu'un accès partiel à S. Dans ce cas, les données qu'il a relevées ne contiennent pas toute l'information sur S.

Quel que soit le cas de figure, l'objectif pour l'opérateur est de parvenir à évaluer les capacités intrinsèques de la variable S à dissocier les deux événements. Les courbes ROC constituent un outil intéressant pour visualiser les forces et faiblesses de S utilisée à cette fin.

Nous avons vu (paragraphe II.2) que la décision sur la présence ou non d'un défaut dans la pièce inspectée était prise après avoir comparé une valeur caractéristique du signal à un seuil. Mais en introduisant ce seuil, on biaise l'évaluation de la performance du capteur dans le sens où le seuil choisi n'est pas forcément le seuil optimal. Plus précisément, il peut arriver que les qualités intrinsèques du signal en fasse un bon outil diagnostique (le signal est bien différent selon qu'il provient d'un défaut ou qu'il ne correspond qu'à un bruit), mais que le seuil soit imposé d'une telle façon que l'on n'en tire pas profit. Nous avons vu d'autre part (paragraphe II.3.B) que POD et PFA sont deux indicateurs indissociables dans l'évaluation du capteur.

L'idée sur laquelle repose la courbe ROC est par conséquent de représenter dans un même repère la PFA et la POD tout en faisant apparaître l'influence de la valeur du seuil sur la performance du capteur.

En modifiant la valeur du seuil de décision, on fait varier les conclusions et par conséquent les valeurs de la POD et de la PFA. En parcourant l'ensemble des seuils possibles, on obtient tous les couples (PFA, POD).

**Construire la courbe ROC revient à représenter dans un même graphique les points  $\{(PFA, POD)\}_{seuil \in \mathcal{S}}$  où  $\mathcal{S}$  représente l'ensemble des seuils.**

La Figure 19 donne un exemple de courbe ROC. La littérature concernant ces courbes comprend plus de cinq mille articles et est par conséquent très riche. Nous ne présenterons donc ici qu'une introduction aux courbes ROC. Pour une revue « élémentaire », on pourra lire l'article de Morin & Al (1998) sur les courbes ROC en biologie médicale [8] ou encore l'introduction aux courbes ROC rédigée par Fawcett



[9]. Pour un tour d'horizon plus général, on consultera le livre de Zhou & Ai (2002) [10], l'ouvrage de M.S. Pepe (2004) [11], ou encore celui de Krzanowski & Hand (2009) [12] sur les courbes ROC pour les données continues (voir partie II.3.C.b.iv). Pour une approche différente appliquée dans le cadre du traitement du signal, on pourra s'intéresser au livre de Van Trees [13].

Le contexte du paragraphe II.3.B et de la Yes-No method est un cas particulier de celui considéré dans le modèle de Berens dans le sens où la taille des défauts n'est pas variable. Pour souligner la distinction, nous n'utiliserons pas les mêmes conventions. Ainsi, la **valeur caractéristique du signal sera notée S** tandis que l'on fera référence au **seuil de décision par  $s_{th}$** .

De façon plus générale, la quantité associée à S est appelée le **score de classification**. D'autre part, on introduit une **variable binaire  $\delta$  indiquant la présence ou non d'un défaut dans la pièce inspectée** :

$$\delta = \begin{cases} 1 & \text{si un défaut est présent} \\ 0 & \text{sinon} \end{cases}$$

Ainsi, la probabilité de détection et la probabilité de fausse alarme s'expriment comme :

$$POD = P(S \geq s_{th} / \delta = 1) \quad PFA = P(S \geq s_{th} / \delta = 0)$$

Muni de ces définitions et conventions, nous pouvons préciser certaines propriétés des courbes ROC :

<b>Propriété 1 :</b> Les points (0,0) et (1,1) appartiennent toujours à la courbe ROC.
<b>Propriété 2 :</b> Le point (PFA, POD) = (0,1) correspond à un détecteur parfait.
<b>Propriété 3 :</b> La courbe ROC est inaltérée si le score de classification (S) subit une transformation monotone.
<b>Propriété 4 :</b> Si $S/\delta = 0$ est stochastiquement plus petite que $S/\delta = 1$ , la courbe ROC est au-dessus de la diagonale de chance.
<b>Propriété 5 :</b> La pente de la courbe ROC correspond au rapport de vraisemblance.
<b>Propriété 6 :</b> Si les densités $f_{S/P}$ et $f_{S/N}$ ont un rapport de vraisemblance monotone (avec S la statistique associée), alors la courbe ROC est concave.
<b>Propriété 7 :</b> L'aire sous la courbe ROC (AUC) est égale à $P(S_p > S_N)$ où $S_N \equiv S/N$ et $S_p \equiv S/P$

Les démonstrations de ces propriétés sont données en Annexes. Pour les propriétés 3 à 5 et 7, on pourra se référer à [12]. La propriété 6 est énoncée dans [14].

## b. Estimations

La définition de la courbe ROC repose sur la probabilité de détection et la probabilité de fausse alarme. Or on peut remarquer que des deux dernières quantités s'expriment comme des fonctions de répartition des lois conditionnelles  $S/\delta = 1$  et  $S/\delta = 0$ , que l'on notera  $S/N$  et  $S/P$  :

$$\begin{aligned} P(S \geq s_{th} / \delta = 1) &= POD_{s_{th}} = 1 - F_{S/P}(s_{th}) \\ P(S \geq s_{th} / \delta = 0) &= PFA_{s_{th}} = 1 - F_{S/N}(s_{th}) \end{aligned} \quad (23)$$

où  $F_{S/P}$  et  $F_{S/N}$  sont respectivement les fonctions de répartition de  $S/N$  et  $S/P$ .

On a vu que la courbe ROC correspondait à l'ensemble des couples (PFA, POD) pour les différentes valeurs du seuil et que la POD et la PFA pouvaient s'exprimer (équation (23)) à l'aide des fonctions de répartition de  $S/P$  et  $S/N$ , on a donc :

$$\text{courbe ROC} \equiv \left\{ \left( 1 - F_{S/N}(s_{th}), 1 - F_{S/P}(s_{th}) \right), s_{th} \in \mathcal{S} \right\} \quad (24)$$

On l'a vu dans la démonstration des propriétés de la courbe ROC, celle-ci peut être vue comme une courbe paramétrée par le seuil. D'après (23), on a  $s_{th} = F_{S/N}^{-1}(1 - PFA)$ , et par suite :

$$POD = 1 - F_{S/P} \left( F_{S/N}^{-1}(1 - PFA) \right) \quad (25)$$

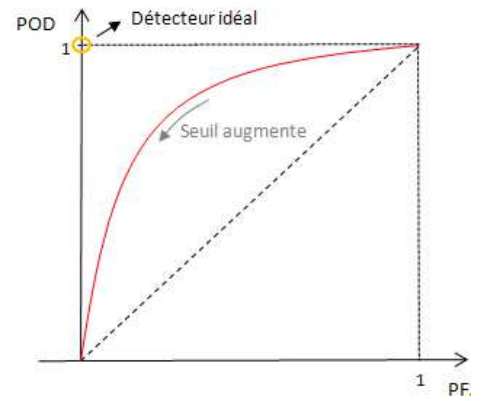


Figure 17 : Exemple de courbe ROC

Il faut savoir que l'estimation de ces courbes dépend du type de données recueillies. Il est naturel de s'adapter au cadre étudié comme l'estimation de la POD s'ajuste au cas succès-échec et au cas signal-réponse. Il convient par conséquent de décrire les différents contextes possibles avant d'aborder le ou les méthodes d'estimation qui leur sont associées.

### i. Cadre commun

- Estimation non-paramétrique – Principe général

Dans tous les cas de figure rencontrés, il est possible de proposer une estimation non paramétrique de la courbe ROC reposant sur la fonction de répartition empirique de  $S_N$  et de  $S_P$ . Le paragraphe suivant expose le principe général de la méthode. De petites adaptations sont nécessaires selon le type de données étudiées. Ces dernières seront précisées dans les paragraphes dédiés aux méthodes propres à chaque genre de données.

De façon « basique », ces estimations reposent sur les estimateurs non paramétriques des fonctions de répartition liées à la valeur caractéristique du signal dans les populations de pièces avec et sans défaut :

$$\hat{F}_{S/N}(x) = \frac{1}{n_0} \sum_{i: \delta_i=0} 1_{\{s_i \leq x\}} \quad \text{et} \quad \hat{F}_{S/P}(x) = \frac{1}{n_1} \sum_{i: \delta_i=1} 1_{\{s_i \leq x\}}$$

où  $n_0 = \text{card}(\{i, \delta_i = 0\})$  et  $n_1 = \text{card}(\{i, \delta_i = 1\})$ .

L'estimation de la courbe ROC consiste alors à remplacer dans l'équation (24), les fonctions de répartition par leur version empirique :

$$\text{courbe } \widehat{ROC} \equiv \left\{ \left( 1 - \hat{F}_{S/N}(s_{th}), 1 - \hat{F}_{S/P}(s_{th}) \right), s_{th} \in \mathcal{S} \right\}$$

ou encore d'après (25) :

$$\widehat{ROC}(x) = 1 - \hat{F}_{S/P} \left( \hat{F}_{S/N}^{-1}(1-x) \right) \quad (26)$$

Le théorème de Glivenko-Cantelli assure la convergence des fonctions de répartition empirique :

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} | \hat{F}_n(x) - F(x) | = 0$$

Une quantité importante de données est nécessaire pour que l'estimation soit relativement juste. Généralement, les échantillons sont de taille faible à modéré. Dans ces circonstances, les courbes ROC obtenues sont des fonctions en escalier avec des paliers très prononcés. Leur allure n'est donc pas lisse.

Afin d'éviter ce désagrément, il est possible d'utiliser des méthodes de lissage par noyaux. Celles-ci permettent de donner une estimation lisse des densités, et par suite, une version des fonctions de répartition empiriques ayant cette même propriété.

Considérons deux échantillons  $(u_1, \dots, u_{n_0})$  et  $(v_1, \dots, v_{n_1})$  issus respectivement des variables  $S/N$  et  $S/P$ . Les estimateurs des densités par les méthodes des noyaux gaussiens sont définis par :

$$f_{S/N}(x) = \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{1}{h_N} \phi \left( \frac{x - u_i}{h_N} \right) \quad \text{et} \quad f_{S/P}(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{h_P} \phi \left( \frac{x - v_i}{h_P} \right)$$

où  $\phi$  est la densité d'une  $\mathcal{N}(0,1)$  et les  $h_k, k \in \{N, P\}$  sont les fenêtres de lissage.

Il suffit alors d'intégrer ces fonctions pour trouver les fonctions de répartition correspondantes :

$$\hat{F}_{S/N}(x) = \int_{-\infty}^x \hat{f}_{S/N}(u) du = \frac{1}{n_0} \sum_{i=1}^{n_0} \int_{-\infty}^x \frac{1}{h_N} \phi \left( \frac{t - u_i}{h_N} \right) dt = \frac{1}{n_0} \sum_{i=1}^{n_0} \Phi \left( \frac{x - u_i}{h_N} \right)$$

où  $\phi$  et  $\Phi$  sont respectivement la densité et la fonction de répartition d'une  $\mathcal{N}(0,1)$ .

On obtient de la même façon  $\hat{F}_{S/P}$ , d'où :

$$\hat{F}_{S/N}(x) = \frac{1}{n_0} \sum_{i=1}^{n_0} \Phi \left( \frac{x - u_i}{h_N} \right) \quad \text{et} \quad \hat{F}_{S/P}(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \Phi \left( \frac{x - v_i}{h_P} \right)$$

Avec cette approche, il convient de fixer correctement les fenêtres  $h_k, k \in \{N, P\}$  régissant le degré de lissage apporté aux fonctions de répartition. D'autre part, les échantillons devront garder une taille conséquente. Pour plus d'information, on pourra consulter [12].

- Estimation paramétrique – Remarques générales

L'approche paramétrique permet d'apporter nos connaissances sur le phénomène ayant généré les données. Les avantages sont multiples. La quantité de données nécessaires est généralement plus faible que pour l'estimation non-paramétrique et l'introduction des paramètres est souvent appréciée à des fins d'interprétations du modèle.

Dans le contexte des courbes ROC, l'estimation paramétrique fait en réalité référence à des procédés dits **fully parametric** que l'on distingue, dans le cadre des données continues, des méthodes d'estimation **semi-paramétriques**. Ces dernières privilégient une approche paramétrique « limitée » en réduisant les contraintes liées aux hypothèses sans pour autant perdre en généralité.

L'approche « fully parametric » correspond à la démarche classique dans le sens où l'on fait des hypothèses paramétriques directement sur les deux distributions d'intérêt  $S/N$  et  $S/P$ . On supposera de plus que les lois des variables appartiennent à une même famille :

$$S/N \sim \mathcal{L}(\theta_N) \quad \text{et} \quad S/P \sim \mathcal{L}(\theta_P) \quad (27)$$

où  $\mathcal{L}$  est une loi continue, et  $\theta$  un vecteur de paramètres.

Les possibilités pour  $\mathcal{L}$  sont nombreuses mais parmi les plus communes on trouve la loi normale, la loi logistique, la loi de Weibull, etc.

En pratique, on s'intéresse particulièrement aux modèles de translation – changement d'échelle (voir le paragraphe sur les bandes de confiance de la partie II.2.D) possédant des propriétés agréables vis-à-vis de la courbe ROC. Comme on l'a vu, les fonctions de répartition de ces modèles s'écrivent à partir d'une fonction de répartition canonique. Dans le cas de l'équation (27), notons  $F$  cette fonction et posons  $\theta = (\mu, \sigma^2)$ . La courbe ROC s'exprime alors en fonction de  $F$  :

$$\text{courbe ROC} \equiv \left\{ \left( 1 - F\left(\frac{t - \mu_N}{\sigma_N}\right), 1 - F\left(\frac{t - \mu_P}{\sigma_P}\right) \right), t \in \mathcal{S} \right\}$$

Posons :

$$y = 1 - F\left(\frac{t - \mu_P}{\sigma_P}\right) = \text{POD} \quad \text{et} \quad x = 1 - F\left(\frac{t - \mu_N}{\sigma_N}\right) = \text{PFA}$$

Alors  $t$  s'écrit en fonction de  $x$  :

$$t = \sigma_N F^{-1}(1 - x) + \mu_N$$

Et on injecte  $t$  dans l'expression de  $y$  :

$$y = 1 - F\left(\frac{\sigma_N}{\sigma_P} F^{-1}(1 - x) - \frac{\mu_P - \mu_N}{\sigma_P}\right) \quad (28)$$

Si de plus, la fonction  $F$  vérifie la propriété de « symétrie » (symétrie vis-à-vis de la loi, cf la loi normale par exemple) suivante :

$$1 - F(x) = F(-x) \quad \forall x \in \mathbb{R}$$

Il est possible de faire apparaître une liaison linéaire entre  $F^{-1}(y)$  et  $F^{-1}(x)$  :

$$y = F\left(\frac{\sigma_N}{\sigma_P} F^{-1}(x) + \frac{\mu_P - \mu_N}{\sigma_P}\right)$$

Soit :

$$\text{ROC}(x) = F\left(\frac{\sigma_N}{\sigma_P} F^{-1}(x) + \frac{\mu_P - \mu_N}{\sigma_P}\right) \quad (29)$$

Cette propriété permet de représenter la courbe ROC sous la forme d'une droite dans un repère muni d'une échelle probit. En se rappelant que  $y = \text{ROC}(x) = \text{POD}$  et  $x = \text{PFA}$ , on a en effet :

$$F^{-1}(\text{POD}) = \frac{\sigma_N}{\sigma_P} F^{-1}(\text{PFA}) + \frac{\mu_P - \mu_N}{\sigma_P}$$

Après changement de paramètres, cette équation devient :

$$F^{-1}(\text{POD}) = a F^{-1}(\text{PFA}) + b \quad \text{où} \quad \begin{cases} a = \frac{\sigma_N}{\sigma_P} \\ b = \frac{\mu_P - \mu_N}{\sigma_P} \end{cases} \quad (30)$$

Tandis que l'équation (29) donne :

$$\text{ROC}(x) = F(a F^{-1}(x) + b) \quad (31)$$

Instinctivement, on serait tenté d'estimer  $a$  et  $b$  par régression linéaire. Cependant, les moindres carrés conventionnels sont inappropriés dans ce cas. En effet, les deux axes représentent des variables dépendantes puisqu'a priori, ni la POD ni la PFA ne sont connues. Ils sont par conséquent sujets à des erreurs. Dans ce cas, utiliser les moindres carrés ou toute autre procédure minimisant la distance verticale des points empiriques à la droite ajustée ne fournit pas la variance des estimateurs. De plus, les points de la courbe ROC peuvent s'avérer corrélés (voir le cas de la rating method) ce qui n'entre pas dans les hypothèses de la régression linéaire classique. Des estimations basées sur le maximum de vraisemblance ont ainsi été mises au point, contenant souvent des procédés itératifs pour lesquels les estimations de  $a$  et  $b$  par moindres carrés peuvent constituer de bonnes valeurs initiales.

L'avantage des modèles *fully parametric* se trouve dans le fait qu'ils permettent d'obtenir une courbe ROC lisse avec une quantité de données moins importante comparé aux modèles non-paramétriques. Ils reposent néanmoins sur des hypothèses fortes qu'il est

nécessaire de vérifier. La robustesse de ces modèles n'est par ailleurs pas assurée, et la propriété d'invariance des courbes ROC n'est pas respectée.

Les modèles semi-paramétriques abordés ci-après permettent de contourner certaines de ces difficultés. Leurs « bonnes » propriétés ont rapidement pris le dessus et les recherches sur les modèles purement paramétriques ont rapidement déclinées.

• Estimation semi-paramétrique – Remarques

Les inconvénients liés aux modèles fully parametric ont motivé l'introduction des modèles semi-paramétriques. Si l'affaiblissement des hypothèses les caractérisant leur permet de couvrir un éventail de situations plus large, cela implique en contrepartie l'apparition potentielle d'un « coût statistique » à travers des *paramètres de nuisance*. Afin de s'en débarrasser, il sera nécessaire de les estimer au préalable ou de les « contourner » par l'intermédiaire de techniques bayésiennes par exemple.

De nombreuses approches semi-paramétriques sont proposées et dépendent fortement du type de données disponible. Comme nous le verrons dans les parties suivantes, les développements les plus importants concernent les données continues (scoring) mais les premières méthodes mises au point traitaient le cas des données binaires et ordinales.

Une sous-catégorie de méthodes semi-paramétriques ayant grandement évolué et prenant une place très largement majoritaire englobe les modèles dits « **distribution free** ». Cette approche exploite la propriété 3 des courbes ROC selon laquelle une transformation monotone n'affecte pas la forme de la courbe. L'hypothèse structurelle consiste donc à supposer qu'il existe une transformation monotone  $h$  des variables  $S/N$  et  $S/P$  telle que :

$$h(S/N) \sim \mathcal{L} \quad \text{et} \quad h(S/P) \sim \mathcal{L}(\theta)$$

où  $\mathcal{L}$  est une loi continue et  $\theta$  un vecteur de paramètres.

Ainsi, dans le cas où (voir [14]) :

- $\mathcal{L} \equiv$  loi normale                    **modèle binormal**
- $\mathcal{L} \equiv$  loi logistique                **modèle à rapport des chances proportionnel**
- $\mathcal{L} \equiv$  loi de Weibull                **modèle à risques proportionnels**

Concernant la transformation  $h$ , un exemple classique est celui où l'on souhaite se replacer dans un cadre de normalité des distributions par une transformation de type Box-Cox. La variable considérée est alors  $S^{(\lambda)}$  obtenue par la transformation suivante :

$$S^{(\lambda)} = \begin{cases} \frac{S^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(S) & \text{si } \lambda = 0 \end{cases}$$

où  $\lambda$  est un paramètre éventuellement à estimer.

Dans le cas où le paramètre  $\lambda$  de la transformation Box-Cox n'est pas connu, il est considéré comme un paramètre de nuisance. Il arrive que certaines modélisations supposent simplement l'existence d'une transformation monotone  $h$  dont la forme n'est pas précisée.

Il est intéressant de remarquer que la structure des modèles *distribution-free* rappelle celle des modèles *fully parametric*. Toutefois, l'introduction de la fonction  $h$  permet, en quelque sorte et sans perdre en généralité, de ramener la distribution de  $S/N$  à la version canonique de la loi  $\mathcal{L}$  ne dépendant pas des paramètres. Les modèles de translation – changement d'échelle joueront là encore un rôle primordial. L'important n'est pas de localiser précisément la distribution de  $S/N$  mais plutôt d'étudier la position relative de  $S/P$  par rapport à celle-ci. C'est d'ailleurs cette idée qui va être grandement exploitée ici, notamment à travers des statistiques faisant intervenir les rangs des observations.

Le modèle le plus étudié et le plus utilisé est de loin le **modèle binormal**, dans lequel on considère une variable  $W$  vérifiant :

$$W/N = h(S/N) \sim \mathcal{N}(0,1) \quad \text{et} \quad W/P = h(S/P) \sim \mathcal{N}(\mu, \sigma^2)$$

La propriété d'invariance de la courbe ROC se traduit alors par :

$$P(S > c/K) = P(h(S) > h(c)/K) = P(W > t/K)$$

où  $t = h(c)$  et  $K \in \{N, P\}$

ii. Cadre de la Yes-No method - Données binaires

Il correspond à une situation où l'accès à  $S$  n'est que partiel.

C'est typiquement le cas pour les données Hit-Miss en CND, où  $S$  correspondrait au signal. Dans ce contexte, les données contiennent uniquement une information binaire : le capteur est considéré comme une boîte noire traitant le signal. Ainsi, les observations se limitent à des réalisations d'une variable indicatrice  $\delta$  indiquant le résultat de l'expérimentation. Elle suppose l'existence d'un seuil  $c$  relatif à  $S$  et tel que :

$$\delta = 1_{\{S > c\}} = \begin{cases} 1 & \text{si } S > c \\ 0 & \text{sinon} \end{cases}$$

Les valeurs prises par  $\delta$  sont généralement associées à une décision de l'opérateur. Dans le cadre du CND par exemple, si ce dernier observe  $\delta = 1$ , il signalera la présence d'un défaut. Il conclura à l'absence de défaut dans le cas contraire.

Pour se placer dans des conditions où l'estimation de la courbe ROC est possible, il est nécessaire de constituer un échantillon vérifiant plusieurs propriétés. Il doit en effet contenir des données provenant d'un bruit ainsi que des données issues d'un signal couplé à du bruit, et cela, pour plusieurs valeurs du seuil :

$$\text{échantillon} = \bigcup_{c_l \in \mathcal{C}} \left\{ \{\delta_i^{N,c_l}\}_{i=1 \dots n_N} \cup \{\delta_j^{P,c_l}\}_{j=1 \dots n_P} \right\}$$

où  $\mathcal{C}$  est l'ensemble des seuils et  $\delta^{k,c_l} = 1_{\{S > c_l/K\}}$  avec  $k \in \{N, P\}$ .

Une fois les données recueillies, une table de confusion similaire à celle du Tableau 1 sera dressée pour chaque valeur du seuil :

Seuil $c_l$	Réalités	
	r+	r-
Décisions		
d+	$x_{l,P} = \sum_{i=1}^{n_P} \delta_i$	$x_{l,N} = \sum_{j=1}^{n_N} \delta_j$
d-	$n_P - x_{l,P}$	$n_N - x_{l,N}$

Tableau 3 : Table de confusion pour le seuil  $c_l$  – Yes/No method

- Estimation non-paramétrique

Comme précisé dans la partie précédente, il est possible d'estimer la courbe ROC de façon non-paramétrique. Il suffit en effet de remarquer que :

$$\hat{F}_{S/N}(c) = \frac{1}{n_N} \sum_{i: \delta_i=0} 1_{\{s_i \leq c\}} = \frac{1}{n_N} \sum_{i: \delta_i=0} (1 - \delta_i^N)$$

$$\hat{F}_{S/P}(c) = \frac{1}{n_P} \sum_{i: \delta_i=1} 1_{\{s_i \leq c\}} = \frac{1}{n_P} \sum_{i: \delta_i=1} (1 - \delta_i^P)$$

On peut aussi éventuellement appliquer les méthodes de lissages précédemment décrites.

- Estimation semi-paramétrique

Dorfman & Alf (1968) [15] ainsi que Ogilvie & Creelman (1968) [16] ont proposé un algorithme d'estimation des paramètres a et b basé sur une estimation par maximum de vraisemblance. Ils ont respectivement traité le cas où  $\mathcal{L}$  est une loi normale et le cas où  $\mathcal{L}$  est une loi logistique.

Etudions le cas de la loi normale :

Pour rappel, dans le contexte de la Yes-No method, bien que l'on considère S comme étant une variable continue, les données recueillies ne sont que binaires. Le modèle *distribution-free* suppose d'autre part qu'il existe une transformation monotone croissante h et une variable latente W telles que :

$$\begin{aligned} S > c &\Leftrightarrow h(S) > h(c) \Leftrightarrow W > t \\ S \leq c &\Leftrightarrow h(S) \leq h(c) \Leftrightarrow W \leq t \end{aligned} \quad \text{et } W/N \sim \mathcal{N}(0,1), \quad W/P \sim \mathcal{N}(\mu_P, \sigma_P^2) \quad (32)$$

Par conséquent, les données observées  $\delta_{l,i}$  vérifient :

$$\delta_{l,i} = 1_{\{s_i > c_l\}} = 1_{\{h(s_i) > h(c_l)\}} = 1_{\{w_i > t_l\}}$$

Le but est donc de remonter à la courbe ROC liée à S en utilisant la structure binaire des données et l'hypothèse de binormalité de W.

Il est important de remarquer que la transformation h est inconnue. Par conséquent, les valeurs des  $t_l$  le sont aussi. Elles font ainsi partie des paramètres du modèle et devront être estimées.

Après avoir effectué les m expériences indépendantes associées aux m valeurs du seuil  $\{c_l, l \in \{1, \dots, m\}\}$ , il est possible de dresser autant de table de confusion. Elles sont identiques à celle exhibée dans le Tableau 3.

Le modèle considéré revient à supposer que les données de chaque population N et P suivent une loi binomiale.

Notons  $x_{l,k}$  l'effectif relatif à case issue du croisement de la première ligne et de la colonne  $k$  ( $k \in \{N, P\}$ ) de la  $l^{ième}$  table de confusion (voir Tableau 3). Alors :

$$x_{l,k} = \text{card}(\{\delta_{i,l} \text{ tq } \delta_{i,l} = 1\}) = \sum_i \delta_{i,l}$$

Utilisons maintenant les propriétés du modèle. L'invariance de la courbe ROC par transformation monotone et les hypothèses (27) donnent les versions théoriques de la POD et de la PFA :

$$\begin{aligned} P_{l,P} &= P(S > c_l/P) = P(W > t_l/P) = 1 - \Phi\left(\frac{t_l - \mu_P}{\sigma_P}\right) = \Phi\left(\frac{\mu_P - t_l}{\sigma_P}\right) = \Phi(b - a t_l) \\ P_{l,N} &= P(S > c_l/N) = P(W > t_l/N) = 1 - \Phi(t_l) = \Phi(-t_l) \end{aligned} \quad (33)$$

avec :

$$a = \frac{1}{\sigma_P} \quad \text{et} \quad b = \frac{\mu_P}{\sigma_P}$$

La vraisemblance est celle relative à 2m lois binomiales :

$$L = \prod_{k \in \{N, P\}} \prod_{l \in \{1, \dots, m\}} C_{n_k}^{x_{l,k}} P_{l,k}^{x_{l,k}} (1 - P_{l,k})^{n_k - x_{l,k}}$$

Et la log-vraisemblance est par conséquent :

$$\log L = \sum_{k \in \{N, P\}} \sum_{l \in \{1, \dots, m\}} x_{l,k} \log P_{l,k} + (n_k - x_{l,k}) \log(1 - P_{l,k}) + \text{termes constants}$$

où les  $P_{l,k}$  sont donnés par (33).

Les paramètres à estimer sont donc  $a, b$  et les  $t_l$ , soit  $m + 2$  paramètres. Il y a ainsi  $m + 2$  équations de vraisemblances non linéaires à résoudre :

$$\frac{\partial \log L}{\partial t_l} = 0 \text{ pour } l = 1, \dots, m ; \quad \frac{\partial \log L}{\partial a} = 0 ; \quad \frac{\partial \log L}{\partial b} = 0$$

Ce système peut être résolu numériquement à l'aide d'une version légèrement modifiée de l'algorithme itératif de Newton-Raphson appelée la *méthode du scoring*. Pour converger, celui-ci demande de « bonnes » valeurs initiales. On propose ci-après un procédé permettant de choisir un ensemble de valeurs initiales.

Chaque table de confusion permet de donner une estimation de la POD et de la PFA (voir Tableau 2 et Tableau 3) :

$$\widehat{POD}_l \triangleq p_{l,P} = \frac{x_{l,P}}{n_P} = \frac{\sum_{i=1}^{n_P} \delta_i}{n_P} \quad \text{et} \quad \widehat{PFA}_l \triangleq p_{l,N} = \frac{x_{l,N}}{n_N} = \frac{\sum_{j=1}^{n_N} \delta_j}{n_N}$$

La courbe ROC « empirique » est donc constituée de  $m$  points :

$$\text{Points de la courbe ROC empirique} \equiv \{(p_{l,N}, p_{l,P}), l : c_l \in \mathcal{S}\} \quad (34)$$

D'autre part, les équations (33) permettent de donner les approximations suivantes :

$$\begin{cases} p_{l,N} \cong \Phi(-t_l) \\ p_{l,P} \cong \Phi(b - a t_l) \end{cases} \Leftrightarrow \begin{cases} \Phi^{-1}(p_{l,N}) \cong -t_l \\ \Phi^{-1}(p_{l,P}) \cong b - a t_l \end{cases}$$

Les valeurs initiales attribuées  $a$  et  $b$  correspondent aux estimations issues de la régression linéaire :

$$\Phi^{-1}(p_{l,P}) = a \Phi^{-1}(p_{l,N}) + b$$

Enfin, les  $-\Phi^{-1}(p_{l,N})$  serviront de valeurs initiales pour les  $t_l$ .

La qualité d'ajustement du modèle peut être mesurée de façon classique à l'aide d'un  $\chi^2$  comparant les effectifs empiriques aux effectifs théoriques :

$$\chi^2(m - 2) = \sum_{l \in \{1, \dots, m\}} \frac{(n_P p_{l,P} - n_P P_{l,P})^2}{n_P P_{l,P}} + \frac{(n_N p_{l,N} - n_N P_{l,P})^2}{n_N P_{l,N}}$$

### iii. Cadre de la Rating Method - Données ordinales

Dans le contexte de la rating method, les données sur  $S$  sont plus riches, mais restent partielles. Elles sont constituées d'observations d'une variable ordinaire  $z$  à  $p + 1$  catégories liée à  $S$  à travers un ensemble de seuils  $(c_k)_{k=1, \dots, p}$  tels que :

$$z = \begin{cases} z^{(1)} & \text{si } S < c_1 \\ z^{(k+1)} & \text{si } c_k < S < c_{k+1} \text{ pour } k < p \\ z^{(p+1)} & \text{si } S > c_p \end{cases}$$

où les  $z^{(l)}$  représenteront les différentes catégories de réponse possibles de l'opérateur (voir le Tableau 4). Ces données contiennent d'avantage d'information et permettent ainsi à l'opérateur de nuancer sa décision : selon la catégorie de la variable  $z$  qu'il aura observée, il indiquera sa décision ainsi que le niveau de certitude qu'il lui accorde.

Afin d'estimer la courbe ROC, il sera toujours nécessaire d'établir un échantillon comportant des données provenant d'un bruit ainsi que des données issues d'un signal couplé à un bruit. La variation de la valeur du seuil est cette fois prise en compte de façon intrinsèque à travers la définition de la variable  $z$  :

$$\text{échantillon} = \left\{ \bigcup_{l \in \{1, \dots, p+1\}} \{z_i^{(l), N}\}_{i=1, \dots, n_N} \right\} \cup \left\{ \bigcup_{l \in \{1, \dots, p+1\}} \{z_j^{(l), P}\}_{j=1, \dots, n_P} \right\}$$

Comme nous le verrons, l'inclusion simultanée de l'ensemble des seuils va introduire une dépendance entre les points de la courbe ROC. La méthode d'estimation de cette dernière devra s'adapter afin de prendre en compte ce phénomène.

Un exemple de table de confusion correspondant à ce cas de figure est présenté dans le Tableau 4 ci-dessous ( $p=5$ ).

Décisions		Niveau de certitude		Réalités	
				r+	r-
d+	Sûr	$S > c_5$			
	Moyen	$c_4 < S < c_5$			
	Pas sûr	$c_3 < S < c_4$			
d-	Pas sûr	$c_2 < S < c_3$			
	Moyen	$c_1 < S < c_2$			
	Sûr	$S < c_1$			

Tableau 4 : Table de confusion – Rating method ( $p=5$ )

- Estimation non-paramétrique

Ce sujet n'a pas été abordé en détails dans cette étude.

- Estimation semi-paramétrique

Dorfman & Alf (1969) [17] ainsi que Ogilvie & Creelman (1968) [16] ont élargi leur démarche au cas des données ordinales. Etudions de la même façon le modèle basé sur la loi normale.

Pour rappel, dans le contexte de la rating method, bien que l'on considère  $S$  comme étant une variable continue, les données recueillies ne sont qu'ordinales. Le modèle *distribution-free* suppose d'autre part qu'il existe une transformation monotone croissante  $h$  et une variable latente  $W$  telles que :

$$\begin{cases} S < c_1 \\ \vdots \\ c_j < S < c_{j+1} \\ \vdots \\ S > c_m \end{cases} \Leftrightarrow \begin{cases} W < t_1 \\ \vdots \\ t_l < W < t_{l+1} \\ \vdots \\ W > t_m \end{cases} \text{ avec : } \begin{cases} W/N \sim \mathcal{N}(0,1) \\ W/P \sim \mathcal{N}(\mu_P, \sigma_P^2) \end{cases} \quad (35)$$

Les données observées  $z_i$  vérifient donc :

$$z_i = \begin{cases} z^{(1),k} & \text{si } W_k < t_1 \\ z^{(l+1),k} & \text{si } t_l < W_k < t_{l+1} \text{ pour } l < m \\ z^{(m+1),k} & \text{si } W_k > t_m \end{cases}$$

où  $W_k = W/k$  avec  $k \in \{N, P\}$

Le but cette fois de remonter à la courbe ROC liée à  $S$  en utilisant la structure ordinale des données et l'hypothèse de binormalité de  $W$ .

Comme précédemment, il est important de remarquer que la transformation  $h$  étant inconnue, les valeurs des  $t_l$  le sont aussi. Elles font ainsi partie des paramètres du modèle et devront être estimées.

Après avoir effectué les  $m$  expériences indépendantes associées aux  $m$  valeurs du seuil  $\{c_l, l \in \{1, \dots, m\}\}$ , on construira la table de confusion associée (voir Tableau 4). Il n'y a donc plus qu'une seule table de confusion dans laquelle tous les seuils apparaissent. En extraire une estimation de la POD et de la PFA pour chaque seuil implique de cumuler certaines quantités.

Le modèle considéré revient à supposer que les données de chaque population  $N$  et  $P$  suivent une loi multinomiale.

Notons  $x_{l,k}$  l'effectif attaché à la case de la table de confusion repérée par la  $l^{\text{ième}}$  ligne avec  $l \in \{1, \dots, m\}$  et la colonne  $k$  avec  $k \in \{N, P\}$ . On a donc :

$$x_{l,k} = \text{card}(\{z_i \text{ tq } z_i = z^{(l,k)}\}) = \sum_i 1_{\{z_i = z^{(l,k)}\}}$$

Utilisons maintenant les propriétés du modèle. Remarquons d'abord que :

$$P(t_l < W < t_{l+1}) = P(W < t_{l+1}) - P(W < t_l)$$

L'invariance par transformation monotone et les hypothèses (32) donnent :

$$\begin{aligned} P_{l,P} &= P(c_l < S < c_{l+1}/P) = P(t_l < W < t_{l+1}) = \Phi\left(\frac{t_{l+1} - \mu_P}{\sigma_P}\right) - \Phi\left(\frac{t_l - \mu_P}{\sigma_P}\right) \\ P_{l,N} &= P(c_l < S < c_{l+1}/N) = P(t_l < W < t_{l+1}) = \Phi(t_{l+1}) - \Phi(t_l) \end{aligned} \tag{36}$$

En posant  $a = \frac{1}{\sigma_P}$  et  $b = \frac{\mu_P}{\sigma_P}$ , il vient :

$$\begin{cases} P_{l,P} = \Phi(a t_{l+1} - b) - \Phi(a t_l - b) \\ P_{l,N} = \Phi(t_{l+1}) - \Phi(t_l) \end{cases}$$

La vraisemblance est celle relative à  $2m$  lois multinomiales :

$$L = \prod_{k \in \{N, P\}} \prod_{l \in \{1, \dots, m\}} \frac{(\sum_l x_{l,k})!}{x_{1,k}! \dots x_{m,k}!} P_{l,k}^{x_{l,k}}$$

On en déduit la log-vraisemblance, qui s'écrit :

$$\log L = \sum_{k \in \{N, P\}} \sum_{l \in \{1, \dots, m\}} x_{l,k} \log P_{l,k} + \text{termes constants}$$

Les paramètres à estimer sont donc  $a, b$  et les  $t_l$ , soit  $m + 2$  paramètres. Il y a ainsi  $m + 2$  équations de vraisemblances non linéaires à résoudre :

$$\frac{\partial \log L}{\partial t_l} = 0 \text{ pour } l = 1, \dots, m ; \quad \frac{\partial \log L}{\partial a} = 0 ; \quad \frac{\partial \log L}{\partial b} = 0$$

Ici encore, afin de résoudre numériquement ces équations, on utilise la méthode du scoring (Newton-Raphson adapté). On détaille ci-après une façon de choisir une valeur initiale pour chacun des paramètres.

Pour chaque valeur du seuil, une estimation empirique de la POD et de la PFA peut être déterminée :

$$\widehat{POD}_{c_l} \triangleq p_{l,P} = \frac{\sum_{j \geq l} x_{j,P}}{n_P} \quad \text{et} \quad \widehat{PFA}_{c_l} \triangleq p_{l,N} = \frac{\sum_{j \geq l} x_{j,N}}{n_N}$$

La courbe ROC « empirique » est donc constituée de  $m$  points qui sont cette fois-ci *corrélés* de par le cumul successif des catégories (équation (37)). On remarquera que la modélisation multinomiale des données permet de s'affranchir de cette corrélation.

$$\text{Points } \textit{corrélés} \text{ de la courbe ROC empirique} \equiv \{(p_{l,N}, p_{l,P}), l : c_l \in \mathcal{S}\} \tag{37}$$

D'autre part, nous avons les mêmes résultats pour  $POD_{c_l}$  et  $PFA_{c_l}$  que dans le cas des données binaires (équations (33)) :

$$\begin{cases} POD_{c_l} = P(W > t_l/P) = \Phi(b - a t_l) \\ PFA_{c_l} = P(W > t_l/N) = \Phi(-t_l) \end{cases}$$

Et par conséquent les mêmes approximations :

$$\begin{cases} p_{l,N} \cong \Phi(-t_l) \\ p_{l,P} \cong \Phi(b - a t_l) \end{cases} \Leftrightarrow \begin{cases} \Phi^{-1}(p_{l,N}) \cong -t_l \\ \Phi^{-1}(p_{l,P}) \cong b - a t_l \end{cases}$$

Pour valeurs initiales des paramètres, on prendra donc les estimations de  $a$  et  $b$  issues de la régression linéaire :

$$\Phi^{-1}(p_{l,P}) = a \Phi^{-1}(p_{l,N}) + b$$

et les valeurs  $-\Phi^{-1}(p_{l,N})$  pour les  $t_l$ .



La qualité d'ajustement du modèle peut être mesurée de façon classique à l'aide d'un  $\chi^2$  comparant les effectifs empiriques aux effectifs théoriques :

$$\chi^2(m-2) = \sum_{l \in \{1, \dots, m\}} \frac{(n_P p_{l,P} - n_P P_{l,P})^2}{n_P P_{l,P}} + \frac{(n_N p_{l,N} - n_N P_{l,N})^2}{n_N P_{l,N}}$$

#### iv. Cadre de la Scoring Method - Données continues :

Dans le cas continu, l'accès à S est direct. L'échantillon est constitué d'observations de cette variable, soit :

$$\text{échantillon} = \{\{S_i\}_{i=1, \dots, n_P}\} \cup \{\{S_i\}_{i=1, \dots, n_N}\}$$

Etant donné que S est à valeurs dans un espace unidimensionnel continu, il est naturel de penser à se ramener au cadre de la rating method ou de la Yes-No method en créant des classes artificielles. Mais en procédant de cette façon, on perd une grande partie de l'information contenue dans les données et transformer une variable continue en une variable catégorielle ou binaire n'est pas anodin. Certains principes doivent en effet être suivis et ceux-ci doivent s'adapter au contexte et aux objectifs de l'étude.

- Estimation non-paramétrique

Si les données recueillies sont continues, l'estimation non-paramétrique de la courbe ROC est directe. Il suffit d'appliquer les procédés décrits dans la partie faisant référence aux techniques non-paramétriques dans le cadre commun.

- Estimation semi-paramétrique

L'estimation semi-paramétrique pour les données continues est certainement celle ayant le plus intéressé les chercheurs. Il en résulte une profusion de techniques exploitant de nombreux principes statistiques.

Avant d'aborder le cas des modèles *distribution-free*, on peut mentionner quelques-unes des approches alternatives. Pour celles-ci, aucune transformation des données n'est envisagée, l'idée consiste juste à essayer de combiner les avantages des modèles paramétriques et des modèles non-paramétriques :

- Une des premières propositions, Li & Ai (1999) [18], repose sur une estimation non paramétrique de la distribution de S/N et sur l'introduction d'hypothèses paramétriques sur la distribution de S/P. Le maximum de vraisemblance est utilisé pour les estimations du modèle paramétrique. Aucune relation fonctionnelle entre les deux distributions n'est considérée.
- Une seconde approche, Quin & Zhang (1997,2003) [19,20], est basée sur une procédure de régression logistique. Les hypothèses sur la forme de la distribution de S/N sont abandonnées. Le principe est de « renverser » l'estimation en modélisant  $P(\delta = 1/S = s)$  conduisant à une expression paramétrique du rapport des vraisemblances :

Modèle logistique avec fonction de lissage r : 
$$P(\delta = 1/S = s) = \frac{e^{\alpha + \beta' r(s)}}{1 + e^{\alpha + \beta' r(s)}} = \psi(s)$$

En utilisant la formule de Bayes et en posant  $\pi = P(\delta = 1) = 1 - P(\delta = 0)$ , le rapport de vraisemblance peut s'écrire :

$$\frac{f(s/\delta = 1)}{f(s/\delta = 0)} = \frac{1 - \pi}{\pi} \frac{\psi(s)}{1 - \psi(s)} = e^{\log(\frac{1-\pi}{\pi}) + \alpha + \beta' r(s)} = e^{\alpha + \beta' r(s)}$$

Cette dernière quantité est appelée le *tilt exponentiel* et :

$$f(s/\delta = 1) = e^{\alpha + \beta' r(s)} \cdot f(s/\delta = 0)$$

On obtient ainsi un modèle semi-paramétrique à deux échantillons :

$$(*) \begin{cases} s_1^N, \dots, s_{n_N}^N \text{ iid de densité } f(s/\delta = 0) \\ s_1^P, \dots, s_{n_P}^P \text{ iid de densité } e^{\alpha + \beta' r(s)} \cdot f(s/\delta = 0) \end{cases}$$

La fonction de répartition  $F_N(s)$  est estimée :

- Soit en calculant la fonction de répartition empirique sur l'échantillon  $s_1^N, \dots, s_{n_N}^N$ .
- Soit en utilisant les deux échantillons et en se basant sur le modèle (\*).

La seconde méthode repose sur une estimation par maximum de vraisemblance des paramètres et aboutit à une forme modifiée de la fonction de répartition empirique avec des poids non uniformes.

En considérant l'échantillon complet  $\{t_1, \dots, t_n\} = \{s_1^N, \dots, s_{n_N}^N, s_1^P, \dots, s_{n_P}^P\}$ , on a ainsi :

$$\tilde{F}_N(t) = \sum_{i=1}^n \tilde{p}_i 1_{\{t_i \leq t\}}$$

Des statistiques de Kolmogorov-Smirnov couplées à une méthode de bootstrap permettent de tester la nullité simultanée des coefficients  $\alpha$  et  $\beta$  et la validité du modèle.

Intéressons-nous maintenant aux principales procédures traitant de l'estimation de la courbe ROC à l'aide des modèles *distribution-free*.

- Hsieh & Turnbull (1996) [14], ont préfiguré la méthode de Metz & Al (détaillée ci-après) mais en signalant que l'adaptation de la procédure de Dorfman & Alf aux données continues en nécessitait une catégorisation entraînant inévitablement une perte d'efficacité.

Ils ont par conséquent proposé une approche différente basée sur un « estimateur de la distance minimale ». Elle consiste à déterminer la courbe paramétrique binormale la plus proche de la courbe ROC empirique :

- Courbe ROC empirique :  $\widehat{ROC}(x) = 1 - \widehat{F}_{S/P}(\widehat{F}_{S/N}^{-1}(1-x))$
- Courbe ROC binormale :  $ROC(x) = \Phi(a \Phi^{-1}(x) + b)$

La divergence entre ces deux courbes est notée  $d(x)$  et est donnée par :

$$d(x) = \widehat{ROC}(x) - ROC(x) = 1 - \widehat{F}_{S/P}(\widehat{F}_{S/N}^{-1}(1-x)) - \Phi(a \Phi^{-1}(x) + b)$$

L'écart quadratique moyen entre les deux courbes correspond à l'intégrale suivante :

$$d^2 = \int_0^1 d^2(x) dx$$

Les estimateurs de la distance minimale de a et b sont par conséquent les valeurs de ces deux paramètres qui minimisent la distance moyenne quadratique  $d^2$  entre les deux courbes.

- Charles E. Metz (Metz & Al 1998) [21] ont présenté un algorithme encore très populaire (malgré la remarque de Hsieh & Turnbull sur le groupement des données et la perte d'efficacité) :

Leur idée consiste à grouper les données continues afin de se replacer dans le cadre de la procédure des Dorfman & Alf pour les données ordinales. Les catégories sont créées à partir des « truth-state runs », c'est-à-dire des statuts réels associés aux scores observés. Cette idée est plus claire sur un exemple. Considérons deux séries d'observations du score S :

$$s_i \text{ avec statut réel } N = \{6,24 ; 1,77 ; 4,61 ; 8,29\}$$

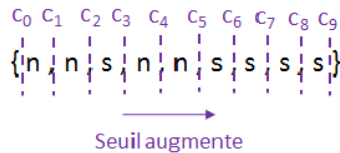
$$s_i \text{ avec statut réel } P = \{12,87 ; 10,22 ; 15,90 ; 5,01 ; 13,35\}$$

Toute l'information pertinente pour l'estimation de la courbe ROC dans un modèle latent est en fait contenue dans les données réduites à l'association du statut réel et du rang de chaque observation. Les deux précédents échantillons sont équivalents, après avoir ordonné les valeurs tout en conservant les statuts associés, à la donnée de l'échantillon :

$$\{n, n, s, n, n, s, s, s, s\}$$

L'avantage de cette représentation est qu'elle est valable autant pour S que pour une transformation monotone  $h(S)=W$  puisque les rangs sont conservés par h.

En se rappelant que POD et PFA sont des probabilités conditionnelles, il est en effet possible de retrouver l'estimation non-paramétrique de la courbe ROC à partir de cet échantillon :



$$\{(PFA_{cl}, POD_{cl})_l = \{ (4/4 ; 5/5), (3/4 ; 5/5), (2/4 ; 5/5), (2/4 ; 4/5), (1/4 ; 4/5), (0/4 ; 4/5), (0/4 ; 3/5), (0/4 ; 2/5), (0/4 ; 1/5), (0/4 ; 0/5) \}$$

On obtient la courbe ROC de la Figure 18 :

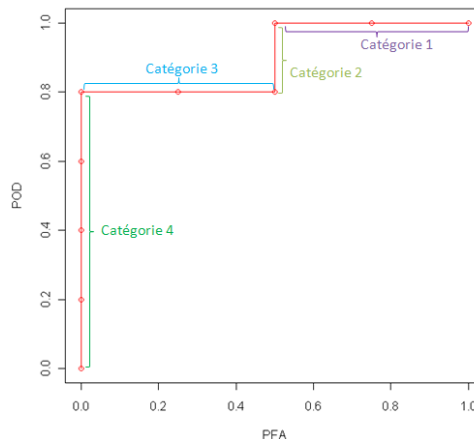


Figure 18 : Courbe ROC empirique et algorithme de C. Metz

Revenons au cadre semi-paramétrique et étudions le modèle binormal. Si l'on considère que les données de score « transformé » sont partitionnées en  $l$  catégorie construites sur les bases de  $l-1$  seuils fixes  $\mathbf{t} = (t_1, \dots, t_{l-1})$  (sans compter  $t_0 = -\infty$  et  $t_l = +\infty$ ), alors les probabilités que le score transformé issu d'un cas réellement négatif et issu d'un cas réellement positif tombe dans la  $i^{\text{ème}}$  catégorie sont respectivement :

$$\begin{cases} P_{i,N} = \Phi(t_i) - \Phi(t_{i-1}) \\ P_{i,P} = \Phi\left(\frac{t_i - \mu_P}{\sigma_P}\right) - \Phi\left(\frac{t_{i-1} - \mu_P}{\sigma_P}\right) = \Phi(a t_i - b) - \Phi(a t_{i-1} - b) \end{cases}$$

Considérons que les effectifs par catégorie des données à disposition sont décrits par :

$$\begin{aligned} \mathbf{u} &= \{u_1, \dots, u_l \mid \sum_i u_i = U\} && \text{pour les cas réellement négatifs} \\ \mathbf{v} &= \{v_1, \dots, v_l \mid \sum_i v_i = V\} && \text{pour les cas réellement positifs} \end{aligned}$$

En pratique les effectifs sont déterminés à partir de la représentation réduite de l'échantillon. Si l'on se réfère à l'exemple précédent, on a :

$$\{n, n, s, n, n, s, s, s\} \equiv \{2n, 1s, 2n, 4s\} \equiv \{\mathbf{u} = \{2,0,2,0\}, \mathbf{v} = \{0,1,0,4\}\}$$

On remarque ainsi que les catégories sont repérables par les paliers successifs observables sur la courbe ROC estimée de façon non paramétrique (voir

La vraisemblance du modèle s'exprime en fonction des  $P_{i,N}$  et  $P_{i,P}$  :

$$L(\mathbf{u}, \mathbf{v}/a, b, \mathbf{t}) = (P_{1,N})^{u_1} \dots (P_{l,N})^{u_l} (P_{1,P})^{v_1} \dots (P_{l,P})^{v_l}$$

D'où l'on déduit la log-vraisemblance :

$$\log L(\mathbf{u}, \mathbf{v}/a, b, \mathbf{t}) = \sum_{i=1}^l u_i \log P_{i,N} + \sum_{i=1}^l v_i \log P_{i,P}$$

Les paramètres à estimer sont donc  $a$ ,  $b$  et  $\mathbf{t} = (t_1, \dots, t_{l-1})$ . Pour cela on utilise le principe de la méthode du scoring détaillée dans la partie consacrée à la rating method. Plus précisément, C. Metz a développé deux algorithmes appelés LABROC4 et LABROC5. Le premier est l'implémentation de la stratégie décrite ci-dessus et fournit de vraies estimations par maximum de vraisemblance. Les valeurs initiales des paramètres jouent un rôle important dans la convergence de l'algorithme. LABROC5 s'applique dans le cas où les échantillons sont de taille importante. Il repose sur une estimation « Quasi-ML » (Quasi Maximum Likelihood) des paramètres, où certaines catégories sont fusionnées de façon astucieuse afin d'accélérer les calculs. On remarquera néanmoins que LABROC4 fait appel à LABROC5 puisque les valeurs finales des paramètres issues de LABROC5 sont utilisées comme valeurs initiales des paramètres pour LABROC4.

Un des désagréments en défaveur de cette méthode est que le nombre de paramètres à estimer augmente avec la taille de l'échantillon. Il se pourrait ainsi que les résultats obtenus ne correspondent pas à des estimations du maximum de vraisemblance. Toutefois, si l'on considère des seuils positionnés de la même façon que pour l'estimation non-paramétrique, alors les estimations obtenues semblent pouvoir être considérées comme des estimations du maximum de vraisemblance.

- Zou & Ai (2000) [22] et Gu & Ghosal (2009) [23] ont aussi adopté une approche s'appuyant sur les rangs.

L'hypothèse de binormalité entraîne que la distribution jointe de  $(S_N, S_P)$  dépend de  $(\mu_P, \sigma_P, h)$ . Mais par invariance, une statistique de rangs aura une distribution ne dépendant que de  $(\mu_P, \sigma_P)$ . L'utilisation de cette propriété conduit à une vraisemblance basée sur les rangs uniquement fonction de  $(\mu_P, \sigma_P)$ . Pour rappel, le modèle binormal consiste en :

$$W_N = h(S/N) \sim \mathcal{N}(0,1) \quad \text{et} \quad W_P = h(S/P) \sim \mathcal{N}(\mu, \sigma^2)$$

Notons  $\mathbf{W}_N = (W_N^1, \dots, W_N^{n_N})$ ,  $\mathbf{W}_P = (W_P^1, \dots, W_P^{n_P})$ , ainsi que  $\mathbf{W} = (\mathbf{W}_N, \mathbf{W}_P) = (Q_1, \dots, Q_n)$  et  $\mathbf{S} = (\mathbf{S}_N, \mathbf{S}_P) = (S_1, \dots, S_n)$

Définissons d'autre part le vecteur des rangs et des labels, ces derniers indiquant le statut réel de chaque élément de l'échantillon (variable binaire) :

$$\mathbf{R} = R(\mathbf{S}) = (R_1, \dots, R_n) \quad \mathbf{L} = L(\mathbf{S}) = L(\mathbf{W}) = (L_1, \dots, L_n)$$

Sous l'hypothèse de binormalité, les rangs et labels associés à  $\mathbf{Q}$  sont les mêmes que ceux associés à  $\mathbf{S}$ . Définissons l'ensemble  $\mathcal{D}$  suivant :

$$\mathcal{D} = \{(\mathbf{w}_N, \mathbf{w}_P) \in \mathbb{R}^{n_N+n_P}, R(\mathbf{w}_N, \mathbf{w}_P) = \mathbf{R}(\mathbf{S}) \text{ et } L(\mathbf{w}_N, \mathbf{w}_P) = \mathbf{L}(\mathbf{S})\}$$

La vraisemblance basée sur les rangs s'en déduit :

$$L(\mu_P, \sigma_P) = P(\mathcal{D} / \mu_P, \sigma_P)$$

Zou & Ai évaluent le maximum de cette vraisemblance à l'aide d'un algorithme de recherche efficace (il ne nécessite pas une complète évaluation de la vraisemblance). L'approche alternative proposée par Gu & Ghosal consiste à ne pas évaluer directement la fonction de vraisemblance. Pour cela, ils ont recouru à un échantillonnage de Gibbs. La moyenne (ou la médiane) de l'a posteriori relatif à la

distribution de  $(\mu_p, \sigma_p)$  est considérée comme l'estimateur de Bayes de la vraisemblance basée sur les rangs (BRL : Bayesian Rank Likelihood estimator).

- Pepe (2000) [24] et Alonzo & Pepe (2002) [25] se sont quant à eux ramenés aux modèles linéaires généralisés. Pour cela, il faut tout d'abord remarquer que :

$$\begin{aligned} ROC(PFA(t)) &= P(S > t/P) = P(F_{S/N}(S) > F_{S/N}(t)/P) \\ &= P(F_{S/N}(S/P) > F_{S/N}(t)) \end{aligned}$$

En notant  $f = PFA = 1 - F_{S/N}(t)$  et  $F_{S/N}(S/P) = PV_p$ , on obtient :

$$ROC(f) = P(1 - PV_p < f)$$

La courbe ROC s'identifie donc à la fonction de répartition de  $1 - PV_p = 1 - F_{S/N}(S/P)$  dont l'estimation s'effectue en deux étapes [26] :

- Estimation de  $F_{S/N}$  à partir des données de la population N soit à l'aide de l'estimateur empirique, soit par l'intermédiaire d'un modèle paramétrique. Les valeurs de  $S/P$  sont standardisées par l'action de  $\hat{F}_{S/N}$  :

$$\widehat{PV}_{P_i} = \hat{F}_{S/N}(S_i/P) \quad \text{où } i = 1, \dots, n_N$$

- Estimation de la fonction de répartition de  $1 - PV_p$ , que l'on notera  $H$  :

$$ROC(f) = P(1 - PV_p < f) = H(f)$$

On suppose alors que  $H$  s'écrit :

$$H(f) = g(\alpha_0 + \alpha_1 g^{-1}(f))$$

Ainsi, dans le cas où  $g \equiv \Phi$ , on retrouve le modèle binormal. Dans le cadre des modèles linéaires généralisés,  $g$  est appelée *la fonction de lien*. Le lien avec ce type de modèle apparaît plus clairement dans la suite de la procédure. Considérons un ensemble discret de points sur l'axe PFA  $\{f_1, \dots, f_k\}$ . Pour chaque observation  $i$  et chaque  $f_k$ , on va enregistrer la valeur de la variable binaire  $U_{ik} = 1_{\{1 - PV_{P_i} \leq f_k\}}$  et de la covariable  $g^{-1}(f_k)$ . En effet,

$$H(f_k) = P(1 - PV_{P_i} < f_k) = E(1_{\{1 - PV_{P_i} \leq f_k\}}) = E(U_{ik}) = g(\alpha_0 + \alpha_1 g^{-1}(f_k))$$

Par conséquent, en ajustant un modèle de régression binaire avec  $g$  pour fonction de lien,  $U$  comme variable de sortie et  $g^{-1}(f)$  comme covariable, on obtient une estimation des paramètres  $\alpha_0$  et  $\alpha_1$ .

On remarquera que, de par la dépendance existant entre les  $U_{ik}$  et de par l'aléa des  $\widehat{PV}_{P_i}$ , l'utilisation d'une procédure de type modèle linéaire généralisé n'est pas totalement justifiée. D'autre part, les estimations obtenues ne sont pas celles du maximum de vraisemblance et n'en possèdent par conséquent pas toutes les qualités (efficacité).

Cette approche autorise l'introduction de covariables. Cette propriété est plus largement abordée dans la partie **Error! Reference source not found.** : **Error! Reference source not found.**. Pour plus de détails, on pourra aussi consulter [27] et [28].

- Cai & Moskowitz (2004) [29] ont proposé deux approches pour l'estimation des paramètres de localisation et d'échelle du modèle binormal pour les courbes ROC. On notera que celles-ci sont valides uniquement dans le cas où le modèle est correctement spécifié. Ils ont donc ajouté à leur étude des procédures de vérification du modèle.

– **Estimateur du maximum du profil de vraisemblance :**

On se positionne dans le contexte classique du modèle binormal où  $h$  est la transformation monotone :

$$W/N = h(S/N) \sim \mathcal{N}(0,1) \quad \text{et} \quad W/P = h(S/P) \sim \mathcal{N}(\mu_p, \sigma_p^2) \equiv \mathcal{N}(\alpha, \beta^2)$$

où, avec les notations et les propriétés associées à l'équation (30) on a :  $\alpha = \frac{b}{a}$  et  $\beta = \frac{1}{a}$

On rappelle que le modèle binormal conduit à supposer que (équation (31)) :

$$\zeta_P(s) = \Phi(a \Phi^{-1}(\zeta_N(s)) + b) \tag{38}$$

$\zeta_K(s)$  est appelée la *fonction de survie* associée à  $S/K$  avec  $K \in \{N, P\}$  et est définie par :

$$\zeta_K(s) = P(S > s/K) = P(W > h(s)/K)$$

On remarque qu'en dérivant  $\zeta_K$  par rapport à  $c$ , on obtient l'opposé de la densité de  $S/K$ , soit  $-f_K$ . Ainsi, en dérivant l'équation (38) par rapport à  $c$ , il vient :

$$-f_P(s) = -\frac{a f_N(s)}{\phi(\Phi^{-1}(\zeta_N(s)))} \phi(a \Phi^{-1}(\zeta_N(s)) + b)$$

Or l'hypothèse de normalité entraîne  $\zeta_N(c) = 1 - P(W \leq h(s)/N) = \Phi(-h(s))$ . Par suite, nous obtenons une expression du rapport de vraisemblance :

$$\frac{f_P(s)}{f_N(s)} = \frac{a \phi(b - a h(s))}{\phi(-h(s))} \quad (39)$$

Nous allons nous appuyer sur cette équation pour écrire la vraisemblance du modèle.

A partir de la fonction de survie, il est possible de définir la *fonction de risque cumulé*, ainsi que le *taux de défaillance* :

- Fonction de risque cumulé :  $\Lambda(s) = -\log \zeta(s)$
- Taux de défaillance :  $\lambda(s) = \frac{d\Lambda(s)}{ds} = -\frac{\zeta'(s)}{\zeta(s)} = \frac{f(s)}{\zeta(s)}$

Ceci nous permet d'écrire  $f_N(c)$  puis  $f_P(c)$  de la façon suivante :

$$f_N(s) = \lambda_N(s) e^{-\Lambda_N(s)} \Rightarrow f_P(c) = \frac{a \phi(b - a h(s))}{\phi(-h(s))} \lambda_N(s) e^{-\Lambda_N(s)}$$

Et la vraisemblance s'en déduit :

$$L(a, b, \Lambda) = \prod_{i=1}^{n_N} \lambda_N(s_i) e^{-\Lambda_N(s_i)} \prod_{i=1}^{n_P} \frac{a \phi(b - a h(s_i))}{\phi(-h(s_i))} \lambda_N(s_i) e^{-\Lambda_N(s_i)}$$

Comme nous l'avons vu, le taux de défaillance correspond à la dérivée du risque cumulé. Cependant, en considérant une fonction de risque cumulé continue, nous ne pouvons pas obtenir les estimateurs du maximum de vraisemblance associés à  $L(a, b, \lambda)$ . Il faut en effet se limiter à une version discrète de  $\Lambda_N$ , auquel cas  $\lambda_N(s_i) = \Delta \Lambda_N(s_i)$  où  $\Delta \Lambda_N(s_i)$  correspond au saut de la fonction de risque cumulé au point  $s_i$ .

Ainsi, afin d'obtenir le profil de vraisemblance, il suffit de remplacer  $\Lambda_N$  par son estimation du maximum de vraisemblance  $\hat{\Lambda}_N$  qui est une fonction croissante en escalier avec un saut en chaque point d'observation  $s_i$ . Dès lors, nous obtenons  $\hat{\lambda}_N(s_i) = \Delta \hat{\Lambda}_N(s_i)$  et  $\hat{h}(s_i) = -\Phi^{-1}(e^{-\hat{\Lambda}_N(s_i)})$  et le profil de vraisemblance est donc :

$$PL(a, b) = L(a, b, \hat{\Lambda})$$

Les estimateurs du maximum du profil de vraisemblance de a et b sont obtenus en résolvant les équations de vraisemblance liées à  $PL(a, b)$ . Cai & Moskowitz démontrent que ces estimateurs sont efficaces et que le biais est asymptotiquement distribué selon une loi normale de moyenne nulle et de matrice de covariance  $\Sigma_{MLE}$  (dont l'expression est complexe). Cette propriété permet de construire des tests de nullité des paramètres basés sur des rapports de vraisemblance distribués comme des  $\chi^2$ , des intervalles de confiance pour chacun d'entre eux ainsi que des régions de confiance.

#### – Estimateur du pseudo maximum de vraisemblance (PLME) :

L'idée de cette méthode est de proposer une estimation alternative de la fonction h. On considère pour cela l'estimation non paramétrique de  $\zeta$  calculée sur l'ensemble des deux échantillons (N et P) :

$$\hat{\zeta}(s) = \frac{1}{n} \sum_{i=1}^n 1_{\{s_i > s\}} + \frac{1}{2} 1_{\{s_i = s\}}$$

On remarque d'autre part que le modèle binormal implique :

$$\begin{aligned} \zeta(s) &= P(S > s/N) \cdot P(N) + P(S > s/P) \cdot P(P) \\ &= \Phi(b - a h(s)) \cdot p_N + \Phi(-h(s)) \cdot p_P \end{aligned}$$

Ces deux dernières propriétés nous permettent d'écrire une équation vérifiée par  $\hat{h}$  pour n'importe quelle valeur de s et des paramètres a et b fixés :

$$\Phi(b - a \hat{h}(s)) \cdot \frac{n_N}{n} + \Phi(-\hat{h}(s)) \cdot \frac{n_P}{n} = \hat{\zeta}(s) \quad (40)$$

En résolvant cette équation, il est possible de déterminer une estimation de  $h(s)$ , fonction de a et b, pour une valeur de s quelconque. Ayant à disposition l'échantillon  $\{s_1, \dots, s_n\}$ , nous pouvons donc trouver  $\{\hat{h}(s_1), \dots, \hat{h}(s_n)\} = \{\hat{h}_1, \dots, \hat{h}_n\}$  à partir du système d'équations basé sur (40). Ces valeurs sont ensuite insérées dans les équations de vraisemblance qui, une fois résolues itérativement, vont donner les estimations de a et b. Alternativement, ces deux étapes peuvent être traitées en résolvant simultanément (40) pour  $\{h_1, \dots, h_n\}$  et les équations de vraisemblance à l'aide de l'algorithme de Newton-Raphson.

Ces procédures sont avantageuses dans le sens où l'on utilise à la fois les données sur les cas et sur les contrôles pour estimer  $h(s)$ .

Cai & Moskowitz ont montré que le biais des estimateurs obtenus est distribué selon une loi normale de moyenne nulle et de matrice de covariance  $\Sigma_{\text{pmle}}$  dont ils donnent l'expression et une méthode d'estimation.

Contrairement à l'approche basée sur le profil de vraisemblance, la procédure du pseudo maximum de vraisemblance présente l'avantage de pouvoir être facilement modifiée de façon à incorporer des covariables :

$$h(S_i/N) \sim \mathcal{N}(\boldsymbol{\gamma}'_N \mathbf{Z}_i^N, 1) \quad \text{et} \quad h(S_j/P) \sim \mathcal{N}(\alpha + \boldsymbol{\gamma}'_P \mathbf{Z}_j^P, \beta^2)$$

où  $Z_N$  et  $Z_P$  sont les vecteurs des covariables associées respectivement au groupe des cas (N) et au groupe des contrôles (P).

La forme de la courbe ROC induite par l'introduction des covariables est :

$$ROC_Z(u) = \Phi(b + a\Phi^{-1}(u) + \boldsymbol{\gamma}'\mathbf{Z}) \quad \text{où} \quad \begin{cases} a = 1/\beta \\ b = \alpha/\beta \\ \boldsymbol{\gamma} = \frac{1}{\beta}(\boldsymbol{\gamma}_P - \boldsymbol{\gamma}_N) \end{cases} \quad \text{et} \quad \mathbf{Z} = (\mathbf{Z}_N, \mathbf{Z}_P)'$$

Le vecteur  $\boldsymbol{\gamma}$  quantifie l'effet des covariables sur la courbe ROC par la différence entre l'effet des covariables dans la population P et dans la population N à un facteur d'échelle  $\beta$  près. Cette approche autorise de plus la comparaison même dans le cas où le niveau des covariables est différent entre les deux populations.

- Erkanli & Al (2006) [29] ont choisi une approche bayésienne. La flexibilité de leur modèle ne repose pas sur l'hypothèse d'existence d'une transformation monotone des données distribuée comme une loi normale, mais sur la description de la distribution du score de classification (le signal) à l'aide d'un modèle de mélange gaussien pour chaque populations N et P :

$$\begin{aligned} S_N/K_N, \theta_{K_N} &\sim \mathcal{N}(\mu_{N,K}, \sigma_{N,K}^2) \quad K_N = 1, \dots, C_N \\ S_P/K_P, \theta_{K_P} &\sim \mathcal{N}(\mu_{P,K}, \sigma_{P,K}^2) \quad K_P = 1, \dots, C_P \end{aligned}$$

où  $C_N$  et  $C_P$  indiquent le nombre de composantes du mélange dans chaque population et où  $\theta_{K_j} = (\mu_{j,K}, \sigma_{j,K})$  avec  $j \in \{N, P\}$ .

Pour des raisons de clarté, ignorons dorénavant les marqueurs N et P. Les résultats suivants seront valables pour les deux populations.

L'approche bayésienne consiste en premier lieu à assigner des distributions a priori sur les paramètres  $K$  et  $\theta_K$  en se basant sur le principe des lois conjuguées. La distribution a posteriori appartiendra ainsi à la même famille de lois et les calculs s'en verront simplifiés.

- un a priori Normal-Gamma sur  $\theta_K$  : la moyenne  $\mu_K$  suit une loi normale indépendamment de l'inverse de la variance  $1/\sigma_K^2$  dont la distribution est modélisée par une loi gamma.
- Un a priori multinomial à  $C$  états sur  $K$  dont les probabilités sont données par :

$$w_1 = R_1 \quad \text{et} \quad w_k = (1 - R_1) \dots (1 - R_{k-1}) R_k \quad \text{pour } k = 2, \dots, C$$

Où les  $R_i$  sont des variables indépendantes de loi  $\beta(1, \alpha)$  et  $R_C = 1$  afin que la somme des  $w_i$  soit égale à 1.

On remarquera que le paramètre  $\alpha$  a une importante influence sur le nombre  $C$  de composantes du mélange et que ce modèle est lié à l'approximation d'un processus de Dirichlet. D'autre part, la densité associée à la distribution du signal dans les populations N et P s'exprime comme celle d'un mélange de lois gaussiennes, soit :

$$f(s/\mathbf{w}, \boldsymbol{\theta}) = \sum_{k=1}^C w_k f(s/\theta_k)$$

où les  $f(s/\theta_k)$  sont des densités de lois normales de moyenne  $\mu_k$  et variance  $\sigma_k$ .

L'espérance de  $f(s/\mathbf{w}, \boldsymbol{\theta})$  par rapport à la distribution a posteriori de  $(\mathbf{w}, \boldsymbol{\theta})$ , c'est-à-dire,  $g(\mathbf{w}, \boldsymbol{\theta}/s)$  avec  $\mathbf{s} = (s_1, \dots, s_n)$ , donne la densité prédictive a posteriori  $f(s/s)$  du futur score  $s$  connaissant les scores précédents  $s_1, \dots, s_n$  :

$$f(s/s) = E_{\mathbf{w}, \boldsymbol{\theta}/s} [f(s/\mathbf{w}, \boldsymbol{\theta})] = \int \dots \int f(s/\mathbf{w}, \boldsymbol{\theta}) g(\mathbf{w}, \boldsymbol{\theta}/s) dw_1 \dots dw_C d\theta_1 \dots d\theta_C$$

A l'aide d'un échantillonnage de Gibbs permettant de simuler des réalisations  $(\mathbf{w}^{(u)}, \boldsymbol{\theta}^{(u)})_{u=1, \dots, U}$  à partir de  $g(\mathbf{w}, \boldsymbol{\theta}/s)$ , on approxime l'espérance de  $f(s/\mathbf{w}, \boldsymbol{\theta})$  par :

$$f(s/s) = E_{\mathbf{w}, \boldsymbol{\theta}/s} [ f(s/\mathbf{w}, \boldsymbol{\theta}) ] \cong \frac{1}{U} \sum_{u=1}^U \sum_{k=1}^c w_k^{(u)} f(s/\theta_k^{(u)})$$

Cette approximation nous permet ensuite de calculer la fonction de répartition associée à  $f(s/s)$  :

$$F(t/s_1, \dots, s_n) = \int_{-\infty}^t f(s/s_1, \dots, s_n) ds$$

La procédure est effectuée sur deux échantillons de scores issus de la population N et de la population P. On obtient par suite la POD et la PFA prédictive a posteriori :

$$\begin{cases} POD(t) = 1 - F_P(t/s_1^{(P)}, \dots, s_{n_P}^{(P)}) \\ PFA(t) = 1 - F_N(t/s_1^{(N)}, \dots, s_{n_N}^{(P)}) \end{cases}$$

Les distributions approximées à l'aide de l'échantillonnage de Gibbs permettent d'évaluer directement les incertitudes liées aux diverses quantités considérées et de donner par exemple un intervalle de confiance pour la courbe ROC. D'autre part, une extension de cette méthode autorise un étiquetage incertain ou même inconnu de chaque élément de l'échantillon : l'appartenance à la population N ou P de chaque pièce inspectée n'est pas connue avec certitude.

### II.3.D. Classification robuste dans le cadre des environnements imprécis – ROCCH

Nous avons jusqu'ici étudié les principales caractéristiques des courbes ROC et nous avons vu qu'elles constituaient un outil efficace de visualisation des performances intrinsèques des traducteurs. La capacité à dissocier bruit et signal est généralement un attribut recherché pour ces instruments.

Les paragraphes de cette partie reposent sur l'article de Provost & Fawcett (2001) [30] concernant la classification robuste en environnements imprécis.

#### a. Courbes ROC – isolement de la performance pure

Les courbes ROC vont naturellement intervenir dans différentes problématiques liées à la détermination d'un seuil optimal ou à la comparaison de traducteurs. La définition même de ces problématiques montre qu'elles reposent sur des critères d'optimalité relatif à la performance du/des traducteur(s) nécessairement soumis à l'appréciation de l'opérateur. Bien que subjectifs, ces critères interviennent inévitablement dans la prise de décision puisqu'ils permettent de prendre en compte des facteurs extérieurs au traducteur dans l'évaluation de ses performances. Ces facteurs vont permettre de juger les capacités du capteur lorsqu'il est plongé dans un certain environnement. En tant que « simple » outil de visualisation se focalisant sur le traducteur, la courbe ROC n'intègre pas ces informations.

On peut se rendre compte de l'importance des facteurs extérieurs en étudiant l'exemple suivant. Replaçons-nous dans le contexte décrit dans la partie II.3.B mais supposons que le nombre de pièces contenant un défaut est bien supérieur au nombre de pièces n'en contenant pas. On se place ainsi dans des *conditions de prévalence* particulières : la distribution a priori des classes est fortement déséquilibrée. Supposons d'autre part que le traducteur se caractérise par une POD élevée et une faible PFA :

Prévalence	Caractéristiques du traducteur
$p(r_+) = 0.001$	$POD = p(d_+/r_+) = 0.99$
$p(r_-) = 0.999$	$PFA = p(d_+/r_-) = 0.01$

On dispose donc a priori d'un excellent traducteur. Mais calculons  $p(r_+/d_+)$  à l'aide de la formule de Bayes :

$$p(r_+/d_+) = \frac{p(d_+/r_+) p(r_+)}{p(d_+/r_+) p(r_+) + p(d_+/r_-) p(r_-)} = \frac{0.99 * 0.001}{0.99 * 0.001 + 0.01 * 0.999} = 0.09$$

*Parmi les pièces détectées positives (présence d'un défaut), seules 9% le sont réellement !* On comprend donc que Si POD et PFA sont de bons indicateurs de la capacité du traducteur à séparer les deux types de signaux, ils ne sont pas suffisants pour apprécier les performances de celui-ci lorsqu'il est introduit dans un environnement particulier. La courbe ROC permet en réalité de séparer la performance pure de classification du traducteur de la distribution des facteurs extérieurs.

Dans le cas où l'on ne s'intéresse qu'à la performance pure, une façon d'ordonner les traducteurs est de les trier selon l'aire sous la courbe ROC qui leur est associée. On obtient ainsi un « résumé », une performance globale, en balayant l'ensemble des valeurs possibles de la POD et de la PFA lorsque le seuil évolue dans son domaine de variation. Dans le paragraphe consacré à la description des courbes ROC, nous avons vu que le capteur idéal se trouvait dans le coin supérieur gauche de l'espace ROC. Un tel capteur possède une aire sous la courbe égale à un (surface de l'espace ROC). A l'inverse, il faut éviter les détecteurs caractérisés par la diagonale de chance et dont l'aire sous la courbe est égale à 0,5. Les traducteurs seront donc notés sur une échelle allant de 0,5 à 1. Cette méthode bien que souvent utilisée,

n'est pas réellement adaptée à la pratique puisque dans des conditions réelles, les coûts de mauvaises classification sont rarement symétriques et il arrive fréquemment que l'on constate un net déséquilibre de la prévalence (voir l'article de Hand D. J. traitant de l'incohérence de l'AUC<sup>1</sup> [31]).

La propriété 7 des courbes ROC impose  $AUC = P(S_p > S_N) = P(S_p - S_N > 0)$ . Dans le cas du modèle paramétrique binormal,  $S_N \sim \mathcal{N}(\mu_N, \sigma_N^2)$  et  $S_p \sim \mathcal{N}(\mu_p, \sigma_p^2)$  de façon indépendante, d'où  $S_p - S_N \sim \mathcal{N}(\mu_p - \mu_N, \sigma_N^2 + \sigma_p^2)$  et :

$$AUC = P(\mathcal{N}(\mu_p - \mu_N, \sigma_N^2 + \sigma_p^2) > 0) = \Phi\left(\frac{\mu_p - \mu_N}{\sqrt{\sigma_N^2 + \sigma_p^2}}\right) = \Phi\left(\frac{b}{\sqrt{1 + a^2}}\right)$$

en effectuant le changement de variable  $a = \frac{\sigma_N}{\sigma_p}$  et  $b = \frac{\mu_p - \mu_N}{\sigma_p}$ .

Pour des résultats généraux sur l'aire sous la courbe ROC, on pourra lire l'article dédié de Bamber (1975) [32] ou encore celui de Hanley & Ai (1982) [33]. Faraggi & Reiser (2002) [34] ont abordé la question de l'estimation de l'aire sous la courbe ROC. On remarquera cependant qu'une partie des articles publiés visant à proposer une procédure d'estimation de la courbe ROC s'accompagne aussi d'une méthode d'estimation de l'AUC. On trouvera d'autre part un résumé de divers problèmes et solutions associées à l'AUC de la courbe ROC (cas continu) dans [12].

### b. Projection de l'information analytique de décision sur l'espace ROC

La partie précédente a montré l'intérêt qu'il peut y avoir à considérer, dans l'évaluation du traducteur, des facteurs ne lui étant pas propre. Nous allons donc voir dans la suite comment ces divers facteurs peuvent être pris en compte à travers un critère d'évaluation qui sera ensuite projeté sur l'espace ROC.

Dans la pratique du contrôle non destructif, les coûts (au sens large) associés à des erreurs de mauvaise décision/classification sont des aspects primordiaux. Les préférences de l'opérateur pour l'un ou l'autre des traducteurs sont en effet implicitement dictées par la répercussion de telles décisions sur le coût global de l'opération. Ce coût global constitue donc un critère sur lequel se repose l'opérateur pour juger un traducteur. En ce sens, il est structuré de façon à incorporer à la fois les décisions et ce qu'elles engendrent.

Ainsi, prévalence et coûts de mauvaise classification sont impliqués dans l'évaluation de l'impact lié à une décision de CND. Une définition analytique possible du coût global consisterait à pondérer les différentes erreurs de décision par les facteurs extérieurs.

On obtiendrait alors le *coût attendu* :

$$\begin{aligned} \text{Coût attendu} &= p(d_-, r_+) \cdot c(d_-, r_+) + p(d_+, r_-) \cdot c(d_+, r_-) \\ &= p(d_-/r_+) \cdot p(r_+) \cdot c(d_-, r_+) + p(d_+/r_-) \cdot p(r_-) \cdot c(d_+, r_-) \\ &= (1 - POD) \cdot p(r_+) \cdot c(d_-, r_+) + PFA \cdot p(r_-) \cdot c(d_+, r_-) \end{aligned}$$

Où  $p(r_+)$  et  $p(r_-)$  correspondent à la prévalence, et :

- $\left\{ \begin{array}{l} c(d_-, r_+) = \text{coût associé à l'erreur "décision négative alors que réalité positive"} \\ c(d_+, r_-) = \text{coût associé à l'erreur "décision positive alors que réalité négative"} \end{array} \right.$

L'expression du coût global permet d'exprimer la POD en fonction de la PFA :

$$POD = \frac{p(r_-) \cdot c(d_+, r_-)}{p(r_+) \cdot c(d_-, r_+) + PFA \cdot \frac{p(r_+) \cdot c(d_-, r_+) - \text{coût attendu}}{p(r_+) \cdot c(d_+, r_-)}}$$

Cette dernière équation est celle d'une droite dans l'espace ROC qui est appelée la **droite d'isoperformance**.

Chaque condition de coût d'erreur et de prévalence correspond à une pente de la droite tandis que le coût attendu est constant sur chacune d'entre elles. Ainsi, chaque distribution de coûts et de prévalence définit une famille de droite d'isoperformance : plus le coût global diminue dans des conditions particulières, plus la performance augmente et plus la droite se translate vers le haut du quadrant de l'espace ROC (voir Figure 21).

Dans le cas où les densités  $f_{S/P}$  et  $f_{S/N}$  ont un rapport de vraisemblance monotone, la courbe ROC est concave. L'optimalité dans des conditions fixées est donc donnée par le point de contact entre la courbe ROC et la droite d'isoperformance qui lui est tangente.

### c. Identification des traducteurs potentiellement optimaux – ROCCH

Une autre problématique que l'on peut aborder dans le cas où l'on dispose de plusieurs capteurs est la sélection de traducteurs. En effet, il peut s'avérer intéressant d'identifier celui possédant les meilleures performances dans des conditions de coûts et de prévalence données. Pour cela, il est possible de s'appuyer sur la procédure décrite dans la partie précédente.

Considérons deux « scénarios » :

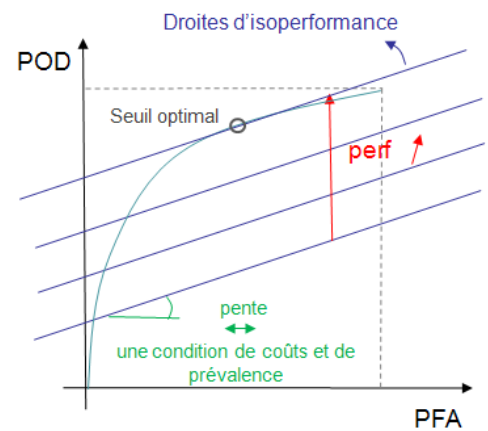


Figure 19 : Droites d'isoperformance dans l'espace ROC

<sup>1</sup> AUC : Area Under Curve, l'aire sous la courbe



$$\begin{cases} \text{Scénario A : } c(d_+, r_-) = c(d_-, r_+) \\ \text{Scénario B : } c(d_+, r_-) = c(d_-, r_+) \end{cases} \text{ avec } p(r_+) = 5 p(r_-)$$

Alors les pentes respectives des droites d'isoperformance sont :

$$\text{pente}(A) = \frac{c(d_+, r_-) p(r_-)}{c(d_-, r_+) p(r_+)} = 5 = \alpha$$

$$\text{pente}(B) = \frac{c(d_+, r_-) p(r_-)}{c(d_-, r_+) p(r_+)} = \frac{1}{5} = \beta$$

On détermine les traducteurs optimaux pour chacun des scénarios en recherchant les tangentes aux courbes ROC (Figure 20).

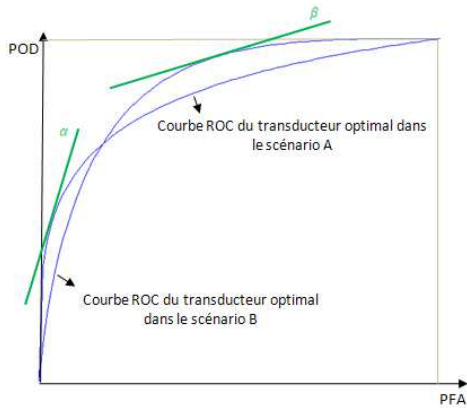


Figure 20 : Optimalité selon le scénario

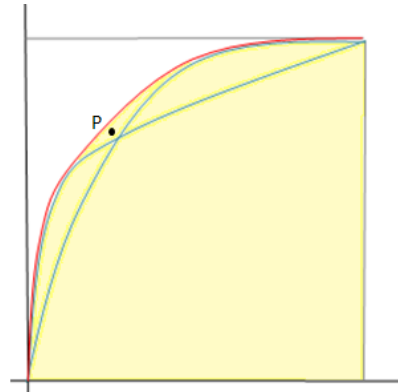


Figure 21 : L'enveloppe convexe ROCCH en jaune

Ce dernier mécanisme permet de mettre en évidence une zone importante de l'espace ROC : **l'enveloppe convexe**. Elle correspond *au plus petite ensemble convexe contenant tous les points des courbes ROC considérées*. Elle est aussi appelée **ROCCH** pour ROC Convex Hull (voir Figure 21).

L'intérêt de cette région se voit de façon assez immédiate. Considérons le point P de la Figure 21. On peut imaginer qu'il correspond à la performance d'un traducteur dont le seuil de décision ne peut être modifié. A priori, étant donné qu'il se trouve au dessus des deux autres courbes ROC, on serait tenté de penser qu'il est optimal (dans le sens « meilleur que les autres ») dans certaines conditions de coûts et de prévalence. Mais sa position, à l'intérieur de la ROCCH, l'en empêche. Il sera en effet toujours possible de translater la droite d'isoperformance de façon à ce qu'elle soit à la fois :

- Tangente à l'une des deux autres courbes ROC.
- Associée à une performance plus élevée (ordonnée à l'origine plus importante)

Une démonstration rigoureuse existe qui, si elle n'est pas très complexe, reste relativement longue. On pourra cependant se persuader de la validité de cette assertion en analysant la Figure 24 : *Le point P ne peut être optimal*.

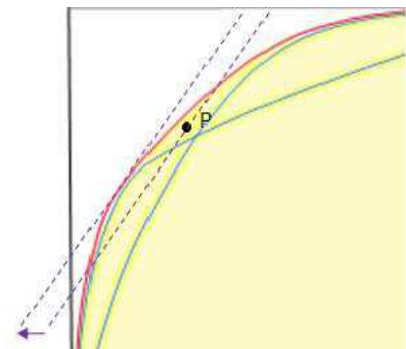


Figure 22 : Le point P ne peut être optimal

On comprend avec ces remarques que le *traducteur optimal appartient à la frontière de la ROCCH*.

D'autre part, les propriétés de l'enveloppe convexe en font un bon outil de sélection préalable des capteurs évitant ainsi d'avoir à sauvegarder une quantité importante de données les concernant :

- Soit ils étendent la ROCCH et ils sont conservés.
- Soit ils n'étendent pas l'enveloppe convexe auquel cas ils ne seront jamais optimaux et ne sont par conséquent pas retenus.

#### d. Sensibilité des résultats aux imprécisions sur l'environnement

Jusque-là, nous avons supposé que la distribution des facteurs extérieurs était bien connue, et nous avons cherché des procédures avec pour objectif, l'identification du meilleur traducteur dans des conditions de coûts et de prévalence fixées. En pratique, il s'avère que la connaissance de cet environnement est bien souvent imprécise. La ROCCH permet alors de retenir un ensemble de candidats pour l'optimalité.

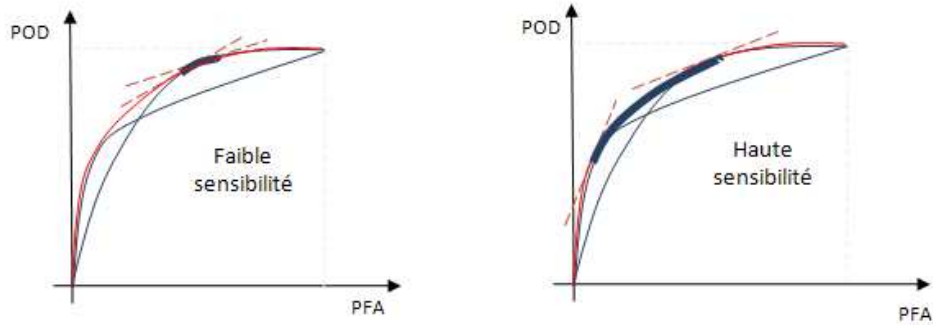


Figure 23 : Sensibilité de la solution optimale aux imprécisions sur l'environnement

Des informations imprécises se traduisent par toute une échelle de pentes possibles pour la droite d'isoperformance. L'ensemble de ces droites s'interceptent avec la frontière de l'enveloppe convexe selon un segment (voir Figure 23). Par conséquent :

- Si ce segment se réduit à un point, alors les conditions environnantes n'influencent pas la sélection du traducteur optimal.
- Si ce segment est de longueur non nulle, sa longueur est un indicateur de la sensibilité de la solution optimale.

Dans le cas où la sensibilité est très importante, on peut imaginer sélectionner autant de traducteurs que de conditions de coûts et de prévalence. On ne dispose alors pas de capteurs dont les capacités de classifications sont **robustes**. Or les traducteurs sont généralement destinés à être utilisés sur une longue période et dans divers contextes, donc dans environnements potentiellement changeant. Dans de telles conditions, spécifier les futurs coûts et distributions est souvent difficile, voire impossible et un manque de robustesse peut s'avérer très handicapant.

#### e. Robustesse – Hybrides de classification basés sur la ROCCH

La notion de robustesse varie selon les domaines dans lesquels elle est recherchée. Les traducteurs sont ici vus comme des instruments de classification binaire. Après inspection, chaque pièce se voit associée à l'une des deux catégories : « contient un défaut » ou « ne contient pas de défaut ». Dans cet exercice :

« Un classifieur est considéré comme robuste si, quelques soient les conditions de coûts et de distributions, sa performance est au moins aussi bonne que le meilleur classifieur pour ces conditions ».

Cette définition est pratique puisqu'elle se limite à la performance des traducteurs connus et étudiés. Cependant, déclarer qu'un classifieur est robuste est plus fort qu'indiquer qu'il est optimal pour un ensemble de conditions spécifiques. En se basant sur les propriétés de la frontière de la ROCCH, il est possible de construire un classifieur hybride robuste.

De façon formelle, on définit tout d'abord un classifieur  $\mu$ -hybride :

Soient  $\mathcal{J}$  l'espace des cas possibles (échantillons),  $\mathcal{C}$  l'espace des ensembles de modèles de classification. Alors, un classifieur  $\mu$ -hybride correspond à la combinaison d'un ensemble  $C \in \mathcal{C}$  de modèles de classification associés à une fonction  $\mu$  telles que :

$$\begin{aligned} \mu : \mathcal{J} \times \mathbb{R} \times \mathcal{C} &\rightarrow \{N, P\} \\ (I, x, C) &\mapsto \mu(I, x, C) \end{aligned}$$

Le réel  $x$  correspond au critère d'optimalité par rapport auquel les classifieurs sont ordonnés. Chaque valeur de  $x$  est associée à une condition de coûts et de prévalence.  $C$  devra quant à lui et dans l'idéal, contenir tous les modèles dont la performance sous certaines conditions est optimale. Le principe du classifieur  $\mu$ -hybride est simple dans le sens où il sélectionne le meilleur classifieur, ou encore, effectue une combinaison entre les meilleurs classifieurs afin de produire les meilleures classifications. Un modèle de classification robuste peut être obtenu en additionnant les propriétés de la ROCCH à celles du classifieur  $\mu$ -hybride :

Le **classifieur ROCCH-hybride** est un classifieur  $\mu$ -hybride où  $C \equiv$  frontière de la ROCCH et où  $\mu$  effectue les classifications en utilisant les classifieurs de  $C$  avec  $x = PFA$ .

Il faut remarquer que le classifieur ROCCH-hybride opère ses sélections sur la base de la PFA. Mais cette propriété combinée à la définition de  $C$  comme la frontière de la ROCCH a permis à Provost & Fawcett de montrer que le classifieur ROCCH-hybride minimise aussi le coût attendu.

Ils ont de plus proposé une généralisation de ce résultat pour les métriques d'évaluation  $f(PFA, POD)$  vérifiant les propriétés suivantes :

$$\frac{\partial f}{\partial POD} \geq 0 \quad \text{et} \quad \frac{\partial f}{\partial PFA} \leq 0$$

Alternativement, il arrive que l'on cherche à résoudre un problème où l'objectif est de maximiser la  $f(PFA, POD)$  par rapport au seuil tout en respectant une contrainte du type  $a POD + b PFA \leq c$  où  $a \geq 0$  et  $b \geq 0$ . Dans le cas où  $f$  vérifie les conditions précédentes, il est possible de montrer que la performance du classifieur ROCCH-hybride est au moins aussi bonne que celle de n'importe quel classifieur préalablement étudié. Les problèmes de type Neyman-Pearson en sont un cas particulier avec  $f(PFA, POD) = POD$  et la contrainte  $PFA \leq PFA_{max}$ .

En pratique, quels sont les avantages du classifieur ROCCH-hybride ?

Le premier atout provient du découplage entre la performance de classification et la distribution des coûts et des classes initié par la courbe ROC et l'enveloppe convexe ROCCH. Cela permet d'aisément éliminer au préalable des classifieurs même si l'on ne connaît pas les futures conditions dans lesquelles ils seront utilisés. Certains ne seront en effet jamais optimaux. On économise ainsi de précieuses ressources en temps et en espace (sauvegarde de données).

Le deuxième atout provient de la structure du ROCCH-hybride qui apporte une certaine flexibilité. La projection de l'information analytique de décision sur l'espace ROC évite, une fois les conditions d'utilisation connues, la mise en place d'une nouvelle série de tests des classifieurs. L'identification de l'optimalité n'est donc pas coûteuse. De plus, cette flexibilité est appréciable même si les conditions de coûts et de prévalence ne sont pas connues avec précision puisqu'elle autorise à moindre coût, le réajustement du classifieur au fil des retours sur l'environnement

### II.3.E. Etiquetage incertain ou impossible

Tous les procédés de construction et d'analyse des courbes ROC proposés jusqu'à maintenant supposent un étiquetage correct de chaque observation de l'échantillon comme appartenant à l'une des populations N ou P. On suppose pour cela l'existence d'un « gold standard », c'est-à-dire d'une méthode de diagnostic de référence considérée comme sûre et que l'on utilise pour étiqueter l'échantillon. Cependant, disposer d'une telle référence reste rare. Il peut même s'avérer impossible, pour diverses raisons, d'associer un label à chaque observation ou encore, que cet étiquetage ne puisse se faire que de façon différée.

Dans le cadre du CND, l'évaluation des méthodes passe généralement par la construction de maquettes devant être détruites a posteriori afin de pouvoir être étiquetées. Le résultat de cette méthode est naturellement soumis aux aléas propres aux manipulations expérimentales et l'examen visuel n'est pas nécessairement un procédé certain dans l'attribution des labels.

L'impact d'une labellisation incorrecte sur l'évaluation des outils de classification est évidemment important. Il convient par conséquent de la prendre en compte dans l'estimation des courbes ROC.

Dans un tel contexte, les méthodes bayésiennes sont bien adaptées (voir [35] pour une introduction aux modèles hiérarchiques bayésiens avec une application à la théorie de la détection du signal). Celles-ci permettent en effet de combiner la connaissance a priori d'un expert et l'information partielle portée par les données. Bien qu'elles aient longtemps été critiquées par une grande partie de la communauté statistique considérant qu'il fallait juste « laisser parler les données », ces techniques ont finalement réussi à s'imposer dans la boîte à outils standard du statisticien. Lorsqu'elle est utilisée, la démarche bayésienne requière toutefois qu'un soin particulier soit apporté à l'a priori de l'expert puisqu'il en conditionne en grande partie le succès final.

Considérons la variable  $L$  comme étant la variable « label » indiquant l'étiquetage de chaque observation. Nous n'allons plus la considérer comme une variable déterministe mais plutôt comme une variable binaire prenant la valeur N ou P avec une certaine probabilité inconnue. Dans le cas où l'on suppose l'existence d'un procédé de référence  $U$  continu, la variable  $L$  pourra être vue comme dérivant d'un seuillage dichotomique de ce procédé.

#### a. Estimation semi-paramétrique bayésienne

##### i. Extension de la méthode d'Erkanli & Al (2006)

Dans la partie II.3.C.b.iv, nous avons étudié la technique bayésienne proposée par Erkanli & Al (2006) [36] permettant d'estimer la courbe ROC dans le cadre d'un étiquetage déterministe des éléments de l'échantillon. Ce procédé s'étend aisément au cas où la variable  $L$  de label est stochastique, assignant de façon imparfaite les observations à la population N ou P.

Considérons la variable  $Z$  latente et binaire indiquant l'appartenance réelle de chaque élément à N ou P. Il est possible de modéliser la relation entre  $L$  et  $Z$  de deux façons :

- Le modèle classique suppose que la loi de  $L$  conditionnellement à  $Z$  est une Bernoulli de paramètre  $\pi$  :

$$L/Z \sim \text{Bernoulli}(\pi) \quad \text{avec} \quad \log\left(\frac{\pi}{1-\pi}\right) = \begin{cases} \beta_0 & \text{si } Z = N \\ \beta_0 + \beta_1 & \text{si } Z = P \end{cases}$$

- Le modèle de Berkson suppose à l'inverse que la distribution de  $Z$  conditionnellement à  $L$  est une Bernoulli de paramètre  $\zeta$  :

$$Z/L \sim \text{Bernoulli}(\zeta) \quad \text{avec} \quad \log\left(\frac{\zeta}{1-\zeta}\right) = \begin{cases} \beta_0 & \text{si } L = N \\ \beta_0 + \beta_1 & \text{si } L = P \end{cases}$$

Pour une spécification complète, ce modèle requière aussi  $\zeta \sim \text{Beta}(a, b)$ .

La variable latente  $Z$  peut alors être simulée à partir de sa distribution a posteriori par l'intermédiaire d'une étape supplémentaire dans l'échantillonnage de Gibbs et les modifications à apporter à la méthode d'Erkanli & Al sont mineures.

##### ii. Méthode de Choi & Al 2006

La méthode suivante a été élaborée par Choi & Al (2006) [37]. Ils ont considéré le cas de deux classifieurs continus dont les scores observés dans la population P et N sont respectivement  $(S_{1P}^{(i)}, S_{2P}^{(i)})_{i=1, \dots, n_P}$  et  $(S_{1N}^{(j)}, S_{2N}^{(j)})_{j=1, \dots, n_N}$ .

- Dans un premier temps, supposons l'existence d'une référence permettant de labelliser correctement les observations

On considère alors les vecteurs :

$$\mathbf{S}_P^{(i)} = \begin{pmatrix} S_{1P}^{(i)} \\ S_{2P}^{(i)} \end{pmatrix} \quad \text{et} \quad \mathbf{S}_N^{(j)} = \begin{pmatrix} S_{1N}^{(j)} \\ S_{2N}^{(j)} \end{pmatrix}$$

Comme dans le modèle binormal, on suppose qu'il existe une transformation monotone croissante de ces scores telle que :

$$\begin{cases} \mathbf{W}_P^{(i)} = \begin{pmatrix} h(S_{1P}^{(i)}) \\ h(S_{2P}^{(i)}) \end{pmatrix} = \begin{pmatrix} W_{1P}^{(i)} \\ W_{2P}^{(i)} \end{pmatrix} \sim \mathcal{N}_2(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P) \\ \mathbf{W}_N^{(j)} = \begin{pmatrix} h(S_{1N}^{(j)}) \\ h(S_{2N}^{(j)}) \end{pmatrix} = \begin{pmatrix} W_{1N}^{(j)} \\ W_{2N}^{(j)} \end{pmatrix} \sim \mathcal{N}_2(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N) \end{cases} \quad \text{où } \boldsymbol{\mu}_K = \begin{pmatrix} \mu_{1K} \\ \mu_{2K} \end{pmatrix} \text{ et } \boldsymbol{\Sigma}_K = \begin{pmatrix} \sigma_{11K}^2 & \sigma_{12K}^2 \\ \sigma_{12K}^2 & \sigma_{22K}^2 \end{pmatrix}$$

Afin d'adopter un point de vue bayésien, il est plus simple de reformuler le modèle en termes de lois marginales de  $\mathbf{W}_P$  et  $\mathbf{W}_N$  pour le classifieur 1 auxquelles on associe les lois conditionnelles  $W_{2P}/W_{1P}$  et  $W_{2N}/W_{1N}$  du classifieur 2 étant donné le classifieur 1. Concernant les distributions conditionnelles, nous avons les propriétés suivantes :

$$\begin{aligned} E(W_{2P}^{(i)}/W_{1P}^{(i)}) &= \mu_{2P} + \rho_P (W_{1P}^{(i)} - \mu_{1P}) \frac{\sigma_{22P}}{\sigma_{11P}} & V(W_{2P}^{(i)}/W_{1P}^{(i)}) &= \sigma_{22P}^2 (1 - \rho_P^2) \\ E(W_{2N}^{(j)}/W_{1N}^{(j)}) &= \mu_{2N} + \rho_N (W_{1N}^{(j)} - \mu_{1N}) \frac{\sigma_{22N}}{\sigma_{11N}} & V(W_{2N}^{(j)}/W_{1N}^{(j)}) &= \sigma_{22N}^2 (1 - \rho_N^2) \end{aligned} \quad \text{et}$$

où  $\rho_P$  et  $\rho_N$  sont donnés par :

$$\rho_N = \frac{\sigma_{12P}}{\sigma_{11P} \sigma_{22P}} \quad \text{et} \quad \rho_P = \frac{\sigma_{12N}}{\sigma_{11N} \sigma_{22N}}$$

L'estimation des paramètres s'effectue à l'aide d'un échantillonnage de Gibbs à partir de leur loi a posteriori. Un a priori informatif sur chaque paramètres est souhaitable, mais à défaut, un a priori non informatif construit sur la base de loi Normales pour les moyennes, de lois gamma pour l'inverse des variances, et enfin de lois uniforme pour les  $\rho_K$  peut suffire.

Les courbes ROC des deux classifieurs peuvent alors être obtenues et correspondent à :

$$ROC_k \equiv \left\{ \left( 1 - \Phi\left(\frac{t - \mu_{kN}}{\sigma_{kkN}}\right), 1 - \Phi\left(\frac{t - \mu_{kP}}{\sigma_{kkN}}\right) \right), t \in ] -\infty, +\infty[ \right\}$$

où k indiquent le classifieur considéré.

- En l'absence de référence, le modèle précédant doit subir une légère modification proche de celle effectuée pour l'extension de la méthode d'Erkanli & Al. On introduit une variable latente  $Z$  binaire indiquant l'appartenance réelle de chaque élément  $i$  de l'échantillon à la population N ou P avec :

$$Z_j \sim \text{Bernouilli}(\pi)$$

La densité de probabilité du score de classification bivariable pour la  $j^{\text{ème}}$  observation est cette fois donnée par :

$$f_{(W_1, W_2)}(w_1^{(j)}, w_2^{(j)}) = \left( f_{(W_1, W_2, P)}(w_1^{(j)}, w_2^{(j)}) \right)^{Z_j} \left( f_{(W_1, W_2, N)}(w_1^{(j)}, w_2^{(j)}) \right)^{1-Z_j}$$

où  $f_{(W_1, W_2, P)}$  et  $f_{(W_1, W_2, N)}$  sont les densités respectives de  $\mathcal{N}_2(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P)$  et  $\mathcal{N}_2(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N)$ .

L'estimation des paramètres se fait par l'intermédiaire d'un échantillonnage de Gibbs. L'expression des diverses densités a posteriori des paramètres découlent de la forme du modèle incluant dorénavant la variable  $Z$ . Cette dernière peut être simulée en ajoutant une étape dans l'échantillonnage. Un a priori sur le paramètre additionnel  $\pi$  est donc nécessaire : Choi & Al proposent une loi Beta dans le cas où aucune information scientifique n'est disponible.

## b. Estimation non-paramétrique bayésienne

L'article de Wang & Al (2007) [38] (voir aussi [12]) contient une démarche non-paramétrique bayésienne pour l'estimation de la courbe ROC pour le cas où  $L$  est une variable stochastique.

- Dans un premier temps intéressons-nous à l'ensemble des paramètres introduits dans le modèle :

Supposons d'une part que les valeurs de  $S$  sont plus élevées dans la population P que dans la population N et d'autre part, que  $K$  seuils  $t_1 < \dots < t_K$  ont été sélectionnés.

A chaque valeur du seuil correspond une PFA et une POD évaluées sur un échantillon de données différent. Notons  $\alpha^{(i)}$  et  $\beta^{(i)}$  les vraies valeurs inconnues de PFA et 1-POD pour les seuils  $(t_i)_{i=1, \dots, K}$ . Lorsque le seuil varie de  $t_1$  à  $t_K$ , on se déplace sur la courbe ROC selon les PFA décroissantes et les 1-POD croissantes : les  $\alpha^{(i)}$  et  $\beta^{(i)}$  vérifient donc :

$$\alpha^{(1)} \geq \dots \geq \alpha^{(K)} \quad \text{et} \quad \beta^{(1)} \leq \dots \leq \beta^{(K)}$$

avec d'autre part,  $\alpha^{(K+1)} = \beta^{(0)} = 0$  et  $\alpha^{(0)} = \beta^{(K+1)} = 1$ .

Etant donné la stochasticité de  $L$ , il est possible de définir une PFA et POD lui étant associées. Nous les noterons  $\alpha$  et  $\beta$  :

$$\alpha = P(P/N) \quad \text{et} \quad \beta = P(N/P)$$

Enfin, notons  $\theta$  la vraie prévalence inconnue et telle que  $\theta = P(P)$  et  $1 - \theta = P(N)$ .

– Attachons-nous maintenant à la description des données :

On suppose que l'échantillon est composé de  $n$  éléments dont  $n_N$  et  $n_P$  d'entre eux sont respectivement étiquetés par  $L$  comme appartenant aux populations N et P.

D'autre part, notons  $x_{iN}, x_{iP}$  le nombre d'éléments respectivement étiquetés N, P par  $L$  mais classifiés par S dans la population P pour le seuil  $t_{i-1}$  et dans la population N pour le seuil  $t_i$ , pour  $i = 2, \dots, K - 1$ .

Pour  $i=1$  : les  $x_{1N}, x_{1P}$  correspondent respectivement aux nombres d'observations attribuées à N, P par L mais classifiées par S dans la population N.

Pour  $i=K$  : les  $x_{KN}, x_{KP}$  correspondent respectivement aux nombres d'observations attribuées à N, P par L mais classifiés par S dans la population P.

Cette description des données associée à la définition des paramètres précédente nous conduit à un modèle multinomial ordonné (voir le chapitre 2 du polycopié de Hurlin C. [39]) dont la vraisemblance est de la forme :

$$L \propto \prod_{i=1}^{K+1} [\theta(\beta^{(i)} - \beta^{(i-1)})(1 - \beta) + (1 - \theta)(\alpha^{(i-1)} - \alpha^{(i)})\alpha]^{x_{iP}} [\theta(\beta^{(i)} - \beta^{(i-1)})\beta + (1 - \theta)(\alpha^{(i-1)} - \alpha^{(i)})(1 - \alpha)]^{x_{iN}}$$

Il y a donc  $2K+3$  paramètres à estimer : les  $K$   $\alpha^{(i)}$ , les  $K$   $\beta^{(i)}$  ainsi que  $\alpha$ ,  $\beta$  et  $\theta$ , mais seulement  $2K + 1$  degrés de liberté. Le modèle ainsi posé n'est donc pas identifiable.

Ce dernier problème peut être contourné si différents groupes d'individus ont des prévalences  $\theta_1, \dots, \theta_G$  ( $G \geq 2$ ) différentes.

Un caractère monotone de la courbe ROC peut être recherché auquel cas la positivité des  $\alpha^{(i)}$  et des  $\beta^{(i)}$  est nécessaire. A cette fin, Wang & Al proposent d'utiliser des a priori de loi de Dirichlet pour ces paramètres et de lois Beta pour  $\alpha$ ,  $\beta$  et les  $\theta_g$ . La construction des distributions a posteriori étant complexe, Wang & Al se sont appuyés sur une méthode d'augmentation de données proposée par Tanner & Wong (1987).

### III. Conclusion

Le modèle Hit/Miss nécessite au préalable le calcul des POD empiriques (proportions de détection) pour chaque taille de défaut. Celles-ci étant empruntées d'aléa, un second temps est dédié à l'ajustement d'une « courbe de régression » dont la forme analytique est supposée être donnée par la fonction de répartition d'une loi logistique ou log-logistique. On remarquera qu'un reparamétrage de ces fonctions s'appuyant sur la proximité des lois logistique et normale permet de se replacer dans le cadre des modèles « probit ». Ces derniers reposent sur la fonction de répartition d'une loi normale. L'estimation des paramètres de la courbe de régression par le principe du maximum de vraisemblance fait apparaître la partie déterministe de la POD empirique et donne la possibilité de mesurer l'influence de la taille du défaut sur la performance de détection.

L'approche Signal/Réponse s'articule quant à elle autour de la modélisation de la relation entre la taille du défaut et l'amplitude maximale du signal qu'elle engendre. Lors de la construction des échantillons de données de signal, des contraintes techniques obligent parfois les opérateurs à ramener les signaux trop importants à une valeur plus faible, ou inversement, les signaux d'intensité moindre à une valeur plus élevée. Ce traitement particulier nécessite lors de la phase d'estimation de la POD, la mise en place d'une procédure capable de gérer les données censurées. Ainsi, si l'hypothèse dominante retenue sous-entend que l'amplitude maximale du signal croît linéairement avec la taille du défaut, l'utilisation d'une régression linéaire « classique » n'est pas appropriée. Afin de contourner cette difficulté, A. Berens a proposé une méthode basée sur un modèle de type « Tobit » encore appelé « régression censurée ». Elle se traduit par le découpage de la vraisemblance en trois parties et par l'utilisation du principe du maximum de vraisemblance pour l'estimation des paramètres. Les différentes hypothèses énoncées au préalable conduisent de façon assez naturelle à une POD s'exprimant comme la fonction de répartition d'une loi normale ou log-normale.

Ces deux types de modélisations suggèrent la résolution d'équations de vraisemblance, opération qui s'effectue généralement par l'intermédiaire de l'algorithme itératif de Newton-Raphson. L'initialisation qu'il requière peut s'avérer handicapante car un choix de paramètres trop éloignés de la solution réelle peut mettre en péril la convergence de l'algorithme. Cet événement se produit parfois en pratique. D'autre part, les hypothèses associées au modèle de Berens semblent a priori très restrictives. Elles ne relient en effet la probabilité de détection qu'à la taille du défaut. Cela n'invalide pas le modèle, mais restreint son champ d'application à des situations très particulières. L'ajout de variables telles que le contraste, le volume du défaut et la distance au traducteur peut être envisagée mais l'augmentation de la quantité de données nécessaire est à prévoir, tout comme l'introduction de phénomènes non linéaires. On remarquera d'autre part que l'hypothèse de linéarité du signal par rapport à la taille du défaut est critiquable au vu des données et des connaissances physiques. Cette dernière remarque a conduit à considérer des alternatives au modèle de Berens, comme l'estimation non paramétrique de la POD ou l'utilisation de la régression polynomiale sur les données « signal/taille de défaut ». La première solution réclamant une trop grande quantité de données a finalement été abandonnée, tandis que la seconde, introduisant un nombre de paramètres trop important, n'a pas non plus été retenue. De plus, chaque méthode alternative réclame la refonte de la procédure de construction de la bande de confiance.

Il est fréquent dans la pratique industrielle d'effectuer des contrôles avec une idée précise de la configuration et du type de défaut que l'on cherche à détecter. Ainsi, malgré les différentes difficultés rencontrées, le modèle de Berens s'avère tout de même opérationnel. Il se présente comme un bon compromis entre quantité de données à recueillir et nombre de paramètres à estimer. Les temps de calcul qu'il réclame se révèlent d'autre part relativement faibles. Une suite logique au développement de ce modèle pourrait être la construction d'un plan d'expérience permettant d'optimiser l'estimation des paramètres.

Dans le cas où le signal est fortement bruité, il peut arriver que l'on commette une erreur en concluant à une détection alors qu'aucun défaut n'est présent dans la pièce : on parle alors de fausse alarme. Ce type d'erreur peut s'avérer aussi dommageable que celui consistant à manquer un défaut (associé à la quantité 1-POD). Dans un tel contexte, l'introduction de la notion de probabilité de fausse alarme (PFA) est indispensable.

L'étude proposée ici a ainsi été l'occasion de mettre en avant le caractère nécessaire et suffisant du couple (*PFA, POD*) dans l'évaluation de la performance intrinsèque d'un capteur à dissocier bruit et signal. Elle a d'autre part permis de mettre en valeur les possibilités offertes par les courbes ROC en tant qu'outil complet de visualisation de cette performance.

Ces courbes consistent en une représentation simultanée de la POD et de la PFA pour les diverses valeurs possibles du seuil de détection. Nous avons exploré les développements auxquels elles sont sujettes, notamment dans le cadre de leur estimation. Sont à privilégier, les méthodes autorisant l'introduction d'une (ou plusieurs) covariable(s). Elles permettent par exemple de mesurer l'impact de la taille du défaut sur la performance d'un capteur, et de se replacer ainsi dans le contexte de la problématique caractérisant l'approche de Berens. L'utilisation de la méthode proposée par Pepe et Pepe & Alonzo (2000 et 2002) est illustrée sur des données simulées à l'aide de la plateforme CIVA. On notera toutefois qu'il aurait été intéressant de mettre en application la méthode de Cai & Moskowitz (2004) fournissant des estimateurs du maximum de vraisemblance des paramètres et fonctionnant avec des tailles d'échantillon assez restreintes. L'implémentation de cette procédure n'a cependant pu être entreprise faute de temps.

Effectuer un parallèle avec l'approche de Berens n'est cependant pas sans conséquence. On se retrouve en effet vite devant des problèmes assez semblables : quantité de données nécessaire, nombre de paramètres, etc. Les procédures adaptées aux courbes ROC proposent cependant leur lot de nouveautés. C'est le cas de l'approche semi-paramétrique (modèle binormal entre autres) qui permet d'assouplir les hypothèses comparé aux modèles totalement paramétriques. De telles idées pourraient éventuellement être réexploitées afin de modifier l'approche POD de Berens.

On remarquera d'autre part que les courbes ROC ne se limitent pas à l'analyse de la performance intrinsèque d'un capteur. De nouveaux concepts ont été bâtis autour d'elles. A travers les droites d'isoperformance, l'enveloppe convexe des courbes (ROCCH) ainsi que sa frontière, il est possible d'introduire des facteurs venant moduler l'appréciation de la performance. On citera par exemple les coûts (au

sens large) associés aux erreurs de détection, ou encore la prévalence (fréquence dans la population générale) des défauts. De plus, des procédures basées sur ces mêmes concepts sont censées permettre :

- d'étudier la sensibilité de la performance d'un capteur face aux éventuelles variations ou incertitudes relatives à ces facteurs.
- de sélectionner parmi les détecteurs existants celui dont la performance est optimale dans un contexte « coûts-prévalence » (l'optimalité étant dans ce cadre définie au sens « meilleur que les capteurs disponible » et non pas au sens absolu).
- d'éliminer un capteur qui n'aura jamais la possibilité d'être optimal.
- de construire un détecteur hybride (ROCCH-hybride) artificiel à partir des capteurs existants en combinant les performances. Il est destiné à rendre les décisions plus robustes face aux incertitudes sur les facteurs de coûts et de prévalence.

Une démarche parfois suivie en pratique consiste à optimiser les réglages d'un détecteur pour atteindre un certain niveau de performance sur différentes configurations mais sans prendre en compte ces facteurs. Cela revient en quelque sorte à isoler les capacités de détections du capteur des conséquences des décisions qu'il permet de prendre. Or ces deux aspects restent intimement liés dans le cadre opérationnel. La recherche d'un détecteur conduisant à des décisions robustes vis-à-vis des incertitudes sur les coûts et la prévalence se justifie donc pleinement. Cela est d'autant plus vrai que ces conditions peuvent fortement varier entre le moment où le capteur est réglé et celui où il est mis en fonction. On ajoutera qu'il est possible, à l'aide de la ROCCH, d'établir un lien entre les courbes ROC et les tests de Neyman-Pearson. Cela permet de rattacher l'étude des courbes ROC à tout un pan des statistiques mathématiques, de la théorie des tests ainsi que de l'optimisation sous contraintes.

Déjà riche d'une abondante littérature, qui plus est en pleine expansion depuis une vingtaine d'années, les courbes ROC offrent donc de nombreuses perspectives dans le domaine de l'évaluation des capteurs en contrôle non destructif. Si certaines procédures n'en sont qu'à leurs balbutiements, elles permettent néanmoins d'introduire un certain nombre de concepts dont l'impact sur les décisions peut s'avérer déterminant.

Les courbes ROC participent ainsi à la réflexion sur les développements à envisager, notamment à propos de la combinaison des résultats issus d'expérimentations et de simulations. Etant utilisées dans des domaines tels que la sociologie ou la psychométrie où l'on étudie les décisions prises par un humain, elles peuvent d'autre part s'avérer intéressantes dans l'étude des contrôles par radiographie.

## Annexes

### Principe du maximum de vraisemblance pour les lois continues

On considère une variable aléatoire  $X$  modélisée à l'aide d'une loi continue  $\mathcal{L}(\theta)$  dépendant d'un vecteur de paramètres  $\theta$  (dimension  $p$ ) a priori inconnus (loi normale de moyenne  $\mu$  et de variance  $\sigma^2$  par exemple). Une telle loi se caractérise entièrement par la donnée d'une densité de probabilité  $f_\theta(x)$ . Le but est de proposer une valeur pour  $\theta$  en s'aidant des observations *indépendantes*  $x_1, \dots, x_n$  recueillies sur  $X$  ( $n$ -échantillon).

Le maximum de vraisemblance consiste à choisir la valeur  $\hat{\theta}$  de  $\theta$  maximisant la probabilité d'observer  $x_1, \dots, x_n$  en supposant que cet échantillon provient effectivement de réalisations indépendantes de  $\mathcal{L}(\theta)$ . Cette dernière probabilité correspond à la vraisemblance de l'échantillon et est donnée par :

$$L(x_1, \dots, x_n, \theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P_\theta(x_i) = \prod_{i=1}^n f_\theta(x_i)$$

Ainsi, le problème à résoudre est le suivant :

$$\hat{\theta} = \underset{\theta}{\text{Argmax}} L(x_1, \dots, x_n, \theta) \quad \text{ou encore} \quad \hat{\theta} = \underset{\theta}{\text{Argmax}} \ln(L(x_1, \dots, x_n, \theta))$$

Le passage au logarithme permet de passer de la maximisation d'un produit à celle d'une somme. De par la croissance et la monotonie de la fonction  $x \mapsto \ln(x)$ , les résultats ne s'en verront pas modifiés.

Ce problème revient à résoudre les équations de vraisemblance :

$$\frac{\partial \ln(L)}{\partial \theta_k} \quad \text{pour } k = 1, \dots, p$$

On vérifiera enfin que  $\hat{\theta}$  est bien un maximum de  $\ln(L)$  (concavité) :

$$\left( \frac{\partial^2 \ln(L)}{\partial \theta^2} \right) \Big|_{\theta=\hat{\theta}} < 0$$

En pratique, les équations de vraisemblance sont résolues par l'algorithme de Newton-Raphson (voir Algorithme de Newton-Raphson dans les Annexes).

$\hat{\theta}_{MV}$  suit asymptotiquement une loi normale de moyenne  $\theta$  (estimateur sans biais) et de variance minimale :

$$\hat{\theta}_{ML} \xrightarrow{\mathcal{L}} \mathcal{N}(\theta, I^{-1}(\theta))$$

où la  $I(\theta)$  représente la matrice d'information de Fisher :

$$I_{ij}(\theta) = -E \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln(L) \right) \quad \text{pour } 1 \leq i, j \leq p$$

### Algorithme de Newton-Raphson et algorithme du scoring – équations de vraisemblance

Il arrive fréquemment que les équations de vraisemblance ne puissent pas être résolues facilement, voire même, qu'elles n'admettent pas de solutions analytiques. Dans ces cas, les algorithmes de résolution numérique tel que celui de Newton-Raphson peuvent être utilisés.

#### • Algorithme de Newton-Raphson

Supposons que l'on dispose d'un échantillon d'observations indépendantes  $\mathbf{X} = (X_1, \dots, X_n)$  d'une variable aléatoire  $X$  admettant une densité de probabilité  $f_\theta(x)$  paramétrée par un vecteur  $\theta$ , inconnu et de dimension  $k$ . Notons  $\hat{\theta}_{MV}$  l'estimateur du maximum de vraisemblance de  $\theta$ .

Cet algorithme utilise *le score*  $V(\theta)$  se définissant comme le gradient de la log-vraisemblance associée à  $\mathbf{X}$ . Sa propriété principale est qu'il s'annule pour  $\theta = \hat{\theta}_{MV}$ , soit :

$$V(\theta) = \nabla_\theta \ln(\mathbf{X}, \theta) \quad \text{et} \quad V(\hat{\theta}_{MV}) = 0$$

L'algorithme repose sur un développement de Taylor de  $V(\theta)$  autour d'une « valeur initiale »  $\theta_0$  :



$$V(\boldsymbol{\theta}) \cong V(\boldsymbol{\theta}_0) - J(\boldsymbol{\theta}_0) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \quad \text{où} \quad J(\boldsymbol{\theta}_0) = - \sum_{i=1}^n \left[ \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^t \ln(f(X_i, \boldsymbol{\theta})) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

avec  $J(\boldsymbol{\theta}_0)$ , la matrice d'information observée en  $\boldsymbol{\theta}_0$ . En utilisant le fait que  $V(\hat{\boldsymbol{\theta}}_{MV}) = 0$ , l'approximation précédente peut se réécrire :

$$\hat{\boldsymbol{\theta}}_{MV} \cong \boldsymbol{\theta}_0 + J(\boldsymbol{\theta}_0)^{-1} V(\boldsymbol{\theta}_0)$$

On exécute donc un algorithme consistant en une série de mise à jour de la valeur du paramètre jusqu'à stabilisation :

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + J^{-1}(\boldsymbol{\theta}_i) V(\boldsymbol{\theta}_i)$$

On considère qu'il y a convergence lorsque cette mise à jour n'engendre plus une modification suffisamment importante de  $\boldsymbol{\theta}_i$  ou bien si le score est assez proche de zéro, c'est-à-dire :

$$\text{si } I^{-1}(\boldsymbol{\theta}_i) V(\boldsymbol{\theta}_i) < \epsilon_1 \quad \text{ou si} \quad V(\boldsymbol{\theta}_i) < \epsilon_2$$

où les quantités  $\epsilon_1$  et  $\epsilon_2$  sont « petites ».

### • Algorithme du Scoring (de Fisher)

L'algorithme du Scoring est une version légèrement modifiée de l'algorithme de Newton-Raphson pour la résolution des équations de vraisemblance. La matrice  $J(\boldsymbol{\theta})$  est remplacée par la matrice de Fisher  $I(\boldsymbol{\theta}) = E[J(\boldsymbol{\theta})]$ . L'algorithme obtenu est appelé l'*algorithme du scoring de Fisher* :

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + I^{-1}(\boldsymbol{\theta}_i) V(\boldsymbol{\theta}_i)$$

## Rappels sur la loi normale, logistique, log-normale et log-logistique

- Le Tableau 5 contient des rappels concernant les propriétés de la loi logistique de paramètres  $\mu$  et  $s$  et de la loi normale de paramètres  $\mu$  et  $\sigma^2$  :

	Loi logistique	Loi normale
Densité	$f(x) = \frac{e^{-\frac{x-\mu}{s}}}{s \left(1 + e^{-\frac{x-\mu}{s}}\right)}$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$
Fonction de répartition	$F(x) = \frac{1}{1 + e^{-\frac{x-\mu}{s}}}$	$F(x) = \Phi(x)$
Espérance	$E(X) = \mu$	$E(X) = \mu$
Variance	$V(X) = \frac{s^2\pi^2}{3}$	$V(X) = \sigma^2$

Tableau 5 : loi logistique de paramètres  $\mu$  et  $s$  - loi normale de paramètres  $\mu$  et  $\sigma$

- La **loi log-logistique** se définit à partir de la loi logistique : on dit que  $X$  suit une loi log-logistique si  $\ln(X)$  est distribuée selon une loi logistique. Autrement dit :

$$X \sim \text{log-logistique} \Leftrightarrow \ln(X) \sim \text{logistique}$$

Ainsi, si  $X \sim \text{log-logistique}$  :

$$P(X \leq x) = P(\ln(X) \leq \ln(x)) = P(Y \leq \ln(x)) \quad \text{où } Y \sim \text{logistique}(\mu, s)$$

Et par conséquent, la fonction de répartition d'une loi log-logistique est donnée par :

$$F_X(x) = \frac{1}{1 + e^{-\frac{\ln(x)-\mu}{s}}} \quad (41)$$

De façon générale, on exprime la fonction de répartition d'une loi log-logistique à partir de paramètres  $\alpha$  et  $\beta$  :

$$F_X(x) = \frac{1}{1 + \left(\frac{x}{\alpha}\right)^{-\beta}}$$

On a donc la relation suivante entre les paramètres de la loi logistique et ceux de la loi log-logistique :

$$\begin{cases} \alpha = e^{-\mu} \\ \beta = \frac{1}{s} \end{cases}, \text{ soit : } \log\text{-logistique}(\alpha, \beta) \equiv \log\text{-logistique}\left(e^{-\mu}, \frac{1}{s}\right)$$

- La **loi log-normale** est définie à partir de la loi normale : on dit que X suit une loi log-normale si  $\ln(X)$  est distribuée selon une loi normale, soit :

$$X \sim \text{log-normale} \Leftrightarrow \ln(X) \sim \text{normale}$$

La fonction de répartition de la loi log-normale est donc donnée par :

$$F_X(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)$$

## Propriété de la matrice de Fisher d'un modèle de translation - changement d'échelle

Soit un n-échantillon  $(x_1, \dots, x_n)$  d'une variable suivant un modèle de translation-changement d'échelle. La log-vraisemblance d'un tel modèle s'écrit :

$$\log L(\underline{x}/\mu, \sigma) = -n \log(\sigma) + \sum_{i=1}^n \log f\left(\frac{x_i - \mu}{\sigma}\right)$$

En posant,  $y = \frac{x - \mu}{\sigma}$  et  $h(z) = \log f(z)$ , on peut l'exprimer sous la forme (voir [41]) :

$$\log L(\underline{y}/\mu, \sigma) = -n \log(\sigma) + \sum_{i=1}^n h(y_i)$$

Et les équations du maximum de vraisemblance s'écrivent donc :

$$\begin{cases} \frac{\partial \log L}{\partial \mu}(\underline{y}/\mu, \sigma) = -\frac{1}{\sigma} \sum_{i=1}^n h'(y_i) = 0 \\ \frac{\partial \log L}{\partial \sigma}(\underline{y}/\mu, \sigma) = -\frac{n}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^n y_i h'(y_i) = 0 \end{cases} \quad (42)$$

où  $h'(z) = \frac{\partial h}{\partial z}(z)$ .

Etant donné que les  $(y_i)_{i=1, \dots, n}$  sont iid, le passage à l'espérance dans les équations de (42) conduit aux relations suivantes :

$$\begin{cases} E\left(\frac{\partial \log L}{\partial \mu}\right) = -\frac{n}{\sigma} E(h'(y)) = 0 \\ E\left(\frac{\partial \log L}{\partial \sigma}\right) = -\frac{n}{\sigma} [1 + E(yh'(y))] = 0 \end{cases} \Leftrightarrow \begin{cases} E(h'(y)) = 0 \\ E(yh'(y)) = -1 \end{cases} \quad (43)$$

De la même façon, on peut écrire les dérivées secondes de la log-vraisemblance, et en utilisant les propriétés mises en évidence dans les équations (43), on obtient :

$$\begin{cases} E\left(\frac{\partial^2 \log L}{\partial \mu^2}\right) = \frac{n}{\sigma^2} E(h''(y)) \\ E\left(\frac{\partial^2 \log L}{\partial \sigma^2}\right) = \frac{n}{\sigma^2} E(h''(y)y^2 + 2h'(y)y + 1) = \frac{n}{\sigma^2} E(h''(y)y^2 - 1) \\ E\left(\frac{\partial^2 \log L}{\partial \mu \partial \sigma}\right) = E\left(\frac{\partial^2 \log L}{\partial \sigma \partial \mu}\right) = \frac{n}{\sigma^2} E(h'(y) + h''(y)y) = \frac{n}{\sigma^2} E(h''(y)y) \end{cases} \quad (44)$$

Les équations (44) permettent d'exprimer la matrice d'information de Fisher de la façon suivante :

$$I(\mu, \sigma) = -\frac{n}{\sigma^2} \begin{bmatrix} E(h''(y)) & E(h''(y)y) \\ E(h''(y)y) & E(h''(y)y^2 - 1) \end{bmatrix} \triangleq -\frac{n}{\sigma^2} \begin{bmatrix} k_0 & -k_1 \\ -k_1 & k_2 \end{bmatrix} \quad (45)$$

La variable  $y$  a par construction une distribution indépendante des paramètres  $\mu$  et  $\sigma$ . D'autre part, la fonction  $h$  s'exprimant en fonction de la densité  $f$  ne dépend pas non plus de ces paramètres. Par conséquent, en s'appuyant sur le théorème de transfert, il est possible

d'affirmer que les différents composants  $k_0, k_1, k_2$  de la matrice de Fisher, correspondant à l'espérance de variables dont la distribution ne dépend pas de  $\mu$  et  $\sigma$ , sont eux-mêmes indépendants de  $\mu$  et  $\sigma$ .

**Remarque :**

Si la fonction  $h$  est paire, on a :

$$\begin{aligned} h(-y) &= h(y) \\ h'(y) &= -h'(-y) \\ h''(y) &= h''(-y) \end{aligned}$$

la fonction  $k(y) = yh''(y)$  est impaire et par conséquent d'intégrale nulle sur  $\mathbb{R}$ . Autrement dit :

$$E(h''(y)y) = k_1 = 0$$

et les termes non diagonaux de la matrice d'information de Fisher sont réduits à 0 :

$$I(\mu, \sigma) = -\frac{n}{\sigma^2} \begin{bmatrix} k_0 & 0 \\ 0 & k_2 \end{bmatrix}$$

La propriété de symétrie sur  $h$  se répercute sur la densité canonique  $f$  :

$$h(-y) = h(y) \Leftrightarrow \log f(-y) = \log f(y) \Leftrightarrow f(-y) = f(y)$$

Par conséquent, dans le cas où  $f$  est paire, on sait par avance que la matrice d'information de Fisher du modèle de translation – changement d'échelle associé est diagonale.

Ainsi, par exemple, dans le cas de la loi normale :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

est une fonction paire et la matrice d'information de Fisher d'un modèle  $\mathcal{N}(\mu, \sigma^2)$  est :

$$I_{\mathcal{N}(\mu, \sigma^2)}(\mu, \sigma) = -\frac{n}{\sigma^2} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

où les termes non diagonaux sont bien nuls.

## Démonstrations des propriétés de la courbes ROC

**Propriété 1 :** Les points (0,0) et (1,1) appartiennent toujours à la courbe ROC.

Lorsque  $s_{th} \leq \min(s)$  :

$$P(S \geq s_{th} / \delta = 1) = P(S \geq s_{th} / \delta = 0) = 1 \quad \text{d'où } POD = PFA = 1$$

Lorsque  $s_{th} \geq \max(s)$  :

$$P(S \geq s_{th} / \delta = 1) = P(S \geq s_{th} / \delta = 0) = 0 \quad \text{d'où } POD = PFA = 0$$

Cela nous permet de remarquer d'autre part que lorsque  $s_{th}$  varie de  $s_{th}^{min} \leq \min(s)$  à  $s_{th}^{max} \geq \max(s)$ , on parcourt la courbe ROC du point (1,1) au point (0,0).

**Propriété 2 :** Le point (PFA, POD) = (0,1) correspond à un détecteur parfait.

Celui-ci discrimine les pièces avec une probabilité nulle de commettre une erreur. Dans la pratique, l'idéal est donc de choisir un seuil de façon à obtenir un capteur le plus proche possible de ce point.

La bissectrice correspond à l'inverse à un ensemble de couples (capteur, seuil) dont les conclusions sont aléatoires.

En effet, pour en chacun des points de la bissectrice on a  $POD = PFA$ , ou de façon équivalente :

$$P(S \geq s_{th} / \delta = 1) = P(S \geq s_{th} / \delta = 0) = P(S \geq s_{th}) = p$$

Autrement dit, le fait qu'un défaut soit présent ou non n'influence pas la décision à laquelle nous mène le détecteur. La probabilité  $p$  de conclure en faveur de la présence d'un défaut est indépendante de  $\delta$ , tout se passe donc comme si cette décision était prise par rapport à la réalisation d'une variable binomiale de paramètre  $p$ . On peut par exemple imaginer décider en lançant une pièce. On dira qu'un défaut est présent si on observe pile, et l'inverse si on observe face. Si la pièce n'est pas pipée, on aura  $p = 1/2$ . Si elle l'est, on aura  $p \in [0,1] \setminus \{1/2\}$ . On parcourt ainsi toute la bissectrice, que l'on appelle aussi la *diagonale de chance*.

**Propriété 3 :** La courbe ROC est inaltérée si le score de classification (S) subit une transformation monotone.

Soit  $T = \varphi(S)$  une transformation croissante monotone de  $S$ . Soit  $t_{th} = \varphi(s_{th})$ , alors :

$$P(T > t_{th}/k) = P(\varphi(S) > \varphi(s_{th})/k) = P(S > s_{th}/k) \quad \forall k \in \{N, P\}$$

Par conséquent, le point associé au seuil  $t_{th}$  de la courbe ROC liée à  $T$  est le même que le point associé au seuil  $s_{th}$  de la courbe ROC liée à  $S$ . Cette remarque étant valable pour n'importe quelle valeur du seuil  $t_{th}$ , on en déduit que les deux courbes ROC sont identiques.

**Propriété 4 :** Si  $S/\delta = 0$  est stochastiquement plus petite que  $S/\delta = 1$ , la courbe ROC est au-dessus de la diagonale de chance.

La comparaison stochastique se traduit en termes de fonctions de répartition de la façon suivante :

$$\forall x, \quad F_S^{\delta=0}(x) \geq F_S^{\delta=1}(x) \quad \text{c'est à dire} \quad P_{\delta=0}(S \leq x) \geq P_{\delta=1}(S \leq x)$$

Or, étant donné que :

$$POD(s_{th}) = P_{\delta=1}(S \geq s_{th}) \quad \text{et} \quad PFA(s_{th}) = P_{\delta=0}(S \geq s_{th})$$

Il vient :

$$\forall s_{th}, \quad POD(s_{th}) \geq PFA(s_{th})$$

Ce qui signifie que pour toute valeur du seuil, le point  $(PFA, POD)$  se trouve au dessus de la bissectrice.

**Propriété 5 :** La pente de la courbe ROC correspond au rapport de vraisemblance.

Pour voir cela, il faut considérer la courbe ROC comme une courbe paramétrée par le seuil. Posons dans ce but :

$$\begin{cases} y(t) = P(S > t / P) = 1 - F_{S/P}(t) \\ x(t) = P(S > t / N) = 1 - F_{S/N}(t) \end{cases} \Rightarrow ROC \equiv \{(x(t), y(t)), t \in S\}$$

Ainsi :

$$\begin{cases} \frac{dy}{dt} = \frac{d}{dt}(1 - F_{S/P}(t)) = -f_{S/P}(t) \\ \frac{dx}{dt} = \frac{d}{dt}(1 - F_{S/N}(t)) = -f_{S/N}(t) \end{cases}$$

D'où :

$$\frac{dy}{dx} = \frac{dy}{dt} \left(\frac{dx}{dt}\right)^{-1} = \frac{f_{S/P}(t)}{f_{S/N}(t)} = \text{Rapport de vraisemblance}$$

Cette propriété permet de faire des liens intéressants avec les tests de Neyman-Pearson basés sur le rapport de vraisemblance.

**Propriété 6 :** Si les densités  $f_{S/P}$  et  $f_{S/N}$  ont un rapport de vraisemblance monotone (avec  $S$  la statistique associée), alors la courbe ROC est concave.

On rappelle qu'une variable  $X$  de loi  $P_\theta$  admettant une densité  $f(x, \theta)$  est dite à rapport de vraisemblance monotone s'il existe une statistique  $T: \mathcal{O}$  (ou  $\mathcal{O}^n$ )  $\rightarrow \mathbb{R}$  telle que la fonction  $g$  définie par :

$$g(T(x)) = \frac{L(x, \theta')}{L(x, \theta)} \quad \text{pour} \quad \theta' > \theta$$

soit une fonction monotone croissante en  $T(x)$ .

L'hypothèse de la propriété 6 revient à considérer que  $X = S$  et  $T = Id$  et en notant  $\mathcal{L}(t)$  le rapport de vraisemblance, que  $\frac{d\mathcal{L}}{dt} \geq 0$ . Par conséquent :

$$\frac{d^2y}{dx^2} = \frac{d\mathcal{L}}{dx} = \frac{d\mathcal{L}}{dt} \frac{dt}{dx} = \frac{d\mathcal{L}}{dt} \left(\frac{dx}{dt}\right)^{-1} = -f_{S/N}(t) \frac{d\mathcal{L}}{dt} \leq 0$$

D'où l'on déduit la concavité de la courbe ROC.

**Propriété 7 :** L'aire sous la courbe ROC (AUC) est égale à  $P(S_p > S_N)$  où  $S_N \equiv S/N$  et  $S_p \equiv S/P$

Pour montrer ce résultat, il faut utiliser les résultats de la démonstration de la propriété 5. En remarquant que l'indépendance de  $S_N$  et  $S_p$  implique que  $P(S > t/P) P(S = t/N) = P(S_p > t, S_N = t)$ , l'aire sous la courbe est donnée par :

$$\begin{aligned}
AUC &= \int_0^1 y(x) dx = \int_{+\infty}^{-\infty} y(t) \frac{dx}{dt} dt = - \int_{+\infty}^{-\infty} P(S > t / P) f_{S/N}(t) dt \\
&= \int_{-\infty}^{+\infty} P(S > t / P) P(S = t / N) dt = \int_{-\infty}^{+\infty} P(S_P > t, S_N = t) dt \\
&= \int_{-\infty}^{+\infty} P(S_P > S_N, S_N = t) dt = P(S_P > S_N)
\end{aligned}$$

La dernière égalité étant obtenue par la formule des probabilités totales.

## Bibliographie

- [1] Wikipedia. [Online]. <http://fr.wikipedia.org/>
- [2] Techniques de l'ingénieur. [Online]. [www.techniques-ingenieur.fr/](http://www.techniques-ingenieur.fr/)
- [3] Charles Hellier, *Handbook of nondestructive evaluation*, McGraw-Hill, Ed., 2003.
- [4] N. Dominguez, V. Feuillard, R. Lebrun, and F. Mangeant, "Fiabilité des CND et estimation de probabilités de détection : problématiques et avancées," *Rapport EADS - CEA*.
- [5] A. P. Berens, "NDE reliability data analysis," *ASM Handbook*, 1989.
- [6] R.C.H Cheng and T.C Iles, "Confidence bands for cumulative distribution functions of continuous random variables," *Technometrics*, 1983.
- [7] R. C. H. Cheng and T. C. Iles, "One-sided confidence bands for cumulative Distribution functions," *Technometrics*, 1998.
- [8] V. Morin, J. F. Morin, M. Mercier, M. P. Moineau, and J. P. Codet, "Les courbes ROC en biologie médicale (Clock around the ROC)," *Immuno-analyse Biologie Spécialisée*, 1998.
- [9] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, 2006.
- [10] X.-H. Zhou, N. A. Obuchowski, and D. K. McClish, *Statistical Methods in Diagnostic Medicine*, Wiley-Interscience, Ed., 2002.
- [11] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, Ed., 2004.
- [12] W. J. Krzanowski and D. J. Hand, *ROC curves for continuous data*, CRC Press, Ed., 2009.
- [13] H. Van Trees, *Detection, Estimation, and Modulation Theory - Part I*, John Wiley & Sons, Ed., 2001.
- [14] W. Hsieh and B.W. Turnbull, "Nonparametric and semiparametric estimation of the receiver operating characteristic curve," *The annals of statistics*, 1996.
- [15] D. Dorfman and E. Alf, "Maximum likelihood estimation of parameters of signal detection theory - A direct solution," *Psychometrika*, 1968.
- [16] J. C. Ogilvie and C. D. Creelman, "Maximum-likelihood estimation of receiver operating characteristic curve parameters," *Journal of mathematical psychology*, 1968.
- [17] D. Dorfman and E. Alf, "Maximum-likelihood estimation of parameters of signal detection theory and determination of confidence intervals - Rating method data," *Journal of mathematical psychology*, 1969.
- [18] G. Li, R. C. Tiwari, and M. T. Wells, "Semiparametric inference for quantile comparison function with applications to receiver operating characteristic curves," *Biometrika*, 1999.
- [19] J. Qin and Biao Zhang, "Using logistic procedures for estimating receiver operating characteristic curves," *Biometrika*, 2003.
- [20] J. Qin and Biao Zhang, "A goodness-of-fit test for logistic regression models based on case-control data," *Biometrika*, 1997.
- [21] C. E. Metz, B. A. Herman, and J.-H. Shen, "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Statistics in medicine*, 1998.

- [22] K.H Zou and D.K. Hall, "Two transformation models for estimating an ROC curve derived from continuous data," *Journal of applied statistics*, 2000.
- [23] J. Gu and S. Ghosal, "Bayesian ROC curve estimation under binormality using a rank likelihood," *Journal of statistical planning and inference*, 2009.
- [24] M. S. Pepe, "An interpretation for the ROC curve and inference using GLM procedures," *Biometrics*, 2000.
- [25] T. A. Alonzo and M. S. Pepe, "Distribution-free ROC analysis using binary regression techniques," *Biostatistics*, 2002.
- [26] M. S. Pepe, G. Longton, and H. Janes, "Estimation and comparison of receiver operating characteristic curves," *The Stata journal*, vol. 9, no. 1, 2009.
- [27] H. Janes and M. S. Pepe, "Accommodating covariates in ROC analysis," *Stata Journal - NIH public access*, 2009.
- [28] H. Janes and M. S. Pepe, "Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve," *Biometrika*, 2009.
- [29] T. Cai and C. S. Moskowitz, "Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test," *Biostatistics*, 2004.
- [30] F. Provost and T. Fawcett, "Robust classification for imprecise environments," *Machine learning*, 2001.
- [31] D. J. Hand, "L'incohérence de l'aire sous la courbe ROC, que faire à ce propos ?," *Revue Modulad*, 2010.
- [32] D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," *Journal of mathematical psychology*, 1975.
- [33] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under receiver operating characteristic (ROC) curve," *Radiology*, 1982.
- [34] D. Faraggi and B. Reiser, "Estimation of the area under the ROC curve," *Statistics in medicine*, 2002.
- [35] J. N. Rouder and J. Lu, "An introduction to bayesian hierarchical models with an application in the theory of signal detection," *Psychonomic bulletin & Review*, 2005.
- [36] A. Erkanli, M. Sung, E. J. Costello, and A. Angold, "Bayesian semi-parametric ROC analysis," *Statistics in medicine*, 2006.
- [37] Y.-K. Choi, W. O. Johnson, M. T. Collins, and I. A. Gardne, "Bayesian inferences for receiver operating characteristic curves in the absence of a gold standard," *Journal of agricultural, biological, and environmental statistics*, 2006.
- [38] C. Wang, B. W. Turnbull, Y. T. Gröhn, and S. S. Nielsen, "Nonparametric estimation of ROC curves based on bayesian models when the true disease state is unknown," *Journal of agricultural, biological, and environmental statistics*, 2007.
- [39] C. Hurlin, *Econométrie des variables qualitatives - Modèles multinomiaux - Chapitre 2*. Université d'Orléan, 2003.
- [40] Robin and et Al, "pROC : an open source package for R and S+ to analyze and compare ROC curves," *BMC bioinformatics*, 2011.
- [41] A. Stuart and K. Ord, *Kendall's advanced theory of statistics - Volume 2*, 5th ed., 1991.
- [42] H. Kang, E. J. Bedrick, S. R. Atlas, and C. L. Willman, "ROC curve of generalized odds-rate model," , The university of New Mexico.
- [43] J. R. Olsen, "Note on the uniqueness of the maximum likelihood estimator for the tobit model," *Econometrica*, 1978.
- [44] C. Hurlin, *Econométries des variables qualitatives - Modèles à variable dépendante limitée - chapitre 3*. Université d'Orléan, 2003.
- [45] Tom Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, 2006.