

Eléments de la théorie de la décision

Paul HONEINE

Université de technologie de Troyes

— 2014 —

Préface

Le présent document constitue le support de cours pour le module “Éléments de la théorie de l’estimation et de la décision”. Ce module est enseigné en Master 2 Recherche “Contrôle Industriel” à l’Université Libanaise, Beyrouth, Liban, dans le cadre d’un double diplôme avec l’Université de technologie de Compiègne, France.

En version *draft*, ce document traite la théorie de la décision, principalement avec la détection binaire, ainsi que les hypothèses multiples. La partie “théorie de l’estimation” est en cours de rédaction.

Ce document n’aurait pas vu le jour sans le soutien de Régis Lengellé de l’Université de technologie de Troyes, France.

Paul Honeine
paul.honeine@utt.fr
Octobre 2014

Table des matières

Préface	iii
Préambule	1
1 Introduction à la théorie de la décision	3
1.1 Formulation du problème	3
1.2 Notations	4
1.3 Règle du maximum de probabilité a posteriori	5
1.4 Exemple : communication binaire	6
1.5 Probabilités de décision conditionnelles	7
2 Les principaux critères de décision	11
2.1 Critère de Bayes	11
2.1.1 Coût moyen de décision	11
2.1.2 Critère de Bayes	12
2.1.3 Performances du détecteur	13
2.1.3.a Probabilités d'erreur	13
2.1.3.b Probabilités de bonne décision	13
2.1.4 Remarque : coût (0,1)	14
2.1.5 Exemple	14
2.2 Critère minimax	15
2.2.1 Définition du critère	15
2.2.2 Etude du graphe de $C(P(S_2))$	16
2.2.3 Critère Minimax	16
2.2.3.a Cas particulier : coûts (0,1)	17
2.3 Critère de Neyman-Pearson	17
2.3.1 Position du problème	17
2.3.2 Critère de Neyman-Pearson	18

TABLE DES MATIÈRES

3	Caractéristiques opérationnelles d'un récepteur	21
3.1	Etude des courbes COR	21
3.2	Utilisation des courbes COR	22
3.2.1	Critère de Neyman-Pearson	22
3.2.2	Critère minimax	22
3.2.3	Critère de Bayes	23
3.3	Démonstrations	23
4	Mesures et hypothèses multiples	29
4.1	Décision avec mesures multiples	29
4.1.1	Critères de décision	30
4.1.1.a	Coût moyen de décision	30
4.1.1.b	Critères de décision	31
4.1.2	Exemple : détection d'un signal constant dans du bruit	31
4.1.3	Exemple : détection de signaux de puissances différentes	33
4.2	Décision avec hypothèses multiples	34
4.2.1	Espace d'observation, espace de décision et coût moyen	34
4.2.2	Critère de Bayes	35
4.2.2.a	Cas général	35
4.2.2.b	Exemple : décision à hypothèses ternaires . .	36
4.2.2.c	Cas particulier des coûts (0.1)	38
4.2.3	Autres critères	38
4.2.3.a	Critère minimax	38
4.2.3.b	Critère de Neyman-Pearson	38
5	Hypothèses composites	41
A	Fonction caractéristique	43

Préambule

En théorie statistique de la décision, trois grands types de problèmes émergent :

La détection de présence / d'absence

Il s'agit de savoir si les mesures observées contiennent ou non un message dû à un évènement recherché. Le message peut être de forme connue (transmission binaire), partiellement connue (sonar actif, pour lequel le signal réfléchi va être filtré par le milieu de propagation et par la réflexion elle-même), inconnue ou aléatoire (sonar passif).

La classification

Lorsqu'un évènement se produit, il peut donner naissance à un signal représentant cet évènement parmi un nombre fini d'évènements possibles. Il s'agit d'une généralisation du problème précédent. A titre d'exemple, en décision radar, il peut être utile de différencier un avion d'un hélicoptère, d'un missile ou de décider de l'absence de réflecteur radar.

L'estimation de paramètres

Lorsque l'on reçoit un message de forme partiellement connue, il est intéressant de pouvoir estimer certains paramètres, comme par exemple l'estimation de la distance et de la vitesse d'une cible en radar.

L'objet de ce document est de présenter les outils de la théorie de la décision statistique pour répondre aux deux premières classes de problèmes. La partie associée à la troisième classe est en cours de rédaction.

Chapitre 1

Introduction à la théorie de la décision binaire à une observation

1.1 Formulation du problème

En traitement du signal, ou en reconnaissance des formes, un processus de décision peut être représenté, en général, de la manière suivante : à partir d'un ensemble d'observations (ou de paramètres ou de caractéristiques), les données observées doivent être attribuées à une classe, en principe, prédéterminée. Par exemple, lors d'un problème de détection radar, le signal reçu est analysé afin de déterminer s'il y a présence de cible ou non. Dans un système de communication binaire, une séquence de 0 et 1 est transmise par un canal. Vu du récepteur, les signaux reçus sont perturbés par du bruit provenant du milieu de transmission et du récepteur lui-même. Celui-ci ne sait pas quel signal représente un 0 et quel signal représente un 1, mais doit effectuer un choix sur le caractère reçu à partir de l'observation, ou des observations, dont il dispose. Pour ces deux cas, il n'y a que deux hypothèses possibles : présence / absence ou 0 / 1. Il s'agit d'hypothèse binaire et donc de décision binaire.

Les deux situations précédentes peuvent être résumées par une source émettant deux signaux possibles à différents instants. Les signaux S_1 et S_2 ont *a priori* respectivement les probabilités $P(S_1)$ et $P(S_2)$ d'avoir été émis. Ces signaux parviennent à l'observateur sous une forme altérée (addition de bruit par exemple), qui en obtient une mesure z d'une variable aléatoire Z .

Connaissant la nature binaire de la source, l'observateur va devoir prendre

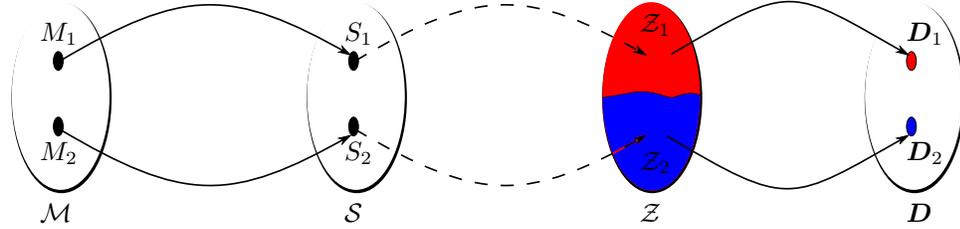


FIGURE 1.1 – Illustration du problème de décision binaire

une décision sur la base de cette mesure. Il prendra la décision D_1 s'il estime avoir reçu le signal S_1 , et prendra la décision D_2 dans l'autre cas. Il devra donc effectuer une partition de l'espace d'observation Z en deux sous ensembles Z_1 et Z_2 . Si l'observation z est dans Z_1 , il prendra la décision D_1 , et vice versa. Tout le problème est donc d'effectuer la partition de l'espace des observations de manière judicieuse, c'est à dire de définir des critères permettant d'effectuer cette partition.

Afin de poser un peu mieux le problème, on considère dans la suite une règle de décision simple pour la traiter.

1.2 Notations

Précisons tout d'abord quelques notations. Soient $P(S_1)$ et $P(S_2)$ les probabilités a priori des signaux S_1 et S_2 . L'observation z est une réalisation d'une variable aléatoire Z qui dépend du signal émis et du canal de transmission. Soit $p(Z)$ la loi de vraisemblance de la variable aléatoire Z . On a :

$$\begin{aligned} p(Z) &= p(Z \text{ et } (S_1 \text{ ou } S_2)) \\ &= p(Z \text{ et } S_1) + p(Z \text{ et } S_2), \end{aligned}$$

car les deux évènements sont disjoints. On notera par la suite l'évènement $(A \text{ et } B)$ par (A, B) . En utilisant la relation de Bayes :

$$P(A, B) = P(A|B) P(B), \quad (1.1)$$

on obtient

$$p(Z) = p(Z|S_1) p(S_1) + p(Z|S_2) p(S_2),$$

où $p(Z|S_1)$ représente la loi de vraisemblance de la variable Z conditionnellement à S_1 , c'est à dire sachant que le signal S_1 a été émis ; de même pour $p(Z|S_2)$.

1.3 Règle du maximum de probabilité a posteriori

Tout d'abord, on étudie la règle de décision suivante : en observant z , une réalisation de Z ,

- D_1 est prise si

$$p(S_1|z) \geq p(S_2|z);$$

- de manière symétrique, la décision D_2 est prise si

$$p(S_2|z) > p(S_1|z).$$

Dans ces expressions, $p(S_j|z)$ désigne la probabilité que le message S_j ait été émis sachant que l'on a observé la valeur z .

En utilisant à nouveau la relation de Bayes, avec

$$p(S_i|z) = \frac{p(z|S_i)P(S_i)}{p(z)}, \quad (1.2)$$

on obtient la règle suivante : la décision D_1 est prise si $\frac{p(z|S_2)}{p(z|S_1)} \leq \frac{P(S_1)}{P(S_2)}$; sinon la décision D_2 est prise. Sous une forme plus compact, on écrit

$$\frac{p(z|S_2)}{p(z|S_1)} \underset{D_1}{\overset{D_2}{\gtrless}} \frac{P(S_1)}{P(S_2)}.$$

Ce critère est appelé *maximum de probabilité a posteriori*. Rappelons qu'il s'agit d'un rapport de probabilité de l'observation z , avec $p(z|S_2)/p(z|S_1)$ (qui est déterministe), que l'on compare au seuil $\lambda = P(S_1)/P(S_2)$.

En considérant le *rapport de vraisemblance* de la variable aléatoire Z , selon

$$\Lambda(Z) = \frac{p(Z|S_2)}{p(Z|S_1)}.$$

En utilisant le rapport de vraisemblance, la règle de décision s'écrit alors

$$\Lambda(Z) \underset{D_1}{\overset{D_2}{\gtrless}} \frac{P(S_1)}{P(S_2)}.$$

Puisque le rapport de vraisemblance $\Lambda(Z)$ est aléatoire, il est caractérisé par sa loi de probabilité, désignée par $p_\Lambda(\cdot)$.

Pour appliquer cette règle de décision, il est indispensable de connaître les probabilités a priori $p(S_1)$ et $P(S_2)$, ainsi que des lois de probabilités de Z conditionnellement à S_1 et S_2 .

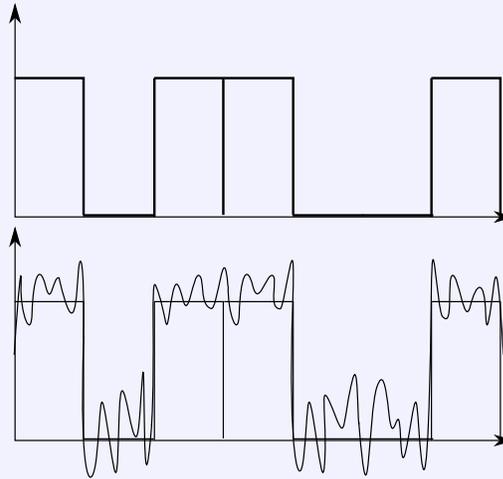


FIGURE 1.2 – Illustration de communication binaire synchrone, avec l’émission d’une séquence de 0 et 1 (figure en haut) et la réception (figure en bas).

1.4 Exemple : communication binaire

En communication binaire synchrone, la source émet un signal binaire, prenant des valeurs 0 ou 1, le basculement se faisant éventuellement aux instants nT , pour $n = 1, 2, \dots$. Voir FIGURE 1.3. Au niveau du récepteur, se superpose au signal un bruit gaussien $b(t)$, supposé tel que $E(b(t)) = 0$ et $\text{Var}(b(t)) = \sigma^2$.

A chaque instant t_n vérifiant $t_n = (n + \frac{1}{2})T$, l’observateur effectue une mesure z , sur la base de laquelle il doit prendre la décision concernant la valeur émise du signal $S_1 = 0$ ou $S_2 = 1$. La règle de décision adoptée est celle du maximum de probabilité a posteriori.

Soit $z(t)$ le signal observé, supposé de la forme $z(t) = x(t) + b(t)$, où $x(t)$ peut prendre les valeurs 0 ou 1. Pour l’observateur, z est une réalisation d’une variable aléatoire Z dont la loi de vraisemblance $p(Z)$ peut s’écrire

$$p(Z) = p(Z|S_1) P(S_1) + p(Z|S_2) P(S_2),$$

avec

$$\begin{aligned} p(Z|S_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2} Z^2\right) \\ p(Z|S_2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2} (Z-1)^2\right). \end{aligned}$$

Le rapport de vraisemblance s'écrit alors

$$\begin{aligned} \Lambda(Z) &= \frac{p(Z|S_2)}{p(Z|S_1)} \\ &= \exp\left(\frac{-1}{2\sigma^2} (1-2Z)^2\right), \end{aligned}$$

d'où la règle de décision

$$\exp\left(\frac{-1}{2\sigma^2} (1-2z)^2\right) \underset{D_1}{\overset{D_2}{\geq}} \frac{P(S_1)}{P(S_2)}.$$

L'utilisation du logarithme de cette expression permet d'écrire

$$\frac{-1}{2\sigma^2} (1-2z)^2 \underset{D_1}{\overset{D_2}{\geq}} \ln \frac{P(S_1)}{P(S_2)},$$

d'où

$$z \underset{D_1}{\overset{D_2}{\geq}} \frac{1}{2} + \sigma^2 \ln \frac{P(S_1)}{P(S_2)}.$$

Par conséquence, l'observation est comparée à la valeur seuil

$$z_s = \frac{1}{2} + \sigma^2 \ln \frac{P(S_1)}{P(S_2)}.$$

L'espace des observations est ici une droite. Cet espace est donc divisé en deux sous-espaces disjoints dans lesquels soit la décision D_1 soit la décision D_2 est prise, selon la valeur de z . Le fait de prendre une décision est donc associé à la réalisation d'une partition de l'espace des observations, ceci étant en fait un résultat très général.

1.5 Probabilités de décision conditionnelles

Comme constaté sur la FIGURE 1.3, il est possible que la décision retenue soit erronée. En effet, il est possible qu'un 1 ait été émis et que z soit inférieur

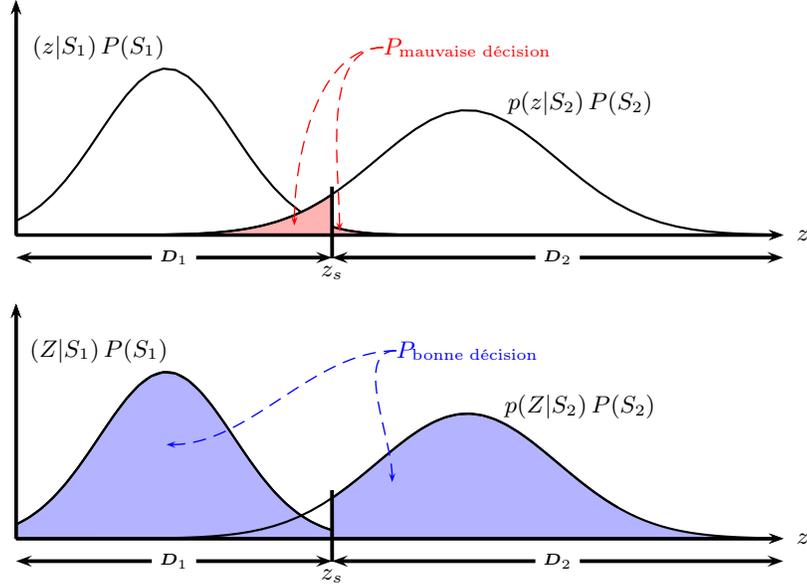


FIGURE 1.3 – Illustration probabilités des bonnes et des mauvaises décisions.

à z_s . De même, il est possible qu'un 0 ait été émis et que z soit supérieur à z_s . Dans la suite, on propose de quantifier les probabilités de bonne décision et d'erreur.

Une bonne décision est prise si on a simultanément :

- S_1 présent et $z \leq z_s$,
- ou S_2 présent et $z > z_s$.

Une mauvaise décision est prise si on a simultanément :

- S_1 présent et $z > z_s$,
- ou S_2 présent et $z \leq z_s$.

On peut donc définir quatre probabilités conditionnelles $P(\mathbf{D}_i|S_j)$, pour $i = 1, 2; j = 1, 2$, où $P(\mathbf{D}_i|S_j)$ représente la probabilité de prendre la décision \mathbf{D}_i lorsque le message émis est S_j . Les probabilités d'erreur conditionnelles sont donc $P(\mathbf{D}_1|S_2)$ et $P(\mathbf{D}_2|S_1)$, et les probabilités conditionnelles de bonne décision sont $P(\mathbf{D}_1|S_1)$ et $P(\mathbf{D}_2|S_2)$. La probabilité d'erreur totale (de mauvaise détection) est donc :

$$P_{\text{err}} = p(\mathbf{D}_1, S_2) + P(\mathbf{D}_2, S_1),$$

soit,

$$P_{\text{err}} = P(\mathbf{D}_1|S_2)P(S_2) + P(\mathbf{D}_2|S_1)P(S_1).$$

De même, la probabilité de bonne décision est donnée par

$$P_{\text{BD}} = P(\mathbf{D}_1, S_1) + P(\mathbf{D}_2, S_2),$$

soit,

$$P_{\text{BD}} = P(\mathbf{D}_1|S_1)P(S_1) + P(\mathbf{D}_2|S_2)P(S_2).$$

Ces probabilités peuvent aussi être exprimées à partir de la loi de vraisemblance de $\Lambda(Z)$, que l'on peut éventuellement déterminer à partir de celle de Z . Soit $p_\Lambda(\cdot|S_i)$ la loi de probabilité de la vraisemblance $\Lambda(Z)$ conditionnellement à S_i (puisque la vraisemblance est une variable aléatoire). On a alors

$$P(\mathbf{D}_1|S_i) \text{ quand } \Lambda(Z) \leq \lambda \mid S_i \quad \text{et} \quad P(\mathbf{D}_2|S_i) \text{ quand } \Lambda(Z) > \lambda \mid S_i,$$

soit

$$P(\mathbf{D}_1|S_i) = \int_{-\infty}^{\lambda} p_\Lambda(\lambda'|S_i) d\lambda' \quad \text{et} \quad P(\mathbf{D}_2|S_i) = \int_{\lambda}^{\infty} p_\Lambda(\lambda'|S_i) d\lambda'.$$

Il est facile de vérifier que $P(\mathbf{D}_1|S_j) + P(\mathbf{D}_2|S_j) = 1$ pour tout $j = 1, 2$.

On peut encore exprimer les probabilités conditionnelles sous une autre forme, en remarquant que l'on a réalisé une partition de l'espace d'observation en deux sous espaces \mathcal{Z}_1 et \mathcal{Z}_2 . \mathcal{Z}_1 est défini par l'ensemble des valeurs z telles que $\Lambda(Z) \leq \lambda$, et \mathcal{Z}_2 par l'ensemble des valeurs z telles que $\Lambda(Z) > \lambda$. Dans ce cas, on a

$$P(\mathbf{D}_i|S_j) = \int_{\mathcal{Z}_i} p(z|S_j) dz. \quad (1.3)$$

On a pour habitude d'appeler $P(\mathbf{D}_2|S_1)$ probabilité de fausse alarme ou *erreur de première espèce*, et $P(\mathbf{D}_1|S_2)$ probabilité de manque ou *erreur de deuxième espèce*.

En revenant à l'exemple précédent en communication binaire synchrone, les différentes probabilités conditionnelles peuvent être évaluée comme

suit :

$$\begin{aligned}
 P(\mathbf{D}_1|S_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{z_s} \exp\left(\frac{-1}{2\sigma^2}z^2\right) dz \\
 P(\mathbf{D}_1|S_2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{z_s} \exp\left(\frac{-1}{2\sigma^2}(z-1)^2\right) dz \\
 P(\mathbf{D}_2|S_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{z_s}^{\infty} \exp\left(\frac{-1}{2\sigma^2}z^2\right) dz \\
 P(\mathbf{D}_2|S_2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{z_s}^{\infty} \exp\left(\frac{-1}{2\sigma^2}(z-1)^2\right) dz
 \end{aligned}$$

En posant

$$\operatorname{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(\frac{-1}{2}u^2\right) du, \quad (1.4)$$

on obtient

$$\begin{aligned}
 P(\mathbf{D}_1|S_1) &= \operatorname{erf}\left(\frac{z_s}{\sigma}\right) \\
 P(\mathbf{D}_1|S_2) &= \operatorname{erf}\left(\frac{z_s+1}{\sigma}\right) \\
 P(\mathbf{D}_2|S_1) &= 1 - \operatorname{erf}\left(\frac{z_s}{\sigma}\right) \\
 P(\mathbf{D}_2|S_2) &= 1 - \operatorname{erf}\left(\frac{z_s+1}{\sigma}\right).
 \end{aligned}$$

Il est facile de vérifier les identités $P(\mathbf{D}_1|S_1) + P(\mathbf{D}_2|S_1) = 1$ et $P(\mathbf{D}_1|S_2) + P(\mathbf{D}_2|S_2) = 1$.

Chapitre 2

Les principaux critères de décision

On a étudié une règle de décision dans le chapitre précédent. Cette règle nécessite la connaissance des lois de probabilités conditionnelles et des probabilités a priori. Par ailleurs, elle ne prend pas en compte la gravité des erreurs. Si l'on considère un problème de détection radar, l'erreur de fausse alarme est moins grave que l'erreur de non détection. Afin de prendre en compte cette remarque, un coût est associé à chacune des décisions ; on introduit ainsi le critère de Bayes. Si les probabilités a priori $p(S_1)$ et $p(S_2)$ sont inconnues, le critère de Bayes devient un critère dit "minimax". Enfin, le critère de Neyman-Pearson concerne le cas où les coûts sont supposés inconnus.

2.1 Critère de Bayes

2.1.1 Coût moyen de décision

Associons à chaque des quatre cas possibles,

$$(D_1, S_1) \quad (D_1, S_2) \quad (D_2, S_1) \quad (D_2, S_2),$$

un coût respectif

$$C_{11} \quad C_{12} \quad C_{21} \quad C_{22}.$$

On suppose bien évidemment qu'un coût de bonne décision est inférieur à celui d'une mauvaise décision, c'est à dire :

$$C_{11} < C_{12}, \quad C_{11} < C_{21} \quad \text{et} \quad C_{22} < C_{12}, \quad C_{22} < C_{21}. \quad (2.1)$$

2.1. Critère de Bayes

Le coût moyen peut alors être calculé selon $C = \sum_{i,j} C_{ij} P(\mathbf{D}_i, S_j)$, ou encore

$$C = \sum_{i,j} C_{ij} P(\mathbf{D}_i | S_j) P(S_j).$$

Ce coût est fonction des C_{ij} que l'on fixe a priori, ainsi que des probabilités conditionnelles $P(\mathbf{D}_i | S_j)$ qui dépendent de la partition de l'espace d'observation que l'on va effectuer.

2.1.2 Critère de Bayes

Le critère de Bayes consiste à déterminer la partition de l'espace des observations qui minimise le coût moyen C . En reprenant de nouveau (1.3), c'est à dire

$$P(\mathbf{D}_i | S_j) = \int_{\mathcal{Z}_i} p(z | S_j) dz,$$

où \mathcal{Z}_i représente le domaine de \mathcal{Z} sur lequel la décision \mathbf{D}_i est prise. En remarquant que les régions \mathcal{Z}_i sont disjointes et couvrent tout l'espace des observations, on a alors pour tout $j = 1, 2$:

$$\sum_i P(\mathbf{D}_i | S_j) = 1.$$

Le coût moyen C s'exprime alors en fonction d'intégrales dépendant uniquement d'une région de l'espace des observations. On obtient alors :

$$\begin{aligned} C &= C_{21}P(S_1) + C_{22}P(S_2) \\ &+ \int_{\mathcal{Z}_1} P(S_2) (C_{12} - C_{22}) p(z | S_2) - P(S_1) (C_{21} - C_{11}) p(z | S_1) dz \end{aligned}$$

La première partie de C , avec les deux termes $C_{21}P(S_1) + C_{22}P(S_2)$, est fixée : c'est le coût fixé. La deuxième partie dépend de \mathcal{Z}_1 , c'est à dire de la partition que l'on va effectuer : c'est le coût contrôlé.

Le problème est de déterminer \mathcal{Z}_1 afin de rendre minimum le coût moyen C . Pour ce faire, on reprend l'hypothèse (2.1) qu'une décision incorrecte coûte plus cher qu'une décision correcte. Dans ce cas, le terme de l'intégrale est la différence de deux termes positifs. Pour rendre le coût minimum, il suffit de prendre la décision \mathbf{D}_1 pour les observations pour lesquelles le premier terme est plus petit que le second et réciproquement, soit

- \mathbf{D}_1 si $P(S_2) (C_{12} - C_{22}) p(z | S_2) < P(S_1) (C_{21} - C_{11}) p(z | S_1)$;
- \mathbf{D}_2 si $P(S_2) (C_{12} - C_{22}) p(z | S_2) > P(S_1) (C_{21} - C_{11}) p(z | S_1)$.

La règle de décision est alors la suivante :

$$\frac{p(z|S_2)}{p(z|S_1)} \underset{D_1}{\overset{D_2}{\gtrless}} \frac{C_{21} - C_{11}}{C_{12} - C_{22}} \frac{P(S_1)}{P(S_2)},$$

soit

$$\Lambda(z) \underset{D_1}{\overset{D_2}{\gtrless}} \lambda,$$

avec une valeur seuil $\lambda = \frac{C_{21} - C_{11}}{C_{12} - C_{22}} \frac{P(S_1)}{P(S_2)}$. Il s'agit d'une règle semblable à celle du maximum de probabilité a posteriori, mais avec une valeur seuil différente. En prenant $C_{21} - C_{11} = C_{12} - C_{22}$, on retrouve cette règle.

2.1.3 Performances du détecteur

Les performances peuvent être étudiées en évaluant les différentes probabilités d'erreur (mauvaise décision) et de bonne décision.

2.1.3.a Probabilités d'erreur

Les probabilités d'erreur conditionnelles $P(D_2|S_1)$ et $P(D_1|S_2)$ et la probabilité d'erreur totale $P_{\text{err}} = P(D_2|S_1)P(S_1) + P(D_1|S_2)P(S_2)$ ont été définies précédemment, et peuvent être évaluées en utilisant le fait que :

$$P(D_2|S_1) = \int_{z_2} p(z|S_1) dz \quad \text{et} \quad P(D_1|S_2) = \int_{z_1} p(z|S_2) dz.$$

Toutefois, ces probabilités d'erreur peuvent être calculées à partir du rapport de vraisemblance, si l'on sait évaluer $p_\Lambda(\lambda'|S_1)$ et $p_\Lambda(\lambda'|S_2)$. On a alors :

$$P(D_2|S_1) = \int_{\lambda}^{\infty} p_\Lambda(\lambda'|S_1) d\lambda' \quad \text{et} \quad P(D_1|S_2) = \int_{-\infty}^{\lambda} p_\Lambda(\lambda'|S_2) d\lambda'.$$

2.1.3.b Probabilités de bonne décision

On a de la même façon :

$$P(D_1|S_1) = \int_{z_1} p(z|S_1) dz \quad \text{et} \quad P(D_2|S_2) = \int_{z_2} p(z|S_2) dz,$$

que l'on peut écrire :

$$P(D_1|S_1) = \int_{-\infty}^{\lambda} p_\Lambda(\lambda'|S_1) d\lambda' \quad \text{et} \quad P(D_2|S_2) = \int_{\lambda}^{\infty} p_\Lambda(\lambda'|S_2) d\lambda'.$$

2.1.4 Remarque : coût (0,1)

Le coût moyen peut être calculé à partir des coûts et des probabilités d'erreur :

$$C = C_{21}P(S_1) + C_{22}P(S_2) + P(S_2)(C_{12} - C_{22})P(\mathbf{D}_1|S_2) - P(S_1)(C_{21} - C_{11})(1 - P(\mathbf{D}_2|S_1)).$$

Si l'on choisit des coûts (0,1), c'est à dire $C_{ii} = 0$ et $C_{ij} = 1$ pour $i \neq j$, la règle de décision devient :

$$\Lambda(z) \underset{\mathbf{D}_1}{\overset{\mathbf{D}_2}{\gtrless}} \frac{P(S_1)}{P(S_2)},$$

et l'on reconnaît la règle de maximum de probabilité a posteriori. Dans ce cas, le coût C devient :

$$\begin{aligned} C &= P(S_1) + P(S_2)P(\mathbf{D}_1|S_2) - P(S_1) + P(S_1)P(\mathbf{D}_2|S_1) \\ &= P(S_2)P(\mathbf{D}_1|S_2) + P(S_1)P(\mathbf{D}_2|S_1) \\ &= P_{\text{err}}, \end{aligned}$$

donc la règle de probabilité maximum a posteriori minimise la probabilité d'erreur totale.

2.1.5 Exemple

On souhaite détecter à un instant t_0 un signal d'amplitude m constante positive dans un bruit $b(t)$ gaussien. Deux possibilités émergent :

- $s(t) = 0$ pour S_1 ;
- $s(t) = m$ pour S_2 .

L'observation est $z = z(t_0) = s(t_0) + b(t_0)$. On suppose que $E(b(t)) = 0$ et $\text{Var}(b(t)) = \sigma^2$, ainsi que $P(S_1) = P(S_2) = \frac{1}{2}$. Le critère choisi est la probabilité d'erreur totale minimum. La solution est donnée par

$$\Lambda(z) \underset{\mathbf{D}_1}{\overset{\mathbf{D}_2}{\gtrless}} \frac{P(S_1)}{P(S_2)},$$

soit

$$\Lambda(z) \underset{\mathbf{D}_1}{\overset{\mathbf{D}_2}{\gtrless}} 1, \tag{2.2}$$

avec

$$\begin{aligned}\Lambda(z) &= \frac{\exp\left(\frac{-1}{2\sigma^2}(z-m)^2\right)}{\exp\left(\frac{-1}{2\sigma^2}(z)^2\right)} \\ &= \exp\left(\frac{-1}{2\sigma^2}(m^2 - 2zm)\right).\end{aligned}$$

En remplaçant cette expression dans (2.2), la règle de décision est

$$2zm - m^2 \underset{D_1}{\overset{D_2}{\geq}} 0,$$

ou encore

$$z \underset{D_1}{\overset{D_2}{\geq}} \frac{m}{2}.$$

Les différentes probabilités conditionnelles peuvent être calculées à partir des lois de $\Lambda(z)$ conditionnellement à S_1 et S_2 , c'est à dire $p(z|S_1)$ et $p(z|S_2)$. On obtient

$$\begin{aligned}P(D_2|S_1) &= 1 - \operatorname{erf}\left(\frac{m}{2\sigma}\right) \\ P(D_1|S_2) &= \operatorname{erf}\left(\frac{-m}{2\sigma}\right) \\ P(D_1|S_1) &= \operatorname{erf}\left(\frac{m}{2\sigma}\right) \\ P(D_2|S_2) &= 1 - \operatorname{erf}\left(\frac{-m}{2\sigma}\right).\end{aligned}$$

où l'expression de la fonction erf est donnée dans (1.4). On peut vérifier facilement que $P(D_2|S_1) = P(D_1|S_2)$ et $P(D_1|S_1) = P(D_2|S_2)$.

2.2 Critère minimax

2.2.1 Définition du critère

Le critère de Bayes suppose la connaissance d'une part les coûts associés aux décisions et d'autre part les probabilités a priori $P(S_1)$ et $P(S_2)$. Supposons que les probabilités a priori ne sont pas connues. Dans ce cas, on procède de la façon suivante :

1. Pour chaque valeur de $P(S_2)$, le coût minimum C est calculé.

2. Parmi toutes les valeurs possibles de $C(P(S_2))$, la valeur de $P(S_2)$ qui maximise $C(P(S_2))$ est retenue.

Ceci correspond à retenir le cas le plus défavorable du coût minimum, d'où le terme *minimax*.

2.2.2 Etude du graphe de $C(P(S_2))$

Le graphe de $C(P(S_2))$ n'est pas quelconque, mais présente toujours une concavité tournée vers le bas. En effet, le coût peut s'écrire :

$$\begin{aligned} C(P(S_2)) = & C_{21}P(\mathbf{D}_2|S_1) + C_{11}(1 - P(\mathbf{D}_2|S_1)) \\ & + P(S_2)\left((C_{22} - C_{11}) + (C_{12} - C_{22})P(\mathbf{D}_1|S_2) \right. \\ & \left. - (C_{21} - C_{11})P(\mathbf{D}_2|S_1)\right). \end{aligned}$$

Il est clair que $C(P(S_2))$ dépend explicitement de façon linéaire de $P(S_2)$, mais il en dépend aussi implicitement par $P(\mathbf{D}_2|S_1)$ et $P(\mathbf{D}_1|S_2)$. En effet, ces probabilités dépendent du seuil λ qui, lui même, dépend de $P(S_2)$.

En fixant $p_2^* = P(S_2)$, on peut alors calculer le seuil λ , ainsi les probabilités $P(\mathbf{D}_2|S_1)$ et $P(\mathbf{D}_1|S_2)$ qui sont alors fixées. La variation **explicite** de $C(P(S_2))$ en fonction de $P(S_2)$ est donc linéaire, et a pour équation :

$$C(P(S_2)) = C_0 + \alpha P(S_2).$$

Montrons maintenant que cette droite est tangente à la courbe $C(P(S_2))$ au point $P(S_2) = p_2^*$. En effet, cette droite passe par le point de coordonnées p_2^* et $C(p_2^*)$. Par ailleurs, elle passe nécessairement au dessus de la courbe $C(P(S_2))$ puisque cette courbe représente le "risque minimum" en fonction de $P(S_2)$. La droite est donc tangente à $C(P(S_2))$, et la courbe a sa concavité tournée vers le bas.

2.2.3 Critère Minimax

L'approche consiste à retenir la valeur de $P(S_2)$ qui maximise la courbe $C(P(S_2))$. Soit celle-ci se trouve en $P(S_2) = 0$ ou $P(S_2) = 1$, soit il existe une valeur $p_{2\max}$ qui la maximise. Dans ce cas, la pente de la tangente doit être nulle, ce qui veut dire que

$$(C_{22} - C_{11}) + (C_{12} - C_{22})P(\mathbf{D}_1|S_2) - (C_{21} - C_{11})P(\mathbf{D}_2|S_1) = 0.$$

C'est l'équation caractéristique du critère minimax. Les coûts étant fixés, cette relation permet d'exprimer par exemple $P(D_2|S_1)$ en fonction de $P(D_1|S_2)$,

$$P(\mathbf{D}_2|S_1) = \frac{(C_{22} - C_{11}) + (C_{12} - C_{22})P(\mathbf{D}_1|S_2)}{(C_{21} - C_{11})},$$

et ainsi déterminer le seuil λ par la résolution de

$$\int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_1) d\lambda' = \frac{(C_{22} - C_{11}) + (C_{12} - C_{22}) \int_{-\infty}^{\lambda} p_{\Lambda}(\lambda'|S_2) d\lambda'}{(C_{21} - C_{11})}$$

2.2.3.a Cas particulier : coûts (0,1)

Si l'on considère des coûts (0,1), l'équation caractéristique devient $P(\mathbf{D}_1|S_2) = P(\mathbf{D}_2|S_1)$. On a alors

$$C = P(\mathbf{D}_1|S_2) = P_e,$$

car

$$\begin{aligned} P_e &= P(\mathbf{D}_1|S_2)P(S_2) + P(\mathbf{D}_2|S_1)P(S_1) \\ &= P(\mathbf{D}_1|S_2)(P(S_2) + P(S_1)) \\ &= P(\mathbf{D}_1|S_2). \end{aligned}$$

2.3 Critère de Neyman-Pearson

2.3.1 Position du problème

Dans la suite, ni les coûts ni les probabilités a priori sont connus. La règle du coût moyen minimum ne peut plus être appliquée, ni la règle minimax. Pour contourner cette difficulté, on propose de raisonner directement sur les probabilités conditionnelles.

Comme précédemment, les probabilités d'erreur sont la probabilité de fausse alarme $P(\mathbf{D}_2|S_1)$ et la probabilité de manque $P(\mathbf{D}_1|S_2)$. En général, on considère la probabilité de fausse alarme $P_F = P(\mathbf{D}_2|S_1)$ et la probabilité de détection $P_D = P(\mathbf{D}_2|S_2) = 1 - P(\mathbf{D}_1|S_2)$. Pour un problème donné, par exemple radar, on souhaite que P_D soit aussi grand que possible et que P_F soit aussi petit que possible. Compte tenu du fait que ces probabilités sont liées, il est judicieux de choisir une valeur de P_F et maximiser P_D . C'est le principe du critère de Neyman-Pearson.

2.3.2 Critère de Neyman-Pearson

Objectif

Se fixer $P(\mathbf{D}_2|S_1) = \alpha$, puis maximiser $P(\mathbf{D}_2|S_2)$.

Procédure

L'objectif est de rendre maximum $P(\mathbf{D}_2|S_2)$, avec la contrainte $\alpha = P(\mathbf{D}_2|S_1)$. En rappelant que le premier est $\int_{\mathcal{Z}_2} p(z|S_2) dz$ et le second $\int_{\mathcal{Z}_2} p(z|S_1) dz$, on a le problème d'optimisation suivant :

$$\begin{aligned} & \text{maximiser } \int_{\mathcal{Z}_2} p(z|S_2) dz \\ & \text{sous contrainte } \alpha = \int_{\mathcal{Z}_2} p(z|S_1) dz. \end{aligned}$$

Il s'agit d'un problème d'optimisation sous contrainte, qu'on peut résoudre en utilisant la technique des multiplicateurs de Lagrange. Pour cela, on souhaite maximiser le Lagrangien associé au problème d'optimisation ci-dessus :

$$J = \int_{\mathcal{Z}_2} p(z|S_2) dz - \lambda \left(\int_{\mathcal{Z}_2} p(z|S_1) dz - \alpha \right).$$

où λ est un paramètre, le multiplicateur de Lagrange, qu'il convient de déterminer. La fonction objective J devant être maximisée, avec

$$J = \lambda \alpha + \int_{\mathcal{Z}_2} \left(p(z|S_2) - \lambda p(z|S_1) \right) dz,$$

le domaine \mathcal{Z}_2 est affecté par les éléments de l'espace d'observation pour lesquels la quantité sous l'intégrale est positive, c'est à dire :

- $z \in \mathcal{Z}_2$ si $p(z|S_2) - \lambda p(z|S_1) > 0$;
- $z \in \mathcal{Z}_1$ si $p(z|S_2) - \lambda p(z|S_1) < 0$.

on a ainsi

$$\frac{p(z|S_2)}{p(z|S_1)} \underset{\mathbf{D}_1}{\overset{\mathbf{D}_2}{\geq}} \lambda.$$

Il s'agit, là encore, d'une règle de décision pour laquelle le rapport de vraisemblance $\Lambda(Z)$ est comparé à une valeur seuil λ . Cette valeur λ sera déterminée par la contrainte, c'est à dire

$$P(\mathbf{D}_2|S_1) = \alpha = \int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_1) d\lambda'.$$

Notons que si $P(\mathbf{D}_2|S_2)$ est maximisée, $P(\mathbf{D}_2|S_1)$ est minimisée. En résumé, le critère de Neyman-Pearson consiste à :

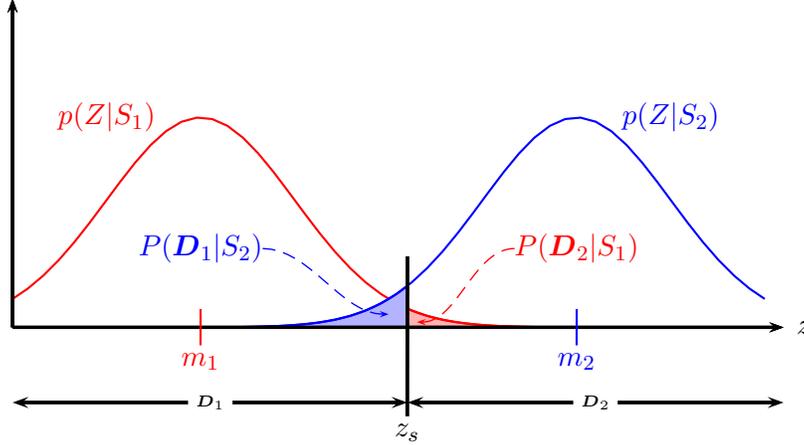


FIGURE 2.1 – Représentation des probabilités d'erreur

1. fixer $P(\mathbf{D}_2|S_1) = \alpha$;
2. en déduire λ ;
3. déterminer la partition de l'espace d'observation ;
4. calculer $P(\mathbf{D}_1|S_2)$ et $P(\mathbf{D}_2|S_2)$.

Une illustration des probabilités d'erreur est donnée dans la FIGURE 2.1.

Reprenons l'exemple de la communication binaire synchrone, avec $\sigma = 1$. Soit $\alpha = P(\mathbf{D}_2|S_2)$.

Le rapport de vraisemblance s'écrit $\Lambda(Z) = \exp\left(Z - \frac{1}{2}\right)$. La règle de décision est alors

$$\exp\left(z - \frac{1}{2}\right) \underset{\mathbf{D}_1}{\overset{\mathbf{D}_2}{\geq}} \lambda.$$

La partition recherchée est donc

$$z \underset{\mathbf{D}_1}{\overset{\mathbf{D}_2}{\geq}} \ln(\lambda) + \frac{1}{2}.$$

Soit $z_s = \ln(\lambda) + \frac{1}{2}$. On a donc

$$P(\mathbf{D}_2|S_1) = \frac{1}{\sqrt{2\pi}} \int_{z_s}^{\infty} \exp\left(-\frac{1}{2}z^2\right) dz = \alpha,$$

d'où $z_s = \text{erf}^{-1}(\alpha)$. On remarque par ailleurs que $\lambda = \exp\left(z_s - \frac{1}{2}\right)$, et qu'il n'est pas nécessaire de le déterminer.

2.3. Critère de Neyman-Pearson

Le seuil z_s étant déterminé, on peut alors calculer

$$P(\mathbf{D}_2|S_2) = \frac{1}{\sqrt{2\pi}} \int_{z_s}^{\infty} \exp\left(-\frac{1}{2}(z-1)^2\right) dz = 1 - \text{erf}(z_s - 1).$$

En reprenant la FIGURE 2.1, on a $m_1 = 0$ et $m_2 = 1$. Exemples de values possibles :

$P(\mathbf{D}_2 S_1)$	z_s	λ	$P(\mathbf{D}_2 S_2)$
0.25	0.675	1.19	0.63
0.2	0.842	1.41	0.60
0.1	1.280	2.18	0.39
0.05	1.645	3.14	0.26
0.01	2.326	6.20	0.09

Chapitre 3

Caractéristiques opérationnelles d'un récepteur

Quelque soit le critère utilisé, il s'agit toujours du rapport de vraisemblance $\Lambda(Z)$ à comparer à un seuil dépendant de la règle de décision choisie. La règle de décision s'écrit toujours

$$\Lambda(z) \underset{\mathbf{D}_1}{\overset{\mathbf{D}_2}{\geq}} \lambda.$$

Pour chaque valeur λ , on peut donc calculer les diverses probabilités conditionnelles. On obtient alors la courbe *Caractéristiques Opérationnelles d'un Récepteur* (COR, ou ROC en anglais pour "Receiver Operating Characteristic"), en représentant $P(\mathbf{D}_2|S_2)$ en fonction $P(\mathbf{D}_2|S_1)$, c'est à dire en éliminant λ entre $P(\mathbf{D}_2|S_2)$ et $P(\mathbf{D}_2|S_1)$.

3.1 Etude des courbes COR

Ayant

$$P(\mathbf{D}_2|S_2) = \int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_2) d\lambda' \quad \text{et} \quad P(\mathbf{D}_2|S_1) = \int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_1) d\lambda',$$

on peut effectuer une étude paramétrique des courbes COR des points $P(\mathbf{D}_2|S_2), P(\mathbf{D}_2|S_1)$.

On peut montrer les propriétés suivantes ; voir Section 3.3 pour les démonstrations :

- $P(\mathbf{D}_2|S_1)$ et $P(\mathbf{D}_2|S_2)$ sont des fonctions décroissantes de λ ;
- la courbe COR passe par les points $(0, 0)$ et $(1, 1)$;

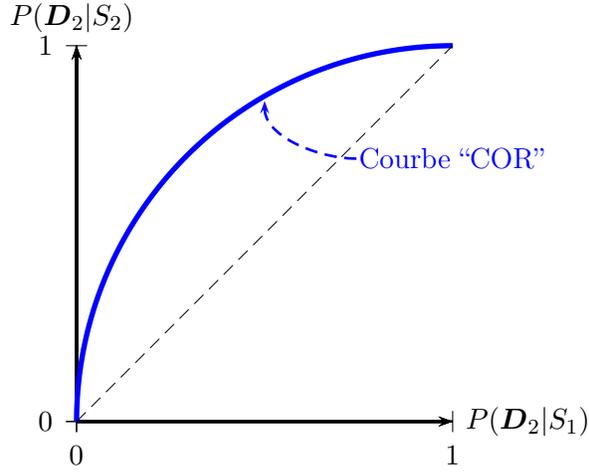


FIGURE 3.1 – Illustration de la courbe COR

- La pente de la courbe COR est λ , c'est à dire $\frac{dP(\mathbf{D}_2|S_2)}{dP(\mathbf{D}_2|S_1)} = \lambda$;
- la courbe est toujours au dessus de la première bissectrice (qui est le pire détecteur) ;
- la concavité est toujours négative.

En conséquence, le graphe possède l'allure illustrée dans la FIGURE 3.1.

3.2 Utilisation des courbes COR

Les courbes COR peuvent être différemment utilisées selon le critère choisi.

3.2.1 Critère de Neyman-Pearson

L'étude de la courbe COR, dans ce cas, permet de choisir le compromis entre probabilité de détection et de fausse alarme. Voir FIGURE 3.2.

3.2.2 Critère minimax

Comme illustré précédemment, le critère minimax est caractérisé, dans le cas où le maximum n'est réalisé ni en $P(S_2) = 0$ ni en $P(S_2) = 1$, par l'équation caractéristique

$$(C_{22} - C_{11}) + (C_{12} - C_{22})P(\mathbf{D}_1|S_2) - (C_{21} - C_{11})P(\mathbf{D}_2|S_1) = 0,$$

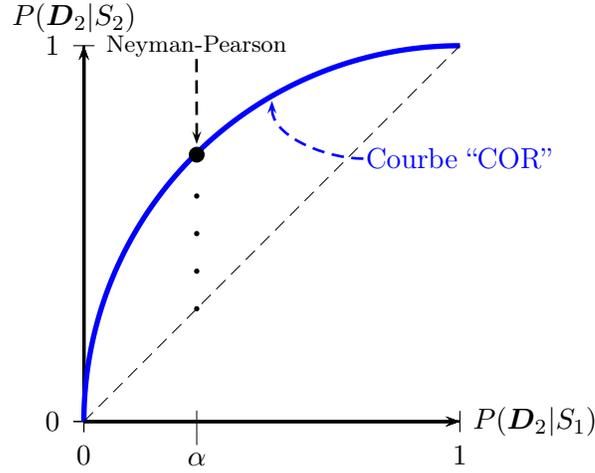


FIGURE 3.2 – Courbe COR et critère de Neyman-Pearson.

ou encore en remplaçant $P(\mathbf{D}_1|S_2)$ par $1 - P(\mathbf{D}_2|S_2)$:

$$(C_{22} - C_{11}) + (C_{22} - C_{12})P(\mathbf{D}_2|S_2) + (C_{11} - C_{21})P(\mathbf{D}_2|S_1) = 0.$$

Les valeurs de $P(\mathbf{D}_2|S_2)$ et $P(\mathbf{D}_2|S_1)$ peuvent donc être déterminées par l'intersection de cette droite avec la courbe COR.

Par ailleurs, dans le cas de coûts (0,1), l'équation caractéristique du critère minimax est donnée par $P(\mathbf{D}_2|S_1) = P(\mathbf{D}_1|S_2)$, soit $1 - P(\mathbf{D}_2|S_2) - P(\mathbf{D}_2|S_1) = 0$. Cette solution est illustrée dans la FIGURE 3.3.

3.2.3 Critère de Bayes

Dans le cas du critère de Bayes, les coûts et les probabilités a priori étant fixées, on se trouve en un point de la courbe COR. Cette courbe peut alors servir à modifier si nécessaire les coûts afin de se déplacer sur la courbe et obtenir des probabilités conditionnelles plus acceptables.

3.3 Démonstrations

On clôt ce chapitre avec les démonstrations des propriétés décrites à la Section 3.1.

Proposition 1. $P(\mathbf{D}_2|S_1)$ et $P(\mathbf{D}_2|S_2)$ sont des fonctions décroissantes de λ

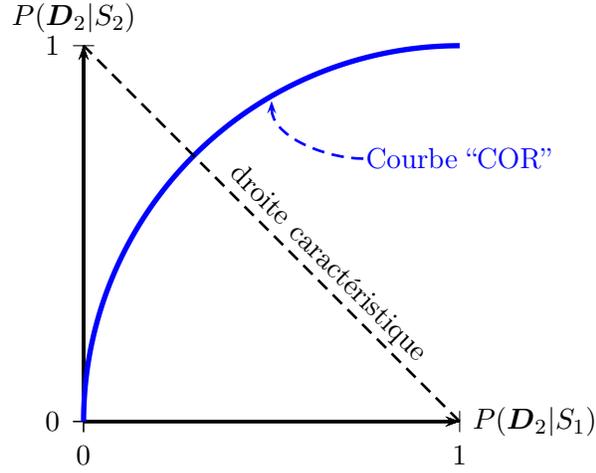


FIGURE 3.3 – Courbe COR et critère minimax dans le cas des coûts (0,1).

Démonstration. La démonstration est intuitive. Puisque $P(\mathbf{D}_2|S_j) = \int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_j) d\lambda'$, pour tout $j = 1, 2$, on a alors

$$\frac{dP(\mathbf{D}_2|S_j)}{d\lambda} = -p_{\Lambda}(\lambda|S_j) \leq 0.$$

■

Proposition 2. *La courbe COR passe par les points (0, 0) et (1, 1).*

Démonstration. Pour tout $j = 1, 2$, on a

$$\lim_{\lambda \rightarrow 0} P(\mathbf{D}_2|S_j) = \lim_{\lambda \rightarrow 0} \int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_j) d\lambda' = \int_0^{\infty} p_{\Lambda}(\lambda'|S_j) d\lambda' = 1.$$

De même,

$$\lim_{\lambda \rightarrow \infty} P(\mathbf{D}_2|S_j) = \lim_{\lambda \rightarrow \infty} \int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_j) d\lambda' = 0.$$

■

Proposition 3. *La concavité est toujours négative et la courbe est toujours au dessus de la première bissectrice, c'est à dire $P(\mathbf{D}_2|S_2) > P(\mathbf{D}_2|S_1)$.*

Démonstration. On montre que $p_{\Lambda}(\lambda|S_2) = \lambda p_{\Lambda}(\lambda|S_1)$, pour tout λ . Pour cela, on utilise le concept de fonction caractéristique, rappelé dans l'Annexe A. La fonction caractéristique de $\Lambda(Z|S_1)$ est désignée par $\phi_1(\omega)$, c'est

à dire

$$\begin{aligned}\phi_1(\omega) &= E(e^{j\omega\Lambda}|S_1) \\ &= \sum_{n=0}^{\infty} j^n E(\Lambda^n|S_1) \frac{\omega^n}{n!},\end{aligned}\quad (3.1)$$

avec

$$\begin{aligned}E(\Lambda^n|S_1) &= \int_R \left(\frac{p(z|S_2)}{p(z|S_1)} \right)^n p(z|S_1) dz \\ &= \int_R \left(\frac{p(z|S_2)}{p(z|S_1)} \right)^{n-1} p(z|S_2) dz \\ &= E(\Lambda^{n-1}|S_2).\end{aligned}\quad (3.2)$$

On a alors

$$\begin{aligned}\frac{d\phi_1(\omega)}{d\omega} &= \sum_{n=0}^{\infty} n j^n E(\Lambda^n|S_1) \frac{\omega^{n-1}}{n!} && \text{[en utilisant (3.1)]} \\ &= j \sum_{n=1}^{\infty} j^{n-1} E(\Lambda^{n-1}|S_2) \frac{\omega^{n-1}}{(n-1)!} && \text{[en utilisant (3.2)]} \\ &= j \sum_{m=1}^{\infty} j^m E(\Lambda^m|S_2) \frac{\omega^m}{m!} \\ &= j \phi_2(\omega) \\ &= j \int_R e^{j\omega\lambda'} p_{\Lambda}(\lambda'|S_2) d\lambda'\end{aligned}$$

Or, en utilisant la définition de $E(e^{j\omega\Lambda}|S_1)$, on a

$$\frac{d\phi_1(\omega)}{d\omega} = \int_R j\lambda' e^{j\omega\lambda'} p_{\Lambda}(\lambda'|S_1) d\lambda'.$$

En combinant ces deux résultats, la relation suivante émerge :

$$\int_R \lambda' e^{j\omega\lambda'} p_{\Lambda}(\lambda'|S_1) d\lambda' = \int_R e^{j\omega\lambda'} p_{\Lambda}(\lambda'|S_1) d\lambda'.$$

Ces deux intégrales sont les transformées de Fourier de $\lambda p_{\Lambda}(\lambda|S_1)$ et $p_{\Lambda}(\lambda|S_2)$, respectivement, ce qui permet d'écrire

$$\lambda p_{\Lambda}(\lambda|S_1) = p_{\Lambda}(\lambda|S_2)\quad (3.3)$$

3.3. Démonstrations

Maintenant, étudions séparément les deux cas $\lambda \geq 1$ et $\lambda < 1$. Pour le premier cas, on a

$$\begin{aligned}
 P(\mathbf{D}_2|S_2) &= \int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_2) d\lambda' \\
 &= \int_{\lambda}^{\infty} \lambda' p_{\Lambda}(\lambda'|S_1) d\lambda' \\
 &\geq \int_{\lambda}^{\infty} \lambda p_{\Lambda}(\lambda'|S_1) d\lambda' \\
 &= \lambda \int_{\lambda}^{\infty} p_{\Lambda}(\lambda'|S_1) d\lambda' \\
 &= \lambda P(\mathbf{D}_2|S_1) \\
 &\geq P(\mathbf{D}_2|S_2),
 \end{aligned}$$

où la première inégalité est due $\lambda' \geq \lambda$ dans l'intégrale. Pour le second cas, on a

$$\begin{aligned}
 P(\mathbf{D}_1|S_2) &= \int_0^{\lambda} p_{\Lambda}(\lambda'|S_2) d\lambda' \\
 &= \int_0^{\lambda} \lambda' p_{\Lambda}(\lambda'|S_1) d\lambda' \\
 &\leq \lambda \int_0^{\lambda} p_{\Lambda}(\lambda'|S_1) d\lambda' \\
 &= \lambda P(\mathbf{D}_1|S_1) \\
 &\geq P(\mathbf{D}_1|S_2),
 \end{aligned}$$

où l'inégalité $\lambda' \leq \lambda$ a été utilisée dans l'intégrale. On peut alors écrire $1 - P(\mathbf{D}_1|S_2) \geq 1 - P(\mathbf{D}_1|S_1)$, ou encore $P(\mathbf{D}_2|S_2) \geq P(\mathbf{D}_2|S_1)$. ■

Proposition 4. *La pente de la courbe COR est λ , c'est à dire*

$$\frac{dP(\mathbf{D}_2|S_2)}{dP(\mathbf{D}_2|S_1)} = \lambda.$$

Démonstration. Puisque $\frac{dP(\mathbf{D}_2|S_j)}{d\lambda} = -p_{\Lambda}(\lambda|S_j)$, pour tout $j = 1, 2$, on a alors

$$\frac{dP(\mathbf{D}_2|S_2)}{dP(\mathbf{D}_2|S_1)} = \frac{\frac{dP(\mathbf{D}_2|S_2)}{d\lambda}}{\frac{dP(\mathbf{D}_2|S_1)}{d\lambda}} = \frac{p_{\Lambda}(\lambda|S_2)}{p_{\Lambda}(\lambda|S_1)} = \lambda,$$

puisque $p_{\Lambda}(\lambda|S_2) = \lambda p_{\Lambda}(\lambda|S_1)$ selon l'expression (3.3) de la Proposition 3. ■

Proposition 5. *La concavité de la courbe COR est toujours négative, c'est à dire*

$$\frac{d^2 P(\mathbf{D}_2|S_2)}{dP(\mathbf{D}_2|S_1)^2} \leq 0.$$

Démonstration. La preuve est directe en utilisant le résultat de la Proposition 4, puisque

$$\frac{d^2 P(\mathbf{D}_2|S_2)}{dP(\mathbf{D}_2|S_1)^2} = \frac{d\lambda}{dP(\mathbf{D}_2|S_1)} = \frac{1}{\frac{dP(\mathbf{D}_2|S_1)}{d\lambda}} = \frac{-1}{p_\Lambda(\lambda|S_1)} \leq 0.$$

■

3.3. Démonstrations

Chapitre 4

Généralisation : décision avec mesures multiples et avec hypothèses multiples

Dans les chapitres précédents, on a considéré le cas des décisions binaires avec une seule mesure, ceci afin de présenter les différents critères de décision. Il va de soi que, d'une façon générale, les situations sont plus complexes. En particulier, la décision n'est généralement pas binaire, mais il faut choisir une parmi M hypothèses. De plus Souvent, une mesure n'est pas suffisante, et l'on doit recourir au traitement de plusieurs mesures pour prendre la décision. Ce sont les deux aspects abordés dans ce chapitre, en commençant par le problème des mesures multiples en décision binaire, puis les décisions avec hypothèses multiples.

4.1 Décision avec mesures multiples

On se place dans le cas où l'observateur dispose de N mesures pour prendre deux décisions possibles \mathbf{D}_1 et \mathbf{D}_2 . Ces N mesures peuvent être :

- N mesures consécutives d'un même paramètre ;
- 1 mesure instantanée de N paramètres ;
- k mesures successives de n_i paramètres, avec $\sum_{i=1}^k n_i = N$.

Soient z_1, z_2, \dots, z_N , les observations, qui sont des réalisations de Z_1, Z_2, \dots, Z_N . On les considère comme les composantes d'un vecteur \mathbf{z} qui est une réalisation d'un vecteur aléatoire \mathbf{Z} de dimension N . L'espace d'observation, qui était l'axe réel dans le cas d'une mesure, est désormais un espace de dimension N . Il représente l'ensemble des réalisations possibles du

vecteur \mathbf{Z} .

De même que dans le cas monodimensionnel, on peut définir la loi de probabilité $p(\mathbf{Z})$, et les lois conditionnelles $p(\mathbf{Z}|S_1)$ et $p(\mathbf{Z}|S_2)$. L'espace d'observations étant un espace N -dimensionnel, prendre une décision revient donc à effectuer une partition de cet espace en deux sous-ensembles \mathcal{Z}_1 et \mathcal{Z}_2 , régions où l'on acceptera respectivement \mathbf{D}_1 et \mathbf{D}_2 dès lors que le vecteur \mathbf{Z} appartiendra à l'une de ces régions. La surface de partition est la surface de décision. Cette surface dépend de la règle de décision choisie.

Rapport de vraisemblance

Par analogie avec le cas d'une observation unique où la décision est prise à partir de

$$\Lambda(z) \underset{D_1}{\overset{D_2}{\gtrless}} \lambda,$$

le cas de mesures multiples concerne une règle de décision de la forme

$$\Lambda(\mathbf{z}) \underset{D_1}{\overset{D_2}{\gtrless}} \lambda.$$

On peut définir un rapport de vraisemblance avec $\Lambda(\mathbf{Z})$.

Probabilité de décision conditionnelle

Les probabilités conditionnelles $P(\mathbf{D}_i|S_j)$ peuvent être définies de manière similaire à celle définie dans les chapitres précédents, avec

$$P(\mathbf{D}_i|S_j) = \int_{\mathcal{Z}_i} p(\mathbf{z}|S_j) d\mathbf{z}, \quad (4.1)$$

avec une intégrale d'ordre N dans l'espace d'observation.

4.1.1 Critères de décision

Reprenant les trois critères de décision précédents, il est aisé de les généraliser au cas de mesures multiples.

4.1.1.a Coût moyen de décision

Le coût moyen de décision est donné par

$$C = \sum_{i,j=1}^2 C_{ij} P(\mathbf{D}_i, S_j),$$

que l'on peut écrire

$$C = \sum_{i,j=1}^2 C_{ij} P(\mathbf{D}_i | S_j) P(S_j).$$

En reprenant (4.1), on obtient

$$C = \sum_{i,j=1}^2 C_{ij} P(S_j) \int_{\mathcal{Z}_i} p(\mathbf{z} | S_j) d\mathbf{z}.$$

4.1.1.b Critères de décision

Si l'on généralise les résultats précédents, on obtient sans difficulté et pour chacun des trois critères :

$$\Lambda(\mathbf{Z}) \underset{D_1}{\overset{D_2}{\geq}} \lambda.$$

où λ dépend de la règle de décision choisie.

4.1.2 Exemple : détection d'un signal constant dans du bruit

Une source peut émettre ou non un signal constant d'amplitude m . On a donc :

- $S_1 : x(t) = 0$ avec $P(S_1) = \frac{1}{2}$;
- $S_2 : x(t) = m$ avec $P(S_2) = \frac{1}{2}$.

L'observateur reçoit le signal $z(t) = x(t) + b(t)$ où $b(t)$ est un bruit blanc stationnaire gaussien de moyenne nulle et de variance σ^2 . En prélevant N échantillons du signal, il doit dire si le signal est présent ou non. Connaissant les probabilités a priori et se fixant des coûts (0,1), on applique la règle de Bayes. Pour une observation z_n , réalisation de la variable aléatoire Z_n , on a

$$\begin{aligned} p(Z_n | S_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2} (Z_n)^2\right) \\ p(Z_n | S_2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2} (Z_n - m)^2\right). \end{aligned}$$

Le bruit étant blanc, les N observations sont indépendantes, on a donc :

$$p(\mathbf{Z}|S_j) = \prod_{n=1}^N p(Z_n|S_1),$$

et donc le rapport de vraisemblance est

$$\begin{aligned} \Lambda(\mathbf{Z}) &= \frac{\prod_{n=1}^N p(Z_n|S_2)}{\prod_{n=1}^N p(Z_n|S_1)} \\ &= \exp\left(\frac{1}{\sigma^2}(m^\top Z - \frac{1}{2}m^\top m)\right) \end{aligned}$$

Compte tenu du critère choisi, la règle de décision est :

$$\Lambda(\mathbf{Z}) \underset{D_1}{\overset{D_2}{\geq}} 1.$$

soit, après réduction :

$$\frac{1}{N} \sum_{n=1}^N Z_n \underset{D_1}{\overset{D_2}{\geq}} \frac{m}{2}.$$

Probabilités de décision conditionnelles

Les probabilités conditionnelles peuvent être calculées, soit à partir de l'intégrale dans l'espace des observations, ici une intégrale au dessous (au dessus) d'un hyperplan, soit à partir du rapport de vraisemblance, ce qui est, en général, la solution la plus simple.

En effet, la règle de décision porte ici sur une combinaison linéaire de variables gaussiennes (la moyenne arithmétique). On sait donc que $\bar{Z} = \frac{1}{N} \sum_{n=1}^N Z_n$ est une variable aléatoire gaussienne dont on peut déterminer aisément les paramètres, conditionnellement à S_1 et S_2 , avec

$$\begin{aligned} - E(\bar{Z}|S_1) &= 0 & Var(\bar{Z}|S_1) &= \frac{\sigma^2}{N}; \\ - E(\bar{Z}|S_2) &= m & Var(\bar{Z}|S_2) &= \frac{\sigma^2}{N}. \end{aligned}$$

Les diverses probabilités peuvent alors être déterminées :

$$P(\mathbf{D}_2|S_1) = \int_1^\infty p_\Lambda(\lambda'|S_1) d\lambda', \quad P(\mathbf{D}_2|S_2) = \int_{-\infty}^1 p_\Lambda(\lambda'|S_2) d\lambda',$$

d'où

$$\begin{aligned} P(\mathbf{D}_1|S_1) &= \operatorname{erf}\left(\frac{m\sqrt{N}}{2\sigma}\right) & P(\mathbf{D}_2|S_1) &= 1 - \operatorname{erf}\left(\frac{m\sqrt{N}}{2\sigma}\right) \\ P(\mathbf{D}_1|S_2) &= \operatorname{erf}\left(-\frac{m\sqrt{N}}{2\sigma}\right) & P(\mathbf{D}_2|S_2) &= 1 - \operatorname{erf}\left(-\frac{m\sqrt{N}}{2\sigma}\right) \end{aligned}$$

4.1.3 Exemple : détection de signaux de puissances différentes

La source peut émettre deux signaux blancs, gaussiens, et tels que

- $E(s(t)|S_j) = 0$;
- $Var(s(t)|S_j) = \sigma_j^2$,

avec $P(S_1) = P(S_2) = \frac{1}{2}$. L'observateur reçoit le signal $z(t) = s(t)$, dont il prélève N échantillons. On prend le critère de Bayes avec coûts (0,1) pour la décision.

Comme précédemment, on exprime le rapport de vraisemblance Λ selon

$$\Lambda(\mathbf{Z}) = \left(\frac{\sigma_1}{\sigma_2}\right)^N \exp\left(\frac{\sigma_2^2 - \sigma_1^2}{2\sigma_1^2\sigma_2^2} \mathbf{Z}^\top \mathbf{Z}\right).$$

La règle de décision est

$$\Lambda(\mathbf{Z}) \underset{D_1}{\overset{D_2}{\geq}} 1,$$

soit

$$\mathbf{Z}^\top \mathbf{Z} \underset{D_1}{\overset{D_2}{\geq}} N \frac{2\sigma_1^2\sigma_2^2}{\sigma_2^2 - \sigma_1^2} \ln\left(\frac{\sigma_2}{\sigma_1}\right),$$

qui est de la forme

$$\sum_{n=1}^N Z_n^2 \underset{D_1}{\overset{D_2}{\geq}} \gamma.$$

Les différentes probabilités peuvent donc être exprimées à partir de la variable $S = \sum_{n=1}^N Z_n^2$. Cette variable S représente une statistique suffisante pour prendre la décision. Pour préciser les différentes probabilités, soit on intègre dans l'espace d'observation (ici à l'intérieur ou à l'extérieur d'une hypersphère de centre l'origine et de rayon $\sqrt{\gamma}$), soit on intègre la loi de probabilité de la variable S . Pour ce faire, il est nécessaire de la déterminer conditionnellement aux deux hypothèses :

- sous S_1 : $\frac{Z_n}{\sigma_1}$ suit une loi normale (0,1) ; de même,
- sous S_2 : $\frac{Z_n}{\sigma_2}$ suit une loi normale (0,1).

Donc, puisque les N variables Z_n étant indépendantes,

- sous S_1 : $\frac{S}{\sigma_1^2}$ suit une loi de khi-2 à N degrés de liberté ;
- sous S_2 : $\frac{S}{\sigma_2^2}$ suit également une loi χ_N^2 .

On peut donc déduire les différentes probabilités :

$$P(\mathbf{D}_2|S_1) = p\left(\frac{S}{\sigma_1^2} > \frac{\gamma}{\sigma_1^2} \mid S_1\right) \quad P(\mathbf{D}_1|S_2) = p\left(\frac{S}{\sigma_2^2} < \frac{\gamma}{\sigma_2^2} \mid S_2\right),$$

que l'on peut déterminer, ainsi que leurs compléments, à partir des tables de la loi de χ^2 .

Remarque : statistique suffisante

Dans les deux exemples précédents, la décision est prise sur une fonction des variables aléatoires Z_1, Z_2, \dots, Z_N . Dans les deux cas, ces fonctions représentent ce que l'on appelle une statistique suffisante pour la prise de décision. Cette notion est à rattacher à la notion d'estimateur exhaustif en statistique.

4.2 Décision avec hypothèses multiples

Dans cette section, la source est maintenant susceptible d'engendrer M messages S_j . Le problème devient donc un problème de décision avec hypothèses multiples. Afin de conserver toute la généralité du problème, on suppose que la décision repose sur N observations, ce qui permet de généraliser les notions introduites précédemment.

4.2.1 Espace d'observation, espace de décision et coût moyen

L'espace d'observation est un espace de dimension N . Le critère de décision choisi permet d'effectuer une partition de cet espace en M sous-espaces \mathcal{Z}_i . L'espace de décision est un espace de dimension M , avec

$$\mathbf{D} = \mathbf{D}_i \text{ si } \mathbf{Z} \in \mathcal{Z}_i.$$

La notion de coût moyen peut être généralisée sans difficulté. Dans le cas d'hypothèses multiples,

$$\begin{aligned}
 C &= \sum_{i,j=1}^M C_{ij} P(\mathbf{D}_i, S_j), \\
 &= \sum_{i,j=1}^M C_{ij} P(\mathbf{D}_i | S_j) P(S_j). \\
 &= \sum_{i,j=1}^M C_{ij} P(S_j) \int_{\mathcal{Z}_i} p(\mathbf{z} | S_j) d\mathbf{z}.
 \end{aligned}$$

où (4.1) a été utilisée.

4.2.2 Critère de Bayes

4.2.2.a Cas général

Les coûts et les probabilités étant connus, on recherche une partition qui minimise le coût moyen C . Celui-ci peut se mettre sous la forme

$$C = \sum_{j=1}^M C_{jj} P(S_j) + \sum_{i=1}^M \int_{\mathcal{Z}_i} \sum_{j=1}^M (C_{ij} - C_{jj}) P(S_j) p(\mathbf{z} | S_j) d\mathbf{z},$$

où la condition $i \neq j$ a été enlevée de la dernière somme, puisque le terme correspondant s'annule. En posant

$$I_i(\mathbf{z}) = \sum_{j=1}^M (C_{ij} - C_{jj}) P(S_j) p(\mathbf{z} | S_j),$$

on obtient

$$C = \sum_{j=1}^M C_{jj} P(S_j) + \sum_{i=1}^M \int_{\mathcal{Z}_i} I_i(\mathbf{z}) d\mathbf{z}.$$

De même que dans le cas d'hypothèses binaires, le premier terme est le coût fixé; il est indépendant de toute partition de l'espace d'observation. Pour minimiser le coût moyen C , on prendra la décision \mathbf{D}_i , c'est à dire que l'on affectera à la région \mathcal{Z}_i les points pour lesquels

$$I_i(\mathbf{z}) < I_j(\mathbf{z}) \quad j = 1, 2, \dots, M, \text{ et } j \neq i.$$

En posant

$$J_i(\mathbf{z}) = \frac{I_i(\mathbf{z})}{p(\mathbf{z}|S_1)},$$

et

$$\Lambda_i(\mathbf{z}) = \frac{p(\mathbf{z}|S_i)}{p(\mathbf{z}|S_1)},$$

où $\Lambda_1(\mathbf{z}) = 1$, on a alors

$$J_i(\mathbf{z}) = \sum_{j=1}^M (C_{ij} - C_{jj}) P(S_j) \Lambda_i(\mathbf{z}).$$

On obtient ainsi

$$\mathcal{Z}_i = \left\{ \mathbf{z} \mid J_i(\mathbf{z}) < J_j(\mathbf{z}), j = 1, 2, \dots, M, j \neq i \right\}.$$

Pour chaque \mathcal{Z}_i , pour $i = 1, 2, \dots, M$, on doit comparer $J_i(\mathbf{z})$ aux $(M-1)$ autres $J_j(\mathbf{z})$. Il y a donc $M(M-1)$ inégalités à vérifier pour l'ensemble des \mathcal{Z}_i , dans un espace de dimension $M-1$, celui des $\Lambda_k(\mathbf{z})$. En fait, et par symétrie, les relations $J_i(\mathbf{z}) = J_j(\mathbf{z})$ définissent donc $M(M-1)/2$ variétés linéaires (hyperplans) dans l'espace des $\Lambda_k(\mathbf{z})$. On admet que ces variétés linéaires passent par un même point et qu'elles définissent M régions de décision, chacune étant déterminée par un système de $M-1$ inégalités.

4.2.2.b Exemple : décision à hypothèses ternaires

Dans le cas d'une décision à hypothèses ternaires, on définit les régions \mathcal{Z}_1 , \mathcal{Z}_2 , et \mathcal{Z}_3 en utilisant les relations précédentes. Par exemple, pour prendre la décision \mathbf{D}_1 il faut que l'observation \mathbf{z} vérifie

$$J_1(\mathbf{z}) < J_2(\mathbf{z}) \quad \text{et} \quad J_1(\mathbf{z}) < J_3(\mathbf{z}),$$

c'est à dire, respectivement,

$$\begin{aligned} (C_{12} - C_{22}) P(S_2) \Lambda_2(\mathbf{z}) + (C_{13} - C_{33}) P(S_3) \Lambda_3(\mathbf{z}) \\ < (C_{21} - C_{11}) P(S_1) + (C_{23} - C_{33}) P(S_3) \Lambda_3(\mathbf{z}), \end{aligned}$$

et

$$\begin{aligned} (C_{12} - C_{22}) P(S_2) \Lambda_2(\mathbf{z}) + (C_{13} - C_{33}) P(S_3) \Lambda_3(\mathbf{z}) \\ < (C_{31} - C_{11}) P(S_1) + (C_{32} - C_{22}) P(S_2) \Lambda_2(\mathbf{z}). \end{aligned}$$

Après simplification et en posant

$$\begin{aligned} A &= P(S2) (C_{12} - C_{22}) & B &= P(S3) (C_{13} - C_{23}) \\ C &= P(S1) (C_{21} - C_{11}) & D &= P(S2) (C_{12} - C_{32}) \\ E &= P(S3) (C_{13} - C_{33}) & F &= P(S1) (C_{31} - C_{11}), \end{aligned}$$

on obtient

$$\mathbf{z} \in \mathcal{Z}_1 \text{ si } \begin{cases} A\Lambda_2(\mathbf{z}) + B\Lambda_3(\mathbf{z}) > C \\ D\Lambda_2(\mathbf{z}) + E\Lambda_3(\mathbf{z}) > F \end{cases}$$

En procédant de même pour les sous-espaces \mathcal{Z}_2 et \mathcal{Z}_3 on obtient :

$$\begin{aligned} \mathbf{z} \in \mathcal{Z}_2 \text{ si } & \begin{cases} A\Lambda_2(\mathbf{z}) + B\Lambda_3(\mathbf{z}) > C \\ (A - D)\Lambda_2(\mathbf{z}) + (B - E)\Lambda_3(\mathbf{z}) > C - F \end{cases} \\ \mathbf{z} \in \mathcal{Z}_3 \text{ si } & \begin{cases} D\Lambda_2(\mathbf{z}) + E\Lambda_3(\mathbf{z}) > F \\ (A - D)\Lambda_2(\mathbf{z}) + (B - E)\Lambda_3(\mathbf{z}) > C - F \end{cases} \end{aligned}$$

Ces relations peuvent être résumées par

$$\begin{aligned} A\Lambda_2(\mathbf{z}) + B\Lambda_3(\mathbf{z}) &\underset{D_1}{\overset{D_2}{\geq}} C \\ D\Lambda_2(\mathbf{z}) + E\Lambda_3(\mathbf{z}) &\underset{D_3}{\geq} F \\ (A - D)\Lambda_2(\mathbf{z}) + (B - E)\Lambda_3(\mathbf{z}) &\underset{D_2}{\overset{D_3}{\geq}} C - F \end{aligned}$$

Les limites de ces domaines sont données dans le plan $(\Lambda_2(\mathbf{z}), \Lambda_3(\mathbf{z}))$ par

$$\begin{aligned} A\Lambda_2(\mathbf{z}) + B\Lambda_3(\mathbf{z}) &= C \\ D\Lambda_2(\mathbf{z}) + E\Lambda_3(\mathbf{z}) &= F \\ (A - D)\Lambda_2(\mathbf{z}) + (B - E)\Lambda_3(\mathbf{z}) &= C - F \end{aligned}$$

Il faut noter que ces trois relations sont linéairement dépendantes, donc elles possèdent un seul point d'intersection. L'important est de constater que pour un problème de décision à hypothèses ternaires, la statistique suffisante est de dimension 2. Ce résultat est généralisable au cas de M décisions, où la statistique suffisante est de dimension $M - 1$.

4.2.2.c Cas particulier des coûts (0,1)

Si l'on suppose que les coûts choisis sont des coûts (0,1), chaque $I_i(\mathbf{z})$ prend la forme :

$$\begin{aligned} I_i(\mathbf{z}) &= \sum_{\substack{j=1 \\ j \neq i}}^M P(S_j) p(\mathbf{z}|S_j) \\ &= \sum_{j=1}^M P(S_j) p(\mathbf{z}|S_j) - P(S_i) p(\mathbf{z}|S_i). \end{aligned}$$

Puisque $\sum_{j=1}^M P(S_j) p(\mathbf{z}|S_j)$ est identique pour tous les $I_i(\mathbf{z})$, la partition est donc définie par

$$\mathcal{Z}_i = \left\{ \mathbf{z} \mid P(S_i) p(\mathbf{z}|S_i) > P(S_j) p(\mathbf{z}|S_j), j = 1, 2, \dots, M, j \neq i \right\},$$

ou encore

$$\mathcal{Z}_i = \left\{ \mathbf{z} \mid p(S_i|\mathbf{z}) > p(S_j|\mathbf{z}), j = 1, 2, \dots, M, j \neq i \right\}.$$

Il s'agit, encore une fois, du critère du maximum de probabilité a posteriori. La probabilité d'erreur associée à un message S_k est donnée par $P_e(S_k) = 1 - P(\mathbf{D}_k|S_k)$, et la probabilité totale d'erreur sera donc son espérance mathématique, soit

$$P_e = \sum_k P_e(S_k) P(S_k).$$

4.2.3 Autres critères

4.2.3.a Critère minimax

Il n'existe pas de solution analytique générale pour le critère minimax. Le problème est donc de déterminer les M probabilités $P(S_j)$ qui rendent maximum le coût moyen minimum de Bayes. En fait, il suffit d'en déterminer $M - 1$ car $\sum_{j=1}^M P(S_j) = 1$.

4.2.3.b Critère de Neyman-Pearson

Dans le cas du critère de Neyman-Pearson, il n'existe pas de solution générale, car il est difficile de choisir les contraintes. En effet, dans le cas de M hypothèses, il y a $M(M - 1)$ probabilités conditionnelles d'erreur et M

probabilités conditionnelles de bonne décision. Une solution éventuelle est de se fixer une borne supérieure pour les probabilités d'erreur et de rechercher une solution maximisant les probabilités de bonne décision

4.2. Décision avec hypothèses multiples

Chapitre 5

Hypothèses composites



Annexe A

Fonction caractéristique

Soit X un variable aléatoire de densité de probabilité $p(x)$, c'est à dire on a $E(g(X)) = \int_{\mathbb{R}} g(x)p(x)dx$, pour toute fonction $g(X)$. La fonction caractéristique de la variable X , désignée par $\phi_X(\omega)$ pour $\omega \in \mathbb{R}$, est la transformée de Fourier de $p(x)$, c'est à dire

$$\phi_X(\omega) = E(e^{j\omega x}) = \int_{\mathbb{R}} e^{j\omega x} p(x) dx.$$

En particulier, on a $\phi_X(0) = \int_{\mathbb{R}} e^0 p(x) dx = 1$. On a

$$\left. \frac{d^n \phi_X(\omega)}{d\omega^n} \right|_{\omega=0} = \int (jx)^n e^{j\omega x} p(x) dx \Big|_{\omega=0} = j^n \int x^n p(x) dx = j^n E(X^n).$$

En conséquence,

$$\begin{aligned} \phi_X(\omega) &= \phi_X(0) + \left. \frac{d\phi_X(\omega)}{d\omega} \right|_{\omega=0} \frac{\omega}{1!} + \left. \frac{d^2 \phi_X(\omega)}{d\omega^2} \right|_{\omega=0} \frac{\omega^2}{2!} + \dots \\ &= \sum_{n=0}^{\infty} \left. \frac{d^n \phi_X(\omega)}{d\omega^n} \right|_{\omega=0} \frac{\omega^n}{n!} \\ &= \sum_{n=0}^{\infty} j^n E(X^n) \frac{\omega^n}{n!}. \end{aligned}$$
