Théorie de l'estimation et de la décision statistique

Paul Honeine

en collaboration avec
Régis Lengellé

– Université de technologie de Troyes –

2013–2014

Quelques références

Decision and estimation theory — J.L. Melsa, D.L. Cohn
éd. McGraw-Hill (1978)

Décision en traitement du signal — P.Y. Arques
éd. Masson (1982)

Statistical signal processing :
Detection, estimation, and time series analysis — L.L. Scharf
éd. Addison-Wesley (1990)

Signaux aléatoires (3 tomes) — B. Picinbono
éd. Dunod (1994)

Fundamentals of statistical signal processing :
Estimation theory (1993), Detection theory (1998) — S.M. Kay
éd. Prentice Hall

Signal : detection and estimation — M. Barkat
éd. Artech-House (2005)

- Estimation
  - de paramètres
  - de signaux/images (et prédiction)
- Détection
- Classification

1. Théorie de l'estimation
   - Problématique
   - Estimateur sans biais a variance minimum
   - Borne de Cramer Rao
   - Recherche de l'estimateur sans biais à variance minimum
   - Méthode du maximum de vraisemblance
   - Méthode des moments
   - Approche Bayésienne
2. Théorie de la décision
   - Introduction
   - Formulation du problème
   - Les principaux critères de décision
     - Critère de Bayes
     - Critère minimax
     - Critère de Neyman Pearson
     - Notion de statistique suffisante
   - Caractéristique opérationnelle d'un récepteur (Définition, Propriétés, Intérêt)
   - Le problème Gaussien
   - Généralisation aux mesures multiples
   - Cas des hypothèses composites
   - Eléments de détection séquentielle
3. Détection à structure imposée
   - Eléments de théorie de l'apprentissage
   - Machines à vecteurs supports

Théorie de l'estimation

Problématique
Estimation ponctuelle

Densité de probabilité :

$$(Z_1, Z_2, \ldots, Z_N) \approx p(z_1, z_2, \ldots, z_N, \theta)$$

- En théorie des probabilités, θ est déterministe
- En statistique, θ est une variable aléatoire

$\hat{\theta}$: estimateur de θ (v. a. fonction des v. a. de l'échantillon)
$\hat{\theta} = g(Z_1, Z_2, \ldots, Z_N)$

Loi de vraisemblance :

$$\begin{aligned} L(Z_1, Z_2, \ldots, Z_N) &= p(z_1, z_2, \ldots, z_N, \theta)\big|_{z_1=Z_1, z_2=Z_2, \ldots, z_N=Z_N} \\ &= p(Z_1, Z_2, \ldots, Z_N, \theta) \end{aligned}$$

Example 0 : ...

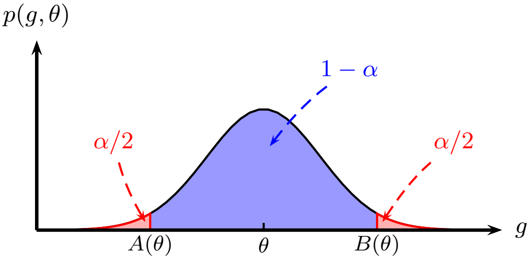
Problématique
Estimation par intervalle de confiance

$$\hat{\theta} = g(Z_1, Z_2, \dots, Z_N) \approx p(g, \theta)$$

$$p(A(\theta) \leq \hat{\theta} \leq B(\theta)) = 1 - \alpha$$

ou encore

$$p(A^{-1}(\hat{\theta}) \leq \theta \leq B^{-1}(\hat{\theta})) = 1 - \alpha$$



Exemple 1 : $Z_i = B_i + \theta$, avec $B_i \approx \mathcal{N}(0, \sigma^2)$, σ étant connu.
$\hat{\theta}_1 = \frac{1}{N}\sum_{i=1}^{N} Z_i$ et $\hat{\theta}_2 = Z_1$

Un estimateur est une variable aléatoire. Ses performances ne peuvent être décrites que statistiquement

Estimateur sans biais à variance minimale
= Minimum variance unbiased (MVU) estimator

Estimateur sans biais à variance minimale
= Minimum variance unbiased (MVU) estimator

Biais : $b(\hat{\theta}) = E(\hat{\theta}) - \theta$

Estimateur sans biais : $E(\hat{\theta}) = \theta$

$\hat{\theta} = g(Z_1, Z_2, \ldots, Z_N)$

$$\begin{aligned} E(\hat{\theta}) &= E(g(Z_1, Z_2, \ldots, Z_N)) \\ &= \iiint g(z_1, z_2, \ldots, z_N)\, p(z_1, z_2, \ldots, z_N, \theta)\, dz_1 dz_2 \ldots dz_N \\ &= \iiint g\, p(g, \theta)\, dg \end{aligned}$$

Exemple : $\hat{\theta}_3 = \frac{1}{2N} \sum_{i=1}^{N} Z_i$

Estimateur sans biais à variance minimale
Motivation : critère d'erreur quadratique moyenne minimale

Definition (Erreur quadratique moyenne)

$$mse(\hat{\theta}) = E\left((\hat{\theta} - \theta)^2\right)$$

$$\begin{aligned} mse(\hat{\theta}) &= E\left((\hat{\theta} - \theta)^2\right) \\ &= E\left((\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2\right) \\ &= var(\hat{\theta}) + b^2(\hat{\theta}) \end{aligned}$$

Exemple : $\hat{\theta}_4 = \frac{a}{N}\sum_{i=1}^{N} Z_i$

L'estimateur MVU existe t'il ?
et, si oui, comment le trouver ?

Estimateur sans biais à variance minimale : estimateur efficace et borne de Cramer-Rao

Retour sur la loi de vraisemblance :

$p(Z,\theta) = p(Z_1, Z_2, \ldots, Z_N, \theta)$
est fonction de θ à $Z = [Z_1 \ Z_2 \ \cdots \ Z_N]^\top$ fixé.

Exemple 1 : ...

Remarque : La loi de vraisemblance porte l'information sur le paramètre à estimer. Plus les variations de celles-ci sont importantes par rapport au paramètre, et plus celui-ci est aisé à estimer.

Exemple : La loi uniforme $\mathcal{U}(0,a)$

Estimateur sans biais à variance minimale

Estimateur efficace et borne de Cramer-Rao

Theorem (Estimateur efficace et borne de Cramer-Rao)

Soit $\hat{\theta}$ un estimateur sans biais de θ. Si $p(Z,\theta)$ vérifie, pour tout θ,

$$E\left(\frac{\partial \ln p(Z,\theta)}{\partial \theta}\right) = 0, \quad \text{alors} \quad var(\hat{\theta}) \geq \frac{-1}{E\left(\frac{\partial^2 \ln p(Z,\theta)}{\partial \theta^2}\right)}.$$

De plus, si et seulement s'il existe $g(\cdot)$ et $I(N,\theta)$ telles que

$$\frac{\partial \ln p(Z,\theta)}{\partial \theta} = I(N,\theta)\left(g(Z) - \theta\right),$$

alors $g(Z)$ est un estimateur sans biais dont la variance atteint la borne. L'estimateur $\hat{\theta}$ est dit efficace.

Exemple 1 : $\hat{\theta}_{eff} = \frac{1}{N}\sum_{i=1}^{N} Z_i$, et $var(\hat{\theta}_{eff}) = \sigma^2/N$

Estimateur sans biais à variance minimale
Estimateur efficace et borne de Cramer-Rao : démonstration

Démonstration (1/3) :
Soit $\hat{\theta}$ un estimateur sans biais de θ.
D'une part, la condition permet d'écrire

$$0 = E\left(\frac{\partial \ln p(Z,\theta)}{\partial \theta}\right) = \int \frac{\partial \ln p(z,\theta)}{\partial \theta} p(z,\theta)dz \overset{①}{=} \int \frac{\partial p(z,\theta)}{\partial \theta} dz \overset{②}{=} \frac{\partial}{\partial \theta} \int p(z,\theta)dz,$$

où $\overset{①}{=}$ est due à la relation $\frac{\partial u(\theta)}{\partial \theta} = u(\theta)\frac{\partial \ln u(\theta)}{\partial \theta}$ car $\frac{\partial \ln u(\theta)}{\partial \theta} = \frac{\partial u(\theta)}{\partial \theta}\frac{1}{u(\theta)}$
et $\overset{②}{=}$ vraie si les bornes ne dépendent pas de θ, donc les supports de p
D'autre part, l'estimateur est sans biais, c'est-à-dire $E(\hat{\theta}) = \theta$, soit en dérivant

$$\begin{aligned}
1 &= \frac{\partial}{\partial \theta} \int f(z)p(z,\theta)dz = \int f(z) \frac{\partial}{\partial \theta} p(z,\theta)dz = \int f(z) \frac{\partial \ln p(z,\theta)}{\partial \theta} p(z,\theta)dz \\
&= \int \left(f(z) - \theta\right) \frac{\partial \ln p(z,\theta)}{\partial \theta} p(z,\theta)dz
\end{aligned}$$

où on peut introduire θ car $\int \theta \frac{\partial \ln p(z,\theta)}{\partial \theta} p(z,\theta)dz = \theta E\left(\frac{\partial \ln p(z,\theta)}{\partial \theta}\right) = 0$

Estimateur sans biais à variance minimale
Estimateur efficace et borne de Cramer-Rao : démonstration

Démonstration (2/3) :

Rappel : inégalité de Cauchy-Schwarz : pour tout $w(z) \geq 0$, on a

$$\left| \int w(z)\, k(z)\, h(z)\, dz \right|^2 \leq \int w(z)\, k^2(z)\, dz \int w(z)\, h^2(z)\, dz,$$

avec égalité si il existe un c tel que $k(z) = c\, h(z)$.

Posons $w(z) = p(z,\theta)$, $k(z) = f(z) - \theta$, et $h(z) = \frac{\partial}{\partial\theta} \ln p(z,\theta)$, alors

$$1 \leq \int p(z,\theta) \big(f(z) - \theta\big)^2 dz \int p(z,\theta) \left(\frac{\partial}{\partial\theta} \ln p(z,\theta)\right)^2 dz$$

Le premier terme vaut $E\Big(\big(f(z) - \theta\big)^2\Big) = var(\hat{\theta})$, et le second $E\left(\left(\frac{\partial}{\partial\theta} \ln p(z,\theta)\right)^2\right)$

$$var(\hat{\theta}) \geq \frac{1}{E\left(\left(\frac{\partial}{\partial\theta} \ln p(z,\theta)\right)^2\right)}$$

Estimateur sans biais à variance minimale
Estimateur efficace et borne de Cramer-Rao : démonstration

Démonstration (3/3) :

Par ailleurs,

$$0 = E\left(\frac{\partial \ln p(z,\theta)}{\partial \theta}\right) = \int \frac{\partial \ln p(z,\theta)}{\partial \theta} p(z,\theta) dz$$

On dérive à nouveau

$$0 = \int \frac{\partial}{\partial \theta}\left(\frac{\partial \ln p(z,\theta)}{\partial \theta} p(z,\theta)\right) dz = \int \frac{\partial^2 \ln p(z,\theta)}{\partial \theta^2} p(z,\theta) + \frac{\partial \ln p(z,\theta)}{\partial \theta} \frac{\partial p(z,\theta)}{\partial \theta} dz$$

Or, le second terme vaut $E\left(\left(\frac{\partial \ln p(z,\theta)}{\partial \theta}\right)^2\right)$ car $\frac{\partial p(z,\theta)}{\partial \theta} = \frac{\partial \ln p(z,\theta)}{\partial \theta} p(z,\theta)$.

Donc

$$E\left(\frac{\partial^2 \ln p(z,\theta)}{\partial \theta^2}\right) = -E\left(\left(\frac{\partial \ln p(z,\theta)}{\partial \theta}\right)^2\right)$$

$$var(\hat{\theta}) \geq \frac{-1}{E\left(\frac{\partial^2 \ln p(z,\theta)}{\partial \theta^2}\right)}$$

Estimateur sans biais à variance minimale
Estimateur efficace et borne de Cramer-Rao : démonstration

Démonstration (3^+/3) :
La borne est atteinte lorsque l'inégalité de Cauchy-Schwarz devient égalité, c'est-à-dire lorsque $h(z) = I(\theta)k(z)$.

$$\frac{\partial \ln p(z,\theta)}{\partial \theta} = I(\theta)\big(f(z) - \theta\big)$$

En dérivant, on obtient

$$\frac{\partial^2 \ln p(z,\theta)}{\partial \theta^2} = -I(\theta) + \frac{\partial I(\theta)}{\partial \theta}\big(f(z) - \theta\big)$$

et en considérant l'espérance de cette expression, on obtient

$$E\Big(\frac{\partial^2 \ln p(z,\theta)}{\partial \theta^2}\Big) = -I(\theta) + \big(E(\hat{\theta}) - \theta\big)$$

où le dernier terme est nul car $\hat{\theta}$ est sans biais.

$$I(\theta) = -E\Big(\frac{\partial^2 \ln p(z,\theta)}{\partial \theta^2}\Big)$$

Estimateur sans biais à variance minimale
Estimateur efficace et borne de Cramer Rao

Theorem (Estimateur efficace et borne de Cramer-Rao, extension au cas vectoriel)

Soit $\hat{\boldsymbol{\theta}}$ un estimateur sans biais de $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \cdots \ \theta_p]^\top$. Si $p(Z, \boldsymbol{\theta})$ vérifie, pour tout $\boldsymbol{\theta}$, $E\left(\frac{\partial \ln p(Z,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = 0$, alors la matrice de covariance $C_{\hat{\boldsymbol{\theta}}}$ de $\hat{\boldsymbol{\theta}}$ vérifie :

$$C_{\hat{\boldsymbol{\theta}}} - I^{-1}(\hat{\boldsymbol{\theta}}) \geq 0, \ \text{où} \ [I(\hat{\boldsymbol{\theta}})]_{i,j} = -E\left(\frac{\partial^2 \ln p(Z,\boldsymbol{\theta})}{\partial \theta_i \, \partial \theta_j}\right).$$

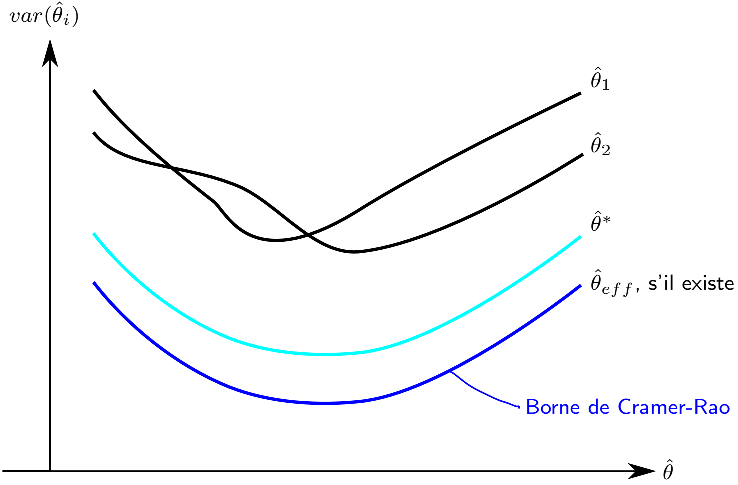
De plus, si et seulement s'il existe une fonction vectorielle $g(\cdot)$ et une matrice $I(N, \theta)$, de taille $p \times p$, telles que

$$\frac{\partial \ln p(Z,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = I(N, \boldsymbol{\theta}) \left(g(Z) - \boldsymbol{\theta}\right),$$

alors $g(Z)$ est un estimateur sans biais dont la variance atteint la borne. L'estimateur $\hat{\boldsymbol{\theta}}$ est dit efficace, avec $C_{\hat{\boldsymbol{\theta}}} = I^{-1}(\hat{\boldsymbol{\theta}})$.

Recherche de l'estimateur sans biais à variance minimale

L'estimateur efficace $\hat{\theta}_{eff}$ n'existe pas toujours ...
Comment faire pour trouver l'estimateur sans biais à variance minimale $\hat{\theta}^*$?



Recherche de l'estimateur sans biais à variance minimale
Notion de statistique suffisante

Example 1 : $\hat{\theta}_1 = \frac{1}{N}\sum_{i=1}^{N} Z_i$
Quelles sont les données utiles pour estimer θ ?

- $Z_1, Z_2, \ldots, Z_N$
- $Z_1 + Z_2, Z_3, \ldots, Z_N$
- $\sum_{i=1}^{N} Z_i$

Chacune de ces statistiques est suffisante pour estimer θ.
Toutefois, $\sum_{i=1}^{N} Z_i$ est de *dimension minimale*.

Definition (statistique suffisante)

Une statistique $T(z_1, z_2, \ldots, z_N)$ est suffisante pour un paramètre θ si $p\left(z_1, z_2, \ldots, z_N \mid T(z_1, z_2, \ldots, z_N), \theta\right)$ est indépendant de θ.

Recherche de l'estimateur sans biais à variance minimale
Notion de statistique suffisante

Theorem (factorisation de Neyman Fisher)

Si $p(Z_1, \ldots, Z_N, \theta) = g(T(Z_1, \ldots, Z_N), \theta)\, h(Z_1, \ldots, Z_N)$, *alors* $T(Z_1, Z_2, \ldots, Z_N)$ *est une statistique suffisante pour* θ

Exemple 1

Theorem (Rao Blackwell Lehmann Scheffé)

Soit $\hat{\theta}$ *un estimateur sans biais de* θ *et* $T(Z_1, Z_2, \ldots, Z_N)$ *une statistique suffisante. Alors* $E(\hat{\theta} \mid T(Z_1, Z_2, \ldots, Z_N))$ *est un estimateur sans biais de variance plus faible que celle de* $\hat{\theta}$*.*
De plus, si la statistique est complète, *alors* $E(\hat{\theta} \mid T(Z_1, Z_2, \ldots, Z_N))$ *est l'estimateur à variance minimale*

Definition (statistique complète)

Une statistique est dite complète s'il existe une seule fonction de celle-ci qui est sans biais.

Recherche de l'estimateur sans biais à variance minimale

Application :
$\tilde{\theta} = E(\hat{\theta} \mid T(Z_1, Z_2, \ldots, Z_N))$ est un estimateur sans biais de θ de variance plus faible que $\hat{\theta}$. Mais $\tilde{\theta} = E(\hat{\theta} \mid T(Z_1, Z_2, \ldots, Z_N))$ est une fonction $g(T(Z_1, Z_2, \ldots, Z_N))$. Si la statistique est complète alors g est unique. $\tilde{\theta}$ est donc unique (indépendant du choix initial de $\hat{\theta}$). C'est donc l'estimateur sans biais de variance minimale (s'il existe). De manière duale, si la statistique est complète, il suffit de trouver la fonction $g(T(Z_1, Z_2, \ldots, Z_N))$ qui soit sans biais.

Exemple 1 :

Méthode du maximum de vraisemblance

Si

- l'estimateur efficace n'existe pas
- le calcul de $E(\hat{\theta} \mid \hat{\theta}_{suff})$ est très difficile
- on ne peut trouver de fonction $g(\hat{\theta}_{suff})$ qui soit sans biais (et unique ...)

alors

- on cherche un estimateur asymptotiquement efficace (attention à la taille de l'échantillon ...)

Definition (estimateur du maximum de vraisemblance)

L'estimateur du maximum de vraisemblance est celui qui maximise la loi de vraisemblance

$$\hat{\theta}_{MV} = \arg\max_{\theta} p(Z, \theta)$$

En général, on maximise plutôt $\ln(p(Z, \theta))$

$$\frac{\partial \ln p(Z, \theta)}{\partial \theta} = 0, \quad \text{et} \quad \frac{\partial^2 \ln p(Z, \theta)}{\partial \theta^2} < 0.$$

Exemple 1 : Appliquer la méthode du maximum de vraisemblance revient à retenir la valeur de θ qui maximise la densité de probabilité des donnés observés.

Méthode du maximum de vraisemblance

Exemple 1 : Appliquer la méthode du maximum de vraisemblance revient à retenir la valeur de θ qui maximise la densité de probabilité des donnés observés.

Exemple 1 : $\hat{\theta}_{MV} = \frac{1}{N}\sum_{i=1}^{N} Z_i$, sur cet exemple, on constate que $\hat{\theta}_{MV}$ est efficace.

En effet, si l'estimateur efficace existe, alors $\frac{\partial \ln p(Z,\theta)}{\partial \theta} = I(\theta)(\hat{\theta}_{eff} - \theta)$. On a alors $\frac{\partial \ln p(Z,\theta)}{\partial \theta} = 0$ pour $\theta = \hat{\theta}_{MV} = \hat{\theta}_{eff}$.

Conclusion :
Si l'estimateur efficace existe, il est donné de manière unique par la méthode du maximum de vraisemblance. Attention, la réciproque n'est pas vraie.

Méthode du maximum de vraisemblance

Theorem

Si les dérivées première et seconde de $\ln p(Z,\theta)$ *existent, et si*

$$E\left(\frac{\partial \ln p(Z,\theta)}{\partial \theta}\right) = 0$$

pour tout θ, *alors* $\hat{\theta} \to \mathcal{N}(\theta, I^{-1}(\theta))$ *lorsque* $N \to \infty$

Theorem

Si $g(\cdot)$ *est bijective, alors l'estimateur du maximum de vraisemblance de* $g(\theta)$ *est* $g(\hat{\theta}_{MV})$, *soit*

$$\widehat{(g(\theta))}_{MV} = g(\hat{\theta}_{MV}).$$

Méthode des moments

Méthode des moments

Supposons qu'il existe une fonction $b(\theta) = E(g(Z)) = \int g(\boldsymbol{z})\, p(\boldsymbol{z}, \theta)\, d\boldsymbol{z}$, où la fonction $b(\cdot)$ est continue et monotone, alors

$$\theta = b^{-1}\left(E(g(Z))\right).$$

La méthode des moments consiste à remplacer le moment théorique $E(g(Z))$ par le moment empirique correspondant $\frac{1}{N}\sum_{i=1}^{N} g(Z_i)$. On en déduit alors un estimateur de θ :

$$\theta = b^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} g(Z_i)\right).$$

Si la fonction $b^{-1}(\cdot)$ est continue, alors l'estimateur obtenu est convergent.

Exemple 1 : $E(Z) = \theta$, alors $\hat{\theta}_{mom} = \frac{1}{N}\sum_{i=1}^{N} Z_i$

Approche Bayésienne

θ est supposé aléatoire de loi de probabilité a priori $p(\theta)$. On a alors

$$p(\theta) = \int p(Z,\theta)\, dZ$$
$$p(Z,\theta) = p(\theta \mid Z)\, p(Z)$$

Par exemple :

$$\begin{aligned} mse(\hat{\theta}) &= \iint (\hat{\theta}-\theta)^2 p(Z,\theta) dZ d\theta \\ &= \left(\iint (\hat{\theta}-\theta)^2 p(\theta|Z) d\theta \right) p(Z) dZ \end{aligned}$$

L'estimateur qui minimise la mse est : $\hat{\theta} = E(\theta|Z)$

On peut aussi imaginer d'autres fonctions *coût* $C(\hat{\theta}-\theta)$:

- $C(\hat{\theta}-\theta) = \left|\hat{\theta}-\theta\right|$, qui donne la médiane de $p(\theta|Z)$
- $C(\hat{\theta}-\theta) = \begin{cases} 0 & \text{si } |\hat{\theta}-\theta| < \delta \\ 1 & \text{si } |\hat{\theta}-\theta| \geq \delta \end{cases}$, qui correspond au mode de $p(\theta|Z)$

Approche Bayésienne

Exemple :

$$p(z_k|\theta) = \begin{cases} \theta \exp(-\theta z_k) & \text{si } z_k \geq 0 \\ 0 & \text{sinon} \end{cases}$$

$$p(\theta) = \begin{cases} \lambda \exp(-\lambda \theta) & \text{si } \theta \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Alors

$$\hat{\theta}_{MAP} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} Z_i + \frac{\lambda}{N}}$$

Théorie de la décision binaire

Théorie de la décision binaire

EXEMPLE INTRODUCTIF 1

Problème :

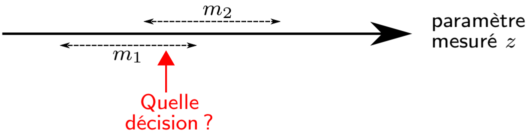
- Reconnaître à l'aide d'une caméra deux types de pièces plates

Méthode :

- Extraction d'un paramètre de l'image (i.e., surface)
- Classer les pièces en fonction du paramètre choisi

Difficulté :

- Pour un type de pièces données, la mesure est aléatoire (fabrication, mesure, ...)



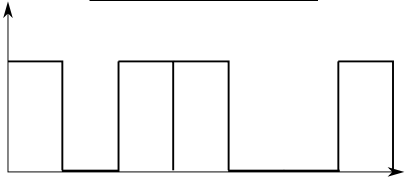
Introduction
Exemple introductif 2

EXEMPLE INTRODUCTIF 2 : Communication binaire synchrone

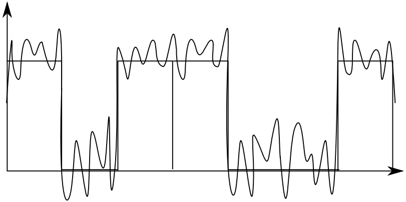
Principe :

signal $\rightarrow$ échantillonnage $\rightarrow$ conversion analogique-numérique

Emission (suite $0-1$)



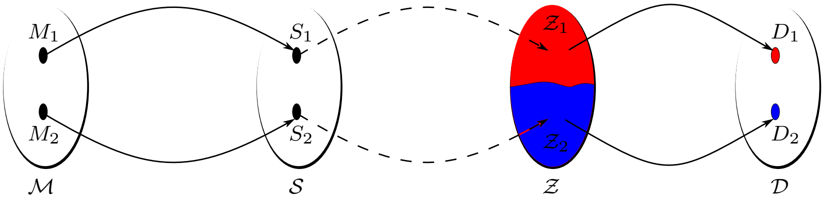
Réception



Décision :

A-t-on reçu 0 ou 1 ?

Formulation du problème



- $\mathcal{M}$: espace des messages, source (les 2 pièces)
- $\mathcal{S}$: espace des signaux (les images)
- $\mathcal{Z}$: espace des observations (les mesures $z \in I\!R$)
- $\mathcal{D}$: espace des décisions (quelle pièce ?)

Comment passer de $\mathcal{Z}$ à $\mathcal{D}$? Quelle partition de $\mathcal{Z}$?

Formulation du problème
Hypothèses et notations

Messages :
Probabilité des messages : $P(M_1), P(M_2)$, avec $P(M_1) + P(M_2) = 1$

Signaux :
Probabilité des signaux : $P(S_i) = P(M_i)$, pour $i = 1, 2$

Probabilités a priori : $P(S_i), P(M_i)$

Observations (mesures) :
Le vecteur des mesures $\boldsymbol{z}$ est la réalisation d'un vecteur aléatoire $\boldsymbol{Z}$

Probabilités conditionnelles des mesures : $p(\boldsymbol{z} \mid S_i)$, pour $i = 1, 2$

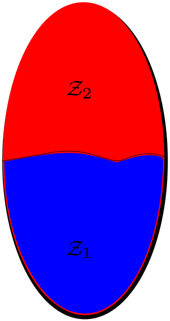
Probabilité des mesures :
$p(\boldsymbol{z}) = p(\boldsymbol{z} \mid S_1)\, P(S_1) + p(\boldsymbol{z} \mid S_2)\, P(S_2)$

Remarque :
$\boldsymbol{Z}$ est un vecteur aléatoire discret ou continue
Si $\boldsymbol{Z}$ est continu, alors $p(\boldsymbol{z} \mid S_i)$ et $p(\boldsymbol{z})$ sont des densités de probabilités, avec $\int_{\boldsymbol{Z}} p(\boldsymbol{z})\, d\boldsymbol{z} = 1$ et $\int_{\boldsymbol{Z}} p(\boldsymbol{z} \mid S_i)\, d\boldsymbol{z} = 1$

Formulation du problème

De l'observation à la décision :

Partition de l'espace des observations, $\boldsymbol{Z}_1$ et $\boldsymbol{Z}_2$,
avec $\boldsymbol{Z}_1 \cup \boldsymbol{Z}_2 = \boldsymbol{Z}$ et $\boldsymbol{Z}_1 \cap \boldsymbol{Z}_2 = \emptyset$



Décision :

- Décision $\boldsymbol{D}_1$ si $\boldsymbol{z} \in \boldsymbol{Z}_1$
- Décision $\boldsymbol{D}_2$ si $\boldsymbol{z} \in \boldsymbol{Z}_2$

Formulation du problème

Probabilités de décision :

- Conditionnelle

$$P(\boldsymbol{D}_i \mid S_j) = \int_{\boldsymbol{Z}_i} p(\boldsymbol{z} \mid S_j)\, d\boldsymbol{z}$$

$$P(\boldsymbol{D}_1 \mid S_j) + P(\boldsymbol{D}_2 \mid S_j) = 1$$

- Totale

$$P(\boldsymbol{D}_i) = \int_{\boldsymbol{Z}_i} p(\boldsymbol{z})\, d\boldsymbol{z}$$

$$P(\boldsymbol{D}_i) = P(\boldsymbol{D}_i \mid S_1)P(S_1) + P(\boldsymbol{D}_i \mid S_2)P(S_2)$$

- Bonnes décisions : $S_i \Leftrightarrow \boldsymbol{D}_i$, pour $i = 1, 2$

$$P_{BD} = P(\boldsymbol{D}_1 \mid S_1)P(S_1) + P(\boldsymbol{D}_2 \mid S_2)P(S_2)$$

- Mauvaises décisions : $S_i \Leftrightarrow \boldsymbol{D}_j$, pour $i \neq j$

$$P_{MD} = P(\boldsymbol{D}_2 \mid S_1)P(S_1) + P(\boldsymbol{D}_1 \mid S_2)P(S_2)$$

Formulation du problème
Etude d'une règle de décision

Principe :
Les mesures étant connues, on choisit l'hypothèse qui a la probabilité maximum

- Décision $\boldsymbol{D}_1$ si $p(S_1 \mid \boldsymbol{z}) > p(S_2 \mid \boldsymbol{z})$
- Décision $\boldsymbol{D}_2$ si $p(S_1 \mid \boldsymbol{z}) < p(S_2 \mid \boldsymbol{z})$

$$\frac{p(S_2 \mid \boldsymbol{z})}{p(S_1 \mid \boldsymbol{z})} \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} 1$$

Critère du maximum de probabilité a postériori
Règle :

$$\frac{p(\boldsymbol{z} \mid S_2)}{p(\boldsymbol{z} \mid S_1)} \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \frac{p(S_1)}{p(S_2)}$$

ceci correspond à comparer le rapport de vraisemblance à un seuil :

$$\Lambda(\boldsymbol{z}) \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \lambda$$

Formulation du problème
Etude d'un problème (classification de pièces)

Hypothèses :

- Paramètres : m_1, m_2, avec $m_1 < m_2$
- Mesures bruitées $z = m_j + v$, avec v une variable aléatoire Gaussienne $\mathcal{N}(0, \sigma^2)$.
- Probabilités a priori : $P(S_1), P(S_2)$

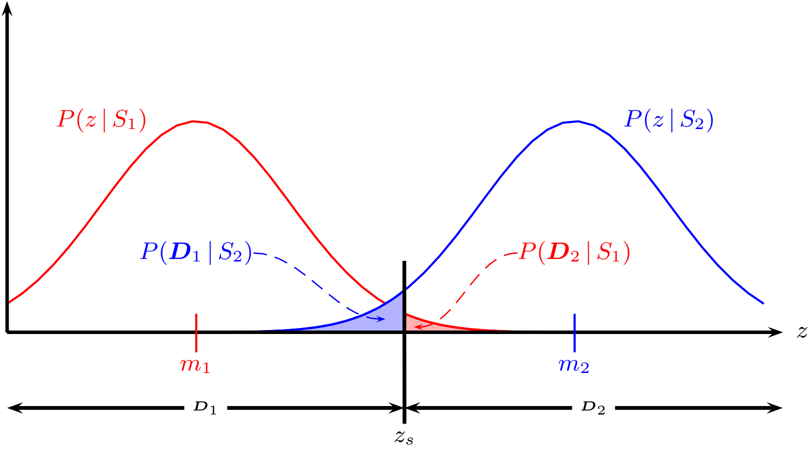
Solution :

$$p(S_2 \mid \boldsymbol{z}) \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} p(S_1 \mid \boldsymbol{z})$$

$$\boldsymbol{z} \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \frac{m_2 + m_1}{2} + \frac{\sigma^2}{m_2 - m_1} \ln \frac{P(S_1)}{P(S_2)}$$

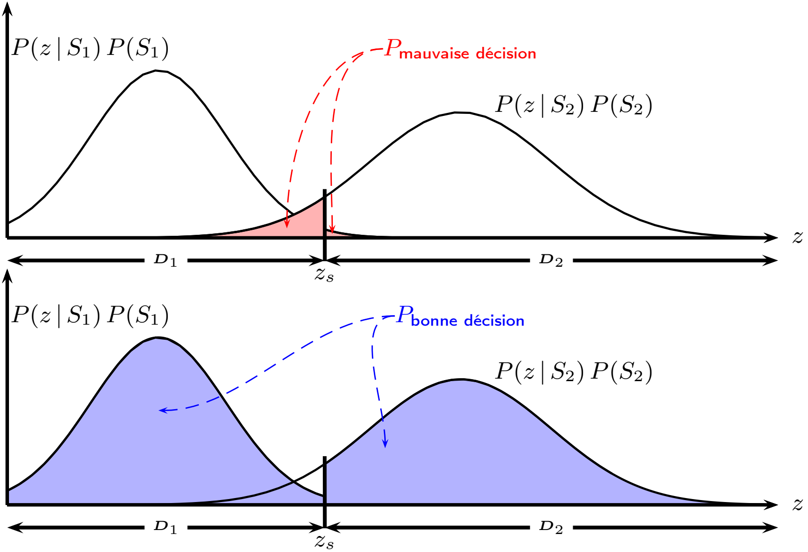
Formulation du problème

Etude d'un problème (classification de pièces)



Formulation du problème

Etude d'un problème (classification de pièces)



Formulation du problème
Etude d'un problème (classification de pièces)

Les probabilités de décision :

$$\begin{aligned} P(\boldsymbol{D}_2 \mid S_i) &= \int_{z_s}^{\infty} p(\boldsymbol{z} \mid S_i)\, d\boldsymbol{z} = 1 - \mathsf{erf}\left(\frac{z_s - m_i}{\sigma}\right) \\ P(\boldsymbol{D}_1 \mid S_i) &= \int_{-\infty}^{z_s} p(\boldsymbol{z} \mid S_i)\, d\boldsymbol{z} = \mathsf{erf}\left(\frac{z_s - m_i}{\sigma}\right) \end{aligned}$$

avec

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

Formulation du problème
Etude du rapport de vraisemblance

- $\boldsymbol{z}$ mesure $\Leftrightarrow$ réalisation de la variable aléatoire $\boldsymbol{Z}$
- $\Lambda(\boldsymbol{z})$ $\Leftrightarrow$ réalisation de la variable aléatoire $\Lambda(\boldsymbol{Z})$

Lois de $\Lambda(\boldsymbol{Z})$:

- Lois conditionnelles $p(\Lambda\ S_i)$
- $p(\Lambda) = p(\Lambda\ S_1)P(S_1) + p(\Lambda\ S_2)P(S_2)$
- ... se déduisent des lois de $\boldsymbol{z}$
- $\Lambda(\boldsymbol{z}) \geq 0$, pour tout $\boldsymbol{z}$

Probabilités de décision :

- Décision $\boldsymbol{D}_1$ si $\Lambda(\boldsymbol{z}) < \lambda$

$$P(\boldsymbol{D}_1 \mid S_j) = \int_0^{\lambda} p_\Lambda(\lambda' \mid S_j) d\lambda'$$

- Décision $\boldsymbol{D}_2$ si $\Lambda(\boldsymbol{z}) \geq \lambda$

$$P(\boldsymbol{D}_2 \mid S_j) = \int_{\lambda}^{\infty} p_\Lambda(\lambda \mid S_j) d\lambda'$$

Critère de Bayes

Hypothèses de base :

- Toutes les décisions ne sont pas équivalentes
  Exemples : détection d'un défaut dans une pièce, détection radar, ...
- On associe un coût à toute décision conditionnelle :
  C_{ij} le coût associé à $\boldsymbol{D}_i$ pour S_j
- $P(S_i)$ et $P(\boldsymbol{z} \mid S_i)$ sont connues, pour $i = 1, 2$
- On définit un *coût moyen* de Bayes

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

Objectif :
Recherche d'une partition de $\mathcal{Z}$ qui rende $\overline{C}$ minimum

Critère de Bayes

Solution :

$$\frac{p(\boldsymbol{z} \mid S_2)}{p(\boldsymbol{z} \mid S_1)} \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \frac{(C_{11} - C_{21})\, p(S_1)}{(C_{22} - C_{12})\, p(S_2)}$$

Pourquoi ?

$$\overline{C} = C_{21}p(S_1) + C_{22}p(S_2) + \int I(\boldsymbol{z})d\boldsymbol{z}$$

coût fixé coût contrôlé

$$I(\boldsymbol{z}) = (C_{11} - C_{21})\, p(\boldsymbol{z} \mid S_1)P(S1) + (C_{12} - C_{22})\, p(\boldsymbol{z} \mid S_2)P(S_2)$$

- $\boldsymbol{Z}_1 = \{\boldsymbol{z} \mid I(\boldsymbol{z}) \leq 0\}$
- $\boldsymbol{Z}_2 = \{\boldsymbol{z} \mid I(\boldsymbol{z}) > 0\}$

FIGURE

Critère minimax

Hypothèses de base :

- Les mêmes que celles du critère de Bayes,
- mais les $P(S_i)$ sont inconnus

Objectif :
Se placer dans le cas le plus défavorable du critère de Bayes

$$\max_{P(S_2)} \overline{C}(P(S_2))$$

Solution :

$$\frac{p(\boldsymbol{z} \mid S_2)}{p(\boldsymbol{z} \mid S_1)} \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \frac{(C_{11} - C_{21})\, p(S_1)}{(C_{22} - C_{12})\, p(S_2)}$$

où $P(S_2)$ est solution de *l'équation caractéristique*

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

Critère minimax

Coût moyen :

$$C(P(S_2)) = C_{21} P(\boldsymbol{D}_2 \mid S_1) + C_{11}\,(1 - P(\boldsymbol{D}_2 \mid S_1))$$
$$+ P(S_2)\Big(\,(C_{22} - C_{11}) + (C_{12} - C_{22})\,P(\boldsymbol{D}_1 \mid S_2) - (C_{21} - C_{11})\,P(\boldsymbol{D}_2 \mid S_1)\Big)$$

Coût minimum :

- Dépend de $P(S_2) \quad (P_2)$
  de façon explicite $\Rightarrow P(S_2)$
  de façon implicite $\Rightarrow \lambda \Leftrightarrow P(\boldsymbol{D}_i \mid S_j)$
- Fixons $P_2 = P_2^*$, alors λ^*, $P(\boldsymbol{D}_1 \mid S_2)$ et $P(\boldsymbol{D}_2 \mid S_1)$ déterminés

Variation explicite $C_e(P_2)$ de $C(P_2)$ en fonction de P_2

$$\overline{C} \quad \Rightarrow \quad \text{variation linéaire par rapport à } P_2$$
$$\Rightarrow \quad \text{passe par } C(P_2^*)$$

FIGURE

Critère minimax

Etude de $C(P(S_2))$:

$$C(P(S_2)) = C_{21} P(\boldsymbol{D}_2 \mid S_1) + C_{11} \left(1 - P(\boldsymbol{D}_2 \mid S_1)\right)$$
$$+ P(S_2) \Big((C_{22} - C_{11}) + (C_{12} - C_{22}) \, P(\boldsymbol{D}_1 \mid S_2) - (C_{21} - C_{11}) \, P(\boldsymbol{D}_2 \mid S_1) \Big)$$

FIGURE

FIGURE

Critère minimax

Cas possibles :

FIGURE

FIGURE

Critère minimax

- La valeur seuil λ est déterminée par l'équation caractéristique,

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

  avec $P(\boldsymbol{D}_1 \mid S_2) = \int_0^{\lambda} p_\Lambda(\lambda' \mid S_2)\, d\lambda'$ et
  $P(\boldsymbol{D}_2 \mid S_1) = \int_{\lambda}^{\infty} p_\Lambda(\lambda' \mid S_1)\, d\lambda'$

- Règle de décision

$$\frac{p(\boldsymbol{z} \mid S_2)}{p(\boldsymbol{z} \mid S_1)} \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \lambda$$

Cas particulier : $C_{ij} = 1 - \delta_{ij}$
$P(\boldsymbol{D}_1 \mid S_2) = P(\boldsymbol{D}_2 \mid S_1)$ et $C = P(\boldsymbol{D}_2 \mid S_1) = P_{MD}$

Critère de Neyman-Pearson

Hypothèses de base :

- Les C_{ij} et les $P(S_i)$ sont inconnus
- Seules les $p(\boldsymbol{z} \mid S_i)$ sont connus

Objectif :
$P(\boldsymbol{D}_2 \mid S_1) = \alpha$ étant imposée

$$\max P(\boldsymbol{D}_2 \mid S_2)$$

Solution :
Recherche d'une partition de $\boldsymbol{Z}$ telle que J est maximum, avec

$$J = P(\boldsymbol{D}_2 \mid S_2) - \lambda\big(P(\boldsymbol{D}_2 \mid S_1) - \alpha\big)$$

où λ est le *multiplicateur de Lagrange*. On a alors

$$\boldsymbol{Z}_2 = \{\boldsymbol{z} \mid p(\boldsymbol{z}|S_2) - \lambda p(\boldsymbol{z} \mid S_1) > 0\}$$

où $\alpha = \int_\lambda^\infty p_\Lambda(\lambda' \mid S_1) d\lambda'$
Règle de décision :

$$\Lambda(\boldsymbol{z}) \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \lambda$$

Critère de Neyman-Pearson

Rappel : optimization avec contraintes

Objectif :

$$\begin{aligned} \text{minimiser} \quad & C = f(x,y) \\ \text{sous la contrainte} \quad & g(x,y) = 0 \end{aligned}$$

Solution :

- Minimisation sans contrainte : $\frac{\partial f(x,y)}{\partial x} = 0$ et $\frac{\partial f(x,y)}{\partial y} = 0$, avec Δx et Δy indépendants
- Minimisation avec contrainte
  FIGURE

  Contrainte : $\boldsymbol{grad} g \boldsymbol{dr} = \mathbf{0}$
  minimisation : $\boldsymbol{grad} f \boldsymbol{dr} = \mathbf{0} \Rightarrow \boldsymbol{grad} f(x,y) = -\lambda \boldsymbol{grad} g(x,y)$, avec λ le *multiplicateur de Lagrange*

Grandeur à minimiser : $f(x,y) + \lambda g(x,y)$

Critère de Neyman-Pearson

Application : communication binaire synchrone

Hypothèse :

- S_1 : signal d'amplitude 0 ; S_2 : signal d'amplitude m
- Bruit additif Gaussien $\mathcal{N}(0, \sigma_v^2)$
- Contrainte $P(\boldsymbol{D}_2 \mid S_1) = \alpha$

Solution :

$$\Lambda(z) = \exp\left(-\frac{m}{2\sigma_v^2}(m - 2z)\right) \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \lambda$$

$$z \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} z_s, \qquad z_s = ?$$

$\Rightarrow P(\boldsymbol{D}_2 \mid S_1) = 1 - \text{erf}(z_s/\sigma_v) = \alpha_0$
$\Rightarrow z_s = \sigma_v \text{erf}^{-1}(1 - \alpha_0)$
$P(\boldsymbol{D}_2 \mid S_2) = 1 - \text{erf}\left(\frac{z_s - m}{\sigma_v}\right)$

Critère de Neyman-Pearson

Application : communication binaire synchrone (suite)

Application numérique : $m = 1$ et $\sigma = 1$

FIGURE

$P(\boldsymbol{D}_2 \mid S_1)$	z_s	λ	$P(\boldsymbol{D}_2 \mid S_2)$
0.25	0.675	1.19	0.63
0.2	0.842	1.41	0.60
0.1	1.28	2.18	0.39
0.05	1.645	3.14	0.26
0.01	2.326	6.2	0.09

Statistique suffisante

FIGURE

Si toute l'information nécessaire à la prise de décision est dans Λ, alors Λ est dite statistique suffisante

Definition (Statistique suffisante)

On appelle *statistique suffisante* toute variable aléatoire ou ensemble de variables aléatoires qui contient toute l'information nécessaire à la prise de décision.

Statistique suffisante

Application : communication binaire synchrone
Détection d'un signal binaire d'amplitude m noyé dans un bruit blanc Gaussien. Sur une période, N échantillons sont prélevés.

S_2 : $z_n = m + v_n$
S_2 : $z_n = v_n$
où $n = 1, 2, \ldots, N$, $E(v_n) = 0$, and $E(v_n v_{n'}) = \sigma_v^2 \delta_{nn'}$

Solution :
Bruit blanc $\Rightarrow$ Echantillons indépendants

$$\Lambda(\boldsymbol{z}) = \exp\left(-\frac{1}{\sigma_v^2}(\boldsymbol{m}^\top \boldsymbol{z} - \frac{1}{2}\boldsymbol{m}^\top \boldsymbol{m})\right)$$

$$\Lambda(\boldsymbol{z}) \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \lambda$$

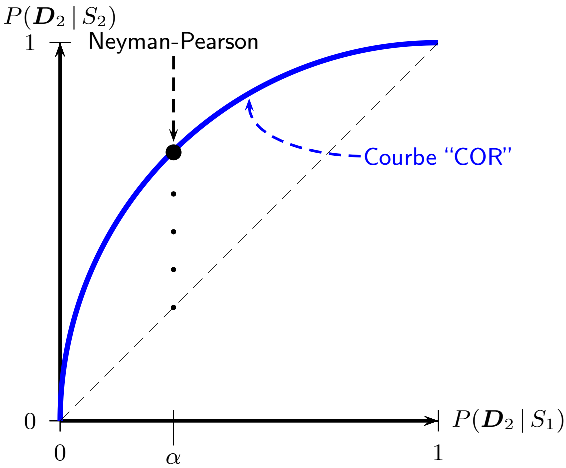
$$\overline{\boldsymbol{z}} = \frac{1}{N}\sum_{n=1}^{N} z_n \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \frac{\sigma_v^2}{Nm}\ln\lambda + \frac{m}{2} = z_s$$

$\overline{\boldsymbol{z}}$ est une statistique suffisante.

Caractéristique opérationnelle d'un récepteur

Definition (Caractéristique Opérationnelle d'un Récepteur (COR))

C'est la représentation de $P(\boldsymbol{D}_2 \mid S_2)$ en fonction de $P(\boldsymbol{D}_2 \mid S_1)$



Caractéristique opérationnelle d'un récepteur

Propriétés des courbes COR :

- $P(\boldsymbol{D}_2 \mid S_j)$ sont des fonctions décroissantes de λ, avec $P(\boldsymbol{D}_2 \mid S_j) = \int_\lambda^\infty p_\Lambda(\lambda' \mid S_j)\, d\lambda'$.
- La courbe COR passe par les valeurs $(0,0)$ et $(1,1)$
- $P(\boldsymbol{D}_2 \mid S_2) > P(\boldsymbol{D}_2 \mid S_1)$
- $\dfrac{dP_D}{dP_F} = \lambda$
- La concavité est toujours négative

Caractéristique opérationnelle d'un récepteur

Application : communication binaire synchrone

- Règle de décision :

$$\overline{z} \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} z_s$$

$\overline{z} = \frac{1}{N}\sum_{n=1}^{N} z_n$ statistique suffisante (suit une loi $\mathcal{N}(m, \sigma_v^2/N)$)

- Probabilité de décision conditionnelle (à partir de $\overline{\boldsymbol{z}}$)

$$P(\boldsymbol{D}_2 \mid S_1) = 1 - \mathrm{erf}\left(\frac{z_s}{\sigma_v/\sqrt{N}}\right) \qquad P(\boldsymbol{D}_2 \mid S_2) = 1 - \mathrm{erf}\left(\frac{z_s - m}{\sigma_v/\sqrt{N}}\right)$$

- Courbes COR :

FIGURE

Caractéristique opérationnelle d'un récepteur

Critère minimax :

FIGURE

Cas particulier : $C_{ij} = 1 - \delta_{ij}$

FIGURE

Critère de Neyman-Pearson :
Les courbes COR permettent de choisir $P(\boldsymbol{D}_2 \mid S_1)$

Caractéristique opérationnelle d'un récepteur

Le problème Gaussien :

Mesures multiples : hypothèses binaires :

$S_1 : p(\boldsymbol{z} \mid S_1) \Rightarrow \mathcal{N}(\boldsymbol{m}_1, \Sigma_1)$ $S_2 : p(\boldsymbol{z} \mid S_2) \Rightarrow \mathcal{N}(\boldsymbol{m}_2, \Sigma_2)$

- Identification de signaux dans un bruit Gaussien coloré
- Tri de pièces mécaniques (plusieurs paramètres)

$$\Lambda(\boldsymbol{z}) = \sqrt{\frac{\det \Sigma_1}{\det \Sigma_2}} \exp\left(-\frac{1}{2}\left((\boldsymbol{z}-\boldsymbol{m}_2)^\top \Sigma_2^{-1}(\boldsymbol{z}-\boldsymbol{m}_2) - (\boldsymbol{z}-\boldsymbol{m}_1)^\top \Sigma_1^{-1}(\boldsymbol{z}-\boldsymbol{m}_1)\right)\right)$$

Règle de décision :

$$(\boldsymbol{z}-\boldsymbol{m}_1)^\top \Sigma_1^{-1}(\boldsymbol{z}-\boldsymbol{m}_1) - (\boldsymbol{z}-\boldsymbol{m}_2)^\top \Sigma_2^{-1}(\boldsymbol{z}-\boldsymbol{m}_2) \underset{\boldsymbol{D}_1}{\overset{\boldsymbol{D}_2}{\gtrless}} \ln\left(\lambda \sqrt{\frac{\det \Sigma_1}{\det \Sigma_2}}\right)$$

Surface de décision : conique dans $\boldsymbol{Z}$

Cas particuliers :

- Bruit identique : $\Sigma_1 = \Sigma_2$
- Mesures indépendantes : Σ_i identiques