Exercice : erreurs de fausse alarme et de non détection

On suppose que l'on dispose d'un paramètre x obtenu par un examen médical qui permet de détecter la présence (classe ω_1) ou l'absence (classe ω_2) d'une maladie. Une valeur de x faible est plutôt signe de la maladie tandis qu'une valeur de x élevée est plutôt signe de son absence. On considère que de ne pas détecter la maladie sur un patient atteint est plus grave que de diagnostiquer la maladie sur un patient sain.

En cas d'erreur de diagnostic, on distingue les probabilités suivantes :
P_{FA} =probabilité de fausse alarme = P(décider ω_1 , ω_2) et
P_{ND} =probabilité de non détection = P(décider ω_2 , ω_1)
Le traitement étant coûteux, il faut faire un compromis pour maintenir le taux de fausse alarme à un niveau raisonnable.

1) Pour modéliser le problème, on introduit deux coûts :
$\lambda(\omega_2 \mid \omega_1)$: coût de non détection
$\lambda(\omega_1 \mid \omega_2)$: coût de fausse alarme

Donner l'expression du risque conditionnel associé à chaque décision. Donner la règle de décision bayésienne en utilisant le rapport de vraisemblance $\dfrac{p(x|\omega_1)}{p(x|\omega_2)}$

Lequel des coûts $\lambda(\omega_1|\omega_2)$ et $\lambda(\omega_2|\omega_1)$ doit être supérieur à l'autre pour ce problème ?

2) On fixe les valeurs des coûts $\lambda(\omega_i|\omega_j)$. On exprime directement la règle de décision bayésienne en fonction du paramètre x et d'un seuil x_0 tel que :

si $x < x_0$ on décide ω_1

si $x > x_0$ on décide ω_2

Donner les expressions des probabilités de fausse alarme (P_{FA}) et de non détection (P_{ND}) en fonction des lois de densités conditionnelles $p(x|\omega_i)$ $i = 1, 2$ et du seuil x_0.

3) On considère maintenant que les lois sont gaussiennes :

$p(x|\omega_1) \approx \mathcal{N}(m_1, \sigma^2)$

$p(x|\omega_2) \approx \mathcal{N}(m_2, \sigma^2)$

En déduire l'expression du point frontière x_0 en fonction de m_1, m_2, σ, des probabilités a priori et des coûts λ. Donner les expressions de P_{FA}, P_{ND} en fonction de σ, x_0 et de la fonction d'erreur complémentaire[1] $erfc$.

Application numérique : $\sigma = 0.5$, prendre un des coûts égal à 1 , l'autre à 4 (cf. question 2), $m_1 = 1, m_2 = 2$, $P(\omega_1) = 0.1; P(\omega_2) = 0.9$; et calculer x_0.

4) On pose $e_{ND} = P_{ND}/P(\omega_1)$ et $e_{FA} = P_{FA}/P(\omega_2)$, probabilités conditionelles d'erreur, et on représente $e_D = 1 - e_{ND}$ en fonction de e_{FA} pour différentes valeurs du point frontière x_0. La courbe obtenue est appelée courbe COR.

On cherche généralement à obtenir une probabilité de non détection faible en acceptant un taux de fausse alarme raisonnable. Dans quelle partie de la courbe COR se situe x_0 ? Comment varie x_0 le long de la courbe COR ?
Peut-on rendre aussi faibles que possible à la fois P_{ND} et P_{FA} ?

1. $erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-t^2} dt$

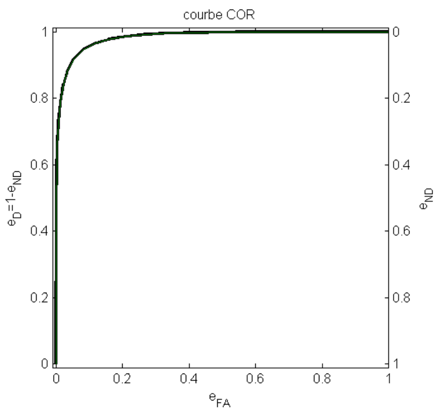

FIGURE 1 – Courbe COR.