

Théorie de la Décision

Tests d'hypothèses. Tests de Bayes. Rapport de Vraisemblance. Statistique suffisante. Tests de Neyman-Pearson. ROC. Tests simples et composés. Tests UMP. Rapport de Vraisemblance Généralisé. Détection d'un signal dans un bruit: le filtre adapté.

3.1 Tests d'hypothèses.

On étudie dans ce chapitre la conception de mécanismes de décision. La figure suivante illustre les problèmes de teste d'hypothèses, pour le cas de décision entre deux alternatives possibles, désignées par H_1 et H_0 .

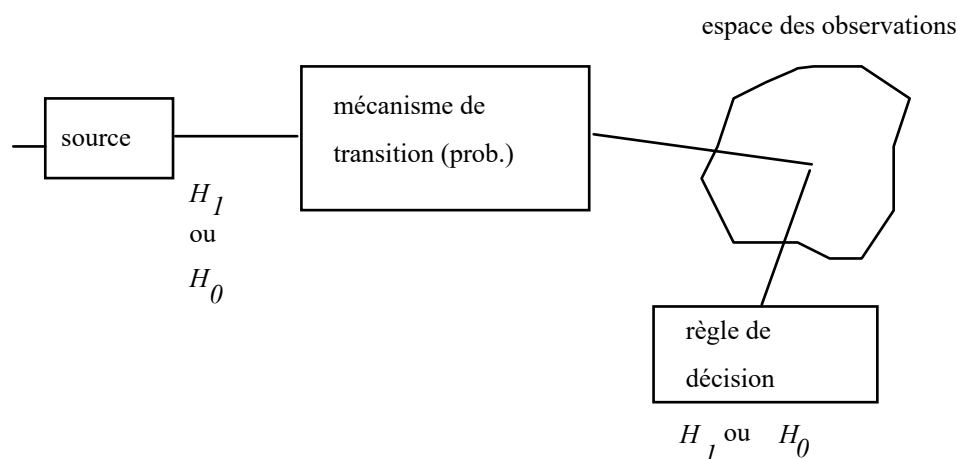


Figure 3.1: Décision entre deux hypothèses alternatives.

Exemples:

3-1 *communications numériques.*

Dans les systèmes de communication numérique, un signal analogique $s(t)$ souffre plusieurs transformations avant d'être effectivement envoyé dans le canal de communication. Il est d'abord *échantillonné* (à une vitesse au moins égale à la fréquence de Nyquist),

$$s_k = s(t - k / f_e),$$

ensuite chaque échantillon est *quantifié*,

$$s_k \rightarrow c_i = q(s_k),$$

et finalement *codé* comme une séquence de symboles choisis dans un alphabet fini (binaire, par exemple) :

$$c_i \rightarrow 0100\dots 1.$$

Le signal original est ainsi transformé dans une série de symboles (binaires dans le cas de l'exemple).

Ce sont ces symboles qui modulent une porteuse (en phase, en amplitude ou en fréquence) pour générer le signal transmis dans le canal de communication. Ainsi, dans le cas où le système de modulation n'a pas de mémoire (absence d'interférence inter-symbolique), le signal $m(t)$ émis pendant l'intervalle de temps correspondant à un symbole est complètement déterminé par le symbole transmis. Par exemple, pour les systèmes de modulation d'amplitude et avec des symboles binaires

$$0 \text{ a } m(t) = -A \cos(\omega_0 t), t \in T_b$$

$$1 \text{ a } m(t) = A \cos(\omega_0 t), t \in T_b$$

Le signal reçu à la sortie du canal de communication est, en général, une version bruité et déformée de $m(t)$:

$$y(t) = H[m(t)] + b(t)$$

où $H[\cdot]$ est l'opérateur de transmission du canal, en général, connu.

Le système de réception doit, en chaque intervalle T_b , faire correspondre au signal reçu, soit un symbole 1, soit un symbole 0. Il doit donc résoudre le problème de décision suivant. Etant donné $y(t)$, décider laquelle des deux hypothèses suivantes est vraie:

$$H_0: y(t) = H[-A \cos(\omega_0 t)] + b(t), t \in T_b$$

$$H_1: y(t) = H[A \cos(\omega_0 t)] + b(t), t \in T_b$$

En référence à la Figure 3.1, on doit, pour cet exemple, effectuer les correspondances suivantes:

- ⇒ *source*: système de codage, qui génère des symboles 0 ou 1;
- ⇒ *mécanisme de transition*: le canal, qui déforme le signal $m(t)$ et ajoute du bruit au message;
- ⇒ *mécanisme de décision*: Application qui fait correspondre à chaque signal observé un des deux symboles:

$$\{y(t) | y(t) = H[\pm A \cos(\omega_0 t)] + b(t), t \in T_b\} \rightarrow \{0,1\}$$

3-2 détection de cibles en radar

Les systèmes de radar utilisent une antenne pour détecter des cibles dans l'espace (en général 3D) autour de l'antenne. Pour chaque cible, on désire savoir sa *position*, sa *direction* et sa *vitesse*. On considère ici le cas simple où on ne détermine pas la vitesse de la cible. L'antenne du radar peut être dirigée de façon à émettre, à chaque instant, une impulsion $s(t)$ seulement dans un cône angulaire étroit autour de la direction θ . S'il existe une cible dans cette direction, le signal émis est réfléchi et reçu avec un temps de retard qui est proportionnel à la distance d entre l'antenne et la cible:

$$r(t) = \alpha s(t - \tau(d)) + b(t)$$

$$\tau(d) = 2 \frac{d}{c}$$

où α dépend de l'attitude de la cible, de ses propriétés de réflexion, de la distance, etc., et c est la vitesse de propagation des ondes électromagnétiques. Pour détecter la cible, le système doit donc résoudre un problème de décision entre les deux hypothèses suivantes:

$$H_0: r(t) = b(t),$$

$$H_1: r(t) = \alpha s(t - \tau) + b(t)$$

où la première hypothèse correspond à l'absence de cible dans la direction θ , et la deuxième à la présence d'une cible à une distance qui est déterminée par τ .

Nota: l'hypothèse H_1 de ce test binaire dépend des deux paramètres α et τ qui ne sont pas connus. Ce type de problèmes de décision, bien plus difficile à résoudre, en général, que les problèmes de décision simple (comme le cas de l'exemple précédant), est connu sous le nom de *tests d'hypothèses composées*.

3-3 reconnaissance de mots

La reconnaissance de mots parmi un dictionnaire est encore un problème de décision: étant donné un segment d'un signal de parole, qu'on admet correspondre à un mot unique, on doit décider de quel mot il s'agit. Ce problème très complexe peut être formalisé comme un problème de décision, avec autant d'hypothèses que de mots. Dans ce cas encore, chaque hypothèse (mot) est *composée*, puisque le signal de parole qui correspond à un mot donné varie avec la personne qui le prononce (homme/femme, région, âge, ...).

Dans tous les exemples présentés, le système de décision est défini par une application de l'espace des observations dans l'ensemble des hypothèses possibles. On désigne cette application par *règle de décision*. Elle détermine, dans l'espace des observations, une partition en sous-ensembles disjoints, chaque sous-ensemble correspondant aux observations qui sont associées à une même hypothèse.

Règle de décision \Leftrightarrow *partition de l'espace d'observations* en régions \mathbf{R}_i associées aux différentes hypothèses: $\mathbf{R}_i \leftrightarrow H_i$

Comme on doit associer une hypothèse à chaque observation possible,

$$\bigcup_i \mathbf{R}_i = \mathbf{S}.$$

Et, comme les hypothèses sont *alternatives*, c'est-à-dire, l'occurrence simultanée de deux hypothèses différentes est impossible, les sous-ensembles \mathbf{R}_i sont disjoints:

$$\mathbf{R}_i \cap \mathbf{R}_j = \emptyset, i \neq j.$$

La règle de décision est facilement décrite en fonction des régions \mathbf{R}_i :

$$\boxed{\mathbf{r} \in \mathbf{R}_i \Rightarrow H_i}$$

où \mathbf{r} représente les observations (qui peuvent être scalaires, vectorielles, où les échantillons d'un signal pris dans un intervalle de temps).

3.2 Tests de Bayes.

Dans cette section, on étudie l'approche Bayésienne aux problèmes de décision, qui est basée sur la connaissance, pour chaque hypothèse H_i , de la probabilité *a priori* pour que cette hypothèse se réalise,

$$P_i = \Pr\{\text{hypothèse } H_i\},$$

et qui associe, à chaque comportement possible du système de décision, un coût (équivalent à une pénalisation ou une récompense):

$$C_{ij} = \text{coût de décider } H_i \text{ quand } H_j \text{ est vraie.}$$

La figure suivante illustre la définition de ces quantités pour un **test binaire** (où on considère que seulement deux hypothèses sont possibles)

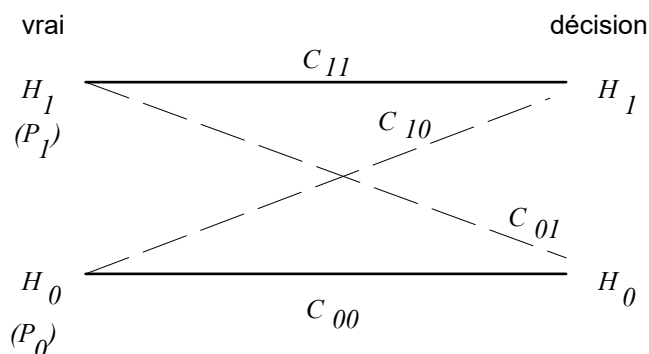


Figure 3.2

Dans la figure précédente, les lignes interrompues représentent les situations d'erreur.

Les tests de Bayes consistent à déterminer les régions de décision \mathbf{R}_1 et \mathbf{R}_0 de façon à *minimiser la valeur moyenne du coût*:

$$C = C_{00} \Pr\{H_0, H_0\} + C_{10} \Pr\{H_1, H_0\} + C_{01} \Pr\{H_0, H_{01}\} + C_{11} \Pr\{H_1, H_1\}$$

Chaque probabilité conjointe qui figure dans cette expression peut être écrite comme:

$$\Pr\{H_i, H_j\} = \Pr\{H_i | H_j\} P_j = \Pr\{r \in R_i | H_j\} P_j = P_j \int_{\mathbf{R}_i} p(\mathbf{r} | H_j) d\mathbf{r}$$

où l'on a exprimé la probabilité de décider H_i quand H_j est vraie comme la probabilité pour que les observations appartiennent à la région R_i où on décide H_i , étant donné que H_j est vraie.

Dans le cas de **tests binaires**, les deux régions de décision sont complémentaires, $R_i = \bar{R}_j$, et on peut donc écrire

$$\int_{R_i} p(\mathbf{r}) d\mathbf{r} = 1 - \int_{R_j} p(\mathbf{r}) d\mathbf{r}, \quad j \neq i; i, j = 0, 1,$$

Avec ce résultat, on peut exprimer le coût de Bayes C en fonction d'une seule région:

$$C = C_{11}P_1 + C_{10}P_0 + \int_{R_0} [P_1(C_{01} - C_{11})p(\mathbf{r}|H_1) - P_0(C_{10} - C_{00})p(\mathbf{r}|H_0)] d\mathbf{r}$$

Les deux premiers termes dans cette expression de dépendent pas des régions de décision, et constituent une pénalisation fixe. Pour minimiser C , il faut donc minimiser l'intégrale. Pour cela, on doit attribuer à R_0 tous les points de l'espace des observations pour lesquels l'intégrand est négatif, ce qui est équivalent à la règle de décision suivante:

$$\frac{p(\mathbf{r}|H_1)}{p(\mathbf{r}|H_0)} > \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} = \gamma,$$

où on a défini le seuil γ :

$$\gamma = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}.$$

On voit donc que le test de Bayes conduit à comparer le rapport entre les densités de probabilités conditionnelles à un seuil (γ). On appelle le rapport des densités conditionnelles dans l'équation précédente le *rapport de vraisemblance*

Notons que les tests de **probabilité d'erreur moyenne minimale** sont un cas particulier des tests de Bayes. Pour les obtenir, il suffit de prendre les valeurs de coût suivantes:

$$C_{00} = C_{11} = 0$$

$$C_{10} = C_{01} = 1$$

pour lesquels le test optimal est

$$\frac{p(\mathbf{r}|H_1)}{p(\mathbf{r}|H_0)} > \frac{P_0}{P_1}$$

Rapport de Vraisemblance.

Le **rapport de vraisemblance** (entre les densités de probabilité conditionnelles correspondant à chaque hypothèse), qui détermine les tests de Bayes, joue un rôle très important dans tous les problèmes de décision statistique et sera représenté par $\Lambda(\mathbf{r})$:

$$\Lambda(\mathbf{r}) = \frac{p(\mathbf{r}|H_1)}{p(\mathbf{r}|H_0)}.$$

En fait, si on considère que les densités conditionnelles résument notre connaissance sur chacune des hypothèses, ce rapport compare directement la vraisemblance des observations sous chacune des hypothèses.

Puisque l'application d'une fonction monotone n'affecte pas la validité d'une inégalité, le test de Bayes est équivalent au test suivant:

$$\begin{array}{c} H_1 \\ \ln \Lambda(\mathbf{r}) > \ln \gamma = \eta, \\ H_0 \end{array}$$

où on a défini le nouveau seuil η :

$$\eta = \ln \gamma.$$

Le logarithme du rapport de vraisemblance est, comme on le verra par la suite, particulièrement simple si les densités conditionnelles appartiennent à une famille exponentielle (dont les densités gaussiennes sont un cas particulier). On appelle $\ln \Lambda(\mathbf{r})$ le *rapport de vraisemblance logarithmique*.

Exemple 3-4.

Soit k une variable aléatoire de Poisson de paramètre λ :

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On observe N échantillons indépendants de k , et on veut décider entre les hypothèses

$$\begin{array}{l} H_1: \lambda = \lambda_1 \\ H_0: \lambda = \lambda_0 \end{array}$$

La densité de probabilité des N observations pour une valeur de λ générique est:

$$P(k_1, \dots, k_N; \lambda) = \frac{e^{-N\lambda}}{\prod_{n=1}^N k_n!} \lambda^{\sum_{n=1}^N k_n}$$

Le rapport de vraisemblance pour ce test est donc

$$\Lambda(k_1, \dots, k_N) = e^{-N(\lambda_1 - \lambda_0)} \left(\frac{\lambda_1}{\lambda_0} \right)^{\sum k_n}$$

3.3 Statistique suffisante.

On appelle *statistique* une application de l'espace des observations dans un autre espace, en général de dimension plus petite que celle de l'espace des observations. Pour des problèmes où les observations prennent valeurs dans un espace de dimension élevée (ou même infinie, comme c'est le cas du problème de communications binaires), il est souvent pratique de formuler le problème de décision en considérant une *statistique* qui est obtenue à partir des observations, au lieu de les traiter directement.

La notion de *statistique suffisante* établit les conditions dans lesquelles on peut faire cette compression de données sans perte d'information. Par définition, $m(\mathbf{r})$ est une **statistique suffisante** si on peut **factoriser** les densités conditionnelles pour chaque hypothèse de la façon suivante:

$$p(\mathbf{r}|H_i) = f_i(m(\mathbf{r}))g(\mathbf{r})$$

où la fonction $g(\mathbf{r})$ dépend des observations mais pas de l'hypothèse i . Dans ce cas, le rapport de vraisemblance dépend des observations \mathbf{r} uniquement à travers la statistique suffisante:

$$\Lambda(\mathbf{r}) = \frac{p(\mathbf{r}|H_1)}{p(\mathbf{r}|H_0)} = \frac{f_1(m(\mathbf{r}))}{f_0(m(\mathbf{r}))}.$$

Exemple 3-5.

Considérons le problème de décision binaire suivant. On effectue N observations indépendantes, r_1, r_2, \dots, r_N , d'une variable aléatoire gaussienne de covariance σ^2 . On sait que la valeur moyenne de la variable, m , peut prendre seulement deux valeurs avec égale probabilité :

$$H_0: m = 0$$

$$H_1: m = a$$

où a est une constante connue. On désire déterminer laquelle des deux valeurs de m correspond aux observations effectuées.

Pour résoudre ce problème on doit d'abord déterminer la fonction de densité des observations sous chaque hypothèse. L'indépendance des observations permet d'écrire:

$$\begin{aligned} p(\mathbf{r}|H_0) &= p(r_1, r_2, \dots, r_N | H_0) = \prod_{i=1}^N p(r_i | H_0) \\ &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(r_i)^2\right\} \\ &= \frac{1}{\sqrt{(2\pi)^N \sigma^N}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^N r_i^2\right\} \end{aligned}$$

Et de la même façon, pour l'autre hypothèse,

$$p(\mathbf{r}|H_1) = \frac{1}{\sqrt{(2\pi)^N \sigma^N}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^N (r_i - m)^2\right\}$$

Le rapport de vraisemblance est le quotient de ces deux densités:

$$\begin{aligned} \Lambda(\mathbf{r}) &= \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^N (r_i - m)^2 - \sum_{i=1}^N r_i^2 \right]\right\} \\ &= \exp\left\{-\frac{1}{2\sigma^2} \left[Nm^2 - 2m \sum_{i=1}^N r_i \right]\right\} \end{aligned}$$

On vérifie donc que la moyenne des observations, $\frac{1}{N} \sum_{i=1}^N r_i$, est une statistique suffisante pour ce problème:

$$\Lambda(\mathbf{r}) = \exp\left\{-\frac{Nm^2}{2\sigma^2}\right\} \cdot \exp\left\{\frac{Nm}{\sigma^2} \frac{1}{N} \sum_{i=1}^N r_i\right\}.$$

Ceci veut dire que le test optimal peut être construit directement à partir de l'analyse de la moyenne, et que, donc, il n'est pas nécessaire de mémoriser toutes les valeurs observées.

Exercice 3-1:

Vérifier que la somme

$$\sum_{n=1}^N k_n$$

de l'exemple 3-4 sur le test sur le paramètre de la variable de Poisson est une statistique suffisante.

3.4 Tests de Neyman-Pearson.

Pour construire le test de Bayes, c'est-à-dire, le coût moyen d'une décision, il faut connaître les probabilités *a priori*, P_i , qui déterminent la valeur du *seuil* auquel le rapport de vraisemblance est comparé. Pour beaucoup d'applications, ces valeurs ne sont pas connues, et on ne peut pas, en conséquence, appliquer l'approche Bayésienne, où encore, même si elles sont connues, le

critère ajusté au problème n'est pas obtenu par des considérations de ce qui se passe pour tout l'ensemble de situations possibles. Les tests de Neyman-Pearson constituent, dans ces cas, une approche alternative.

Au lieu de considérer l'occurrence de chaque hypothèse comme un phénomène aléatoire, et d'optimiser la performance moyenne, les tests de Neyman-Pearson agissent directement sur les mesures de performance suivantes:

$$P_D = \Pr\{\text{décider } H_1 | H_1\}$$

$$P_F = \Pr\{\text{décider } H_1 | H_0\}$$

$$P_M = \Pr\{\text{décider } H_0 | H_1\}$$

P_D est appelée la **probabilité de détection**, P_F la **probabilité de fausse alarme**, et P_M est la **probabilité de non-détection** ("miss"). Cette terminologie est dérivée des systèmes radar, où l'hypothèse H_1 est normalement associée à la présence d'une cible, et l'hypothèse H_0 à l'absence de cible.

On veut, usuellement, avoir une valeur de P_D le plus grande possible, et, en même temps, une valeur de P_F la plus petite possible. Cependant, ces deux contraintes ne sont pas indépendantes, et sont même contradictoires: pour augmenter P_D on doit augmenter R_1 , et pour minimiser P_F on doit diminuer R_1 . Les tests de Neyman-Pearson correspondent à *maximiser la probabilité de détection*, avec une *contrainte fixe sur la probabilité de fausse alarme* :

$$\boxed{\max P_D \quad \text{sous la contrainte } P_F = \alpha' \leq \alpha}$$

La solution est obtenue à travers l'utilisation des multiplicateurs de Lagrange, et est encore une fois, donnée par le rapport de vraisemblance:

$$\boxed{\begin{array}{c} H_1 \\ \Lambda(\mathbf{r}) > \lambda \\ \Lambda(\mathbf{r}) < \lambda \\ H_0 \end{array}}$$

où maintenant le seuil λ est déterminé de façon à vérifier la contrainte:

$$P_F = \int_{\lambda}^{\infty} p(\Lambda | H_0) d\Lambda = \alpha .$$

Pour arriver à cette solution on minimise la fonction F , qui est obtenue en ajoutant à la fonction à minimiser (la probabilité de non-détection), le multiplicateur de Lagrange fois la contrainte (si la contrainte est vérifiée, on aura minimisé la fonction originelle):

$$\begin{aligned} F &= P_M + \lambda [P_F - \alpha'] \\ &= \int_{R_0} p(\mathbf{r} | H_1) d\mathbf{r} + \lambda \left(\int_{R_1} p(\mathbf{r} | H_0) d\mathbf{r} - \alpha' \right) \\ &= \lambda(1 - \alpha') + \int_{R_0} (p(\mathbf{r} | H_1) - \lambda p(\mathbf{r} | H_0)) d\mathbf{r} \end{aligned}$$

Il résulte immédiatement de cette expression que la règle de décision optimale est encore la comparaison du rapport de vraisemblance avec un seuil (λ), qui doit être déterminé de façon à

satisfaire la contrainte. Ceci est obtenu en définissant R_0 comme l'ensemble des points \mathbf{r} où la fonction intégrée est négative.

Exercice 3-2:

Dans un problème de décision binaire, une des deux hypothèses suivantes est vraie:

$$\begin{aligned} H_1: r = s + n \\ H_0: r = n \end{aligned}$$

où les variables aléatoires s et n sont indépendantes,

$$p(\mathbf{s}) = \begin{cases} a e^{-as}, & \mathbf{s} \geq 0 \\ 0, & \mathbf{s} < 0 \end{cases} \quad p(\mathbf{n}) = \begin{cases} b e^{-bn}, & \mathbf{n} \geq 0 \\ 0, & \mathbf{n} < 0 \end{cases}$$

1. Montrer que le rapport de vraisemblance est proportionnel à l'observation:

$$\Lambda(\mathbf{r}) = k \mathbf{r}.$$

2. Déterminer le seuil pour le test de Bayes optimal, en fonction des coûts et des probabilités *a priori*.

3. Déterminer le seuil du test de Neyman-Pearson en fonction de la probabilité de fausse alarme.

3.5 Tests Minimax.

Les tests **minimax** sont une autre façon de résoudre des problèmes de décision quand on ne connaît pas les probabilités *a priori*, et correspondent à choisir *le test qui a la meilleure performance dans le cas le plus défavorable*.

Pour dériver ce test, on commencera par analyser l'influence du choix du seuil sur le risque de Bayes. Admettons alors qu'une certaine valeur du seuil de décision est fixée, soit η_0 . Une fois que le seuil est fixé, la performance du test, obtenu en comparant le rapport de vraisemblance avec lui, est déterminée:

$$\begin{aligned} P_F(\eta_0) &= \Pr\{\Lambda(\mathbf{r}) > \eta_0 | H_0\} \\ P_M(\eta_0) &= \Pr\{\Lambda(\mathbf{r}) < \eta_0 | H_1\} = (1 - P_D(\eta_0)) \end{aligned}$$

Le coût de Bayes est donc

$$C(\eta_0) = C_{11}P_1P_D(\eta_0) + C_{01}P_1P_M(\eta_0) + C_{00}P_0(1 - P_F(\eta_0)) + C_{10}P_0P_F(\eta_0).$$

Si on utilise dans cette équation les relations

$$P_D(\eta_0) + P_M(\eta_0) = 1,$$

et

$$P_0 + P_1 = 1$$

on obtient une expression qui dépend uniquement de P_1 :

$$\begin{aligned} C(\eta_0) &= C_{00}(1 - P_F(\eta_0)) + C_{10}P_F(\eta_0) + \\ &P_1[(C_{11} - C_{00}) + (C_{01} - C_{11})P_M(\eta_0) - (C_{10} - C_{00})P_F(\eta_0)] \end{aligned}$$

De cette équation, on peut conclure que le coût de Bayes associé à ce test -- pour cette valeur fixe du seuil -- est une fonction linéaire de P_1 .

Ce test est le test optimal si la relation suivante est vérifiée:

$$\eta_0 = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} = \frac{(1 - P_1)(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

ce qui est vrai pour

$$P_1^0 = \frac{1}{\eta_0(C_{01} - C_{11}) / (C_{10} - C_{00}) + 1}.$$

Pour toutes les autres valeurs de $P_1 \neq P_1^0$, le test optimal (qui correspond à un seuil différent) aura un coût inférieur. La figure suivante montre la variation du coût du test optimal, et du coût pour un seuil déterminé, en fonction de la valeur de la probabilité *a priori* :

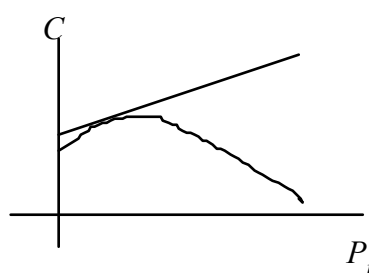


Figure 3.3: Risque de Bayes pour le test fixe et pour le test optimal.

On déduit donc que le coût pour un seuil fixe est tangent à la courbe qui décrit le coût du test optimal pour chaque valeur possible de P_1 .

Les tests **minimax** correspondent à choisir le test optimal pour la situation où le coût est maximale, ce qui est obtenu en prenant le test qui a un *coût constant pour toutes les valeurs possibles de la probabilité a priori*. Pour obtenir ce test, on doit donc choisir un seuil qui conduit à des valeurs de P_M, P_F tels que

$$\boxed{C_{11} - C_{00} + (C_{01} - C_{11})P_M - (C_{10} - C_{00})P_F = 0}.$$

Pour des coûts de Bayes qui sont différents de zéro seulement pour les situations d'erreur, le test minimax est défini par

$$\frac{P_M}{P_F} = \frac{C_{10}}{C_{01}}.$$

Exemple 3-6.

On sait qu'une pièce est soit standard (PF), soit à deux cotés face (FF). On doit décider sur le type de la pièce avec un seul tirage. Le coût d'une décision fausse est 1 et celui d'une décision correcte est 0. La décision est prise selon la règle suivante:

- ⇒ si on observe P on décide (PF)
- ⇒ si on observe F on prend une décision aléatoirement: (PF) avec probabilité ρ , et (FF) avec une probabilité $(1 - \rho)$.

Montrer que le coût maximal est minimisé pour $\rho = \frac{1}{3}$.

La valeur moyenne du coût de Bayes est, dans ce cas

$$C = \rho P_{FF} + (1 - \rho)(1 - P_{FF}) \frac{1}{2}$$

où on a admis que la pièce standard est équilibrée (égale probabilité pour face et pile). Cette expression est linéaire en P_{FF} . Le maximum va donc se trouver sur un des limites, $P_{FF} = 0$ ou $P_{FF} = 1$, selon le signe du coefficient qui multiplie P_{FF} et le coût est constant si ce coefficient est nul.

La valeur maximale du coût est obtenue pour

- $P_{FF} = 1$ si $\rho > \frac{1}{3}$, et dans ce cas $C_{\max} = \rho > \frac{1}{3}$
- $P_{FF} = 0$ si $\rho < \frac{1}{3}$, et dans ce cas $C_{\max} = \frac{1 - \rho}{2} > \frac{1}{3}$
- indépendante de P_{FF} si $\rho = \frac{1}{3}$, est dans ce cas $C = \frac{1}{3}$.

Cet exemple confirme l'analyse que nous avons fait précédemment, c'est à dire, que le test minimax (qui minimise le pire cas) est obtenu quand le risque devient indépendant de la valeur de la probabilité *a priori*.

3.6 Receiver Operating Characteristic (ROC)

La performance des tests d'hypothèses est jugée à travers la paire (probabilité de détection, probabilité de fausse alarme), ou, dans le cas des tests Bayesiens, par le coût de Bayes qui est aussi une fonction de (P_D, P_F) .

On introduit ici quelque nomenclature usuelle en statistique.

On appelle *taille d'un test binaire* à la probabilité avec laquelle on choisit l'hypothèse H_1 quand l'hypothèse H_0 est vraie:

$$\alpha = \Pr\{\text{décider } H_1 \mid H_0 \text{ est vraie}\} = \Pr[\Lambda(r) > \eta \mid H_0] = P_{FA},$$

c'est-à-dire, la taille d'un test est la probabilité de fausse alarme, P_{FA} . On appelle encore à α la *probabilité d'erreurs du type 1*.

La *puissance* d'un test est la probabilité de décider correctement l'hypothèse H_1 , c'est-à-dire, la probabilité de détection, P_D :

$$\beta = \Pr\{\text{décider } H_1 \mid H_1 \text{ est vraie}\} = \Pr[\Lambda(r) > \eta \mid H_1]$$

La probabilité *d'erreurs du type 2* est la probabilité d'une hypothèse erronée quand H_1 est vraie, et est égale à 1 moins la puissance du test.

On désire, naturellement, avoir un *test de petite taille et le plus puissant possible*. Comme on l'a étudié, le test plus puissant d'une taille donnée est un test de vraisemblance : le test de Neyman-Pearson.

Un test est dit *non-biaisé* si sa puissance est plus grande que sa taille. Cela veut dire qu'en moyenne on décide plus probablement H_1 quand H_1 est vraie que quand c'est hypothèse H_0 est vraie.

On appelle ROC (*Receiver Operating Characteristic*) d'un test la courbe de l'évolution de la puissance du test (probabilité de détection) en fonction de sa taille (probabilité de fausse alarme). En effet, ces deux probabilités sont conjointement déterminées par le choix du seuil, γ .

Ces courbes possèdent plusieurs propriétés:

1. La courbe $P_D(P_F)$ est *convexe*

$$P_D(\alpha P_{F_1} + (1 - \alpha) P_{F_2}) \geq \alpha P_D(P_{F_1}) + (1 - \alpha) P_D(P_{F_2}).$$

2. Tous les tests de vraisemblance ont un ROC qui sont au-dessus de la ligne $P_D = P_F$ (cela veut dire qu'ils son *non-biaisés*).

3. La pente de la courbe en un point particulier est égale à la valeur du seuil auquel on doit comparer le rapport de vraisemblance pour obtenir les probabilités de détection et de fausse alarme qui lui correspondent.

Dem.

$$P_D = \int_{\eta}^{\infty} p(\Lambda \mid H_1) d\Lambda$$

$$P_F = \int_{\eta}^{\infty} p(\Lambda \mid H_0) d\Lambda$$

Si on dérive les deux équations et que l'on prend leur quotient:

$$\frac{dP_D}{d\eta} = -p(\Lambda|H_1), \quad \frac{dP_F}{d\eta} = -p(\Lambda|H_0) \Rightarrow \frac{dP_D}{dP_F} = \frac{dP_D/d\eta}{-p(\Lambda|H_0)}$$

Mais,

$$P_D(\eta) = \int_{R_1} p(\mathbf{r}|H_1) d\mathbf{r} = \int_{R_1} \Lambda(\mathbf{r}) p(\mathbf{r}|H_0) d\mathbf{r}$$

où R_1 est la région de décision pour l'hypothèse 1, c.a.d., où $\Lambda(\mathbf{r}) > \eta$. Alors, par un changement de variable, on peut écrire la dernière intégrale directement comme une moyenne qui utilise la densité de probabilité du rapport de vraisemblance:

$$P_D(\eta) = \int_{R_1} \Lambda(\mathbf{r}) p(\mathbf{r}|H_0) d\mathbf{r} = \int_{\eta}^{\infty} \Lambda p(\Lambda|H_0) d\Lambda \Rightarrow \frac{dP_D}{d\eta} = -\eta p(\Lambda|H_0)$$

Si on utilise ce résultat dans l'expression de $\frac{dP_D}{dP_F}$, il résulte facilement

$$\frac{dP_D}{dP_F} = \eta.$$

De cette démonstration, on peut déduire que la courbe du ROC est non-décroissante, car η est toujours positif.

4. Le test minimax est déterminé par l'intersection de la ligne

$$C_{11} - C_{00} + (C_{01} - C_{11})(1 - P_D) - (C_{10} - C_{00})P_F = 0$$

et le ROC.

3.7 Tests simples et composés

Les tests binaires étudiés sont des tests **simples**. Ce nom dérive du fait que sous chaque hypothèse la densité de probabilité des observations est complètement connue, ce qui permet d'écrire le rapport de vraisemblance.

Cependant, dans de nombreuses situations, la densité de probabilité sous chaque hypothèse dépend de paramètres qui ne sont pas connus, c'est-à-dire, il existe plusieurs densités de probabilité qui sont des candidates possibles pour décrire les observations sous au moins une des hypothèses.

Exemple 3-7

$$H_1: r = s + n$$

$$H_0: r = n$$

où n est une variable aléatoire de densité $p_n(N)$ connue et s est une constante déterministe inconnue.

Si on admet connu s , les densités associées à chaque hypothèse sont dans ce cas:

$$P_r(R|H_1, s) = p_n(R - s)$$

$$p_r(R|H_0) = p_n(R)$$

Cependant, s n'est pas connu, et donc on a un nombre infini de possibles densités de probabilité qui peuvent décrire les observations sous hypothèse H_1 . On dit alors que H_1 est une hypothèse composée, car on peut lui associer plusieurs densités, selon la valeur de s .

Dans ces cas on appelle le paramètre inconnu *paramètre de nuisance*, ce qui traduit le fait qu'on est intéressé seulement à prendre une décision et pas à connaître la valeur de s .

On peut formaliser généralement ces situations en introduisant un vecteur de paramètres de nuisance, θ , dont dépendent les densités des observations sous chaque hypothèse:

$$H_1: p(R|H_1, \theta)$$

$$H_0: p(R|H_0, \theta)$$

Dans ces situations, les définitions précédentes de *taille* et *puissance* sont :

$$\alpha = \sup_{\theta} \Pr\{H_1|H_0, \theta\}$$

$$\beta(\theta) = \Pr\{H_1|H_1, \theta\}$$

c'est-à-dire, la taille est, par définition, la valeur supérieure de la taille du test pour toutes les valeurs possibles de θ -- c'est-à-dire le pire cas possible -- et la puissance est une fonction du vecteur de paramètres inconnus.

Pour les tests composés, on désire déterminer un test qui soit optimal (dans le sens de Neyman-Pearson) pour toutes les valeurs possibles du paramètre inconnu, c'est-à-dire que, parmi tous les tests avec une taille égale ou inférieure à α , soit le plus puissant pour toutes les valeurs possibles du paramètre inconnu. On appelle à un test avec cette propriété le test *uniformément le plus puissant* (UMP).

Si on prend n'importe quel test, sa performance est naturellement limitée par la performance du test qui est défini pour la vraie valeur du paramètre inconnu, c'est-à-dire, il aura toujours une performance inférieure ou égale au test optimal pour la vraie valeur du paramètre. La représentation de la probabilité de détection du test optimal en fonction de la valeur inconnue du paramètre, pour différentes valeurs de probabilité de fausse alarme, est appelée la **fonction de puissance**. Cette courbe établit une limite à la performance de tous les tests.

On dit qu'un test est **uniformément le plus puissant** si, pour toutes valeurs de θ , la probabilité de détection qui correspond à une probabilité de fausse alarme (taille) fixée est maximale. La condition nécessaire et suffisante pour l'existence d'un test uniformément le plus puissant est la suivante: le rapport de vraisemblance et le seuil doivent pouvoir être déterminés sans connaître θ .

Exemple 3-8:

Soit k le nombre d'événements dans une expérience, décrit par une loi de Poisson de paramètre θ

$$P_{\theta}(k) = \frac{e^{-\theta}}{k!} \theta^k$$

On va construire le test uniformément le plus puissant pour tester les hypothèses composées

$$H_1: \theta > \theta_0$$

$$H_0: \theta \leq \theta_0$$

Le rapport de vraisemblance pour tester l'hypothèse composée $H_1: \theta > \theta_0$ contre l'hypothèse simple $H_0: \theta = \theta_0$ est

$$\frac{P(k|\theta = \theta_1 < \theta_0)}{P(k|\theta = \theta_0)} = \left(\frac{\theta_1}{\theta_0}\right)^k e^{-(\theta_1 - \theta_0)}.$$

Comme le rapport de vraisemblance est monotone en k , on peut écrire le test directement en fonction de l'observation:

$$\begin{array}{c} H_1 \\ k > \\ < \\ H_0 \end{array} \gamma = \ln \eta \frac{\theta_1 - \theta_0}{\ln(\theta_1 - \theta_0)}$$

Le seuil γ peut être déterminé de façon à ce que la taille du test soit α :

$$\alpha = \sum_{k > \gamma} e^{-\theta_0} \frac{\theta_0^k}{k!} = \Pr[k > \gamma | H_0].$$

Cette taille donne bien le pire cas possible, car pour $\theta < \theta_0$ la taille sera plus petite que α (vérifier):

$$\alpha = \sum_{k>\gamma} e^{-\theta_0} \frac{\theta_0^k}{k!} = \sup_{\theta \leq \theta_0} \sum_{k>\gamma} e^{-\theta} \frac{\theta^k}{k!}.$$

Le test est bien le plus puissant - celui qui a une plus grande probabilité de détection pour une valeur donnée de θ_1 - parmi tous les tests avec une taille inférieure ou égale à α .

3.8 Rapport de Vraisemblance Généralisé.

Le rapport de vraisemblance généralisé est une approche à la résolution de problèmes de décision quand on ne peut pas construire un test uniformément le plus puissant.

Cette procédure consiste à associer à chaque hypothèse H_i une densité unique, parmi toutes les densités possibles, en prenant celle qui maximise la vraisemblance des données, et à l'utiliser pour construire un "rapport de vraisemblance réduit":

$$\Lambda_g(\mathbf{r}) = \frac{\max_{\theta} p(\mathbf{r}|H_1, \theta)}{\max_{\theta} p(\mathbf{r}|H_0, \theta)} \begin{matrix} > & H_1 \\ < & H_0 \end{matrix} \gamma$$

Exercice 3-3:

On dispose de N mesures d'une même quantité inconnue h . On ne sait pas avec lequel de deux instruments, I_1 ou I_2 , les mesures ont été prises. On sait que le premier instrument n'introduit pas de biais, et a des erreurs de mesures gaussiennes, de variance σ_1^2 , et que le deuxième instrument introduit un biais connu b , et des erreurs de mesure de variance $\sigma_2^2 > \sigma_1^2$. Déterminer le test de Maximum de Vraisemblance pour identifier l'instrument utilisé.

3.9 Détection d'un signal dans un bruit: le filtre adapté

Considérons le problème de décision suivant:

$$\begin{aligned} H_1: r_k &= s_k + n_k, & k = 1, 2, \dots, K \\ H_2: r_k &= n_k, & k = 1, 2, \dots, K \end{aligned}$$

où

$\Rightarrow \{n_k\}_{k=1}^K$ sont des variables aléatoires i.i.d. $N(0, \sigma_n^2)$

$\Rightarrow \{s_k\}_{k=1}^K$ sont des valeurs connues qui vérifient la condition de normalisation

$$\sum_{k=1}^K s_k^2 = E.$$

Ce modèle représente le problème de la détection d'un signal connu, d'énergie E , dans du bruit blanc Gaussien.

On détermine le test qui **minimise la probabilité d'erreur**, en admettant connues les probabilités *a priori* P_0 et P_1 - situation usuelle en communications.

Pour écrire le rapport de vraisemblance, on doit d'abord déterminer la densité des observations sous chaque hypothèse. L'indépendance des observations implique

$$p_{r_1, \dots, r_k}(R_1, \dots, R_k | H_1) = \prod_{k=1}^K p_{r_k}(R_k | H_1)$$

Mais

$$p_{r_k}(R_k | H_1) = N(s_k, \sigma_n^2) = \frac{1}{\sqrt{2\pi\sigma_n}} e^{-\frac{1}{2\sigma_n^2}(R_k - s_k)^2}$$

et donc

$$p_{r_1, \dots, r_k}(R_1, \dots, R_k | H_1) = \frac{1}{(\sqrt{2\pi\sigma_n})^K} e^{-\frac{1}{2\sigma_n^2} \sum_{k=1}^K (R_k - s_k)^2}$$

Pour l'autre hypothèse

$$p_{r_1, \dots, r_k}(R_1, \dots, R_k | H_0) = \frac{1}{(\sqrt{2\pi\sigma_n})^K} e^{-\frac{1}{2\sigma_n^2} \sum_{k=1}^K R_k^2}$$

Le rapport de vraisemblance est donc

$$\Lambda(R_1, \dots, R_k) = e^{-\frac{1}{2\sigma_n^2} \sum_{k=1}^K (s_k^2 - 2s_k R_k)}$$

Le test optimal (qui minimise la probabilité d'erreur) est

$$e^{-\frac{1}{2\sigma_n^2} \sum_{k=1}^K (s_k^2 - 2s_k R_k)} \begin{matrix} H_1 \\ > \frac{P_0}{P_1} \\ < \frac{P_0}{P_1} \\ H_0 \end{matrix}$$

ou, d'une façon équivalente,

$$\begin{matrix} H_1 \\ \sum_{k=1}^K s_k R_k > \frac{\sqrt{E}}{2} + \frac{\sigma_n^2}{\sqrt{E}} \ln \frac{P_0}{P_1} = \gamma^* \\ < \frac{\sqrt{E}}{2} + \frac{\sigma_n^2}{\sqrt{E}} \ln \frac{P_0}{P_1} = \gamma^* \\ H_0 \end{matrix}$$

où on a défini le signal normalisé $s'_k = \frac{s_k}{\sqrt{E}}$, où $E = \sum_{k=1}^K s_k^2$ est l'énergie du signal transmis.

On vérifie que le détecteur optimal effectue le produit interne du signal reçu avec une réplique (normalisée) du signal émis. Cette opération peut s'interpréter de plusieurs façons:

1- filtre adapté

Cette interprétation considère le produit interne comme un échantillon pris à la sortie d'un filtre dont la réponse impulsionnelle $h_{adap}(n)$ est déterminée par le signal émis:

$$h_{adap}(n) = s_{K-n}$$

c'est-à-dire,

$$L(R) = (r * h_{adap})(K),$$

où la sortie du filtre à un instant générique est donnée par

$$(r * h_{adap})(p) = \sum_{q=p-K}^p r_q h_{adap}(q-p) = \sum_{q=p-K}^p r_q s_{K-p+q}$$

ou encore

$$(r * h_{adap})(p) = \sum_{q=0}^K r_q s_q.$$

La figure suivante illustre le cas d'un signal émis triangulaire, et la réponse impulsionnelle du filtre du détecteur optimal correspondant.

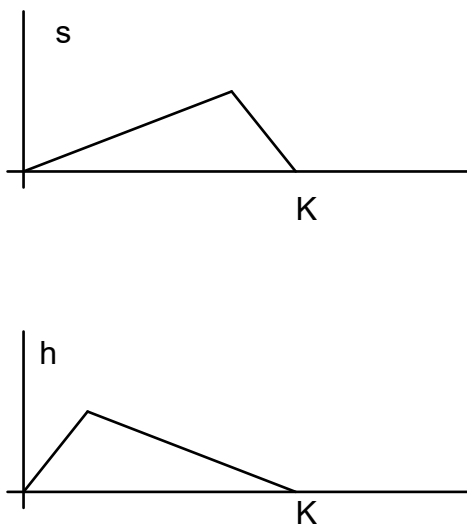


Figure 3.4: Signal et réponse impulsionnelle du filtre.

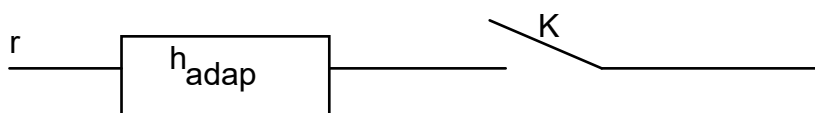


Figure 3.5: Détecteur optimal (configuration filtre adapté).

2- Corrélateur

Ce même détecteur peut être interprété comme la sortie d'un bloc qui effectue la corrélation du signal reçu avec une réplique du signal émis, voir la figure suivante.

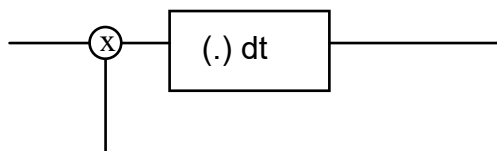


Figure 3.6: Récepteur corrélateur.

Calcul de la probabilité d'erreur

On détermine par la suite la probabilité d'erreur du détecteur optimal. Pour ce calcul, on va admettre que *les deux hypothèses sont également probables*, et donc que le test optimal est simplement

$$L(r) = \sum_{k=1}^K s'_k R_k \begin{array}{l} H_1 \\ > \frac{\sqrt{E}}{2} \\ < \frac{\sqrt{E}}{2} \\ H_0 \end{array}$$

Pour calculer la probabilité d'erreur,

$$P_e = P_0 \Pr \left[L(R) > \frac{\sqrt{E}}{2} | H_0 \right] + P_1 \Pr \left[L(R) < \frac{\sqrt{E}}{2} | H_1 \right],$$

on doit déterminer la densité de probabilité de $L(R)$ sous chaque hypothèse. Puisque $L(R)$ est une combinaison linéaire de v.a. Gaussiennes, il est aussi une v.a. Gaussienne. Il reste donc à calculer sa moyenne et sa variance.

Hypothèse H0.

Pour la moyenne,

$$E[L(R)|H_0] = E \left[\sum_{k=1}^K s'_k R_k | H_0 \right] = \sum_{k=1}^K s'_k E[R_k | H_0] = 0.$$

La v.q.m. est donnée par

$$E[L(R)^2 | H_0] = E \left[\sum_{k=1}^K s'_k R_k \sum_{n=1}^K s'_n R_n | H_0 \right] = \sum_{k=1}^K s'_k s'_n E[R_n R_k | H_0] = \sum_{k=1}^K s'_k s'_n \sigma_n^2 \delta_{l-k} = \sigma_n^2$$

et donc la variance est

$$Var[L(R)|H_0] = \sigma_n^2.$$

Hypothèse H1

Pour l'autre hypothèse on obtient, d'une façon similaire,

$$E[L(R)|H_1] = E \left[\sum_{k=1}^K s'_k R_k | H_1 \right] = \sum_{k=1}^K s'_k E[R_k | H_1] = \sqrt{E},$$

et

$$Var[L(R)|H_1] = \sigma_n^2.$$

La probabilité d'erreur est donc (en utilisant la symétrie du problème)

$$P_e = \int_{\sqrt{E}/2}^{\infty} p(L(R)|H_0) dL(R) = \frac{1}{\sqrt{2\pi}\sigma_n} \int_{\sqrt{E}/2}^{\infty} e^{-\frac{L^2}{2\sigma_n^2}} dL.$$

En utilisant la définition de la **fonction d'erreur complémentaire**,

$$erfc(x) = \frac{1}{\sqrt{\pi}} \int_x^{\infty} e^{-u^2} du,$$

l'expression précédente peut être écrite simplement comme,

$$P_e = erfc\left(\frac{\sqrt{E}/2}{\sqrt{2}\sigma_n}\right).$$

EXERCICES

1. Les densités qui décrivent les observations dans un problème de décision binaire sont:

$$H_1 : p_r(R) = \frac{1}{2} \exp(-|R|)$$

$$H_2 : p_r(R) = \frac{1}{\sqrt{2\pi}} \exp(-R^2 / 2)$$

a) Calculer le rapport de vraisemblance.

b) Le test est

$$\begin{array}{c} H_1 \\ \Lambda(R) > \eta \\ < \\ H_0 \end{array}$$

Déterminer les régions de décision pour plusieurs valeurs de η .

2. Considérer le problème de test d'hypothèses suivant. Etant données K observations indépendantes

$$H_1 : r_i \text{ est Gaussien, moyenne } 0 \text{ et variance } \sigma_1^2, \quad i = 1, \dots, K$$

$$H_0 : r_i \text{ est Gaussien, moyenne } 0 \text{ et variance } \sigma_0^2, \quad i = 1, \dots, K$$

où $\sigma_0 < \sigma_1$.

a) Déterminer le rapport de vraisemblance.

b) Supposons que le seuil est η :

$$\begin{array}{c} H_1 \\ \Lambda(R) > \eta. \\ < \\ H_0 \end{array}$$

Montrer que $m(r) = \sum_{i=1}^K R_i^2$ est une statistique suffisante. Déterminer le seuil γ pour le test

$$\begin{array}{c} H_1 \\ m(r) > \gamma \\ < \\ H_0 \end{array}$$

en fonction de $\gamma, \sigma_0, \sigma_1$.

c) Déterminer une expression pour les probabilités d'erreur.

d) Tracer le **ROC** pour $K=1, \sigma_1^2 = 2, \sigma_0^2 = 1$

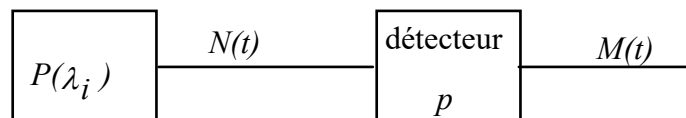
e) Quelle est la valeur du seuil pour le critère *minimax* si $C_M = C_F$ et $C_{00} = C_{11} = 0$?

3. Définir le détecteur optimal (probabilité d'erreur minimale) pour un système de transmission ternaire, qui transmet avec une probabilité égale les symboles $(-1, 0, 1)$, avec la codification suivante:

$$\begin{array}{l} -1: m_k = -s_k, \quad k = 1, \dots, K \\ 0: m_k = 0, \quad k = 1, \dots, K \\ 1: m_k = s_k, \quad k = 1, \dots, K \end{array}$$

Le signal reçu est la superposition du signal transmis avec du bruit blanc Gaussien, de moyenne nulle et covariance σ_n^2 . Calculer la probabilité d'erreur du détecteur optimal. Vérifier que pour des symboles également probables, et avec le critère de probabilité d'erreur minimale, le détecteur optimal est le détecteur de Maximum de Vraisemblance.

4. Considérer un système de communication optique:



Un laser transmet pendant l'intervalle $[0, t[$ un nombre $N(t)$ de photons qui suit une loi de Poisson de paramètre qui dépend du symbole binaire transmis: $\lambda = \lambda_0$ pour un symbole 0, et $\lambda = \lambda_1$ pour un symbole 1. Chaque photon est détecté avec une probabilité p . A partir de l'observation du nombre de photons détectés, $M(t)$, on doit décider quel symbole a été transmis.

a) Déterminer la loi de probabilité de $M(t)$.

b) Déterminer le rapport de vraisemblance pour ce problème.

c) Tracer le ROC pour $\lambda_0 = 1$ et $\lambda_1 = 2$, et pour une valeur de t choisie de façon que la probabilité de détection soit 0.5 pour une probabilité de fausse alarme égale à 0.01.

5. Considérer le problème de détection d'un signal connu dans du bruit blanc, pour lequel on a vérifié que le détecteur de probabilité d'erreur minimale est le filtre adapté, mais où on admet maintenant que le signal est reçu avec un facteur multiplicatif inconnu:

$$H_1: r_k = m s_k + n_k, \quad k = 1, \dots, K, \quad m > 0$$

$$H_0: r_k = n_k, \quad k = 1, \dots, K$$

a) Déterminer le rapport de vraisemblance pour ce problème.

b) Vérifier qu'un test uniformément le plus puissant existe. Spécifier le test.