

Test Bayésien pour le Problème du Test des Hypothèses Multiples avec le Coût Quadratique

Jian ZHANG¹, Lionel FILLATRE², Igor NIKIFOROV¹

¹ICD - LM2S - Université de Technologie de Troyes
12, rue Marie Curie - BP 2060 - 10010 Troyes - France

²Laboratoire I3S - UMR7271 - UNS CNRS
CS 40121 - 06903 Sophia Antipolis - France

jian.zhang@utt.fr, lionel.fillatre@i3s.unice.fr
igor.nikiforov@utt.fr

Résumé – Cet article s’intéresse à la construction d’un test Bayésien pour un problème de décision entre hypothèses multiples lorsque la fonction de perte est quadratique. Ce critère quadratique est spécialement adapté aux problèmes de classification pour lesquels le coût d’une erreur de décision varie suivant les hypothèses concernées. L’originalité de cet article est d’associer à chaque hypothèse une étiquette vectorielle unique définie dans un espace Euclidien. Les erreurs de décision sont comparées en utilisant une norme vectorielle. Le test proposé permet de minimiser la somme des probabilités de fausse classification pondérées avec l’écart quadratique entre les hypothèses au sens de la norme utilisée. Le test Bayésien pour la fonction de perte quadratique est calculé de façon explicite, puis des bornes théoriques sur ses performances statistiques sont proposées. Enfin, il est établi que le rapport entre les probabilités de classification erronée de ce nouveau test et celles du test Bayésien pour la fonction de perte 0–1 est borné par une constante dépendant du nombre des hypothèses.

Abstract – This paper focuses on the construction of a Bayesian test for a multiple hypothesis testing problem when the loss function is quadratic. This quadratic criterion is especially suitable for the classification problems where the cost of a decision error varies with the concerned hypotheses. This paper originally relates each hypothesis with a unique label vector defined in an Euclidean space and the decision errors are compared using a vectorial norm. The proposed test minimizes the sum of the misclassification probabilities weighted with the squared difference between the hypotheses in the sense of the used norm. The Bayesian test for the quadratic loss function is calculated explicitly and then the theoretical limits on its statistical performance are proposed. Finally, it is established that the ratio between the misclassification probabilities of the new test and those of the Bayesian test for the 0-1 loss function is bounded by a constant which depends on the number of hypotheses.

1 Motivation et contribution

Cet article étudie la construction d’un Test entre Hypothèses Multiples (THM) [3, 4] dans le cadre d’une approche Bayésienne avec l’utilisation d’une fonction de pondération quadratique. Concrètement, considérons un vecteur aléatoire $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ composé de n observations indépendantes. Les hypothèses à tester H_1, \dots, H_n sont de la forme

$$H_k : X_k = \Delta + \xi_k \text{ et } X_i = \xi_i, \forall i \neq k, \quad (1)$$

où le biais $\Delta > 0$, connu, représente l’anomalie/cible à détecter et $\xi_i \sim N(0, \sigma^2)$ suit une loi normale de moyenne nulle et de variance σ^2 connue. Les variables ξ_i sont mutuellement indépendantes. Chaque hypothèse H_i est associée à une étiquette vectorielle (identifiant) unique $\theta_i \in \mathbb{R}^q$ qui caractérise l’hypothèse. L’ensemble des étiquettes est noté $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$. Un THM est une fonction $\delta(X) : \mathbb{R}^n \mapsto \{1, \dots, n\}$ telle que H_j est acceptée lorsque $\delta(X) = j$. Le THM étudié dans cet article s’appuie sur deux éléments principaux : 1) la probabilité *a priori* $p_i > 0$ de l’hypothèse H_i est connue pour tout $1 \leq i \leq n$ avec $\sum_{i=1}^n p_i = 1$ et 2) une erreur de décision est

pondérée au moyen de la quantité

$$L^Q(\theta_i, \theta_{\delta(X)}) = \|\theta_i - \theta_{\delta(X)}\|_2^2 \quad (2)$$

où $\theta_i \in \Theta$ désigne l’étiquette de l’hypothèse vraie et $\theta_{\delta(X)}$ désigne l’étiquette décidée par le THM. De cette façon, l’espace des hypothèses est doté d’une métrique qui permet de comparer les hypothèses. Le test Bayésien recherché est le test qui minimise le risque de Bayes associé aux probabilités p_i et à $L^Q(\theta_i, \theta_{\delta(X)})$.

Illustrons l’intérêt de cette approche sur un exemple. Ce type de THM est particulièrement bien adapté aux problèmes de localisation d’intrusion dans un Réseau de Capteurs Sans Fil (RCSF). Dans le RCSF illustré sur la figure 1, la présence d’une intrusion au niveau d’un capteur provoque l’enregistrement d’un biais Δ par ce capteur. Il s’agit alors de détecter la mesure affectée par ce biais tout en minimisant la distance à parcourir pour se rendre jusqu’au capteur ayant perçu l’intrusion. Les capteurs peuvent être différemment éloignés les uns des autres. Le fait de prendre une mauvaise décision et de se rendre jusqu’à un capteur très éloigné du lieu exacte de l’intrusion peut représenter un coût important. Dans cet exemple, l’étiquette θ_i

correspond typiquement à la position géographique du i -ième capteur et peut même contenir des informations sur le coût du déplacement pour se rendre jusqu'à lui. Historiquement, [2] a publié la première solution au problème du THM pour la fonction de perte 0–1 donnée par

$$L^{0-1}(\theta_i, \theta_{\delta(X)}) = \begin{cases} 1 & \text{si } \theta_i \neq \theta_{\delta(X)}, \\ 0 & \text{si } \theta_i = \theta_{\delta(X)}, \end{cases} \quad (3)$$

pour tout $\theta_i \in \Theta$. Bien que $L^{0-1}(\theta_i, \theta_{\delta(X)})$ puisse être facilement manipulée, elle n'est pas appropriée pour certaines applications comme la localisation d'intrusion dans un RCSF puisque la perte causée par une localisation erronée ne dépend pas de la distance entre l'emplacement détecté de la cible et son emplacement réel. Cependant, changer la fonction de perte a des conséquences considérables sur la difficulté à obtenir un test optimal. Les contributions principales de cette étude sont les suivantes :

1. Le test Bayésien qui minimise la fonction de perte quadratique est calculé sous une forme explicite,
2. Le risque de Bayes est exprimé en fonction des probabilités de classification erronée et les performances statistiques asymptotiques du test Bayésien proposé, qui sont très difficiles à calculer sous forme exacte, sont bornées de façon analytique,
3. Lorsque le Rapport Signal-sur-Bruit (RSB) devient infiniment grand, il est démontré que le rapport entre les mêmes probabilités de classification erronée du test proposé et du test associé à la fonction de perte 0–1 tend vers une constante déterminée par le nombre des hypothèses.

Il est intéressant de noter qu'il existe une bijection naturelle entre les ensembles finis $\{H_1, \dots, H_n\}$ et Θ . De ce fait, chaque test $\delta(X)$ peut s'interpréter comme un estimateur Bayésien discret défini par $\hat{\theta}(X) = \theta_{\delta(X)}$.

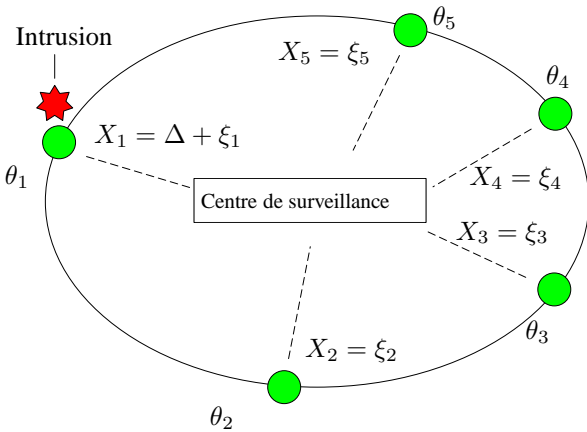


FIGURE 1 – Localisation d'une intrusion dans un RCSF avec un centre de surveillance. Le i -ième capteur est à l'emplacement géographique θ_i . Il mesure la valeur moyenne Δ en présence d'une intrusion.

2 Test Bayésien multi-hypothèses

2.1 Risque de Bayes et test Bayésien

Dans le cadre bayésien [1], la qualité d'un test $\delta(X)$ est évaluée avec le risque de Bayes $R(\theta, \delta(X))$:

$$R(\theta, \delta(X)) = \sum_{i=1}^n \int_{\mathbb{R}^n} L(\theta_i, \theta_{\delta(x)}) \phi(x, \theta_i) dx \quad (4)$$

où $\phi(x, \theta_i)$, $x \in \mathbb{R}^n$ et $\theta_i \in \Theta$, représente la fonction de densité conjointe de (X, θ) et $L(\theta_i, \theta_{\delta(x)})$ est la fonction de perte. La valeur de $L(\theta_i, \theta_{\delta(x)})$ est le coût de décider $\theta_{\delta(x)}$ lorsque le paramètre réel est θ_i (voir les exemples (2) et (3)). Le risque de Bayes est la valeur moyenne de la fonction de perte par rapport à la distribution conjointe du vecteur d'observation X et de la variable aléatoire θ . Le test qui minimise le risque de Bayes (4) est défini comme le test bayésien $\hat{\delta}(X)$:

$$\hat{\delta}(X) = \arg \min_{\delta(X) \in \mathcal{K}} R(\theta, \delta(X)) \quad (5)$$

où \mathcal{K} désigne l'ensemble des tests $\delta(X) : \mathbb{R}^n \mapsto \{1, \dots, n\}$.

2.2 Résultats généraux sur le test Bayésien

Sous l'hypothèse H_k donnée par (1), X_1, \dots, X_n sont indépendantes, $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n$ sont identiquement distribuées avec une densité Gaussienne commune $\varphi_0(x)$, donnée par

$$\varphi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

tandis que X_k admet la densité Gaussienne $\varphi_1(x) = \varphi_0(x - \Delta)$. De ce fait, la fonction de densité de probabilité conjointe $f_k(x)$ du vecteur $X = (X_1, X_2, \dots, X_n)$ est :

$$f_k(x) = \varphi_1(x_k) \prod_{i=1, i \neq k}^n \varphi_0(x_i)$$

où $x = (x_1, \dots, x_n)$. La fonction de densité de probabilité conjointe $\phi(x, \theta_i)$ de (X, θ) est donc donnée par :

$$\phi(x, \theta_i) = p_i f_i(x). \quad (6)$$

Par ailleurs, soit $f(x)$ la densité marginale de X :

$$f(x) = \sum_{i=1}^n p_i f_i(x) > 0, \quad \forall x \in \mathbb{R}^n.$$

La probabilité *a posteriori* $\pi(\theta_i|x)$ de θ_i étant donné l'observation x est définie par

$$\pi(\theta_i|x) = \frac{\phi(x, \theta_i)}{f(x)}$$

pour tout $\theta_i \in \Theta$ et pour tout $x \in \mathbb{R}^n$. Le risque de Bayes $R(\theta, \delta(X))$ s'écrit alors

$$R(\theta, \delta(X)) = \int_{\mathbb{R}^n} \left[\sum_{i=1}^n L(\theta_i, \theta_{\delta(x)}) \pi(\theta_i|x) \right] f(x) dx. \quad (7)$$

Il est ensuite immédiat [1] de montrer que le test Bayésien optimal est donné par

$$\hat{\delta}(X) = \arg \min_{\delta(X) \in \mathcal{K}} \sum_{i=1}^n L(\theta_i, \theta_{\delta(X)}) \pi(\theta_i | X). \quad (8)$$

La qualité d'un THM $\delta(X)$ est généralement caractérisée par l'ensemble de valeurs

$$\alpha_{i,j} = \Pr_i(\delta(X) = j) \quad (9)$$

pour tout $1 \leq i \neq j \leq n$, qui représente la probabilité de classification erronée de H_j quand H_i est l'hypothèse vraie. La proposition suivante montre que le risque de Bayes est directement lié à la probabilité $\alpha_{i,j}$.

Proposition 1 *Le risque de Bayes $R(\theta, \delta(X))$ du test $\delta(X)$ pour tester les hypothèses H_1, \dots, H_n données par (1) avec une fonction de perte arbitraire $L(\theta_i, \theta_{\delta(X)})$ et les probabilités a priori p_1, p_2, \dots, p_n satisfait*

$$R(\theta, \delta(X)) = \sum_{i=1}^n \sum_{j=1}^i [p_i \alpha_{i,j} L(\theta_i, \theta_j) + p_j \alpha_{j,i} L(\theta_j, \theta_i)] \quad (10)$$

où $\alpha_{i,j}$ est donnée par (9).

2.3 Test Bayésien pour la perte 0-1

Le théorème suivant est une extension directe d'un résultat établi par [2] dans le cadre de la fonction de perte $L^{0-1}(\theta_i, \theta_{\delta(X)})$.

Théorème 1 *Le test bayésien $\hat{\delta}^{0-1}(X)$ pour tester les hypothèses H_1, \dots, H_n données par (1) avec la fonction de perte 0-1 et les probabilités a priori p_1, p_2, \dots, p_n est donné par*

$$\hat{\delta}^{0-1}(X) = \arg \max_{1 \leq k \leq n} \left(X_k + \frac{\sigma^2}{\Delta} \ln p_k \right). \quad (11)$$

Selon (10), le risque de Bayes de $\hat{\delta}^{0-1}(X)$ est

$$R^{0-1}(\theta, \hat{\delta}^{0-1}(X)) = \sum_{i=2}^n \sum_{j=1}^{i-1} (p_i \alpha_{i,j}^{0-1} + p_j \alpha_{j,i}^{0-1}) \quad (12)$$

où $\alpha_{i,j}^{0-1}$ sont les erreurs de classification pour $\hat{\delta}^{0-1}(X)$.

3 Test Bayésien quadratique

3.1 Test Bayésien pour la perte quadratique

Le théorème suivant (dont la démonstration est omise par manque de place) calcule le test Bayésien pour la fonction de perte $L^Q(\theta_i, \theta_{\delta(X)}) = \|\theta_i - \theta_{\delta(X)}\|_2^2$.

Théorème 2 *Le test bayésien $\hat{\delta}^Q(X)$ pour tester les hypothèses H_1, \dots, H_n données par (1) avec la fonction de perte quadratique $L^Q(\theta_i, \theta_{\delta(X)})$ et les probabilités a priori p_1, p_2, \dots, p_n est donné par*

$$\hat{\delta}^Q(X) = \arg \min_{1 \leq m \leq n} \sum_{k=1, k \neq m}^n p_k \|\theta_k - \theta_m\|_2^2 \exp\left(\frac{\Delta X_k}{\sigma^2}\right). \quad (13)$$

Selon (10), le risque de Bayes de $\hat{\delta}^Q(X)$ est

$$R^Q(\theta, \hat{\delta}^Q(X)) = \sum_{i=2}^n \sum_{j=1}^{i-1} (p_i \alpha_{i,j}^Q + p_j \alpha_{j,i}^Q) \|\theta_i - \theta_j\|_2^2 \quad (14)$$

où $\alpha_{i,j}^Q$ sont les erreurs de classification pour $\hat{\delta}^Q(X)$.

3.2 Performance asymptotique des tests

Les probabilités de classification erronées pour $\alpha_{i,j}^Q$ sont très difficiles à calculer de façon analytique. Le théorème 3 (dont la démonstration est omise par manque de place) propose donc des bornes pour évaluer ces probabilités. Il établit que $\alpha_{i,j}^{0-1}$ et $\alpha_{i,j}^Q$ ont la même borne inférieure et la même borne supérieure, désignées respectivement par $P_{i,j}^{\text{inf}}$ et $P_{i,j}^{\text{sup}}$. Soit λ le RSB défini par $\lambda = \frac{\Delta}{\sigma}$. Soient $\Phi(\cdot)$ la fonction de répartition de la loi normale centrée réduite et $r = \min_{1 \leq i \neq j \leq n} \|\theta_i - \theta_j\|_2^2$, resp. $R = \max_{1 \leq i \neq j \leq n} \|\theta_i - \theta_j\|_2^2$, la distance minimale, resp. maximale, entre les étiquettes.

Théorème 3 *Les tests bayésiens $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^Q(X)$ vérifient, pour tout $1 \leq i \neq j \leq n$,*

$$\begin{aligned} P_{i,j}^{\text{inf}} &\leq \alpha_{i,j}^{0-1} \leq P_{i,j}^{\text{sup}}, \\ P_{i,j}^{\text{inf}} &\leq \alpha_{i,j}^Q \leq P_{i,j}^{\text{sup}}, \end{aligned}$$

$$\begin{aligned} P_{i,j}^{\text{inf}} &= \Phi\left(\nu_{j,i}^1 - \frac{\lambda}{\sqrt{2}}\right) \prod_{k=1, k \neq j, k \neq i}^n \Phi(\nu_{j,k}^1), \\ P_{i,j}^{\text{sup}} &= 1 - \Phi^{n-2}\left(\nu_{i,j}^2 + \frac{\lambda}{\sqrt{2}}\right) \prod_{k=1, k \neq j, k \neq i}^n \Phi\left(\nu_{i,k}^2 + \frac{\lambda}{\sqrt{2}}\right), \\ \nu_{i,k}^\ell &= \frac{C^\ell + \ln \frac{p_i}{p_k}}{\lambda \sqrt{2}}, \quad \ell = 1, 2, \quad k = 1, \dots, n, \quad i \neq k, \\ C^1 &= \ln \frac{r}{(n-1)(R-r) + R}, \\ C^2 &= \ln \frac{r}{(n-2)(R-r) + R}. \end{aligned}$$

À partir du théorème 3, il est clair que r , R et n affectent conjointement $P_{i,j}^{\text{inf}}$ et $P_{i,j}^{\text{sup}}$. Cependant, le corollaire suivant montre que le rapport entre $\alpha_{i,j}^{0-1}$ et $\alpha_{i,j}^Q$ est borné par une constante dépendant uniquement de n lorsque $\lambda \rightarrow +\infty$.

Corollaire 1 *Lorsque $\lambda \rightarrow +\infty$, les bornes $P_{i,j}^{\text{inf}}$ et $P_{i,j}^{\text{sup}}$ des probabilités d'erreurs des tests $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^Q(X)$ vérifient*

$$\begin{aligned} P_{i,j}^{\text{inf}} &\underset{\lambda \rightarrow \infty}{\sim} \left(\frac{1}{2}\right)^{n-2} \Phi\left(-\frac{\lambda}{\sqrt{2}}\right), \\ P_{i,j}^{\text{sup}} &\underset{\lambda \rightarrow \infty}{\sim} 1 - \Phi^{2n-4}\left(\frac{\lambda}{\sqrt{2}}\right), \\ \frac{P_{i,j}^{\text{sup}}}{P_{i,j}^{\text{inf}}} &\underset{\lambda \rightarrow \infty}{\sim} (n-2) \cdot 2^{n-1} \end{aligned}$$

où $f(t) \underset{t \rightarrow \infty}{\sim} g(t)$ signifie que $\lim_{t \rightarrow +\infty} f(t)/g(t) = 1$ pour $f(t) > 0$ et $g(t) > 0$.

Soit $C_{\text{seuil}} = (n-2) \cdot 2^{n-1}$ pour $n \geq 3$. Le corollaire 1 conduit immédiatement à

$$\frac{1}{C_{\text{seuil}}} \leq \frac{\alpha_{i,j}^{0-1}}{\alpha_{i,j}^Q} \leq C_{\text{seuil}}$$

lorsque $\lambda \rightarrow +\infty$.

4 Résultats numériques

Cette section considère l'exemple de la localisation d'intrusion illustrée sur la figure 1 en supposant que toutes les probabilités *a priori* p_i sont égales.

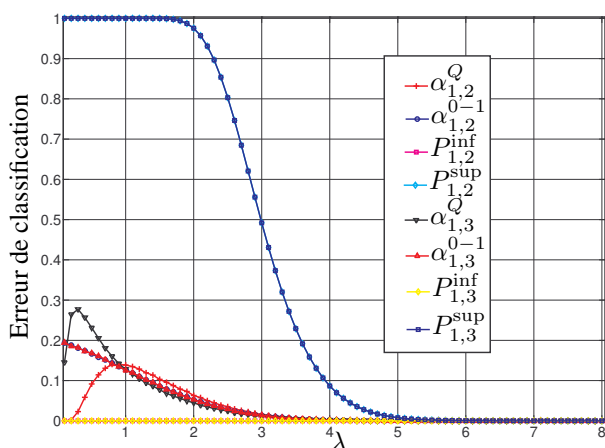


FIGURE 2 – Estimation de $\alpha_{1,2}$ et celle de $\alpha_{1,3}$ pour les tests $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^Q(X)$ en fonction de λ . Les bornes $P_{1,2}^{\text{inf}}$, $P_{1,2}^{\text{sup}}$, $P_{1,3}^{\text{inf}}$ et $P_{1,3}^{\text{sup}}$ sont tracées en fonction de λ .

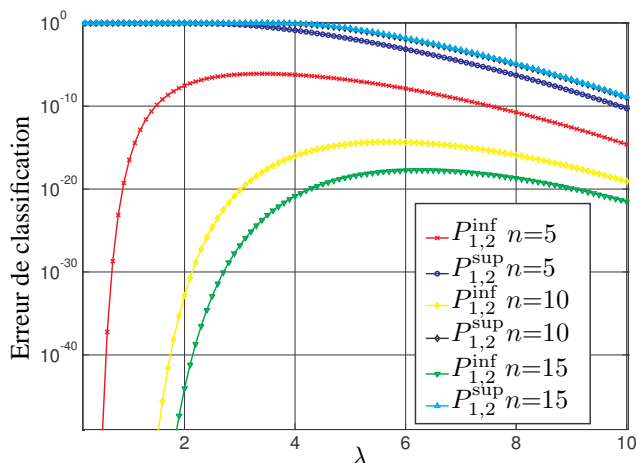


FIGURE 3 – Illustration du resserrement des bornes $P_{1,2}^{\text{inf}}$ et $P_{1,2}^{\text{sup}}$ lorsque $\lambda \rightarrow +\infty$ et évolution des bornes $P_{1,2}^{\text{inf}}$ et $P_{1,2}^{\text{sup}}$ lorsque n vaut respectivement 5, 10 et 15.

La figure 2 présente la comparaison entre $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^Q(X)$ pour les probabilités $\alpha_{1,2}$ et $\alpha_{1,3}$. Les autres probabilités se comparent de façon similaire. Les probabilités sont estimées à partir d'une simulation de Monte Carlo. Les bornes inférieure et supérieure $P_{i,j}^{\text{inf}}$ et $P_{i,j}^{\text{sup}}$ sont tracées pour illustrer la pertinence du théorème 3. Pour l'exemple considéré, les étiquettes θ_1 , θ_2 et θ_3 vérifient $\|\theta_1 - \theta_2\|_2^2 < \|\theta_1 - \theta_3\|_2^2$, il est donc naturel que $\alpha_{1,2}^Q > \alpha_{1,3}^Q$ pour le test quadratique $\hat{\delta}^Q(X)$. Par contre, le test usuel $\hat{\delta}^{0-1}(X)$ ne tient pas du tout compte de cet écart métrique puisque $\alpha_{1,2}^{0-1} = \alpha_{1,3}^{0-1}$. Par conséquent, on peut en déduire qu'il existe une relation compensatrice entre la probabilité de classification erronée $\alpha_{i,j}^Q$ et l'écart métrique $\|\theta_i - \theta_j\|_2^2$ afin de minimiser le risque quadratique de Bayes.

Afin d'illustrer le resserrement des bornes et leur utilité, l'ordonnée de la figure 3 est tracée en échelle logarithmique. Les bornes $P_{1,2}^{\text{inf}}$ et $P_{1,2}^{\text{sup}}$, lorsque n vaut, respectivement, 5, 10, 15, sont également tracées pour vérifier la relation entre les bornes et le nombre des hypothèses n . Le rapport de $P_{1,2}^{\text{sup}}$ sur $P_{1,2}^{\text{inf}}$ tend vers une constante.

5 Conclusion

Cet article propose un test Bayésien pour classifier des hypothèses multiples. Ce test minimise l'erreur quadratique moyenne de classification des étiquettes vectorielles attribuées aux hypothèses. Les performances statistiques de ce test sont étudiées de façon analytique et comparées au test usuel qui ne tient pas compte des différences métriques entre les hypothèses.

Références

- [1] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer, 2010.
- [2] T. S. Ferguson. *Mathematical statistics. A decision theoretic approach*. Academic Press, New York and London, 1967.
- [3] L. Fillatre et I. Nikiforov. Asymptotically uniformly minimax detection and isolation in network monitoring. *IEEE Transactions on Signal Processing*, 60(7) :3357–3371, July 2012.
- [4] L. Fillatre. Constrained Epsilon-Minimax Test for Simultaneous Detection and Classification. *IEEE Transactions on Information Theory*, 57(12) :8055–8071, 2011.