# Détection et Classification Minimax d'Observations Discrètes

Lionel Fillatre

*Université Côte d'Azur, CNRS, I3S, France*
*CS 40121 - 06903 Sophia Antipolis Cedex, France*
*Email : lionel.fillatre@i3s.unice.fr*

**Résumé.** Cet article s'intéresse au calcul d'un test minimax de niveau contraint entre plusieurs hypothèses impliquant des observations discrètes et une fonction de perte arbitraire. Le test minimax de niveau contraint minimise le risque de classification maximum et il garantit simultanément que la probabilité de rejeter l'hypothèse nulle, appelé le niveau du test, est bornée par une valeur donnée. Ce type de test est particulièrement adapté au problème de détection et de classification simultanée. Cet article montre que le test minimax de niveau contraint est la solution d'un problème de programmation linéaire qui calcule le risque maximum de classification et la distribution a priori la plus défavorable. Le test minimax égalise les risques de classification dont les probabilités a priori sont strictement positives. Le test est appliqué au problème de détection et d'estimation ponctuelle discrète qui consiste à détecter puis estimer la valeur d'un paramètre appartenant à un ensemble fini.

**Mots-clés.** Test d'hypothèses, Détection et classification, Programmation Linéaire

**Abstract.** This paper deals with a multihypothesis testing framework for calculating the optimal minimax test with discrete observations and an arbitrary loss functions subject to a constrained false alarm probability. The minimax test minimizes the maximum classification risk while the probability to reject the null hypothesis must satisfy a given false alarm level. It is well adapted to the problem of simultaneous detection and classification. It is shown that the constrained minimax test is the solution of a linear programming problem which computes the worst case classification risk and the worst case prior distribution. The minimax test equalizes the classification risks whose prior probabilities are strictly positive. The minimax framework is applied to the simultaneous detection and discrete point estimation problem which consists in detecting and inferring the value of a parameter belonging to a finite set.

**Keywords.** Hypothesis Testing, Detection and Classification, Linear Programming

## 1 Introduction

Le problème de détection et de classification simultanée d'un paramètre inconnu dans des mesures bruitées apparaît souvent en ingénierie. Il s'agit donc en premier lieu de détecter la présence du paramètre et, s'il est présent, d'estimer sa valeur en exploitant une fonction de perte qui mesure l'impact d'une erreur d'estimation. Les probabilités a priori d'apparition des paramètres sont supposés inconnus. Le test minimax devient alors très pertinent contrairement au test de Bayes qui nécessite la connaissance d'informations a priori. La conception de tests minimax avec un niveau contraint entre plusieurs hypothèses composées d'un nombre fini de paramètres a été initiée par Baygün et Hero (1995). Ces premiers résultats ne considèrent pas une fonction de perte arbitraire mais juste les probabilités d'erreur. De plus, le test minimax proposé dépend de certains coefficients inconnus, à savoir les poids et le seuil de détection, qui sont très difficiles à calculer en pratique. Cet article propose une extension de ces travaux en utilisant une fonction de perte arbitraire et en calculant de façon automatique les coefficients nécessaires à la construction du test.

Il y a deux tendances dans la litérature pour construire des tests minimax. La première tendance, bien exposée dans Ferguson (1967); Berger (1985); Lehman (1986), consiste à calculer de façon analytique le test minimax . De nombreux travaux existent mais cela reste toujours difficile de construire un test minimax pour une situation spécifique, par exemple en absence de certaines propriétés d'invariance. La seconde tendance consiste à calculer le test de façon numérique comme dans Barankin (1951); Dantzig et Wald (1951); Weiss (1961); Schaafsma (1970); Baumann (1968); Krafft (1970); Krafft et Schmitz (1970); Nelson (1966); Goldenshluger et al. (2015). Tous ces travaux s'intéressent aux tests entre deux hypothèses. Dans le cas de plusieurs hypothèses, il y a nettement moins de résultats. Citons par exemple Nelson (1966); Krafft et Schmitz (1970); Chang et Davisson (1990). Enfin, en ajoutant une contrainte sur le niveau du test, le nombre de travaux devient extrêmement faible comme souligné dans Baygün et Hero (1995). Cet article s'inscrit précisément dans cet axe de recherche.

# 2 Test minimax de niveau $\alpha$

Soit $X$ une variable aléatoire discrète prenant des valeurs $\boldsymbol{x}$ dans un espace fini $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ où $\boldsymbol{x}_i \in \mathcal{A}^n$, $\mathcal{A}$ est l'alphabet fini de codage et $n$ est la taille de $\boldsymbol{x}$. La variable $X$ admet la loi de probabilité $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ où $\boldsymbol{\theta} \in \mathbb{R}^p$ est un vecteur de paramètres. Nous disposons de $K+1$ paramètres, i.e., $\boldsymbol{\theta} \in \Theta = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K\}$. Nous souhaitons tester les hypothèses simples $\mathcal{H}_0$, $\mathcal{H}_1$,. . . ,$\mathcal{H}_K$ telles que $\mathcal{H}_k : \{X \sim p_{\boldsymbol{\theta}_k}\}$. Un problème de décision statistique entre plusieurs hypothèses est équivalent au jeu statistique $(\Theta, \Psi, w)$ avec une expérience qui implique la variable aléatoire $X$ dont la loi de probabilité $p_{\boldsymbol{\theta}}$ dépend de l'état $\boldsymbol{\theta} \in \Theta$. Au vue d'une mesure $X = \boldsymbol{x}$, nous souhaitons choisir une action $d(\boldsymbol{x}) \in \Psi = \{\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K\}$ où $d : \mathcal{X} \mapsto \Psi$ est la décision. La perte est une quantité aléatoire positive $w(\boldsymbol{\theta}, d(X))$ où $w : \Theta \times \Psi \mapsto [0, +\infty)$ est la fonction de perte et son espérance, appelée la fonction de risque, est $R(\boldsymbol{\theta}, d) = \mathbb{E}^X_{\boldsymbol{\theta}}[w(\boldsymbol{\theta}, d(X))]$. Ce papier s'appuie plus précisément sur les tests randomisés. Le problème de décision est alors le jeu statistique $(\Theta, \Psi^*, w)$ où $\Psi^*$ contient toutes les lois de probabilité sur $\Psi$. Les tests randomisés de niveau $\alpha$ sont définis comme suit.

**Définition 1.** *Un test randomisé de niveau $\alpha$ pour tester $\mathcal{H}_0$, $\mathcal{H}_1$,...,$\mathcal{H}_K$ est une variable aléatoire discrète $\boldsymbol{\delta}(\boldsymbol{x}) : \mathcal{X} \mapsto \Psi$ qui admet une loi de probabilité notée $\underline{\boldsymbol{\delta}}(\boldsymbol{x})=(\delta_0(\boldsymbol{x}), \delta_1(\boldsymbol{x}), \ldots, \delta_K(\boldsymbol{x})) \in \Psi^*$ tel que $R(\boldsymbol{\theta}_0, \boldsymbol{\delta}) \leq \alpha$. L'ensemble des tests randomisés de niveau $\alpha$ est noté $\mathcal{D}^*_\alpha$.*

La fonction de risque $R(\boldsymbol{\theta}, \boldsymbol{\delta})$, aussi notée $R(\boldsymbol{\theta}, \underline{\boldsymbol{\delta}})$, se ré-écrit

$$R(\boldsymbol{\theta}, \boldsymbol{\delta}) = R(\boldsymbol{\theta}, \underline{\boldsymbol{\delta}}) = \sum_{i=1}^{m} \sum_{j=0}^{K} p_{\boldsymbol{\theta}}(\boldsymbol{x}_i)\delta_j(\boldsymbol{x}_i) w(\boldsymbol{\theta}, \boldsymbol{\psi}_j). \tag{1}$$

Ce papier considère que la fonction de perte vérifie $w(\boldsymbol{\theta}_0, \boldsymbol{\psi}_0) = 0$, $w(\boldsymbol{\theta}_0, \boldsymbol{\psi}_j) = 1$, $\forall j \geq 1$. De cette façon, $R(\boldsymbol{\theta}_0, \boldsymbol{\delta})$ est la probabilité de fausse alarme du test $\boldsymbol{\delta}$.

**Définition 2.** *Un test randomisé $\boldsymbol{\delta}^* \in \mathcal{D}^*_\alpha$ est un test minimax de niveau $\alpha$ entre les hypothèses $\mathcal{H}_0$, $\mathcal{H}_1$,...,$\mathcal{H}_K$ si*

$$\max_{\boldsymbol{\theta} \in \Theta \setminus \boldsymbol{\theta}_0} R(\boldsymbol{\theta}, \boldsymbol{\delta}^*) = \inf_{\boldsymbol{\delta} \in \mathcal{D}^*_\alpha} \max_{\boldsymbol{\theta} \in \Theta \setminus \boldsymbol{\theta}_0} R(\boldsymbol{\theta}, \boldsymbol{\delta}).$$

Les théorèmes d'existence des tests minimax doivent être adaptés à la présence d'une contrainte sur le niveau du test. Cette extension est résumée dans le théorème suivant.

**Théorème 1.** *Pour un problème de décision $(\Theta, \Psi^*, w)$ avec $0 < \alpha < 1$, il existe un test minimax $\boldsymbol{\delta}^*$ de niveau $\alpha$.*

*Démonstration.* La preuve s'inspire de Ferguson (1967) et Baygün et Hero (1995). □

Le théorème suivant montre que le calcul d'un test minimax de niveau contraint revient à résoudre un problème de programmation linéaire.

**Théorème 2.** *Le test $\boldsymbol{\delta}^*(\boldsymbol{x})$ est un test minimax de niveau $\alpha$ pour tester $\mathcal{H}_0$, $\mathcal{H}_1$,...,$\mathcal{H}_K$ si et seulement s'il existe $\gamma^* \geq 0$ tel que $(\underline{\boldsymbol{\delta}}^*, \gamma^*) \in \Psi^* \times \mathbb{R}^+$ est une solution du programme linéaire où la forme linéaire $b(\underline{\boldsymbol{\delta}}, \gamma) = \gamma$ est minimisée par rapport à $(\underline{\boldsymbol{\delta}}, \gamma) \in \mathcal{K}$ défini par :*

$$\begin{aligned}
\delta_i(\boldsymbol{x}) &\geq 0, \quad && \forall \boldsymbol{x} \in \mathcal{X}, \forall 0 \leq i \leq K, && (2)\\
-\delta_i(\boldsymbol{x}) &\geq -1, && \forall \boldsymbol{x} \in \mathcal{X}, \forall 0 \leq i \leq K, && (3)\\
\sum_{i=0}^{K} \delta_i(\boldsymbol{x}) &\geq 1, && \forall \boldsymbol{x} \in \mathcal{X}, && (4)\\
-\sum_{i=0}^{K} \delta_i(\boldsymbol{x}) &\geq -1, && \forall \boldsymbol{x} \in \mathcal{X}, && (5)\\
\alpha - R(\boldsymbol{\theta}_0, \underline{\boldsymbol{\delta}}) &\geq 0, && && (6)\\
\gamma - R(\boldsymbol{\theta}_k, \underline{\boldsymbol{\delta}}) &\geq 0, && \forall k \in 1, \ldots, K. && (7)
\end{aligned}$$

*Démonstration.* La preuve s'inspire de Schaafsma (1970). □

Le programme linéaire, dit problème primal, du Théorème 2 peut facilement s'écrire sous forme matricielle et être résolu numériquement. Il est immédiat de vérifier que le problème dual consiste à chercher un vecteur $\boldsymbol{z} \in \mathbb{R}^{(m+1)(K+1)+2m}$ de la forme $\boldsymbol{z} = (\boldsymbol{v}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{q})$ où $\boldsymbol{v} \in \mathbb{R}^{m(K+1)}$, $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^m$ et $\boldsymbol{q} \in \mathbb{R}^{K+1}$. La classe des solutions faisables pour le problème dual est notée $\mathcal{L}$. Le théorème suivant, qui donne la forme du test minimax, est le résultat principal de l'article.

**Théorème 3** (Test minimax de niveau $\alpha$)**.** *Soit $\boldsymbol{y}^* = (\boldsymbol{\delta}^*, \gamma^*) \in \mathcal{K}$ et $\boldsymbol{z}^* = (\boldsymbol{v}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{q}^*) \in \mathcal{L}$ des solutions du problème primal et du problème dual. Pour tout $1 \leq i \leq m$ et $0 \leq j \leq K$, soit $g_j^*(\boldsymbol{x})$ la fonction*

$$g_j^*(\boldsymbol{x}_i) = \sum_{k=0}^{K} q_k^* p_{\boldsymbol{\theta}_k}(\boldsymbol{x}_i) w(\boldsymbol{\theta}_k, \boldsymbol{\psi}_j) \tag{8}$$

*où $q_k^*$ est donné par $\boldsymbol{q}^*$. Le test minimax $\boldsymbol{\delta}^*$ de niveau $\alpha$ vérifie*

$$\delta_j^*(\boldsymbol{x}_i) = \begin{cases} 1 & si \;\; g_j^*(\boldsymbol{x}_i) < \min_{\ell \neq j} g_\ell^*(\boldsymbol{x}_i), \\ 0 & si \;\; g_j^*(\boldsymbol{x}_i) > \min_{\ell \neq j} g_\ell^*(\boldsymbol{x}_i). \end{cases} \tag{9}$$

*En cas d'égalité, i.e., $g_j^*(\boldsymbol{x}_i) = \min_{\ell \neq j} g_\ell^*(\boldsymbol{x}_i)$, la décision est choisie de façon aléatoire entre les $0 \leq \delta_j^*(\boldsymbol{x}_i) \leq 1$ fournies par la solution du problème primal. La distribution discrète $(q_1^*, \ldots, q_K^*)$ déduide de $\boldsymbol{q}^*$ est la distribution a priori la plus défavorable des hypothèses $\mathcal{H}_1, \ldots, \mathcal{H}_K$. Par ailleurs, le test $\boldsymbol{\delta}^*$ vérifie $R(\boldsymbol{\theta}_0, \boldsymbol{\delta}^*) = \alpha$ et $\max_{1 \leq k \leq K} R(\boldsymbol{\theta}_k, \boldsymbol{\delta}^*) = \gamma^*$.*

*Démonstration.* La preuve s'inspire de Schaafsma (1970). ☐

# 3 Application : détection et estimation ponctuelle

Cette section s'intéresse à un exemple qui consiste à détecter puis estimer la valeur moyenne d'échantillons normaux si cette moyenne est non nulle. Les données initiales sont continues mais elles sont quantifiées pour calculer le test minimax. Soit $f_\theta(u)$ le densité de probabilité d'une distribution normale univariée de moyenne $\theta \in \mathbb{R}$. L'écart-type $\sigma = 0.5$ est connu. Quatre moyennes sont possibles : $\theta_0 = 0$ (hypothèse nulle), $\theta_1$, $\theta_2 = 1.5$ and $\theta_3 = 2.5$. La valeur de $\theta_1$, qui vérifie $\theta_0 < \theta_1 < \theta_2$, varie au cours des simulations. Les données réelles sont quantifiées en utilisant $N = 500$ couples $(x_i, C_i)$ tels que les valeurs de quantification $x_i$ sont $x_i = a + (i-1)\Delta$ avec $\Delta = (b-a)/(N-1)$ et les cellules de quantifications $C_i$ sont uniformément réparties sur un intervalle $[a, b]$ de $\mathbb{R}$ où les densités de probabilité sont non négligeables. Le classifieur $\boldsymbol{\delta}(u)$ prend sa décision à partir de la donnée quantifiée $x_i = Q(u)$ quand $u \in C_i$. La distribution $p_{\theta_i}(x)$ de l'hypothèse $\mathcal{H}_i$ est obtenue en quantifiant la densité $f_{\theta_i}(u)$. L'ensemble des actions est égal à l'ensemble des paramètres, i.e., $\Psi = \Theta = \{\theta_0, \theta_1, \theta_2, \theta_3\}$. La fonction de perte est $w(\theta_i, \theta_j) = |\theta_i - \theta_j|$, $\forall \theta_i, \theta_j \in \Theta \setminus \boldsymbol{\theta}_0$. Le niveau prescrit est $\alpha = 0.01$.

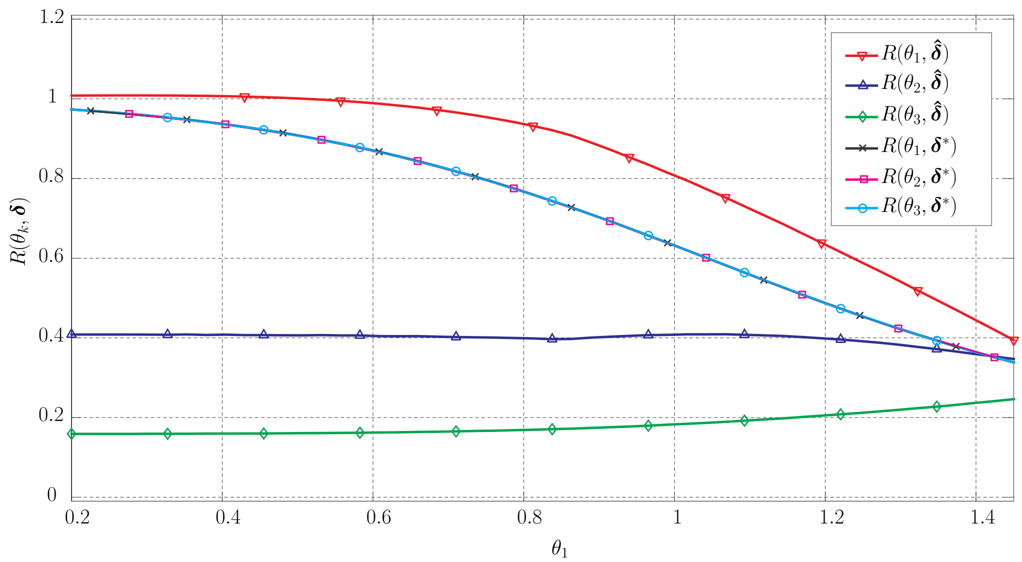


FIGURE 1 – Risques $R(\theta_k, \boldsymbol{\delta})$ pour le test RVG $\hat{\boldsymbol{\delta}}$ et le test minimax $\boldsymbol{\delta}^*$ en fonction de $\theta_1$.

La figure 1 montre $R(\boldsymbol{\theta}_k, \boldsymbol{\delta})$ pour le test RVG (Rapport de Vraisemblance Généralisé) $\hat{\boldsymbol{\delta}}(u)$ et le test minimax $\boldsymbol{\delta}^*(x)$ de niveau $\alpha$ en fonction du paramètre $\theta_1$. Le paramètre $\theta_1$ est le plus proche du paramètre $\theta_0$ de l'hypothèse nulle. La modification de $\theta_1$ permet donc de contrôler le niveau de difficulté du problème de détection et également de modifier la séparabilité des hypothèses alternatives. Le test RVG est un test de Bayes, avec une distribution uniforme des hypothèses alternatives, appliqué aux données continues. Les performances du test RVG sont calculées par intégration numérique et le seuil est choisi pour satisfaire le niveau $\alpha$. Le test minimax égalise les fonctions de risque alors que le test RVG est pénalisé par l'hypothèse $\mathcal{H}_1$ qui est la plus proche de l'hypothèse $\mathcal{H}_0$.

# 4 Conclusion

Ce papier a montré qu'un test minimax de niveau contraint entre plusieurs hypothèses était obtenu en solutionnant un problème de programmation linéaire. La solution du problème fournit le test mais également la distribution a priori la plus défavorable. Ce test peut être utilisé pour traiter des données continues préalablement quantifiées.

# Références

[1] Barankin, E. W. (1951). On systems of linear equations, with applications to linear programming and the theory of tests of statistical hypotheses. *University of California*

*publications in statistics*, 1(8) :161–214.

[2] Baumann, V. (1968). Eine parameterfreie theorie der ungünstigsten verteilungen für das testen von hypothesen. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 11(1) :41–60.

[3] Baygün, B. and Hero, A. O. (1995). Optimal simultaneous detection and estimation under a false alarm constraint. *IEEE Trans. Inf. Theory*, 41(3) :688–703.

[4] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer.

[5] Chang, C.-I. and Davisson, L. (1990). Two iterative algorithms for finding minimax solutions. *IEEE Trans. Inf. Theory*, 36(1) :126–140.

[6] Dantzig, G. B. and Wald, A. (1951). On the Fundamental Lemma of Neyman and Pearson. *The Annals of Mathematical Statistics*, 22(1) :87–93.

[7] Ferguson, T. (1967). *Mathematical Statistics : A Decision Theoretic Approach*. Academic Press.

[8] Goldenshluger, A., Juditsky, A., and Nemirovski, A. (2015). Hypothesis testing by convex optimization. *Electronic Journal of Statistics*, 9(2) :1645–1712.

[9] Krafft, O. (1970). Programming methods in statistics and probability theory. In Rosen, J., Mangasarian, O., and Ritter, K., editors, *Nonlinear Programming*, pages 425–446. Academic Press.

[10] Krafft, O. and Schmitz, N. (1970). A symmetrical multiple decision problem and linear programming. *Operations Research Verfahren*, 7 :126–149.

[11] Lehman, E. (1986). *Testing Statistical Hypotheses, Second Edition*. Chapman & Hall.

[12] Nelson, W. (1966). Minimax solution of statistical decision problems by iteration. *Ann. Math. Statist.*, 37(6) :1643–1657.

[13] Schaafsma, W. (1970). Most stringent and maximin tests as solutions of linear programming problems. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(4) :290–307.

[14] Weiss, L. (1961). *Statistical Decision Theory*. McGraw-Hill series in probability and statistics. McGraw-Hill.