Compte-rendu du TP n°1
La classification bayésienne

Arnaud Aillaud

8 octobre 2022

Table des matières

1 Une première idée des enjeux du problème **2**
1.1 Question 1 2
1.2 Question 3 3
1.3 Question 4 4

2 Apport des méthodes bayésiennes de segmentation **6**
2.1 Question 2 6
2.2 Question 4 7

Liste des tableaux

1 Comparaison des taux d'erreur moyens pour chaque signal en fonction des paramètres de bruit 4
2 Comparaison des taux d'erreur moyens avec classification bayésienne pour chaque signal en fonction des paramètres de bruit 6
3 Comparaison des taux d'erreur moyens pour chaque méthode en fonction des probabilités de chaque classe 8

1 Une première idée des enjeux du problème

1.1 Question 1

1. c. ***Bruiter le signal signal.npy avec un bruit gaussien, puis segmenter le signal bruité suivant le critère du maximum de vraisemblance et afficher sur un même graphique les courbes du signal original, du signal bruité et du signal segmenté.***

Avec le fichier **signal.npy** et les paramètres de la première ligne du tableau de la question 4., on obtient le graphe suivant :

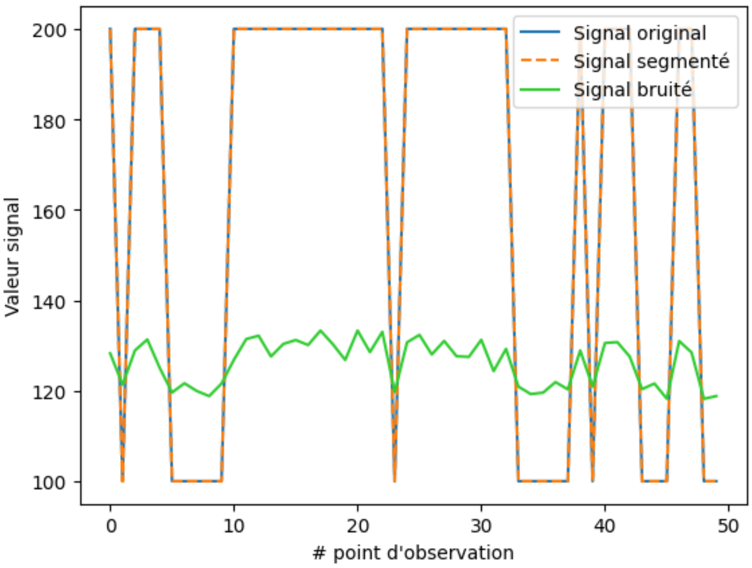
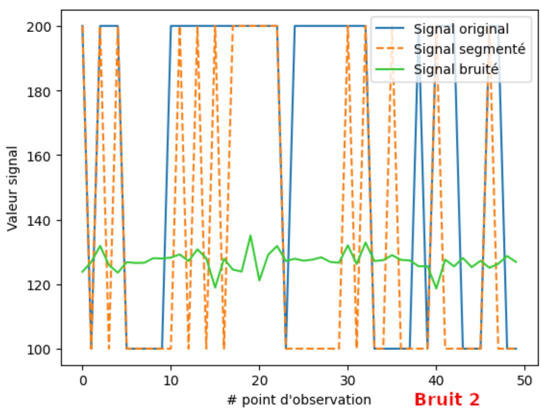
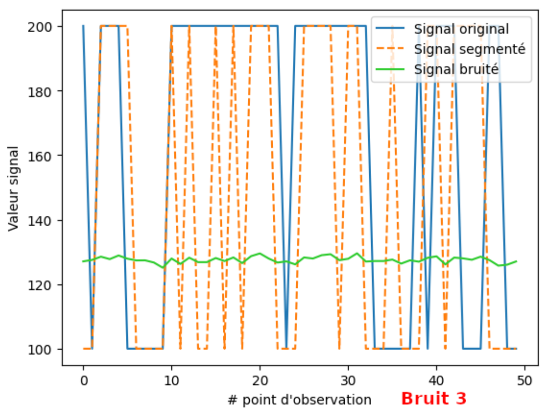


FIGURE 1 – Reconstruction du signal pour m_1=120, m_2=130, σ_1=1 et σ_2=2

Dans ce cas, le signal est parfaitement reconstruit car les paramètres des bruits gaussiens utilisés sont suffisamment différents. A titre de comparaison, on obtient les figures suivantes en utilisant les paramètres de bruit des autres lignes du tableau de la question 4, toujours avec le même signal initial





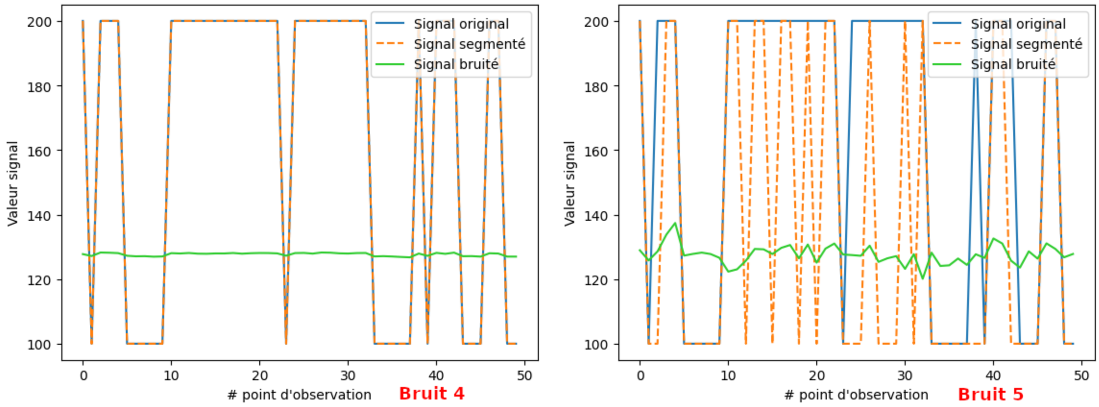


FIGURE 2 – Reconstruction des signaux pour m_1=127, $m_2 \in [127, 128, 128, 128]$, $\sigma_1 \in [1, 1, 0.1, 2]$ et $\sigma_2 \in [5, 1, 0.1, 3]$

Excepté pour le signal reconstruit à partir du bruit de la 4ème ligne du tableau, on observe des erreurs de classification sur les différents graphes (signal segmenté $\neq$ signal original). Les paramètres des densités décrivant le bruit ont donc sans surprise un impact très important sur la qualité du signal reconstruit.

1.2 Question 3

3. a. ***Programmer le calcul de cette erreur moyenne et tracer son évolution au cours des itérations successives.***

Toujours avec le **signal.npy**, et les parmètres m_1=120, m_2=130, σ_1=1 et σ_2=2, on obtient la courbe suivante pour un nombre d'itérations de 2000

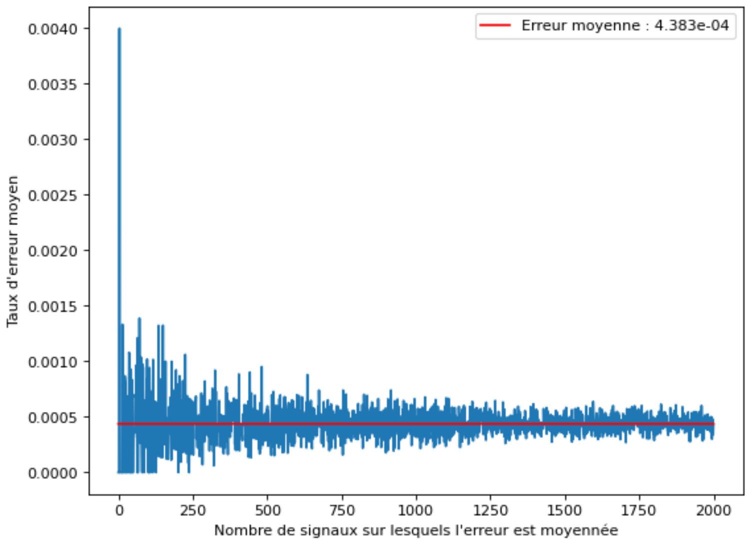


FIGURE 3 – Evolution du taux d'erreur moyen avec les paramètres m_1=120, m_2=130, σ_1=1 et σ_2=2

b. ***Que constate-t-on lorsque T devient très grand ?***

On constate que les variations de taux d'erreur diminuent avec l'augmentation de T et que le **taux d'erreur moyen semble converger vers la moyenne arithmétique des taux d'erreurs moyens** ($\sim 4\text{x}10^{-4}$ dans ce cas).

c. ***Comment expliquer ce phénomène ?***

Ce phénomène s'explique par **la loi des grands nombres**.
En effet, dans le contexte du TP, la fonction L suivante définit une fonction de perte associée à la stratégie $\hat{s}$ choisie :

$$L(\omega_i, \omega_j) = \begin{cases} 0 \text{ si } \omega_i = \omega_j \\ 1 \text{ si } \omega_i \neq \omega_j \end{cases}$$

Le taux d'erreur calculé pour un signal de taille n est ainsi défini comme suit :

$$\begin{aligned} taux &= \frac{1}{n} \sum_{i=1}^{n} L(\hat{s}(y_i), x_i) \\ &\xrightarrow[n \to +\infty]{} \mathbb{E}[L(\hat{s}(Y), X)] \qquad \text{d'après la loi des grands nombres} \end{aligned}$$

d. ***Comment l'interpréter en termes de niveau de bruit ?***

La valeur observée correspond donc à la "perte moyenne" liée à la stratégie $\hat{s}$.
Qualitativement, en termes de niveau de bruit, si cette valeur est **faible**, le signal est donc **peu bruité**. Réciproquement, une valeur **élevée** correspond à un **signal très bruité**.
Le taux d'erreur moyen est donc un bon candidat pour juger de la qualité de la stratégie de classification (mais pas de son optimalité).

1.3 Question 4

4. ***Tester la méthode avec les 6 signaux mis à votre disposition et les bruits du tableau ci-dessous. Présenter les résultats de segmentation dans un tableau récapitulatif et s'en servir pour déterminer ce qu'est un fort ou un faible niveau de bruit.***

Bruits				**Signaux**					
m_1	m_2	σ_1	σ_2	signal	signal1	signal2	signal3	signal4	signal5
120	130	1	2	$4.38\text{x}10^{-4}$	$4.05\text{x}10^{-4}$	$4.02\text{x}10^{-4}$	$4.05\text{x}10^{-4}$	$4.04\text{x}10^{-4}$	$4.04\text{x}10^{-4}$
127	127	1	5	0.207	0.178	0.176	0.177	0.177	0.176
127	128	1	1	0.308	0.309	0.309	0.309	0.309	0.309
127	128	0.1	0.1	$2.12\text{x}10^{-7}$	$4.06\text{x}10^{-7}$	$2.92\text{x}10^{-7}$	$2.90\text{x}10^{-7}$	$2.86\text{x}10^{-7}$	$3.00\text{x}10^{-7}$
127	128	2	3	0.429	0.384	0.381	0.382	0.382	0.381

TABLE 1 – Comparaison des taux d'erreur moyens pour chaque signal en fonction des paramètres de bruit

Les résultats pour les 6 signaux avec les 5 paramètres de bruit proposés sont présentés dans le tableau Table 1 ci-dessus. Le taux d'erreur moyen a été calculé avec T = 2000 pour le fichier signal.npy, 1000 pour signal1.npy et 500 pour les 4 autres.
Un premier constat est que le taux d'erreur moyen varie peu en fonction du signal original pour un couple de paramètres de bruit donné, la valeur de taux d'erreur moyen obtenue est principalement déterminée par les paramètres de bruit. Les signaux ségmentés à partir des signaux bruités 1 et 4 ont des taux d'erreur très faibles ($\sim 10x^{-4}$ et $10x^{-7}$), alors que ceux obtenus à partir des signaux bruités 2, 3 et 5 ont des niveaux de bruit forts. Un fort niveau de bruit est observé quand les densités de bruit utilisées ont des "paramètres proches". L'écart entre les moyennes ne suffit pas à caractériser cette "proximité", comme le montrent le bruit 4 (moyennes très proches, mais écart-type très faible) et le bruit 2 (moyennes identiques, mais n'est pas le signal le plus bruité grâce aux différences d'écart-type).

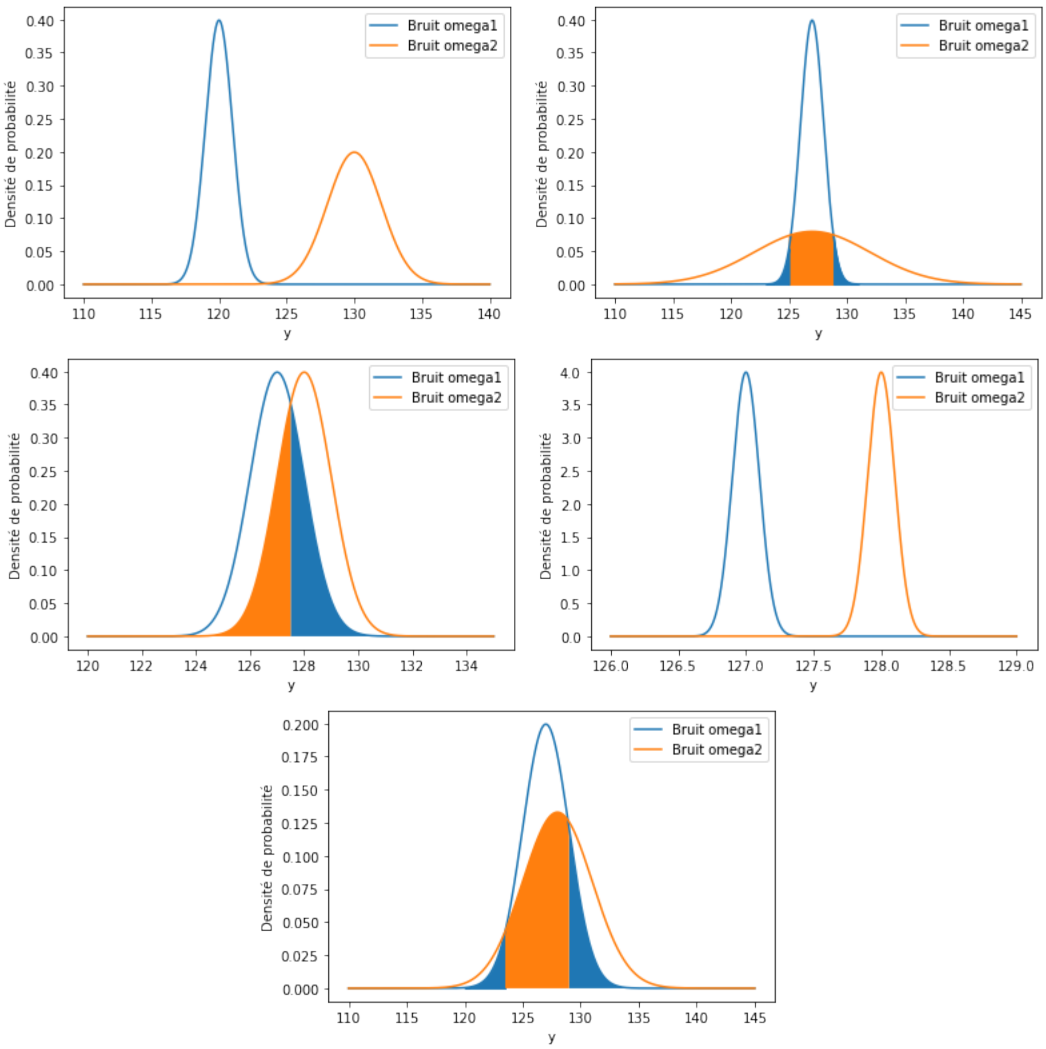
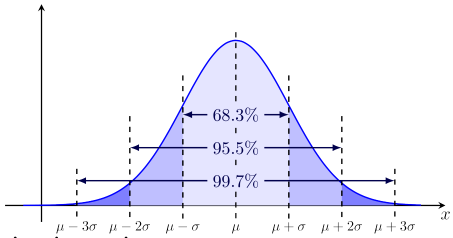


FIGURE 4 – Représentation des densités de bruit pour les différentes valeurs de paramètres données dans le tableau (bruit 1 en haut à gauche → bruit 5 en bas)

Visuellement, un y_i est mal classifié s'il est dans un intervalle dont l'aire sous la courbe (pondérée par la probabilité a priori) est colorée sur la figure ci-dessus : les zones colorées en orange correspondent aux erreurs de classification sur ω_2 ($f_2(y) < f_1(y)$ donc $\hat{s}(y) = cl1$), celles en bleu aux erreurs de classification sur ω_1 ($f_1(y) < f_2(y)$ donc $\hat{s}(y) = cl2$). Un signal est donc peu bruité si l'aire sous les courbes sur l'intervalle d'intersection des densités de probabilité des deux bruits est faible (et réciproquement fortement bruité si cette aire est importante).
Les intervalles où la densité de probabilité de f_1 et f_2 est maximale sont disjoints pour les bruits 1 et 4, ce qui explique le faible niveau de bruit observé.

Pour une gaussienne de moyenne μ et d'écart-type σ, on sait que :

- $\sim$ 68.3% de la densité est inclue dans l'intervalle $[\mu - \sigma, \mu + \sigma]$
- $\sim$ 95.5% de la densité est inclue dans l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$
- $\sim$ 99.7% de la densité est inclue dans l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$



Les bruits ayant des densités gaussiennes, on peut proposer la caractérisation suivante :

$[m_1 - 3\sigma_1,\ m_1 + 3\sigma_1] \cap [m_2 - 3\sigma_2,\ m_2 + 3\sigma_2] = \{\varnothing\}$ $\Rightarrow$ Faible niveau de bruit (intervalles à 99.7% disjoints)

$[m_1 - \sigma_1,\ m_1 + \sigma_1] \cap [m_2 - \sigma_2,\ m_2 + \sigma_2] \neq \{\varnothing\}$ $\Rightarrow$ Fort niveau de bruit (intervalles à 68.3% s'intersectent)

Avec cette caractérisation, on trouve bien que les signaux bruités avec les bruits 1 et 4 sont faiblement bruités, et ceux bruités avec les bruits 2, 3 et 5 sont fortement bruités.

2 Apport des méthodes bayésiennes de segmentation

2.1 Question 2

2. ***Tester la méthode avec les mêmes signaux et les mêmes bruits que précédemment. Présenter les résultats dans un tableau. Comparer et commenter.***

Bruits				**Signaux**					
m_1	m_2	σ_1	σ_2	signal	signal1	signal2	signal3	signal4	signal5
120	130	1	2	4.35x10^{-4}	4.02x10^{-4}	4.03x10^{-4}	4.05x10^{-4}	4.04x10^{-4}	4.03x10^{-4}
127	127	1	5	0.199	0.178	0.176	0.177	0.177	0.176
127	128	1	1	0.281	0.309	0.309	0.309	0.309	0.309
127	128	0.1	0.1	2.69x10^{-7}	2.13x10^{-7}	3.30x10^{-7}	2.82x10^{-7}	2.84x10^{-7}	2.87x10^{-7}
127	128	2	3	0.360	0.384	0.381	0.382	0.382	0.381

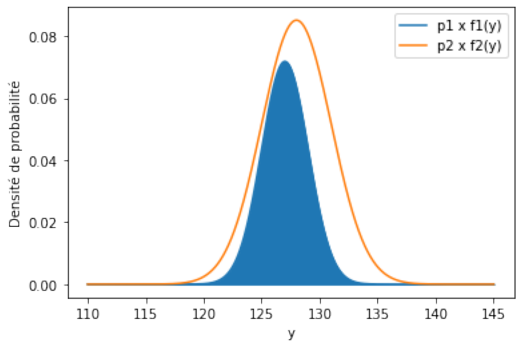
Table 2 – Comparaison des taux d'erreur moyens avec classification bayésienne pour chaque signal en fonction des paramètres de bruit

On remarque peu de différences entre les valeurs obtenues avec la méthode du Maximum de Vraisemblance (MV) et celles obtenues avec la méthode du Maximum de Vraisemblance a Posteriori (MAP).
Les principales différences observées sont une **diminution du niveau de bruit pour le premier signal** avec presque tous les bruits (valeurs représentées en orange dans le tableau ci-dessus).
Pour expliquer ces résultats, il suffit de calculer les probabilités p_1 et p_2 pour chacun des signaux

Signal	p_1	p_2
signal.npy	0.36	0.64
signal1.npy	0.492	0.508
signal2.npy	0.50106	0.49894
signal3.npy	0.49778	0.50222
signal4.npy	0.49854	0.50146
signal5.npy	0.50126	0.49874

La principale différence entre les deux méthodes est la pondération des $p(y|\omega_i)$ par les probabilités p_i dans la méthode MAP. Or, à part pour signal.npy, **ces probabilités sont très proches de 0.5 pour les autres signaux** : elles n'influencent donc que très légèrement la valeur retournée par la stratégie $\hat{s}$. Ainsi, il est tout à fait cohérent d'observer des résultats très proches entre méthode MV et MAP pour ces signaux.

Remarque : le taux d'erreur correspondant au bruit 5 pour le premier signal est égal à la probabilité p_1. En effet, après pondération des gaussiennes avec les probabilités a priori respectives, la courbe représentant la probabilité du bruit appliqué à la classe 1 se trouve entièrement incluse sous la courbe représentant la probabilité du bruit appliqué à la classe 2 (cf figure ci-jointe). Ainsi, la stratégie bayésienne classera renverra toujours la classe 2 dans ce cas, et le taux d'erreur est donc la probabilité d'avoir en entrée un signal correspondant à la classe 1.



2.2 Question 4

4. ***Comparer et commenter les résultats obtenus en segmentant, avec les deux méthodes et les bruits de la question I.4, cinq signaux simulés avec simul2 en faisant varier p1 et p2.***

Les valeurs du tableau suivant ont été obtenues en calculant la moyenne des taux d'erreurs moyens pour 1000 itérations sur des signaux de 500 valeurs. La méthode notée "MV" dans le tableau correspond au maximum de vraisemblance, celle notée "MAP" correspond au maximum de vraisemblance a posteriori.

Méthode	**Bruits**				**Probabilités p1 / p2**				
	m_1	m_2	σ_1	σ_2	0.1 / 0.9	0.3 / 0.7	0.5 / 0.5	0.7 / 0.3	0.9 / 0.1
MV	120	130	1	2	$5.2\text{x}10^{-4}$	$4.6\text{x}10^{-4}$	$4.1\text{x}10^{-4}$	$3.4\text{x}10^{-4}$	$2.9\text{x}10^{-4}$
MAP	120	130	1	2	$3.4\text{x}10^{-4}$	$4.3\text{x}10^{-4}$	$4.0\text{x}10^{-4}$	$3.2\text{x}10^{-4}$	$1.6\text{x}10^{-4}$
MV	127	127	1	5	0.264	0.220	0.176	0.133	0.089
MAP	127	127	1	5	0.100	0.202	0.176	0.121	0.047
MV	127	128	1	1	0.309	0.309	0.308	0.309	0.309
MAP	127	128	1	1	0.099	0.253	0.308	0.253	0.098
MV	127	128	0.1	0.1	$2.8\text{x}10^{-7}$	$2.2\text{x}10^{-7}$	$3.0\text{x}10^{-7}$	$2.5\text{x}10^{-7}$	$2.8\text{x}10^{-7}$
MAP	127	128	0.1	0.1	$1.0\text{x}10^{-7}$	$1.9\text{x}10^{-7}$	$3.0\text{x}10^{-7}$	$2.9\text{x}10^{-7}$	$1.7\text{x}10^{-7}$
MV	127	128	2	3	0.517	0.449	0.381	0.313	0.245
MAP	127	128	2	3	0.100	0.300	0.381	0.265	0.096

Table 3 – Comparaison des taux d'erreur moyens pour chaque méthode en fonction des probabilités de chaque classe

On remarque que **pour** $p_1 = p_2 = 0.5$**, les résultats obtenus avec les 2 méthodes sont identiques**, ce qui est cohérent avec les observations faites à la question précédentes.

En outre, on remarque que les niveaux de bruit obtenus avec la méthode MAP sont **toujours** inférieurs à ceux obtenus avec la méthode MV, et que cette différence augmente avec l'augmentation de la différence entre les probabilités p_1 et p_2. Ceci s'observe particulièrement bien avec le bruit 3 par exemple :

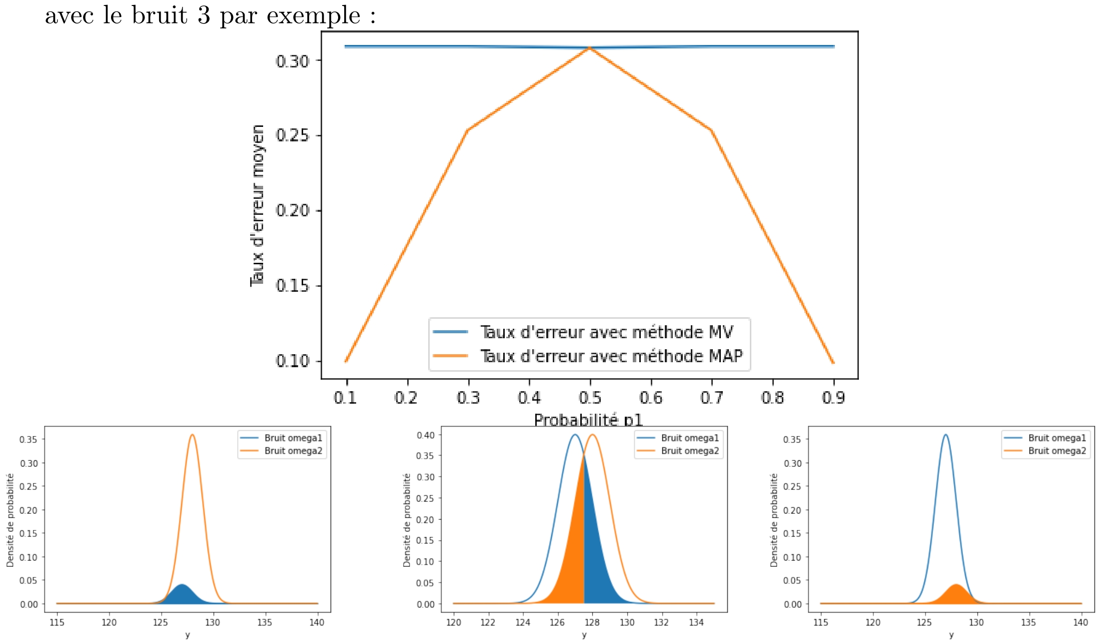


Figure 5 – Comparaison des taux d'erreur obtenus avec méthode MV et méthode MAP pour différentes valeurs de p1 pour le bruit 3

Les 3 gaussiennes sous la figure suivante présentent les densités du bruit 3 pour les valeurs de $p_1 \in [0.1, 0.5, 0.9]$ avec la méthode MAP.
De même que pour la remarque faite à la question 2., on observe que le taux d'erreur tend vers la probabilité p_1 (resp. p_2) lorsque la courbe de la densité du bruit associée à ω_1 (resp. ω_2) est entièrement située sous la courbe de la densité du bruit associée à ω_2 (resp. ω_1). Ce résultat s'observe également pour le bruit 5.

De manière générale, le niveau de bruit est inférieur en utilisation la méthode MAP. Ceci confirme donc que **la méthode MAP est plus optimale que la méthode MV**, et illustre bien **l'apport du contexte** que représente la probabilité a priori.