ESTIMATION DU TAUX D'ERREUR BINAIRE PAR LA METHODE DU NOYAU.

OUATTARA Ardjouma, BAMOUNI Kevin, UO/LANIBIO/MASTER 1 MAIME

October 28, 2015

1 Introduction

La probabilité d'erreur binaire PEB, qui littéralement désigne la probabilité qu'un bit transmit soit erroné a la réception, est une valeur qui permet d'évaluer les performances d'un système de communication numérique, de connaitre le canal afin d'y apporter d'éventuelle amélioration pour assurer une qualité optimale de communication. Cependant pour des systèmes évolués et assez complexes il est difficile voir quasiment impossible de déterminer avec exactitude la PEB, pour cela l'on utilise une estimation, le TEB (taux d'erreur binaire). La méthode la plus connu et la plus basique de calcul du TEB est la méthode dite de Monte Carlo MC, qui consiste à simuler l'émission d'un nombre N de bit connu, dans un canal bruité, dénombrer le nombre d'erreur e à la réception et à faire le rapport e/n, pour obtenir le TEB. Cette méthode au procédé simple a la principale conséquence d'imposer un cout de calcul important, car pour obtenir une valeur plus proche de la PEB, il faut que N tende vers $+\infty$. A titre d'exemple, pour un canal qui impose une PEB de 10^{-6}, pour obtenir un TEB à 10^{-1} près par la méthode de MC, il faudrait simuler $N = 108$ bit transmit avec un nombre e d'erreur de 10^2.

Dans le but d'évaluer le TEB de manière plus rapide et au moins aussi efficace que MC, de nouvelles méthodes ont été proposées dont la méthode du noyau que nous étudierons. Elle est basée essentiellement sur l'estimation, par une technique itérative et non-paramétrique, de la fonction densité de probabilité des valeurs observées au récepteur.

2 Approche du problème

Soit la chaine de communication suivante :

Soit $(b_i)_{1\leq i\leq N}$ les bits modulés à transmettre, X_i correspondants aux b_i atténués (après passage dans le canal), la réception souple et les $\hat{b}_i$ correspondants à la décision dure.

Le principe de la décision dure est le suivant :

$$\begin{cases} X_i \gneq 0 \Rightarrow \hat{b}_i = +1 \\ X_i \lneq 0 \Rightarrow \hat{b}_i = -1 \end{cases}$$

Soit $f_X(x)$ la fonction densité de probabilté des $(X_i)_{1\leq i\leq N}$.

Soit π_+et π_- la probabilité que le bit transmit b_i, soit un $+1$, respectivement un -1 c'est à dire :

$$\begin{cases} \pi_+ = P\left[b_i = +1\right] \\ \pi_- = P\left[b_i = -1\right] \end{cases}$$

où $\pi_+ + \pi_- = 1$.

π_+ et π_- ne sont pas connu au récepteur.

La probabilité d'erreur binaire (PEB) correspond essentiellement à décider un $+1$ à la décision dure alors que le bite transmit était un -1, et vice versa.

Ainsi la PEB P_e peut être modélisé par :

$P_e = P\left[\hat{b}_i \neq b_i\right]$

$P_e = P\left[(X \lneq 0),(b_i\text{=+1})\right] + P\left[(X \lneq 0),(b_i\text{=-1})\right]$

$P_e = P\left[X \lneq 0/b_i\text{=+1}\right] \times P\left[b_i = +1\right] + P\left[X \lneq 0/b_i\text{=-1}\right] \times P\left[b_i = -1\right]$

$P_e = P\left[X \lneq 0/b_i = +1\right] \times \pi_+ + P\left[X \gneq 0/b_i\text{=-1}\right] \times \pi_-$

En utilisant les densités on peut écrire :

$$P_e = \pi_+ \int_{-\infty}^{0} f_X^{b_+}(x)\,dx + \pi_- \int_{0}^{+\infty} f_X^{b_-}(x)\,dx \tag{1}$$

où $f_X^{b+}(x)$ est la fonction de densité conditionnel de X tel que $b_i = +1$ et $f_X^{b-}(x)$ est la fonction de densité conditionnel de X tel que $b_i = -1$.

Au vu de cette expression de la PEB, on peut en déduire que pour trouver la PEB, on trouve les expressions des fonctions de densité de probabilités conditionnelles des X_i avant de les intégrer et d'en faire la somme.

Cependant nous ne disposons pas de l'expression algébrique des fonctions de densités conditionnelles qui dépendent, du modèle du canal de communication et de la configuration du recepteur. Donc il est très difficile voir impossible de trouver la loi de distribution des X_i donc de leur densité de probabilités conditionnelles.

C'est dans le but d'apporter une solution à ce problème, qui est de trouver les expressions des densités, qu'intervient la méthode du noyau. Méthode qui va permettre d'estimer donc les densités conditionnelles $f_X^{b+}(x)$ et $f_X^{b-}(x)$. Cette technique est non paramétrique, du fait qu'aucune hypothèse n'est faite sur la loi vu la difficulté d'apprentissage des paramètres, et se base donc essentiellement sur la répartition (spatiales) des observations X_i.

3 Estimation non-paramétrique de la fonction de densité : Méthode du noyau.

Pososns $C = \{X_1, \ldots, X_N\}$, Supposons les classes, $C_- = \{X_i, 1 \leq i \leq N_- \diagup b_i = -1\}$ et $C_+ = \{X_i, 1 \leq i \leq N_+ \diagup b_i = +1\}$ tels que $C = C_- \cup C_+$ et $N = N_- + N_+$.

En utlisant la méthode du noyau, on peut alors écrire une estimation de chaque fonction densité de probabilité en utilisant les classes C_- et C_+, comme suit :

$$\hat{f}_{X,N+}^{b+}(x) = \frac{1}{N_+ \times h_{N+}} \sum_{X_i \in C_+} K\left(\frac{x - X_i}{h_{N+}}\right) \tag{2}$$

$$\hat{f}_{X,N-}^{b-} = \frac{1}{N_- \times h_{N-}} \sum_{X_i \in C_-} K\left(\frac{x - X_i}{h_{N-}}\right) \tag{3}$$

où les h_{N_+} et h_{N_-}sont appellés paramètres de lissage dépendant de la taille des échantillons N_-et N_+.

$K\left(.\right)$est une quelconque fonction de densité de probabilité paire et régulière de variance 1 et de moyenne *nulle*.

Dans la suite, nous utiliserons le noyau gaussien :

$$K(x) = \frac{1}{2\sqrt{\pi}}\exp(\frac{-x^2}{2}) \tag{4}$$

Dans la méthode du noyau, le choix des paramètres de lissage est très important du fait qu'ils influeent directement sur l'exactitude, la précision de l'estimation des fonctions de densité. Donc il convient de trouver le paramètre de lissage dit optimal.

D'après (doc), il est montré que si $h_N{\to}0$ lorsque $N\ {\to}\infty$, alors l'estimateur de la densité $\hat{f}_{X,N}(x)$ est assymptotiquement non-biaisé. il est aussi montré que si $h_N{\to}0$ et $Nh_N\ {\to}\infty$lorsque $N\ {\to}{+}\infty$ alors l'Erreur Quadratique Moyenne (EQM) tend vers0.

(h_{N+-}désigne h_{N_+} ou h_{N_-}, N_{+-} désigne N_-ou N_+, f_X^{b+-} désigne f_X^{b+} ou f_X^{b-})

$\lim_{N\to+\infty} E\left[\left(\hat{f}_{X,N}(x) - f_X(x)\right)^2\right] = 0$

En outre le paramètre de lissage optimal s'obtient en minimisant l'Intégration de l'Erreur Quadratique Moyenne (IEQM).

d'apres (doc) une approximation de IEQM est donné par la formule suivant: $IEQM = \frac{M(K)}{Nh_N} + \frac{J\left(f_X\right)h_N^4}{4}$.

Ainsi en minimisant IEQM pécédent, ont obtient que le paramètre de lissage optimal sera :

$$h_N^* = N^{-\frac{1}{5}}\left(J\left(f_X^{b+-}\right)\right)^{-\frac{1}{5}}\left(M\left(K\right)\right)^{+\frac{1}{5}}, \tag{5}$$

où $M\left(K\right) = \int_{-\infty}^{+\infty} K^2(x)\,dx$ et $J\left(f_X^{b+-}\right) = \int_{-\infty}^{+\infty}\left\{\left(f_X^{b+-}\right)''(x)\right\}^2 dx$

Comme h_N^* dépend de la densité conditionnelle inconnu f_X^b, alors il a été suggéré d'utilisé le noyau gaussien dont le paramètre $M\left(K\right)$ est donné par :

$M(K) = \frac{1}{2\sqrt{\pi}}$,

et dans le cas où nous supposons que f_X^{b+-} suit une loi normale de paramètres (m_{+-}, σ_{+-}^2) alors

$J\left(f_X^{b+-}\right) = \frac{3}{8\sqrt{\pi}\sigma_{+-}^5}$,

(m_{+-}, σ_{+-}^2) peuvent être respectivement la moyenne et la variance de C_+ou C_-.

Ainsi le paramètre de lissage optimal h_{N+-}^*, en utilisant un noyau gaussien et une densité de probabilité conditionnelle gaussienne, est :

$$h_{N_{+-}}^* = \left(\frac{4}{3N_{+-}}\right)^{\frac{1}{5}} \sigma_{+-} \tag{6}$$

4 Estimation de la PEB : le TEB.

Comme nous avons pu le voir, l'expression du paramètre de lisssage orptimal dépend de N_{+-}, qui est le nombre de X_i reçu sachant que le bit qui a été transmit est un $+1$ respectivement un -1. Supposant les $b_i$inconnus à la réception, il se pose un problème de detection. Car il faudrait connaître le nombre de $+1$ et de -1 transmit sans pour autant connaître le message exact d'origine. Ainsi, le calcul du taux d'erreur binaire qui sera proposé devra effectuer la classification des X_i dans les deux classes C_- et C_+ (évoqués au paragraphe précédent) puis l'estimation des fonctions de densité conditionnelles et des probabilités dites "à priori" π_+ et π_-selon une méthode itérative qui utilisera les probabilités dites "à postériori" $P\left[b_i = +1 \diagup X_i\right]$ et $P\left[b_i = -1 \diagup X_i\right] \forall i = 1, ..., N$. Soit T le nombre d'itération ($t = 1, ..., T$), qui est choisi de telle sorte que à $t = T$, $\left|h_{N+-}^t - h_{N+-}^{t-1}\right| \approx 0$. A chaque itération t, on estime toutes les valeurs de $P\left[b_i = +1 \diagup X_i\right]$ et $P\left[b_i = -1 \diagup X_i\right] \forall i = 1, ..., N$ en utilisant les valeurs de π_+ et π_- et les fonctions de densités estimées à l'itération $t-1$. En d'autres termes on utilise les probabilité à priori pour calculer les probabilités à postériori et à chaque itération on calcul deux fonction de probabilités conditionnelles en fonction de celles de l'itération précédente. Notons q'à l'itération t=1, on ne dispose d'aucune information autre que les X_i brutes reçu, on les utilise donc pour calculer les valeurs initiales de π_+ et π_-.

Soit $\theta = (\pi_+, N_+, h_{N_+}, \pi_-, N_-, h_{N_-})$. la valeur de θ est itérativement calculée, et à l'itération T , on obtien une valeur estimée qui sera utilisée pour calculer le TEB en utulisant la formule suivante :

$$\hat{p}_{e.N} = \frac{\pi_+}{N_+} \sum_{i \in C_+} Q\left(\frac{X_i}{hN_+}\right) + \frac{\pi_-}{N_-} \sum_{i \in C_-} Q\left(-\frac{X_i}{hN_-}\right) \tag{7}$$

Avec $Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-t^2/2\right) dt$. Cette formule est obtenue en introduisant les expressions des fonctions de densités conditionnelles $f_X^{b_+}(x)$ (2) et $f_X^{b_-}(x)$ (3) et celles de θ dans la formule analytique de la PEB (équation (1)).

Pour résumer, afin de pouvoir calculer le TEB $\hat{p}_{e.N}$, nous devons tout d'abord estimer à l'aide d'une méthode itérative $\theta = (\pi_+, N_+, h_{N_+}, \pi_-, N_-, h_{N_-})$.

Tout d'abord il faudra trouver les classes $C_- = \{X_i, 1 \leq i \leq N_- \diagup b_i = -1\}$ et $C_+ = \{X_i, 1 \leq i \leq N_+ \diagup b_i = +1\}$ (méthode de classification) puis en déduire les paramètres N_- et N_+.

Ensuite il faudra estimer, itérativement, les probabilités “à priori” π_+et π_-. Il faut noter que dans un premier temps on utilise les probabilités π_+et π_- pour estimer $P\left[b_i = +1 \diagup X_i\right]$ et $P\left[b_i = -1 \diagup X_i\right]$, puis dans un second temps on utilise $P\left[b_i = +1 \diagup X_i\right]$ et $P\left[b_i = -1 \diagup X_i\right]$ pour estimer π_+et π_- (ceux qui nous interressent dans l'estimation de θ).

En utilisant les paramètres pécédemment calculés on estime le paramètre de lissage optimal h en utilisant sa formule (équation (6)).

Après avoir obtenu l'estimation de θ on calcul la valeur estimée du TEB en utilisant sa formule. (équation(7)).

5 Classification des $X_1, ..., X_N$ et estimation de θ en utilisant un algorithme espérance-maximisation (EM) Stochastique.

Une définition tirée de wikipédia nous dit que: *L'algorithme espérance-maximisation (en anglais Expectation-maximisation algorithm, souvent abrégé EM), proposé par Dempster et al. (1977), est une classe d'algorithmes qui permettent de trouver le maximum de vraisemblance des paramètres de modèles probabilistes lorsque le modèle dépend de variables latentes non observables.*

(Référence: A.P. Dempster, N.M. Laird et Donald Rubin, « Maximum Likelihood from Incomplete Data via the EM Algorithm », Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no 1,1977, p. 1–38 (JSTOR 2984875))

L'algorithme EM effectue de manière itérative des estimations et des maximisations afin d'obtenir une estimation du maximum de vraissemblance lorsque les données sont jugées incomplètes ou érronées. le EM stochastique introduit une notion d'aléatoire dans l'algorithme Em.

Nous voulons classer les receptions souples $X_1, ..., X_N$ en deux classes tout en supposant que les bits de l'information b_i transmis initialements sont totalement inconnus. Nous utiliserons donc l'algorithme EM stochastique, qui d'après sa définition permettrait d'estimer θ.

l'algorithme se déroule essentiellement en 3 étapes :

A. L'estimation

On estime les probabilités, à chaque itération, "à postériori" $P\left[b_i = +1 \diagup X_i\right]$ et $P\left[b_i = -1 \diagup X_i\right]$ en utilisant les probabilités "à priori"π_+et π_- en utilisant la formule suivante :

$$
\begin{cases}
\rho_{i_+}^{(t)} = P\left[b_i = +1 \diagup X_i\right] & = \frac{\pi_+^{(t-1)} \times \hat{f}^{b+}_{X,N_+^{(t-1)}}(x_i)}{\pi_+^{(t-1)} \times \hat{f}^{b+}_{X,N_+^{(t-1)}}(x_i) + \pi_-^{(t-1)} \times \hat{f}^{b-}_{X,N_-^{(t-1)}}(x_i)} \\
\rho_{i_-}^{(t)} = P\left[b_i = -1 \diagup X_i\right] & = \frac{\pi_-^{(t-1)} \times \hat{f}^{b-}_{X,N_-^{(t-1)}}(x_i)}{\pi_+^{(t-1)} \times \hat{f}^{b+}_{X,N_+^{(t-1)}}(x_i) + \pi_-^{(t-1)} \times \hat{f}^{b-}_{X,N_-^{(t-1)}}(x_i)}
\end{cases}
\tag{8}
$$

On initialise π_+et π_-en utilisant les X_i:

$$
\begin{cases}
\pi_+^0 & = \frac{|X_i \gneq 0|}{N} \\
\pi_-^0 & = \frac{|X_i \lneq 0|}{N}
\end{cases}
\tag{9}
$$

B. La maximisation

A cette étape commence l'estimation de θ.

A chaque itération t, On obtient une estimation de θ en maximisant l'espérance conditionnelle $Q(\theta^{(t)})$ du logarithme de la vraissemblance. De la maximisation de $Q(\theta^{(t)})$ on en déduit l'estimation des probabilités "à priori" :

$$
\begin{cases}
\pi_+^t & = \frac{1}{N} \sum_{i=1}^{N} \rho_{i_+}^{(t)} \\
\pi_+^t & = \frac{1}{N} \sum_{i=1}^{N} \rho_{i_-}^{(t)}
\end{cases}
\tag{10}
$$

C. La classification

L'estimation des paramètres $N_+, h_{N_+}, N_-, h_{N_-}$et des fonctions de densité conditionnelles passent par la classification des X_i en $C_-, et\, C_+$.

La technique de EM stochastique, combine la vraisemblance et les probabilités $U_1^{(t)}, ..., U_N^{(t)}$ d'une variable aléatoire U de loi uniforme sur $[0,1]$ pour effectuer la classification selon le procédé suivant:

$$
\begin{cases}
C_+^{(t)} & = \left\{ X_i : \rho_{i_+}^{(t)} \geq U_i^{(t)} \right\} \\
C_-^{(t)} & \qquad = \bar{C_+^t}
\end{cases}
\tag{11}
$$

De la classification on en déduit les autres paramètres:

- $N_+^{(t)} = |C_+^{(t)}|$et $N_-^{(t)} = |C_-^{(t)}|$
- Les paramètres de lissage optimaux $h_{N_{+-}}^{(t)}$ sont estimés à l'aide de la formule (6).
- Les fonctions de densités conditionelles sont estimées à l'aide des formules (2) et (3).