Sur un problème d'estimation pour des processus de Poisson composés et filtrés

Alfred O. Hero III
Dept. of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109-2122, USA
email: hero@eecs.umich.edu

Résumé

Les processus de Poisson composés et filtrés forment une classe de modèles très utile pour certaines applications en traitement du signal, traitement de l'image, et télécommunications. Une des premières applications de ce type de modèle en traitement de l'image a été proposée par Bernard Picinbono en 1955 pour la distribution des grains d'argent dans un film photographique. Ici on introduit un modeèle de Picinbono généralisé dont l'objectif est d'estimer les paramètres du processus de Poisson filtré en présence de lissage spatial et de bruit additif Gaussien. En posant le problème de l'estimation dans le contexte de la théorie de l'information, on est conduit à une représentation du modèle par la composition d'un canal Poissonnien et d'un canal Gaussien. Cette composition mène naturellement à un estimateur paramétrique du type "expectation-maximization" (EM) et à une borne du type "distortion-rate" sur l'erreur d'estimation.

Abstract

Compound and filtered Poison processes are useful models for many applications in signal processing, image processing, and communications. One of the earliest imaging applications of these models was proposed by Bernard Picinbono in a 1955 paper on silver dye photographs. In this paper we treat a generalized model with the primiary objective being to estimate parameters of the filtered Poisson process in the presence of spatial smoothing and additive Gaussian noise. By imbedding the estimation problem into the context of information theory we decompose the model into the cascade of a discrete event Poisson process channel and a continuous Gaussian waveform channel. This naturally leads to a expectation-maximization (EM) type estimation algorithm and a distortion-rate lower bound on estimation error.

Remerciements

Une grande partie de ce travail a été effectuée à l'Ecole Normale Supérieure de Lyon pendant la période 1.99 - 3.99 lors de la mission sabbatique de l'auteur. L'auteur veut donc remercier le Laboratoire de Physique pour son accueil et le CNRS pour le financement de son poste temporaire à l'ENS-Lyon.

1 Introduction

Cet article traite un problème de l'estimation paramétrique à partir d'une mesure qui provient d'un processus de Poisson filtré en présence d'un bruit Gaussien additif. Notamment dans le domaine du signal on peut citer les travaux de Gatti et Svelto en détection des particules sous-atomiques nucléaires [17]; ceux de O'Reilly en détection de paquets de photons en communication optique [31]; ceux de Dorfman en analyse des signaux neuronaux [11]; ceux de Bruckstein, Shan et Kailath en localisation des échos acoustiques composés [7], ceux de Mendel en analyse des signaux sismiques [29]; et ceux de Faure en révérberation acoustique sous-marine [14].

Ce modèle est aussi pertinent dans diverses applications du traitement de l'image. On peut citer les travaux de Amoss et Davidson [1] et de Wernick et Morris [39] sur la détection d'une image ayant une intensité faible; ceux de Salomon et Glavich [35] sur la localisation et poursuite des étoiles en astronomie optique; ceux de Kazovsky [28] sur l'estimation de la position d'un faisceau de lumière sur un réseau de photo-détecteurs; et ceux de Clinthorne, Rogers, Shao et Koral [8] sur la localisation de la photo-scintillation induite par l'intéraction d'un rayon gamma avec un crystal de sodium.

Dans son papier de 1955 [32], qui précède les papiers cités ci-dessus d'au moins une quinzaine d'années, Picinbono considérait un modèle pour la transparence d'un film photographique formé d'une superposition de grains d'argent de taille variable et dont le nombre et les positions de ces grains suivent une loi de Poisson homogène. Il est particulièrement intŕessant pour ce numéro spécial de la revue Traitement du Signal, de noter que, à ma connaissance, ce papier était le premier à proposer un modèle de Poisson filtré dans le domaine de l'image.

Ce travail de Picinbono est le point de départ pour notre article. On donne une extension du modèle de Picinbono qui permettra l'étude du comportement des estimateurs de la distribution de la taille des grains en présence de lissage spatial et de bruit additif Gaussien. Les résultats de Picinbono présentés dans [32] sur la forme de la moyenne et de la covariance spatiales du modèle s'y montreront essentiels pour notre étude.

2 Un Modèle Général

Soit $\Theta = [\Theta_1, \ldots, \Theta_p]^T$ un vecteur de variables aléatoires (v.a.) prenant des valeurs $\theta = [\theta_1, \ldots, \theta_p]^T$ dans un sous-ensemble de $\mathbb{R}^p$ et ayant une densité $f_\Theta(\theta)$. Notre but sera de développer un estimateur de type maximum à posteriori (MAP) de Θ et d'analyser les bornes sur ses performances. L'estimation de Θ sera basée sur une observation d'une image $Y = \{Y(u) : u \in I\}$ qui est composée d'une image S, dite le signal, et une image W, dite le bruit. Ici I dénote le support de l'image I, supposé carré $I = [-a, a] \times [-a, a]$, et $|I|$ dénotera sa superficie $|I| = 4a^2$. Le signal S est engendré par un processus ponctuel marqué $dM = \{dM(u) : u \in I\}$, dont la distribution dépend de Θ, et un lissage dû à l'acquisition optique. Le processus dM crée N points $\{t_i\}^N$ en I ainsi que N points, appelés des "marques", $\{R_i\}_{i=1}^N$ dans un espace de marques $(0, \infty)$.

On suppose que, en conditionnant sur $\Theta = \theta$ et N, les points $\{U_i\}_{i=1}^N$ et les marques $\{R_i\}_{i=1}^N$ sont mutuellement indépendants et indépendamment distribués (i.i.d.) ayant des densités marginales $f_{U|\Theta}(u|\theta)$ et $f_{R|\Theta}(r|\theta)$, respectivement. On suppose aussi que N est un v.a. de Poisson de paramètre $E[N|\Theta] = E[N] = \Lambda > 0$ indépendant de Θ.

Ces hypothèses impliquent que, conditionnés sur Θ, les points U_i et les marques R_i forment un processus de Poisson marqué avec intensités $\lambda f_{U|\Theta}(u|\theta)$, sur $u \in I$, et $\lambda f_{R|\Theta}(r|\theta)$, sur $r \in (0, \infty)$, respectivement, où $\lambda = \Lambda/|I|$ est l'intensité spatiale du processus. Le processus ponctuel dM est ainsi dit "un processus de Poisson marqué et composé." La distribution jointe de (dM, Θ) ayant une forme analytique bien connue, le problème de l'estimation de Θ à partir de dM est relativement facile à étudier, e.g., voir [21, 5, 4, 30, 38, 15].

Le problème général qui nous concerne ici est l'estimation de Θ a partir d'une observation Y dérivée par filtrage, lissage, et bruitage du processus de Poisson marqué et composé dM:

$$Y(u) = S(u) + W(u), \quad u \in I \tag{1}$$

où W est un bruit blanc spatial Gaussien de moyenne nulle indépendant de S et ayant pour densité spectrale $N_o/2$, et $S(u)$ est un signal ayant la représentation générale

$$S(u) = h(u) \star g(u; dM). \tag{2}$$

Ici $h(u)$ est une fonction de lissage spatial, qu'on suppose sphériquement symétrique ($h(u) = h(\|u\|)$), "$\star$" dénote la convolution spatiale, et $g(u;dM)$ est un filtrage du processus dM.

La fonction de filtrage $g(\bullet;dM)$ spécifie la façon dont les points et les marques de dM sont insérés dans l'image Y. Dans certaines applications, on pourrait raisonnablement proposer que la fonction de filtrage produise une superposition linéaire de N fonctions de Dirac dans le rectangle I:

$$g(u;dM) = \sum_{i=1}^{N} R_i \delta(u - U_i). \tag{3}$$

Ce modèle linéaire est une généralisation à deux dimensions du modèle de "shot noise," qui est pertinent dans les applications de photo-détection [17, 25, 20], de détection de potentiels neurologiques [36], et d'autres applications [29, 14]. En deux dimensions, nous avons adopté le modèle (3) de g en [23] pour un problème d'estimation de la position θ sur un détecteur optique spatial d'un faisceau de lumière faible et d'intensité connue. Dans le contexte d'un détecteur CCD, la mesure Y est le champ électrique induit à la sortie du détecteur, $\{U_i\}$ sont les positions de photons incidents sur le détecteur, ayant l'intensité $\lambda(u|\Theta) = \lambda(u - \Theta)$, $\{R_i\}$ sont les charges électriques déposées par chacun de ces photons, présumées indépendantes des $\{U_i\}$, et W est un bruit dû à l'électronique.

Par contre, dans ce papier on considère un cas de filtrage nonlinéaire pour lequel $g(u;dM)$ a la représentation

$$g(u;dM) = \prod_{i=1}^{N} D\left(\frac{u - U_i}{R_i}\right), \tag{4}$$

où $D(u)$ est la fonction indicatrice d'un disque de surface π et de centre zéro

$$D(u) = \begin{cases} 1, & \|u\| \leq 1 \\ 0, & \|u\| > 1 \end{cases},$$

et U_i et R_i dénotent, respectivement, la position et le rayon de chaque disque. Il est important de noter que $g(u;dM)$ est une fonction binaire qui prend la valeur 1 s'il existe au moins un disque couvrant le point u et la valeur 0 dans le cas contraire.

Dans la Figure 1, on montre trois réalisations de l'image S et Y pour une densité $f_{R|\Theta}(r|\theta)$ linéaire de la forme

$$f_{R|\Theta}(r|\theta) = \begin{cases} \theta 2r + (1-\theta), & r \in [0,1] \\ 0, & r \notin [0,1] \end{cases}, \tag{5}$$

où $\theta \in [-1, 1]$. Chaque ligne de la Figure 1 correspond à un choix différent de θ. Pour $\theta = 0$ les rayons R_i sont uniformes. Pour $\theta = -0.8$ et $\theta = 0.8$ ces rayons ont des densités plus concentrées vers 0 et 1, respectivement.

La fonction $1 - g(u; dM)$ est la fonction de transparence $T(u)$ introduite par Picinbono en 1955 [32] pour un modèle de la distribution des grains d'argent dans un film photographique. Le modèle (4), qui est aussi connu sous les noms de "modèle Booléen" et "processus de recouvrement" [19], a été proposé plus tard pour d'autres applications rélatives à la biologie et aux sciences de la vie [10], à la stéréologie [13], et à la balistique [12]. Bien que dans certains domaines d'application les densités $f_{U|\Theta}$ et $f_{R|\Theta}$ puissent toutes les deux dépendre du paramètre $\Theta = \theta$, on se limitera dans cet article au cas plus simple où f_U est une densité uniforme sur I et on cherchera à estimer un attribut de la densité $f_{R|\Theta}$, e.g. la surface moyenne $\pi E[R_i^2|\Theta]$ des disques. Le problème général de l'estimation du parametre Θ de $f_{U|\Theta}$ s'appelle la granulomètrie et plusieurs estimateurs ad hoc ont été proposés, voir par exemple [33, 37].

Finalement, remarquons que le lissage de l'image $g(\bullet; dM)$ et l'addition de bruit, prescrits par notre modèle général (1), nous permettent de prendre en compte les effets de la transcription physique de l'image, e.g. par un procédé d'enregistrement électronique, mécanique, ou chimique. Malheureusement, le bruitage et le lissage rendent impossible une représentation analytique de la distribution jointe de Y, Θ et donc l'estimation optimale de Θ est beaucoup plus difficile à partir du modèle (1) qu'à partir des mesures directes dM du processus ponctuel. Même dans le cas plus simple d'une seule dimension, on doit se contenter d'une approximation de l'estimateur optimal $\hat{\Theta}$; e.g. voir [24, 20, 3, 27, 18, 2].Le but de notre article sera de montrer que, malgré cette difficulté de représentation analytique, les moments d'ordre deux du pair (Y, dM) sont suffisants pour construire des approximations simples de l'estimateur MAP-EM et pour borner l'erreur de l'estimation.

3 Représentation statistique par un canal d'information

Les mesures Y sont reliées aux paramètres Θ par la densité conditionnelle $f_{Y|\Theta} = \{f_{Y|\Theta}(y|\theta)\}_{y,\theta}$ dont $l(\theta) = \ln f_{Y|\Theta}(y|\theta)$ est appelée la fonction de log-vraisemblance. Puisque Θ est un vecteur de paramètres aléatoires, il existe une interprétation de Shannon très utile qui associe $f_{Y|\Theta}$ aux probabilités de transition d'un canal de mesure C (voir figure 2). Dans le langage de la théorie de l'information de Shannon, l'entrée de ce canal Θ est la v.a. produite à la *source* et la sortie de ce canal Y est la v.a. produite à la *destination*. Supposons que X soit une v.a. quelconque. L'identité de Bayes nous donne: $f_{Y|\Theta}(y|\theta) = \int_X f_{Y|X,\Theta}(y|x,\theta) f_{X|\Theta}(x|\theta)dx$. Quand la v.a. X est définie telle que $f_{Y|X,\Theta}(y|x,\theta)$ ne dépende pas de θ, l'identité de Bayes affirme que le canal C peut être décomposé en une cascade de deux canaux C_1 et C_2 dont les probabilités de transition sont spécifiées par $f_{X|\Theta}$ et $f_{Y|X}$, respectivement. Dans le langage de l'algorithme EM, dont on reparlera dans la suite, un tel X est un jeu de données complètes qui porte plus d'information sur Θ que Y.

Dans le contexte du modèle (1), un jeu de données complètes naturelle est le processus ponctuel marqué dM, ce qui donne la décomposition illustrée dans la Figure 2. Dans ce cas, puisque, étant donné dM, la partie signal S de $Y = S + W$ est déterministe, la probabilité de transition de C_2 est spécifié par $f_{Y|X,\Theta}(y|x,\theta) = f_W(y - s)$ qui est évidemment indépendante de θ. D'autre part la probabilité de transition de C_1 est spécifiée par $f_{dM|\theta}$ qui correspond à un canal du type purement Poissonnien. Comme on le verra plus loin, cette décomposition C_1C_2 de C permettra le calcul d'une borne analytique sur la capacité C qui sera utile pour séparer les zones d'opération en régime où l'estimation est limitée par les statistiques de dM et en régime où elle est limitée par les statistiques engendrées par la formation de l'image Y à partir de dM.

4 Un estimateur MAP du type EM

Rappelons que l'algorithme EM peut être écrit de la façon suivante:

Initialisation: θ^0, $k = 0$

Pour $k = 1, \ldots$

- [E Step] `Estimer`

$$Q(\theta, \hat{\theta}^k) = E[\ln f_{X|\Theta}(\underline{X}|\theta)|Y, \Theta = \hat{\theta}^k] \tag{6}$$

- [M Step] `Maximiser`

$$\hat{\theta}^{k+1} = \text{argmax}_\theta \left\{ Q(\theta|\hat{\theta}^k) + \ln f_\Theta(\theta) \right\} \tag{7}$$

L'algorithme EM est basé sur la définition d'un jeu de données complètes X qui est traditionnellement défini par une fonction g qui détermine X par la relation $Y = g(X)$ [9]. Cependant, pour notre application ainsi que pour d'autres en traitement du signal, il sera plus simple d'utiliser la définition équivalente [26]: X est un jeu de données complètes si X contient plus d'information sur Θ que Y dans le sens précis que $f_{Y|X,\Theta} = f_{Y|X}$. En vue de la décomposition du canal de mesure C montrée dans la Figure 2, il est clair que la définition $X = dM$ produit un jeu de données complètes valide.

La mise en oeuvre exacte de l'algorithme EM est rendue difficile sinon impossible par le fait que l'espérance $Q(\theta|\overline{\theta}) = E[\ln f_{X|\Theta}(\underline{X}|\theta)|Y, \Theta = \overline{\theta}]$ ne prend pas une forme analytique. Cette difficulté a été la motivation d'un algorithme EM approché introduit dans un autre article de l'auteur pour les processus de Poisson filtrés evoluant dans une dimension [3]. Ce dernier algorithme repose sur le fait que pour deux v.a. U et V, l'espérance conditionelle $E[U|V]$ est l'estimateur optimal de U dans le sens que $E[(U - \hat{U})^2]$ est minimisée pour $\hat{U} = E[U|V]$. Donc, $Q(\theta|\overline{\theta})$ étant un estimateur optimal du log de la fonction de vraisemblance $l_X(\theta) = \ln f_{X|\Theta}(X|\theta)$, $Q(\theta|\overline{\theta})$ admet une approximation polynômiale d'ordre q: $Q(\theta|\overline{\theta}) \approx \hat{Q}^{(q)}(\theta|\overline{\theta}) = \hat{l}_X^{(1)}(\theta) + \ldots + \hat{l}_X^{(q)}(\theta)$. Cette approximation est donnée par la projection de $l_X(\theta)$ sur l'espace linéaire $\mathrm{span}\{\mathrm{Y}^{(1)}, \ldots, \mathrm{Y}^{(q)}\}$ où $Y^{(k)} = \{\Pi_{j=1}^k Y(z_j) : z_1, \ldots, z_k \in I\}$.

Dans cet article, on se limitera à une approximation de $Q(\theta|\overline{\theta})$ d'ordre $q = 1$, c'est-à-dire linéaire en Y. Définissons les fonctions

$$\begin{aligned}
\bar{S}_{\overline{\theta}}(u) &= E[S(u)|\Theta = \overline{\theta}], \;\; u \in I. \\
\bar{l}_{\theta,\overline{\theta}} &= E[l_X(\theta)|\Theta = \overline{\theta}] \\
K_{\theta,\overline{\theta}}(u) &= \mathrm{cov}(l_X(\theta), S(u)|\Theta = \overline{\theta}), \;\; v \in I \\
K_{\overline{\theta}}(u,v) &= \mathrm{cov}(S(u), S(v)|\Theta = \overline{\theta}), \;\; u, v \in I.
\end{aligned}$$

Avec ces définitions on a la forme générale et bien connue de l'estimateur linéaire optimal de $l_X(\theta)$

$$\hat{Q}^{(1)}(\theta|\overline{\theta}) = \bar{l}_{\theta,\overline{\theta}}(u) + \int_{-a}^{a}\int_{-a}^{a} du \; K_{\theta,\overline{\theta}}(u) \int_{-a}^{a}\int_{-a}^{a} dv \; \mathcal{K}_{\overline{\theta}}^{-1}(u,v)(Y(v) - \bar{S}_{\overline{\theta}}(u)) \tag{8}$$

où $\mathcal{K}_{\overline{\theta}}^{-1}(u,v) = K_{\overline{\theta}}(u,v) + N_o/2\ \delta(u-v)$ est l'opérateur inverse associé à la fonction de covariance de Y.

Rappelons que $Y(u) = h(u) \star (1 - T(u)) + W(u)$, où $T(u)$ est la fonction de transparence introduite par Picinbono [32]. Il suit des résultats de Picinbono sur la moyenne et la covariance spatiale de T que

$$\begin{aligned}
\bar{S}_{\overline{\theta}}(u) &= H(0)\left(1 - \exp\left\{-\pi\lambda m_2(\overline{\theta})\right\}\right) \\
K_{\overline{\theta}}(u,v) &= h(u) \star \operatorname{cov}(T(u), T(v)|\overline{\theta}) \star h(v), \\
\operatorname{cov}(T(u), T(v)|\overline{\theta}) &= \exp\left\{-2\pi\lambda m_2(\overline{\theta})\right\}\left[\exp\left\{\lambda\int_0^\infty r^2 B\left(\frac{\|u-v\|}{2r}\right) f_{R|\Theta}(r|\overline{\theta})dr\right\} - 1\right] \qquad (9)
\end{aligned}$$

où $m_2(\theta) = \int_0^\infty r^2 f_{R|\Theta}(r|\theta)dr$ est le moment conditionnel d'ordre 2 des rayons $\{R_i\}$ des disques, $B(t)$ est la superficie de la lentille d'intersection de deux disques de rayon 1 séparés par une distance t, et $H(\omega)$ est la transformée de Fourier du filtre de lissage spatial $h(u)$. Dans notre cas, où $f_{R|\Theta}$ est linéaire de la forme (5), on peut exploiter une approximation excellente de la fonction de covariance

$$\operatorname{cov}(T(u), T(v)|\overline{\theta}) = e^{-\pi\lambda m_2(\overline{\theta})}\left(1 - e^{-\pi\lambda m_2(\overline{\theta})}\right)\left(1 + \frac{\|u-v\|}{2}\right)^{\lambda+2}. \qquad (10)$$

Par une méthode de calcul similaire à celle de Picinbono, on obtient des expressions pour la moyenne conditionnelle de $l_X(\theta)$ et la corrélation conditionnelle entre $l_X(\theta)$ et Y

$$\begin{aligned}
\overline{l}_{\theta,\overline{\theta}} &= H(0)\lambda|I|\int_0^\infty f_{R|\Theta}(r|\overline{\theta}) \ln f_{R|\Theta}(r|\theta)dr \\
K_{\theta,\overline{\theta}}(u) &= H(0)\lambda|I|\exp\left\{-\pi\lambda m_2(\overline{\theta})\right\}\int_0^\infty \left[1 - \pi r^2/|I|\right] f_{R|\Theta}(r|\overline{\theta}) \ln f_{R|\Theta}(r|\theta)dr. \qquad (11)
\end{aligned}$$

En remplaąnt les identités (9) et (11) dans l'expression (8), et en appliquant l'identité de Parseval, on aboutit à une forme simple pour la fonction $\hat{Q}^{(1)}$

$$\begin{aligned}
\hat{Q}^{(1)}(\theta|\overline{\theta}) &= \Lambda\int_0^\infty (1 + r^2 q(Y,\overline{\theta})) f_{R|\Theta}(r|\overline{\theta}) \ln fR|\Theta(r|\theta)dr \\
q(Y,\overline{\theta}) &= H(0)\pi\lambda e^{-\pi\lambda m_2(\overline{\theta})}\frac{1}{|I|}\int_{-\infty}^\infty\int_{-\infty}^\infty \frac{\mathrm{FT}_{\mathrm{Y}-\overline{\mathrm{S}}_{\overline{\theta}}}(\omega)}{|H(\omega)|^2\Phi_T(\omega|\overline{\theta}) + N_o/2}d\omega \qquad (12)
\end{aligned}$$

où $\mathrm{FT}_{\mathrm{Y}-\overline{\mathrm{S}}_{\overline{\theta}}}(\omega)$ dénote la transformée de Fourier 2D de $Y(u) - \overline{S}_{\overline{\theta}}(u)$ sur le domaine $u \in I$, et $\Phi_T(\omega|\overline{\theta})$ dénote la densité spectrale conditionnelle de T, c'est-à-dire la transformée de Fourier 2D de la fonction $\operatorname{cov}(T(u), T(0)|\overline{\theta})$ (Eq. (10)) par rapport à u.

L'équation (12) accomplit le "E Step" de notre algorithme EM approché. Une forme analytique de la maximisation requise par le "M Step" peut être obtenue dans le cas d'une densité conditionnelle $f_{R|\Theta}$ exponentielle ou linéaire.

4.1 Disques à densité de rayon exponentielle

Soit $f_{R|\Theta}(r|\theta) = \theta e^{-\theta r}$, $r > 0$, $\theta \in [0, \infty]$. On obtient facilement une équation pour la valeur de θ qui maximise $\hat{Q}^{(1)}(\theta|\overline{\theta})$:

$$\theta = \overline{\theta}\, \frac{\overline{\theta}^2 + 2!\, q(Y, \overline{\theta})}{\overline{\theta}^2 + 3!\, q(Y, \overline{\theta})}.$$

En remplaąnt θ par l'itéré θ^{k+1} et $\overline{\theta}$ par l'itéré θ^k, on obtient une récurrence pour l'algorithme EM approché qui cherche à maximiser la fonction de vraisemblance non pénalisée, i.e. $\ln f_\Theta(\theta) =$ constante dans (7). Ceci correspond à l'utilisation commune d'un "improper prior" dans l'algorithme MAP. Notons que le nouvel itéré $\theta = \theta^{k+1}$ est obtenue par l'application d'un facteur correcteur à l'ancien itéré $\overline{\theta} = \theta^k$. Ce facteur augmente (réduit) $\overline{\theta}$ quand les résidus filtrés $q(Y, \overline{\theta})$ sont négatifs (positifs). La récurrence EM approché a des points fixes spécifiés par la condition $q(Y, \theta) = 0$. Autrement dit, l'algorithme s'arrête quand θ^k rend $q(Y, \theta^k)$ égal à zéro. Notons aussi qu'il est nécessaire que $\inf_{\overline{\theta}>0} |\overline{\theta}^2 + 6q(Y, \overline{\theta})|$ soit supérieur a zéro pour assurer la stabilité de l'algorithme.

4.2 Disques à densité de rayon linéaire

Soit $f_{R|\Theta}(r|\theta)$ donné par la forme linéaire (5) et supposons Θ uniforme sur $[-1, 1]$. La valeur de θ qui maximise $\hat{Q}^{(1)}(\theta|\overline{\theta})$ doit satisfaire

$$\int_0^1 (1 + r^2 q(Y, \overline{\theta})(r - \tfrac{1}{2}) \frac{\overline{\theta} 2r + (1 - \overline{\theta})}{\theta 2r + (1 - \theta)} dr = 0, \tag{13}$$

L'intégrale peut être calculée explicitement et on obtient une équation nonlinéaire en θ. Cette équation ne permet pas d'obtenir une forme explicite de la solution θ et donc on est contraint à utiliser les méthodes numériques. Cependant, on déduit directement de la forme intégrale (13) que les itérés θ^k ont le même comportement en fonction de $q(Y, \overline{\theta})$ que dans le cas d'une densité exponentielle.

On a mis en oeuvre les algorithmes EM approchés dans les cas exponentiel et linéaire présentés ci-dessus. Sur la base de nos simulations, on a constaté que les estimateurs convergeait très rapidement dans les deux cas (en moins de dix itérations) et que la limite était comparable à l'estimateur construit par la "méthode des moments." Plus de détails sur le comportement expérimental de ces estimateurs itératifs seront présentés dans un autre article.

5 Une borne inférieure sur l'erreur d'estimation

Pour un estimateur $\hat{\Theta} = [\hat{\Theta}_1, \ldots, \hat{\Theta}_p]^T$ fonction de l'observation Y, définissons le critère d'erreur

$$\mathrm{MSE} = \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{E}[(\Theta_{\mathrm{i}} - \hat{\Theta}_{\mathrm{i}})^2].$$

Pour développer une borne inférieure sur MSE, il serait naturel de songer à la borne de Cramer-Rao (CR). Or, comme on le sait déjà, la densité $f_{Y|\Theta}$ ne prend pas une forme analytique et la borne de CR n'est pas calculable. Une alternative présentée ici est d'utiliser la théorie de l'information pour développer une borne de Shannon sur MSE. Cette approche a été utilisée dans le cas de modèle linéaire (Eq. (3)) par l'auteur [20, 23]. Ici on donnera seulement un rappel bref de la méthode de construction de ces bornes. En ce qui concerne notre application, l'importance de notre approche est qu'elle peut fournir des bornes analytiques qui dépendent seulement de la divergence de Kullback sur la famille de densités $\{f_{dM|\Theta}(dM|\theta)\}_\theta$, et de la densité spectrale du signal S.

5.1 La Borne de Shannon

Étant données deux variables aléatoires V et Z, V étant "la source" (la variable que l'on veut estimer) et Z étant "la destination" (les observées), définissons l'information mutuelle $I(V;Z) = E[\ln P_{Z|V}(Z|V)/P_Z(Z)]$. $P_Z(\bullet)$ est la distribution de Z et $P_{Z|V}(\bullet|V)$ est la distribution conditionnelle qui décrit le canal, c'est-à-dire la probabilité de transition de l'entrée V et de la sortie Z. Soit $\rho(V,Z)$ une fonction de distorsion entre la source V et une estimée $\hat{V}(Z)$, et soit $\overline{\rho} = E[\rho(V,Z)]$ la moyenne de la distorsion. Ici $\rho(V,Z)$ est la somme des carrés de l'erreur et pour ce cas $\overline{\rho}$ est égal à MSE. La théorie de Shannon [16] préscrit que pour n'importe quelle valeur d de $\overline{\rho}$ la capacité C du canal reliant V et Z ne peut pas être inférieure à la fonction $R_\rho(d)$, dite "rate-distortion function"

$$\inf_{P_{Z|V}:\overline{\rho}\leq d} I(V,Z) := R_\rho(d) \leq C := \sup_{P_V} I(V,Z)$$

La fonction $R_\rho(d)$ est toujours strictement décroissante sur le domaine $d < d_{max}$ ou d_{max} est la somme des variances *a priori* des composants de V. Donc, $R_\rho^{-1}(\bullet)$ étant l'inverse de la fonction $R_\rho(d)$ sur $d < d_{max}$, dite "distortion-rate function," on a pour $\mathrm{MSE} \leq \mathrm{d_{max}}$:

$$d = \mathrm{MSE} \geq \mathrm{R}_\rho^{-1}(\mathrm{C}). \tag{14}$$

L'expression (14) bornant la MSE peut être simplifiée en utilisant la borne de Shannon sur R_ρ et la théorème de Shannon sur la capacité de la composition de deux canaux. La borne de Shannon prend la forme [6]: $R_\rho(d) \leq H(V) - \frac{1}{2}\ln(2\pi de)$, où $H(V) = E[-\ln P_V(V)]$ est l'entropie de Shannon. Le "data processing theorem" de Shannon dit que C étant la capacité d'un canal composé de la cascade de deux canaux ayant pour capacités C_1 et C_2 [16]

$$C \leq \min\{C_1, C_2\}.$$

Cette borne dit simplement que la capacité totale du canal composé est dominée par le canal le plus faible. En appliquant ces deux bornes de Shannon à l'inégalité (14) on obtient facilement la forme finale de la borne de Shannon sur le MSE

$$\mathrm{MSE} \geq \frac{1}{2\pi\mathrm{e}}\ \mathrm{e}^{2\mathrm{H(V)}}\ \mathrm{e}^{-2\min\{\mathrm{C_1},\mathrm{C_2}\}}. \tag{15}$$

La quantité $\frac{1}{2\pi e}\ e^{2H(V)}$ est appelée le "entropy power" qui est égal à la variance *a priori* de V quand V est une variable Gaussienne.

6 Application au processus de Poisson filtré

Identifions les données $Y = Z$ et les paramètres $\theta = V$ dans la borne de Shannon (15). La capacité C est spécifiée par la distribution conditionnelle $P_{Y|\Theta}$ que l'on sait ne pas être calculable. Cependant, on peut facilement traiter les capacités C_1 et C_2 dans la décomposition de C décrite dans la figure 2.

6.1 Une borne sur la capacité C_1

A cause de l'indépendance mutuelle de $\{U_i\}$ et l'ensemble $\{\Theta, \{R_i\}\}$, on démontre facilement l'identité $I(\Theta, dM) = I(\Theta, \{R_i\}_{i=1}^N) = H(\{R_i\}_{i=1}^N) - H(\{R_i\}_{i=1}^N|\Theta)$, qui est la différence entre l'entropie du processus des marques $\{R_i\}_{i=1}^N$ et son entropie conditionelle. En utilisant le fait que le processus de Poisson a la plus grande entropie parmi les processus ponctuels ayant la même intensité, on arrive à une borne supérieure sur la capacité C_1 similaire à l'expression obtenue en [21, Lemme 4]

$$C_1 \leq C_1^* = \Lambda \sup_{f_\Theta} \int f_\Theta(\theta) \int dr\, f_{R|\Theta}(r|\theta) \ln \frac{f_{R|\Theta}(r|\theta)}{f_R(r)} d\theta \tag{16}$$

$$f_R(r) \stackrel{\text{def}}{=} \int f_\Theta(\theta) f_{R|\Theta}(r|\theta) d\theta,$$

L'expression C_1^* est simplement la capacité d'un canal purement Poissonien. C_1^* est égal à la moyenne de la distance de Kullback entre la densité conditionnelle $f_{R|\Theta}(r|\theta)$ et la densité marginale $f_R(r)$. La capacité C_1 est nulle quand $f_{R|\Theta}(r|\theta)$ est constante en θ et donc identique a $f_R(r)$. Dans ce cas dM (et de même pour Y) ne porte aucune information sur Θ.

6.2 Une borne sur la capacité C_2

Il est assez simple d'obtenir une borne explicite sur C_2 en utilisant le fait suivant: la capacité d'un canal avec sortie continue n'est jamais supérieure à celle d'un canal Gaussien qui livre un processus ayant la même covariance à la sortie. Ceci suit directement de l'identité $I(Y|dM) = H(Y) - H(Y|dM)$ et du fait bien connu que l'entropie $H(Y)$ est maximisée par l'entropie d'un processus Gaussien:

$$\begin{aligned} C_2 &= \sup_{p_{dM}} \{H(Y) - H(Y|dM)\} \\ &= \sup_{p_{dM}} \{H(Y) - H(W|dM)\} \\ &= \sup_{p_{dM}} \{H(Y)\} - H(W) \\ &\leq H(\mathcal{N}) - H(W), \end{aligned} \tag{17}$$

où $\mathcal{N}$ est un processus Gaussien de moyenne nulle ayant pour fonction de covariance $\text{cov}(Y(u), Y(v))$, $u, v \in I$. Pour obtenir la deuxième ligne de (17) on a bénéficié des faits suivants: conditionné sur

dM, le signal S est déterministe; dM et W sont des v.a indépendantes; et $H(W)$ ne dépend pas de p_{dM}. Puisque $H(W)$ est l'entropie du bruit blanc Gaussien, l'expression finale de (17) est la capacité C_2^* d'un canal Gaussien. En plus, puisque Y est un champ homogène, on a recours à l'expression classique pour la capacité d'un canal spatial Gaussien quand I est grand par rapport à la distance de décorrelation de S [34]:

$$C_2^* = \frac{|I|}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln \left(1 + \frac{\Phi_S(\omega)}{N_o/2} \right) d\omega \tag{18}$$

où $\Phi_S(\omega)$ est la densité spectrale de S.

6.3 La borne appliquée au modèle de Picinbono

On se limitera ici au cas où $f_{R|\Theta}(r|\theta)$ est de forme linéaire comme énoncé en (5), et où f_Θ est une densité uniforme. Pour appliquer la borne au modèle lissé et bruité de Picinbono, il suffit de calculer les capacités C_1^* et C_2^* pour ce choix de $f_{R|\Theta}(r|\theta)$.

6.3.1 Capacité C_1^*

Puisque $f_{R|\Theta}(r|\theta)$ est définie seulement pour $\theta \in [-1, 1]$, il faut faire la maximisation en (16) sur l'espace de densités $f_\Theta(\theta)$ supportées sur cette intervalle. Definissons l'entropie conditionelle de R_i pour $\Theta = \theta$

$$H(R|\theta) = -\int f_{R|\Theta}(r|\theta) \ln f_{R|\Theta}(r|\theta) dr.$$

Avec ceci on peut répresenter l'information mutuelle de la faon suivante

$$I(\Theta, dM) = H(R|\mu) - \int H(R|\theta) f_\Theta(\theta) d\theta$$

où $\mu = E[\Theta] = \int \theta f_\Theta(\theta) d\theta$. La maximization de $I(\Theta, dM)$ sur f_Θ donne la capacité C_1^*. On réduira ce problème de maximisation à une suite de deux maximisations sous-jacentes. D'abord on maximise le Lagrangien $L(f_\Theta)$ pour une valeur de μ fixe:

$$L(f) = -\int H(R|\theta) f(\theta) + \xi_1 \int f(\theta) d\theta + \xi_2 \int \theta f(\theta) d\theta$$

où ξ_1 et ξ_2 sont des facteurs à déterminer pour que la fonction maximisante $f^*_{\xi_1,\xi_2}$ de $L(f)$ satisfasse les deux contraintes: $\int f(\theta) d\theta = 1$ et $\int \theta f(\theta) d\theta = \mu$. Soit f^*_μ cette fonction maximisante qui satisfait ces contraintes. Il suffit par la suite de maximiser $I(\Theta, dM) = H(R|\mu) - \int H(R|\theta) f_\mu^*(\theta) d\theta$ sur μ. Avec des "water pouring arguments" standards pour ce genre de problème, on arrive par ce procédé à une expression simple pour la capacité

$$C_1^* = \Lambda a \tag{19}$$

où $a \approx 0.0698$, et une expression pour la densité $f_\Theta = f_\Theta^*$ qui atteint cette capacité

$$f_\Theta^*(\theta) = 1/2 - a/3 + a\theta^2.$$

Cette densité, montrée en Figure 3, est optimale dans le sens où elle maximise l'information mutuelle entre Θ et dM et, par conséquent, parmi toutes les densités de Θ possibles, elle devrait permettre une estimation de Θ avec le minimum de variance. Ceci est en accord avec l'intuition que dans l'ensemble de densités conditionnelles possibles, $\{f_{R|\Theta}(r|\theta) : \theta \in [-1, 1]\}$, c'est la densité uniforme ($\theta = 0$) qui produit des R_i ayant la plus grande variance. Le même comportement en fonction de θ serait attendu des estimateurs de Θ. Donc, f_Θ^* défavorise les valeurs de θ proches de 0, cas pour lesquels la variance de l'estimateur de Θ serait la plus grande.

6.3.2 Capacité C_2^*

L'expression (18) dépend de la densité spectrale Φ_S du signal S. Le calcul de cette quantité est immédiat en utilisant les expressions (11), l'identité $\text{cov}(U, V) = E[cov(U, V|\Theta)] + E[(E[U|\Theta] - E[U])(E[V|\Theta] - E[V])]$, et la forme supposée linéaire de $f_{R|\Theta}$. Puisque la fonction de covariance conditionelle $K_\theta(u, 0)$ de S a une symétrie sphérique (voir Eq. (9)), $\Phi_S(\omega)$ bénéficie aussi de cette symétrie en ω. Définissons la fonction $p(u) = (1 - \|u\|/2)^{2+\lambda}$ et sa transformée de Fourier $P(\|\omega\|)$. En faisant un changement de coordonnées rectangulaire$\to$polaire, on obtient une expression simple:

$$C_2^* = \pi |I| \int_0^\infty \rho \ln \left(1 + \kappa \, \frac{|H(\rho)|^2 P(\rho)}{N_o/2} \right) d\rho \tag{20}$$

où, $M_\theta(t) = E[e^{t\theta}]$ étant la fonction caractéristique de f_Θ,

$$\kappa = e^{-\pi\lambda/3} \left[M_\theta(-\pi\lambda/3) - M_\theta^2(-\pi\lambda/6) e^{-\pi\lambda/3} \right].$$

Dans la Figure 4, on montre C_1^* et C_2^* en fonction de l'intensité λ pour le cas de $f_\Theta(\theta)$ uniforme sur $[-1,1]$ et les paramètres I, N_o, σ cités en Figure 1. Rappelons que c'est le minimum de C_1^* et C_2^* qui détermine la borne de Shannon (15). Notons que la capacité C_1^* est croissante en λ et tend vers l'infini: l'estimation de Θ basée sur les mesures directes de dM bénéficie toujours de plus de points. Par contre la capacité C_2^* prend un maximum et décroît à zéro pour grand λ: pour un estimateur de Θ basé sur les mesures Y, un nombre de plus en plus grand d'occlusions entre disques produit une dégradation de performance. On peut observer une dégradation similaire dans le rapport signal/bruit. Cette dégradation pour λ grand est contraire à ce qu'on a constaté pour le modèle linéaire (3) étudié en [20, 22, 23].

En général, la borne de Shannon nous permet de séparer la performance en deux régimes d'opération, l'un pour $C_1^* < C_2^*$, que l'on appellera le régime limité par les statistiques Poissonniennes, et l'autre pour $C_2^* < C_1^*$, que l'on appellera le régime limité par les statistiques Gaussiennes. Donc on peut constater sur la Figure 4 qu'il y a trois régimes en fonction de λ. En mesurant Y, c'est seulement pour les valeurs de λ ni trop petites ni trop grandes qu'on peut espérer atteindre le régime Poissonnien de performance. Les frontières de ces régimes dépendent de la superficie $|I|$ de l'image, du niveau de bruit N_o, de la largeur 2σ de $h(u)$, et de la densité *a priori* f_Θ.

7 Conclusion

Dans cet article on a étudié un problème d'estimation pour une version bruitée et lissée d'un modèle d'image introduit par Bernard Picinbono en 1955. Pour ce problème, on a présenté des algorithmes d'estimation de type EM ainsi que des bornes de type Shannon sur l'erreur d'estimation et la capacité du canal de mesure. La détermination de la valeur pratique de ces résultats nécessitera un travail de simulation non négligeable pour déterminer l'efficacité et la fiabilité de nos estimateurs.

References

[1] J. Amoss and F. Davidson, "Detection of weak optical images with photon counting techniques," *Applied Optics*, vol. 11, pp. 1793–1799, 1972.

[2] C. Andrieu and P. Duvaut, "Bayesian estimation and detection of shot noise processes using reversible jumps," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc.*, pp. 3681–3684, Munich, Germany, 1997.

[3] N. Antoniadis and A. O. Hero, "Time delay estimation for filtered Poisson processes using an EM-type algorithm," *IEEE Trans. on Signal Processing*, vol. 42, no. 8, pp. 2112–2123, 1994.

[4] I. Bar-David, "Information in the time of arrival of a photon packet: capacity of PPM channels," *J. Opt. Soc. Am.*, vol. 63, No. 2, pp. 166–170, Feb. 1973.

[5] I. Bar-David, "Minimum-mean-square-error estimation of photon pulse delay," *IEEE Trans. on Inform. Theory*, pp. 326–330, May 1975.

[6] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs NJ, 1971.

[7] A. M. Bruckstein, T. J. Shan, and T. Kailath, "The resolution of overlapping echos," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol. ASSP-33, pp. 1357–1368, Dec. 1985.

[8] N. H. Clinthorne, W. L. Rogers, L. Shao, and K. Koral, "A hybrid maximum likelihood position computer for scintillation cameras," *IEEE Trans. Nuclear Science*, vol. NS-37, no. 2, pp. 658–663, 1990.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Society, Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[10] P. J. Diggle, "Binary mosaics and the spatial pattern of heather," *Biometrics*, vol. 37, pp. 531–539, 1981.

[11] L. J. Dorfman, "The distribution of conduction velocities (DCV) in peripheral nerves: A review," *Muscle Nerve*, vol. 7, pp. 2–11, 1984.

[12] A. R. Eckler, "A survey of coverage problems associated with point and area targets," *Technometrics*, vol. 11, pp. 561–589, 1969.

[13] H. Elias, *Stereology*, Springer, Berlin, 1967.

[14] P. Faure, "Theoretical models of reverberation noise," *J. Acoust. Soc. Am.*, vol. 36, pp. 259–268, 1964.

[15] R. M. Gagliardi and S. Karp, *Optical Communications*, Wiley, New York, 1976.

[16] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.

[17] F. Gatti and V. Svelto, "Review of theories and experiments of resolving time with scintillation counters," *Nuclear Instruments and Methods*, vol. 43, pp. 248–268, 1966.

[18] J. A. Gubner, "Computation of shot-noise probability distributions and densities," *SIAM J. Sci. Comput.*, vol. 17, pp. 750–761, May 1996.

[19] P. Hall, *Introduction to the theory of coverage processes*, Wiley, New York, 1988.

[20] A. O. Hero, "Timing estimation for a filtered Poisson process in Gaussian noise," *IEEE Trans. on Inform. Theory*, vol. 37, no. 1, pp. 92–106, Jan. 1991.

[21] A. O. Hero, "Lower bounds on estimator performance for energy invariant parameters of multidimensional Poisson processes," *IEEE Trans. on Inform. Theory*, vol. 35, pp. 843–858, July 1989.

[22] A. O. Hero, "Recovering photon-intensity information from continuous photo-detector measurements," in *Proceedings of the 25-th Conference on Information Sciences and Systems*, pp. 643–648, Johns Hopkins University, Baltimore, MD, Mar. 1991.

[23] A. O. Hero, "Theoretical limits for optical position estimation using imaging arrays," in *Actes du Colloque GRETSI*, pp. 793–796, Juan-les-Pins, France, Sept. 1991.

[24] A. O. Hero, N. Antoniadis, N. H. Clinthorne, and W. L. Rogers, "Optimal and sub-optimal post-detection timing estimators for PET," *Proc. of IEEE Nuclear Science Symposium*, vol. NS-37, no. 2, pp. 725–729, April 1990.

[25] A. O. Hero, N. Antoniadis, N. H. Clinthorne, W. L. Rogers, and G. D. Hutchins, "Optimal and sub-optimal post-detection timing estimators for PET," *IEEE Trans. Nuclear Science*, vol. NS-37, no. 2, pp. 725–729, 1990.

[26] A. O. Hero and J. A. Fessler, "A recursive algorithm for computing CR-type bounds on estimator covariance," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 1205–1210, July 1994.

[27] T. T. Kadota, "Approximately optimum detection of deterministic signals in Gaussian and compound Poisson noise," *IEEE Trans. on Inform. Theory*, vol. 34, pp. 1517–1527, Nov. 1988.

[28] L. Kazovsky, "Beam position estimation by means of detector arrays," *Opt. Quantum Electron.*, vol. 13, pp. 201–208, 1981.

[29] J. M. Mendel, "White noise estimators for seismic data processing in oil exploration," *IEEE Trans. Automatic Control*, vol. AC-22, no. 5, pp. 694–706, Oct. 1977.

[30] P. N. Misra and H. W. Sorenson, "Parameter estimation in Poisson processes," *IEEE Trans. on Inform. Theory*, pp. 87–90, Jan. 1975.

[31] J. J. O'Reilly, "Generating functions and bounds in optical communications," in *Problems of randomness in communications engineering*, K. Cattermole and J. O'Reilly, editors, chapter 7, pp. 119–133, Wiley, New York, Nov. 1987.

[32] B. Picinbono, "Modèle statistique suggéré par la distribution de grains d'argent dans les films photographiques," *Comptes Rendus de l'Académie des Sciences*, vol. Séance du 6 Juin, pp. 2206–2208, 1955.

[33] B. D. Ripley, *Spatial statistics*, Wiley, New York, 1981.

[34] D. J. Sakrison and V. R. Algazi, "Comparison of line-by-line and two-dimensional encoding of random images," *IEEE Trans. on Inform. Theory*, vol. IT-17, no. 4, pp. 386–398, July 1971.

[35] P. Salomon and T. Glavich, "Image signal processing in sub-pixel accuracy star trackers," in *Proc. Soc. Phot-Opt. Instrum. Eng.*, pp. 64–74, 1980.

[36] R. Schoonhoven, D. F. Stegeman, and J. P. C. de Weerd, "The forward problem in electroneurography – I: A generalized volume conductor model," *IEEE Trans. on Biomed. Eng.*, vol. BME-33, pp. 327–334, 1986.

[37] J. Serra, *Image analysis and mathematical morphology*, Academic Press, New York, 1982.

[38] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, Springer-Verlag, New York, 1991.

[39] M. N. Wernick and G. M. Morris, "Image classification at low light levels," *J. Opt. Soc. Am.*, vol. 3, pp. 2179–2187, 1986.

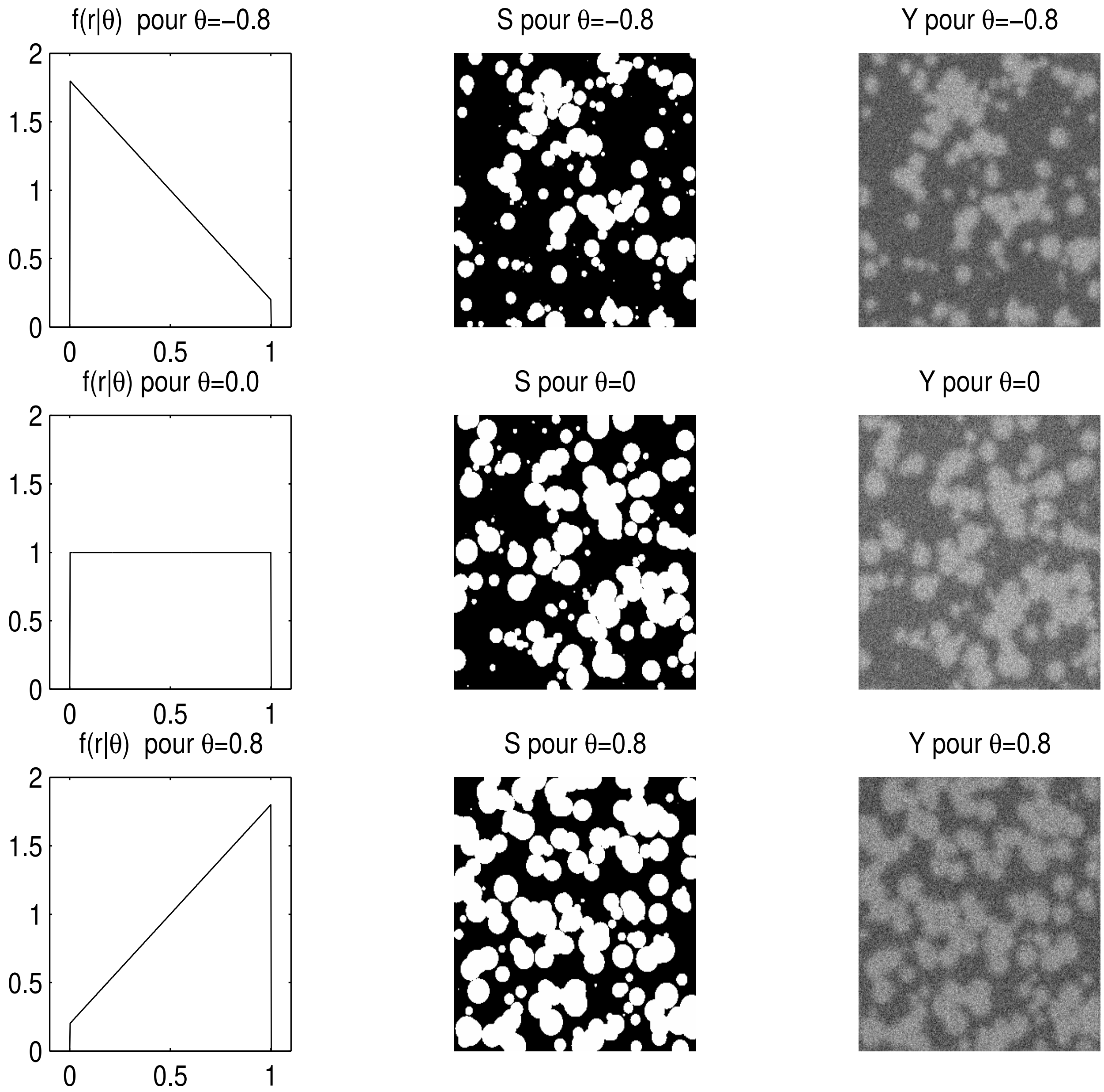

Figure 1: *Trois réalisations de l'image S et Y pour la densité des rayons (en mm) des disques montrée dans la première colonne. La taille de l'image est 20×20 mm et le niveau gris est tel que le noir correpond à la valeur minimum de l'image et le blanc à la valeur maximum. L'intensité est $\lambda = 0.5$, ce qui correspond à une moyenne de $\Lambda = 200$ disques dans chaque image. Pour toutes les images dans la troisième colonne, le rapport signal/bruit est 3dB et la fonction de lissage spatial $h(u)$ est de forme Gaussienne symétrique de largeur $2 * \sigma$ fixée à 0.66 mm.*

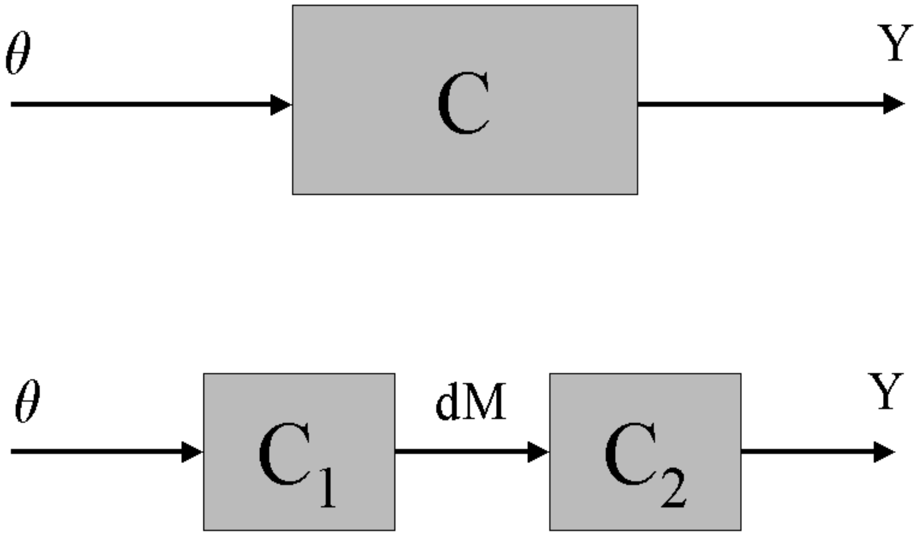

Figure 2: *(a) Représentation statistique de Y comme sortie d'un canal d'information C avec entrée Θ. (b) Décomposition du canal C en 2 canaux C_1 et C_2.*

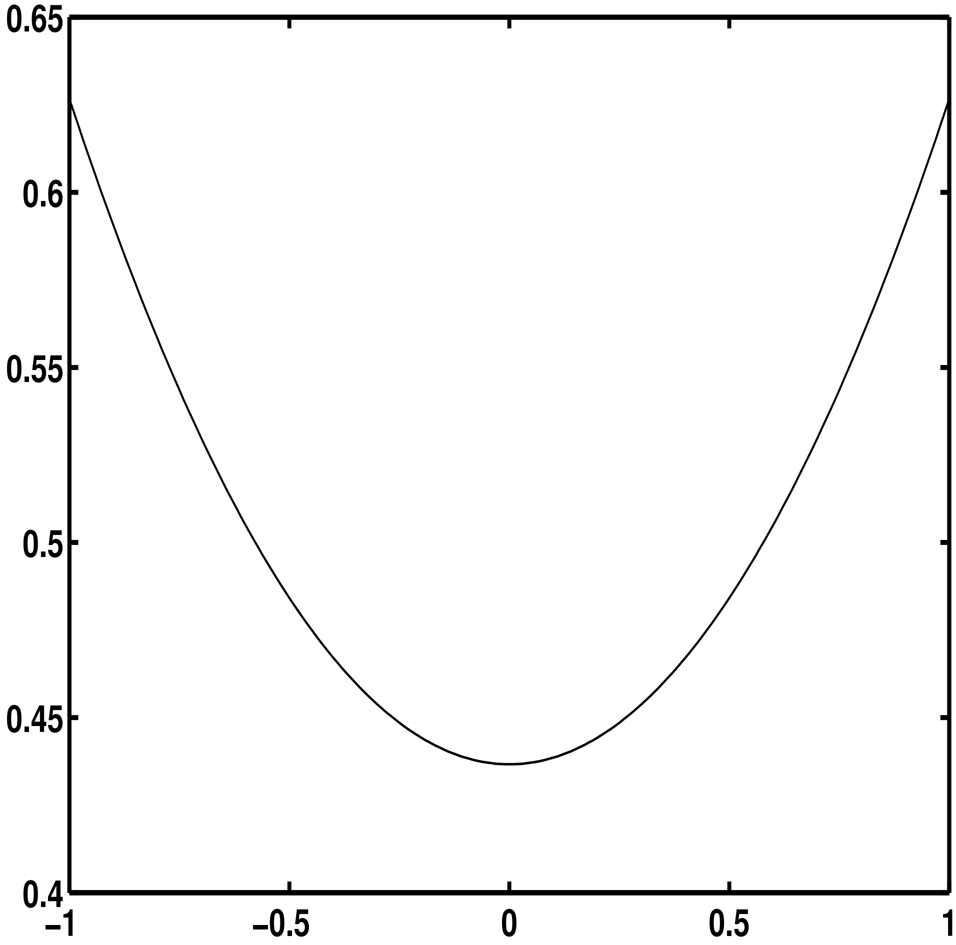

Figure 3: *La densité* f_Θ^* *qui maximise l'information mutuelle* $I(\Theta, dM)$ *pour le cas où* $f_{R|\Theta}$ *est linéaire.*

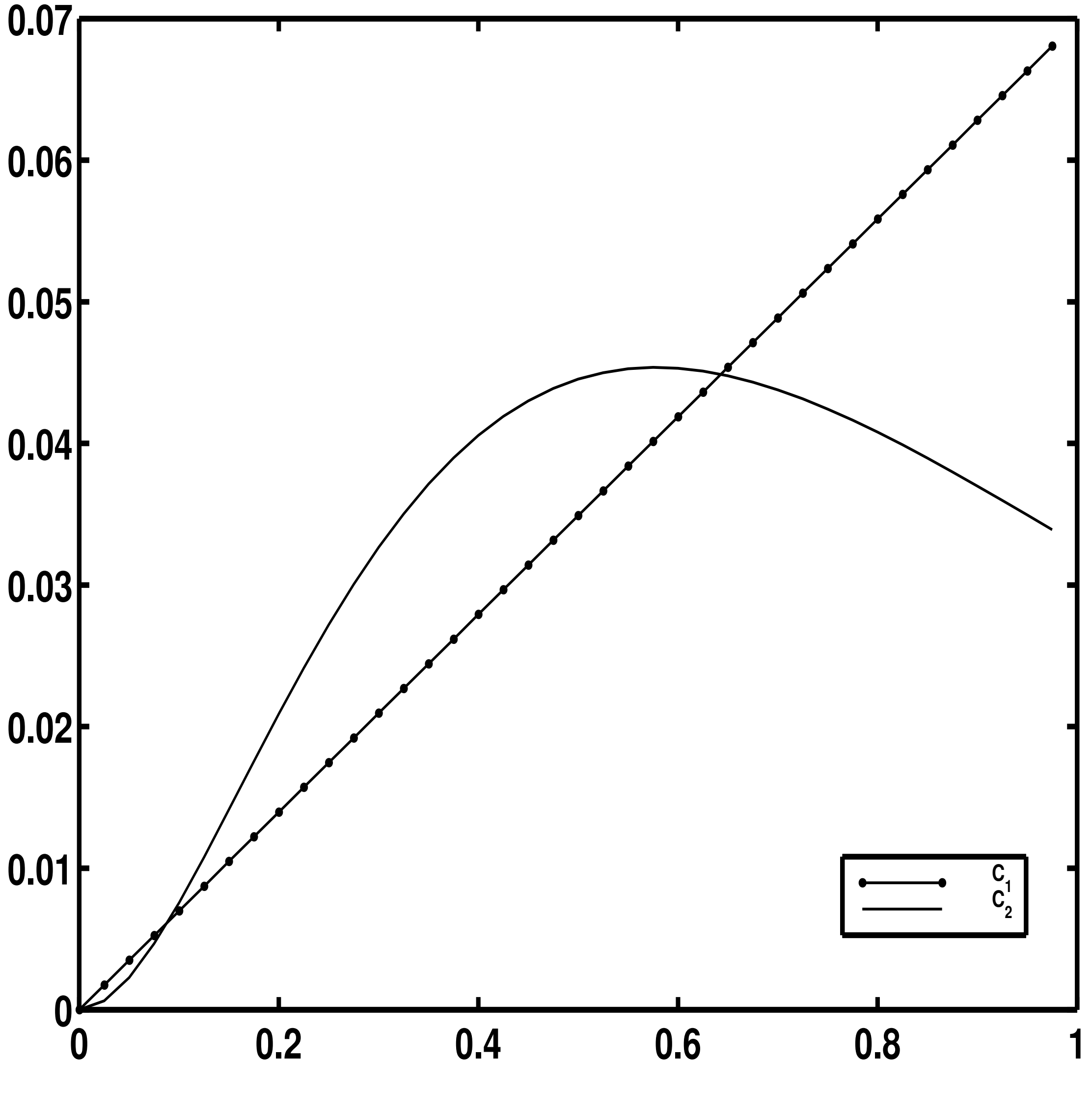

Figure 4: *Les capacités C_1^* et C_2^* en fonction de l'intensité λ pour le cas où $f_{R|\Theta}$ est linéaire, $f_\Theta(\theta)$ est uniforme sur $[-1, 1]$, et les mêmes paramètres que ceux de la Figure 1.*