République Algérienne Démocratique et populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mouloud Mammeri de Tizi-Ouzou

Faculté de Génie Electrique et Informatique
Département Automatique

MEMOIRE DE MAGISTER

En Automatique

Option : **Traitement d'Images et Reconnaissance de Formes**

Présenté par :

LARBI Kahina

ingénieur U.M.M.T.O.

Segmentation d'images basée sur la modélisation statistique d'histogrammes

Mémoire soutenu le / /2012 devant le jury d'examen composé de :

DIAF Moussa	Professeur à l'U.M.M.T.O.	**Président**
HAMMOUCHE Kamal	M.C.A. à l'U.M.M.T.O.	**Rapporteur**
HADDAB Salah	M.C.A. à l'U.M.M.T.O.	**Examinateur**
LOUNI Hamid	M.C.A. à l'U.M.M.T.O.	**Examinateur**

Remerciements

Ce présent travail a été effectué au sein du Laboratoire Robotique et Vision de la faculté Génie Electrique et Informatique de l'université Mouloud Mammeri de Tizi-Ouzou.

Je tiens en premier lieu à adresser mes plus vifs remerciements à mon directeur de mémoire, Monsieur Hammouche Kamal pour m'avoir initié à effectuer ce travail. Je lui exprime ma profonde gratitude pour m'avoir fait profiter de ses connaissances, mais aussi de ses méthodes de travail, et surtout de sa rigueur scientifique. Sans sa disponibilité permanente, son soutien et ses conseils, ce travail n'aurait pas pu aboutir.

Ma profonde gratitude s'adresse à Monsieur Diaf Moussa professeur à l'UMMTO pour m'avoir donné la chance de travailler dans son équipe de recherche. Je le remercie aussi d'avoir accepté de présider le jury de ce mémoire.

Je tiens aussi à remercier les membres de jury, Monsieur Haddab Salah Maitre de Conférences (A) à l'UMMTO et Monsieur Louni Hamid Maitre de Conférences (A) à l'UMMTO pour l'honneur qu'ils me font en participant au jury, et qui ont pris la peine de lire avec soin ce mémoire pour juger son contenu.

Merci enfin, à ma très chère famille, particulièrement à mes parents qui m'ont toujours soutenu et cru en moi. Nul remerciement ne pourra exprimer ma reconnaissance envers mon mari qui a su m'accompagné tout au long de ces années.

Merci à Dieu qui m'a donné la force d'aller jusqu'au bout de cette thèse

SOMMAIRE

Chapitre 3 : Seuillage paramétrique basé sur l'approximation de l'histogramme par un mélange de différentes distributions.

Introduction générale

Introduction générale

Au cours de la dernière décennie, le domaine de traitement d'images s'est énormément développé et un grand nombre de travaux ont été effectués dans différents domaines d'applications tels que le domaine médical, la télédétection, etc.

Dans un système de traitement d'images, la segmentation d'images est l'opération la plus importante car elle conditionne la qualité de l'interprétation d'une image. Un bon résultat de segmentation ne permet pas forcément une bonne interprétation, mais nous ne pouvons pas obtenir une bonne interprétation à partir d'un mauvais résultat de segmentation.

La segmentation d'images a pour but de déterminer les régions d'une image cohérentes, au sens d'un critère fixé a priori. De nombreux critères de segmentation existent ; suivant le domaine d'application et le type d'images traitées, le critère prendra en compte le niveau de gris, la texture, la couleur ou le mouvement. Plusieurs approches de segmentation sont apparues depuis quelques années. Certaines d'entres elles cherchent à délimiter les régions homogènes par leurs contour (approche contour) alors que d'autres cherchent à retrouver les régions homogènes (approche région).

Lorsqu'on cherche à extraire les objets contenus dans l'image de son fond, les techniques de segmentation appartenant à l'approche région débouchent sur une catégorie de méthodes dite méthodes de segmentation par seuillage. Ces méthodes consistent à délimiter les niveaux de gris des objets par des valeurs appelés seuils. Ces seuils sont localisés dans les vallées de l'histogramme situés entre les modes de l'histogramme. Chaque mode forme une classe de pixels dont les niveaux de gris sont similaires et correspond à des zones de même luminance dans l'image. Ces seuils sont alors déterminés à partir de l'histogramme en optimisant un critère (méthodes non paramétriques) ou en approximant l'histogramme par un mélange de fonctions de densité de probabilité (méthodes paramétriques).

Dans ce mémoire, nous nous intéressons au seuillage paramétrique basé sur la modélisation statistique d'histogrammes en approximant l'histogramme de l'image par un mélange de distributions statistiques. Ces distributions sont définies par des lois paramétriques pour chaque classe. Les méthodes classiques de seuillage paramétrique supposent l'existence d'une seule famille de lois, commune à toutes les classes. Généralement, cette famille est considérée de type Gaussien. Or cette hypothèse n'est pas toujours vérifiée et la loi suivie par les niveaux de gris des pixels peut varier d'une région à une autre, et que pour une même image peuvent coexister plusieurs classes des lois différentes **[1]**. Dans ce mémoire, nous proposons

une méthode de seuillage paramétrique qui consiste, en premier lieu, d'identifier la fonction de densité de probabilité de chaque classe puis de déterminer le seuil à partir des paramètres des fonctions de densités de probabilités ainsi identifiées.

Ce mémoire est principalement scindé en trois chapitres.

Dans le premier chapitre, nous présenterons une brève revue de méthodes de segmentation d'images en niveaux de gris. Nous décrirons les différentes approches et quelques méthodes de segmentation par seuillage d'histogrammes.

Le deuxième chapitre est consacré à l'étude des distributions ou des lois de probabilités. Plusieurs lois de distributions usuelles sont présentées. Nous exposerons également quelques méthodes d'estimation des paramètres des distributions ainsi quelques méthodes de tests d'adéquation utilisés en statistique.

Dans le troisième chapitre, nous présenterons une méthode de seuillage par modélisation des histogrammes. Cette méthode consiste à approximer l'histogramme bimodal de l'image par deux distributions définies par des lois de probabilité différentes appartenant à une famille de distributions composées de huit lois de probabilité: Gaussienne, Log-normale, Chi2, Bita, Gamma, Exponentielle, Rayleigh et Weibull. La sélection du seuil optimal est effectuée selon les deux modèles de distribution identifiées. Ce chapitre est ainsi divisé en deux parties. La première partie décrit toutes les étapes pour le calcul du seuil à partir d'un histogramme bimodal approximé par une combinaison linéaire de deux distributions. La deuxième partie est réservée aux tests et à la présentation des résultats obtenus sur plusieurs images à niveaux de gris par l'approche proposée.

Chapitre 1

Etat de l’art sur la segmentation d’images

1.1 Introduction

La segmentation d'images constitue une étape essentielle en traitement d'images. De nombreuses méthodes ont été proposées dans la littérature, dont la plus part sont basées sur le seuillage.

Dans ce premier chapitre, nous présenterons les différentes approches de la segmentation d'images, ensuite, nous nous intéresserons aux méthodes de seuillage en général et le seuillage paramétrique en particulier.

1.2 Définition de la segmentation

La segmentation d'images consiste à regrouper les pixels des images qui partagent une même propriété pour former des régions connexes.

Zucker **[2]** définit la segmentation d'image comme le partitionnement de l'ensemble des pixels d'une image I en n sous ensembles appelées régions $R_i, i = 1, \dots, n : I = \{R_1, R_2, \dots, R_n\}$ telle que aucune région ne doit être vide, l'intersection entre deux régions doit être vide et l'ensemble des régions doit recouvrer toute l'image. Une région est un ensemble de pixels connexes ayant des propriétés communes qui les différencient des pixels des régions voisines.
Cette définition se traduit mathématiquement par les relations suivantes :

$$\bigcup_{i=1}^{n} R_i = I \qquad avec \qquad R_i \cap R_j = f \quad i \neq j$$

$$et \quad \begin{cases} P(R_i) = vrai & \forall\, i = 1,\dots,\ n \\ P(R_i \cup R_j) = faux & R_i \quad adjacente \quad à \quad R_j \end{cases}$$

$P(.)$ désigne un prédicat d'homogénéité.

1.3 Différentes approches de la segmentation d'images

Il existe une multitude de méthodes de segmentation qu'on peut regrouper en deux grandes catégories **[3]** :

- Segmentation fondée sur les contours (approche contour).
- Segmentation fondée sur les régions (approche région).

Le schéma de la figure (1.1) nous donne une taxinomie de ces différentes approches.

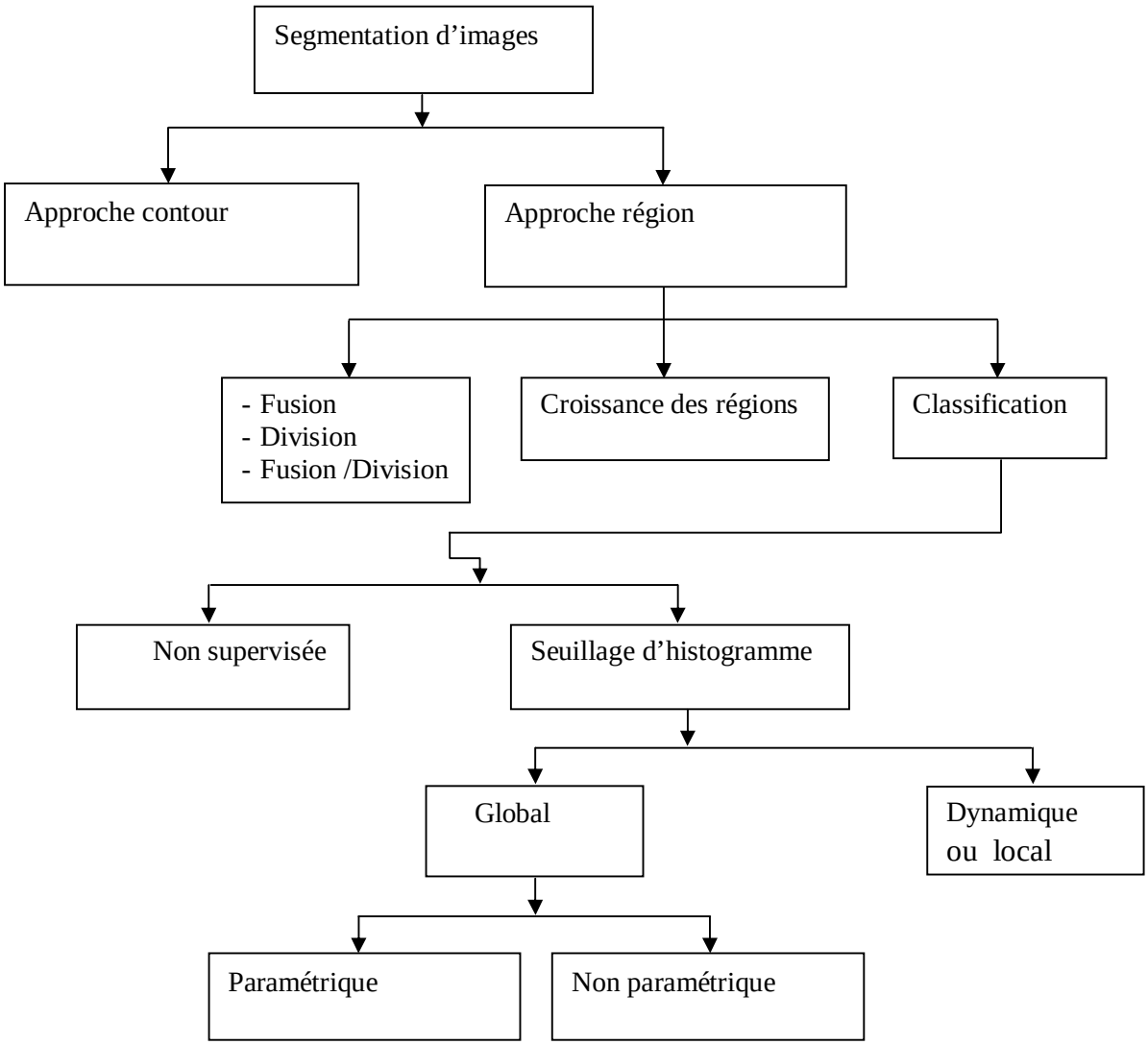


Figure1.1 : Approches de la segmentation d'images

1.3.1 Approche contour

Les premiers modèles de segmentation cherchent à extraire les contours des objets présents dans l'image. Ils s'appuient sur la détection des changements abrupts de la fonction de luminance ou de niveau de gris.

L'application de détecteurs de contours sous la forme de filtres dérivateurs ou reposant sur des critères d'optimalité permet d'obtenir les contours des objets présents dans la scène. Parmi ces filtres dérivateurs, on peut citer les opérateurs de gradient et Laplacien, comme ceux de Roberts **[4]**, de Prewitt **[5]**, de Sobel **[6]**, et de Kirsh**[7]**. Comme filtre optimal, on peut citer les

filtres de Canny**[8]**, de Deriche**[9]** et celui de Shen et Castan **[10], [11]**. Ce genre de techniques est peu exploitables car elles peuvent donner leurs contours non fermés et restent sensible au bruit. Une autre alternative à la détection de contours est proposée par les contours actifs (snakes en anglais) **[12]**. Cette méthode consiste à initialiser une courbe et faire évoluer cette courbe jusqu'à ce qu'elle coïncide avec le contour de l'objet ou de la région à détecter.

1.3.2 Approche région

L'approche région cherche à regrouper les pixels en régions homogènes. Elle se caractérise par la mesure d'uniformité des régions construites dans l'image. Ces régions sont construites en évaluant la similarité entre les pixels ou entre un pixel et ceux d'une même région. On distingue les méthodes par croissance de régions, par division-fusion et par classification.

1.3.3.1 Segmentation par croissances de régions

Ce type de segmentation permet de sélectionner un pixel ou un ensemble de pixels de l'image, appelé germe, autour duquel on fait croître une région. Les régions sont construites en ajoutant successivement à chaque germe les pixels qui lui sont connexes et qui vérifient un critère de similarité. La croissance s'arrête lorsque tous les pixels ont été traités.

La littérature en traitement d'images est riche en méthodes de segmentation par croissance de régions **[13], [14]**.

Trémeau et Borel **[15]** proposent un algorithme de segmentation qui combine une croissance de régions suivie d'un processus de fusion de régions. Cet algorithme procède par un balayage séquentiel de l'image et considère le premier pixel comme un germe. Il tente alors de faire croître ce germe le plus longtemps possible en y agrégeant les pixels voisins.

L'avantage des méthodes de croissance de régions est de préserver la forme de chaque région de l'image. Cependant une mauvaise sélection des pixels de départ, un choix de critère de similarité, aussi qu'un ordre mal adapté selon lequel les pixels voisins sont examinés, peuvent entraîner des phénomènes de sous segmentation ou de sur segmentation.

1.3.3.2 Segmentation par division /fusion

Ce type de méthode consiste à diviser l'image, considérée comme une région initiale, en régions de plus en plus petites. Le principe consiste à tester d'abord le critère d'homogénéité retenu sur l'image entière. Si le critère est valide, l'image est considérée comme segmentée ;

sinon, l'image est découpée en zones plus petites et la méthode est réappliquée sur chacune des zones nouvellement obtenues.

La division peut se faire en quatre parties, en six parties, en polygones, etc. La méthode la plus connue est la méthode de *quadtree* **[16]** où chaque zone est divisée par 4. L'inconvénient de ces méthodes est que deux parties adjacentes peuvent vérifier le même critère sans avoir été regroupées dans la même région.

Pour éviter ce problème, une procédure de fusion des petites régions similaires au sens d'un prédicat de regroupement est appliquée.

La fusion de régions est principalement fondée sur l'analyse d'un graphe d'adjacence de régions qui analyse une image présegmentée **[17]**, constituée d'un ensemble de régions. C'est une structure de données constituée d'un graphe non-orienté dont chaque nœud représente une région et chaque arête représente une adjacence entre deux régions. Le procédé consiste à fusionner deux nœuds reliés par une arrête à condition qu'ils respectent un critère de fusion. A titre d'exemple, on peut citer les méthodes de Schettini **[18]**, Saarinen **[19]**, Trémeau et Colantoni **[20]**.

1.3.3.3 Segmentation par classification

Ce type de méthode considère une région comme un ensemble de pixels connexes appartenant à une même classe. Elles supposent donc que les pixels qui appartiennent à une même région possèdent des caractéristiques similaires et forment un nuage de points dans l'espace des attributs. La classification consiste à retrouver ces nuages de points qui correspondent aux classes des pixels présentes dans l'image.

1.4 Segmentation d'images par classification des pixels

La classification peut se faire de deux manières: la première suppose l'existence de certains pixels dont l'appartenance aux classes est connue à priori, elle est très peu utilisée en segmentation car elle nécessite l'intervention de l'utilisateur. La seconde dite non supervisée (clustering), vise à regrouper automatiquement des pixels de l'image en classes sans aucune connaissance préalable sur l'appartenance des pixels aux classes. Comme méthode de classification non supervisée, on peut citer l'algorithme K-means et sa version floue (algorithme Fuzzy C-means), ainsi que l'algorithme d'Estimation-Maximisation (EM).

1.4.1 Algorithme K_*means*

C'est l'un des algorithmes les plus connu en classification non supervisée. Il vise à produire un partitionnement des pixels de manière à ce que les pixels d'une même classe soient semblables et les pixels issus de deux classes différentes soient dissemblables. L'idée principale est de définir K centroïdes, un pour chaque classe $\{C_k\}_{1\leq k\leq K}$. Chaque classe C_k est ainsi caractérisée par son centre noté μ_k et le nombre d'éléments N_k.

L'algorithme k-*means* dans sa formulation originale cherche à minimiser une fonction de coût global définie par : $J = \sum_{k=1}^{K} \sum_{(x,y),i\in C_k} (f(x,y) - \mu_k)^2$

où $f(x,y)$ représente le niveau de gris du pixel de coordonnées(x,y).

Il se déroule selon les étapes suivantes :

1. Initialisation de chaque centre μ_k .
2. Pour chaque pixel(x,y), calculer la distance $d(f(x,y),\mu_k)$ aux différents centres des classes μ_k, et affecter à la classe la plus proche $C_l = \arg min_k\, d(f(x,y),u_k)$ avec $d(f(x,y,\mu_k) = |f(x,y) - \mu_k|)$
3. Mise à jour de nombre de pixels et des centres μ_k des classes; $\mu_k = \frac{\sum_{(x,y)\in C_k} f(x,y)}{N_k}$
4. Arrêt si $N_k = N_{k+1}\ \ \forall (x,y) \in C_k$, sinon retour à l'étape 2.

Le principal inconvénient de cette méthode est que la classification finale dépend du choix de la partition initiale. Le minimum global n'est pas obligatoirement atteint, on est seulement certain d'obtenir la meilleure partition à partir de la partition de départ choisie **[21]**.

De nombreuses variantes peuvent être rencontrées. Par exemple, au lieu de calculer le centre des classes, après avoir affecté tout les pixels, les centres de gravité peuvent être calculés immédiatement après chaque affectation. La méthode des K-means a été généralisée sous l'appellation de la "*méthode des nuées dynamiques*" **[22].** Au lieu de définir une classe par un seul point (son centre de gravité), elle est définie par un groupe de points (noyau de classe).

Un autre algorithme proposé dans la littérature et qui est issu de l'algorithme *K-means* est l'algorithme ISODATA **[23]**. L'avantage de ce dernier est qu'il permet de regrouper les pixels sans connaître a priori le nombre exact de classes présentes dans l'image. Ce nombre pourra être modifié au cours des itérations.

Une version floue de l'algorithme K-means appelée Fuzzy C-means est également très populaire. Cet algorithme nécessite la connaissance préalable du nombre de classes et génère les classes par un processus itératif en minimisant une fonction objective. Il permet d'obtenir une partition floue de l'image en donnant à chaque pixel un degré d'appartenance (compris entre 0 et 1) à une classe donnée. La classe à laquelle est associé un pixel est celle dont le degré d'appartenance sera le plus élevé.

L'algorithme Fuzzy C-means possède les mêmes inconvénients que l'algorithme *K-means* à savoir la sensibilité à la répartition initiale et le choix du nombre de classes.

1.4.2 Algorithme Estimation-Maximisation

Les méthodes de classification non supervisée citées précédemment sont qualifiées de méthodes déterministes car elles n'utilisent pas les notions de statistique. D'autres méthodes de classification non supervisée ont été proposées dans un cadre statistique. Le principe de ces méthodes consiste à estimer la fonction de densité de probabilité sous jacente à l'ensemble des données à classer et assimiler chaque mode de cette fonction à une classe. Sous l'hypothèse paramétrique, ces méthodes consistent à fixer, ***a priori***, un modèle aux fonctions de densités de probabilités conditionnelles de chaque classe. La fonction densité de probabilité en un point est alors composée d'un mélange de K composantes ou fonctions de densité de probabilité conditionnelle pondérées par leurs probabilités ***a priori***. Les paramètres du modèle relatifs à chaque classe et les probabilités ***a priori*** des classes constituent les paramètres du mélange que l'on cherche à identifier à partir de l'ensemble des observations à analyser. En l'absence de toute information pouvant nous aider à choisir le modèle de ces fonctions, on fait, généralement, appel à la loi gaussienne pour sa facilité de manipulation sous forme mathématique et parce qu'elle suit beaucoup d'exemples naturels de distributions. L'estimation de ces paramètres est assurée, généralement, par l'algorithme itératif proposé par Dempster, Laird et Rubin et connu sous le nom de « *Estimation-Maximisation* » (EM) **[24]**.

Dans l'algorithme EM, la densité de probabilité $f(X_i)$ en un point X_i est décrite par un modèle de mélange. Le principe consiste à décomposer cette densité en une somme de K composantes $f(X_i/\Theta_k)$ conditionnellement aux paramètres Θ_k correspondant aux K classes. Il s'agit alors d'estimer les paramètres $\Theta_k (k = 1,2, \dots, K)$ à partir d'un échantillon X. Ces densités de probabilités $f(X_i/\Theta_k)$ peuvent aller du modèle le plus simple aux distributions les plus complexes. Les proportions P_k entre les différentes composantes représentent les probabilités ***a***

priori des différentes classes notées $f(C_k)$. En général, ces proportions sont également inconnues et doivent être estimées sous les contraintes :

$$P_k \in]0\ 1] \quad \text{et} \quad \sum_{k=1}^{K} P_k = 1 \tag{1.1}$$

Soit $\Theta = [P_1, P_2, \dots, P_k, \Theta_1, \Theta_2, \dots, \Theta_K]$, le vecteur de paramètres à estimer. La fonction densité de probabilité $f(X_i)$ en un point X_i est donnée par la relation suivante:

$$f(X_i/\Theta) = \sum_{k=1}^{K} P_k f(X_i/\Theta_k) \tag{1.2}$$

L'estimation des paramètres d'un mélange est effectuée suivant l'algorithme Estimation-Maximisation (EM) qui est basé sur la maximisation de la loi de vraisemblance. La loi de vraisemblance d'un ensemble d'échantillons X_n d'une variable aléatoire X relativement au modèle de paramètre Θ s'écrit :

$$L(\Theta) = f(X/\Theta) \tag{1.3}$$

Sous l'hypothèse que les données de l'ensemble d'apprentissage X_n sont des réalisations indépendantes du vecteur aléatoire X, la loi de vraisemblance se réécrit en un produit de probabilités :

$$L(\Theta) = \prod_{i=1}^{n} f(X_i / \Theta) \tag{1.4}$$

Dans le cas de modèles de mélange, cette équation se met sous la forme :

$$L(\Theta) = \prod_{i=1}^{n} \sum_{k=1}^{K} P_k f(X_i / \Theta_k) \tag{1.5}$$

$$Log\left(L(\Theta)\right) = \sum_{i=1}^{n} Log \sum_{k=1}^{K} P_k f(X_i / \Theta_k) \tag{1.6}$$

La solution de cette équation par la méthode d'estimation du maximum de vraisemblance équivaut à la recherche des racines de l'équation suivante :

$$\frac{\partial \mathrm{Log}\left(\mathrm{L}(\Theta)\right)}{\partial \Theta} = 0 \tag{1.7}$$

Dans le cas des mélanges gaussiens, les Θ_k représentent les moyennes et les matrices de covariance de la $k^{ième}$ classe. Les paramètres qui annulent les dérivées sont données par les équations suivantes, pour $k = 1,2, \dots, K$.

$$P_k = \frac{1}{n}\sum_{i=1}^{n} f(C_k / X_i),\ \overline{X}_k = \frac{\sum_{i=1}^{n} f(C_k / X_i) X_i}{\sum_{i=1}^{n} f(C_k / X_i)} \text{ et } \Sigma_k = \frac{\sum_{i=1}^{n} f(C_k / X_i)\left(X_i - \overline{X}_k\right)\left(X_i - \overline{X}_k\right)^T}{\sum_{i=1}^{n} f(C_k / X_i)} \tag{1.8}$$

où $f(C_k/X_i)$ représente l'estimation de la probabilité *a posteriori* d'être en présence d'une observation X_i de la classe C_k. Elle est obtenue par la formule de Bayes :

$$f(C_k / X_i) = \frac{f(X_i / C_k) P_k}{\sum_{j=1}^{K} f(X_i / C_j) P_j} \tag{1.9}$$

En pratique, l'algorithme EM (Fig.1.2) est devenu quasiment incontournable pour l'identification d'un mélange. Il fournit de bons résultats, même s'il n'échappe pas aux principaux problèmes des algorithmes de classification, tels que sa sensibilité aux conditions initiales et une convergence lente vers un optimum éventuellement local. De plus, la connaissance a priori du nombre de classes est exigée. Cependant, l'hypothèse paramétrique ne peut être satisfaisante que si, effectivement, la distribution des données suit la loi choisie.

1- Initialisation des paramètres

2- Etape d'estimation

Pour $k = 1, \dots, K$; $i = 1, \dots, n$

Calculer les probabilités *a posteriori* $f(C_k/X_i)$, par l'équation 1.9

3- Etape de Maximisation

Pour $k = 1, \dots, K$

Calculer les valeurs de P_k, $\overline{X}_k$ et Σ_k à l'aide des équations 1.8

4- Répéter les étapes 2 et 3 jusqu'à convergence.

Figure 1.2 : Algorithme EM dans le cas d'un mélange Gaussien

Lorsque les objets qui composent l'image peuvent être distingués par leurs niveaux de gris, la segmentation par classification des pixels peut être alors abordée par des techniques de seuillage.

1.4.3 Segmentation par seuillage

Le seuillage est une technique de segmentation très populaire à cause de sa facilité de mise en œuvre et sa rapidité. Elle permet d'extraire les objets du fond de l'image. Dans le cas le plus classique, les pixels de l'image sont classés en deux classes par l'intermédiaire d'un niveau de

gris S appelé seuil. La première classe regroupe les pixels du fond et la deuxième classe regroupe les pixels de l'objet.

Soit l'image $I(M \times N)$, supposons que $f(x, y)$ représente le niveau de gris d'un pixel de coordonnées $(x, y), 0 \le x \le M, 0 \le y \le N$ et S est le seuil choisi. Les pixels de l'objet sont ceux dont le niveau de gris est inférieur à S et les pixels dont le niveau de gris est supérieur à S appartiennent au fond. L'image segmentée G est définie pour chaque pixel de coordonnées (x, y) par : $g(x, y) = \begin{cases} 1 & si \quad f(x, y) > s \\ 0 & si \quad f(x, y) \le s \end{cases}$

Selon Horaud **[25]**, il existe trois grandes techniques de seuillage: global, local et dynamique. Le seuil S peut être alors considéré comme une fonction sous forme de : $S = t(p(x, y), f(x, y))$ où $p(x, y)$ représente des propriétés locales du pixel (x, y). Si S ne dépend que du niveau de gris $f(x, y)$ du pixel, le seuillage est dit global, s'il dépend en plus de $p(x, y)$ le seuillage est dit local et si S dépend à la fois de (x, y) , de $p(x, y)$ et de $f(x, y)$ le seuillage est dit dynamique ou bien adaptatif.

Dans le premier cas, un seul seuil S est définit pour tous les pixels de l'image, alors que dans les deux derniers cas, on définit pour chaque pixel un seuil $S(x, y)$. Le problème de seuillage revient alors à chercher le bon seuil S.

Notons également qu'il existe des méthodes de seuillage global qui utilise l'information locale pour déterminer un seuil global. Ces méthodes sont généralement basées sur des histogrammes bidimensionnels **[26-29]**.

1.4.3.1 Seuillage dynamique ou local

Pour le seuillage dynamique, la classification d'un pixel dépend non seulement du son niveau de gris mais aussi de ses informations locales c'est-à-dire des niveaux des gris des pixels voisins. On définit alors un seuil pour chaque pixel selon sa position.

Dans cette famille de méthodes, le calcul de seuil peut se faire en considérant une fenêtre de voisinage de taille $W \times W$ centrée autour d'un pixel qu'on fera glisser tout au long de l'image. Le seuil dépendra alors du pixel et de l'information extraite à partir de son voisinage. Parmi ces méthodes, on peut citer la méthode de Niblack ou celle de Bernsen **[30]**.

La technique de seuillage local par niveau logique (LLT) proposée par Kamel et Zhao **[31]**, est basée sur la comparaison du niveau de gris d'un pixel avec les niveaux de gris moyen de quelques pixels voisins.

D'autres techniques consistent à subdiviser l'image en des petites fenêtres. Pour chacune d'entre elles, on calcule le seuil en utilisant l'une des méthodes de seuillage global. Ces méthodes ont été étudiées par plusieurs auteurs : **[32-39]**.

Le seuillage adaptatif convient aux images dont le fond n'est pas uniforme. C'est-à-dire que des variations d'éclairements sont présentes dans l'image.

1.4.3.2 Seuillage global

Dans les méthodes de seuillage global, un seuil unique est calculé pour tous les pixels de l'image. Ces méthodes reposent sur l'exploitation de l'histogramme de toute l'image qui caractérise la distribution des niveaux de gris. En général, une méthode de seuillage consiste à déterminer la valeur optimale du seuil S^* en se basant sur un certain critère.

Les méthodes de seuillage globales peuvent être réparties en deux grandes catégories :

- **les méthodes non paramétriques :** Ces méthodes permettent de trouver le seuil optimal de segmentation sans aucune estimation de paramètres. Généralement, ces méthodes sont basées sur l'optimisation de critères statistiques.
- **les méthodes paramétriques :** Ces méthodes supposent que les niveaux de gris des différentes classes de l'image suivent une certaine fonction de densité de probabilité. Généralement, ces fonctions de densités de probabilité sont supposées suivre un modèle Gaussien. En partant d'une approximation de l'histogramme de l'image par une combinaison linéaire de Gaussiennes, les seuils optimaux sont localisés à l'intersection de ces dernières.

1.4.3.2.1 Méthodes de seuillage non paramétrique

Ces méthodes consistent à déterminer le seuil optimal S à partir de l'histogramme de l'image. La méthode la plus connue est sans doute la méthode d'Otsu **[40]**. Celle-ci tente de segmenter l'image en 2 classes en maximisant un critère de séparabilité entre classes. L'opération de seuillage est vue comme une séparation des pixels d'une image en deux classes

C_0 et C_1 (objet et fond) à partir d'un seuil S. Ces deux classes sont désignées en fonction du seuil :

$C_0 = \{0,1,...,S\}$ *et* $C_1 = \{S+1,...,L-1\}$ où L est le nombre de niveaux de gris.

Soient : s_w^2 : la variance intraclasse , s_b^2 : la variance interclasse et s_t^2 : la variance totale.

Le seuil optimum S^* peut être déterminé en maximisant un des trois critères suivants :

$$l = \frac{s_b^2}{s_w^2} \qquad h = \frac{s_b^2}{s_t^2} \qquad k = \frac{s_t^2}{s_w^2}$$

Ces trois critères sont équivalents, mais le plus simple à utiliser est h .

Le seuil optimum S^* est défini par : $S^* = argmax\ (\eta)$

Cette expression mathématique signifie que S^* est le seuil optimum qui maximise le critère.

Les variances précédentes sont définies par :

$$s_t^2 = \sum_{i=0}^{L-1} (i - m_t)^2 p_i \quad \text{avec} \quad m_t = \sum_{i=0}^{L-1} i p_i$$

$$s_b^2 = p_S q (m_1 - m_2)^2 \qquad avec \qquad p_S = \sum_{i=0}^{S} h_i / N \qquad q = 1 - p_S$$

$$m_1 = \frac{m_t - m_s}{1 - p_t} \quad ; \quad m_2 = \frac{m_s}{p_t} \quad ; \quad m_s = \sum_{i=0}^{S} i p_i$$

$h(i)$: étant l'effectif d'apparition du niveau de gris i dans l'image et N le nombre de pixels de l'image. $p_i = \frac{h(i)}{N}$ correspond à la probabilité d'apparition du niveau de gris i.

D'autres méthodes de seuillage sont basées sur l'entropie de l'histogramme. On parle alors de seuillage entropique. Parmi ces méthodes, on peut citer les méthodes de Pun **[41]**, Kapur **[42]**, Johansen et Bille **[43]**, de cross entropie **[44]**, d'entropie de Renyi **[45]**, etc.

1.4.3.2.2 Méthodes de seuillage paramétrique

Soit h_g $g \in [0, L-1]$ l'histogramme de l'image, où L est le nombre de niveaux de gris. Cet histogramme peut être approché par un mélange de distributions où chaque distribution correspond à une classe : $\hat{h}_g = \sum_{k=1}^{K} P_k f(g/C_k)$

où P_k est la probabilité à priori de la classe C_k et $f(g/C_k)$ est la fonction de densité de probabilité de la classe C_k correspondant à la $k^{ième}$ distribution. Chaque distribution est caractérisée par des paramètres qu'on notera θ_k. Le problème de seuillage paramétrique consiste alors à estimer les paramètres (P_k, θ_k) de chaque distribution, en utilisant soit l'algorithme EM **[46]**, soit en minimisant l'erreur totale suivante :

$$J_1(\theta) = \frac{1}{L}\sum_{g=0}^{L-1}\left(h_g - \hat{h}_g\right)^2 \qquad (1.10)$$

Par l'intermédiaire d'algorithmes d'optimisation standards **[47], [48]** ou métaheuristiques **[49-51]**.

Généralement, toutes les distributions sont considérées du même type et Gaussienne. L'expression analytique d'une fonction de densité de probabilité Gaussienne est donnée par :

$$f_k(g/q_k) = f(g/m_k, s_k) = \frac{1}{s_k\sqrt{2p}} \exp\left[-\frac{(g-m_k)^2}{2(s_k)^2}\right] \qquad (1.11)$$

Où $q_k = (m_k, s_k)$ représente un vecteur dont les composantes sont : la moyenne et l'écart-type respectivement.

Dans le cas bimodal, Kittler et Illingworth considèrent l'histogramme comme un mélange de deux distributions Gaussiennes. L'une correspond à la classe "fond" et l'autre à la classe "objet " **[52]**. Ils déterminent le seuil optimal en résolvant l'équation suivante :

$$P_1 \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(g-\mu_1)^2}{2\sigma_1^2}} = P_2 \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(g-\mu_2)^2}{2\sigma_2^2}} \quad , \forall\, g. \qquad (1.12)$$

Ou en minimisant le critère suivant :

$$J(s) = 1 + 2[P_1(s)log\sigma_1(s) + P_2(s)log\sigma_2(s)] - 2[P_1(s)logP_1(s) + P_2(s)logP_2(s)] \qquad (1.13)$$

avec $P_1(s) = \sum_{i=0}^{s} h(i)$ $P_2(s) = \sum_{i=s+1}^{L-1} h(i)$

$$\sigma_1^2(s) = \frac{\sum_{i=0}^{s}(i-\mu_1(s))^2}{P_1(s)} \quad ; \quad \sigma_2^2(s) = \frac{\sum_{i=s+1}^{L-1}(i-\mu_2(s))^2}{P_2(s)} \tag{1.14}$$

$$\mu_1(s) = \frac{\sum_{i=0}^{s} i*h(i)}{P_1(s)} \quad ; \quad \mu_2(s) = \frac{\sum_{i=s+1}^{L-1} i*h(i)}{P_2(s)} \tag{1.15}$$

1.5 Conclusion

Deux grandes approches pour la segmentation d'images sont définies dans la littérature, il s'agit de l'approche contour et l'approche région. L'approche région contient un grand nombre de méthodes dont la plus part d'entre elles se basent sur la classification des pixels.

Les méthodes de classification des pixels consistent à affecter à chaque pixel une classe qui définit les régions à extraire de l'image. Elles se basent soit sur l'extraction des classes d'une manière non supervisée, soit sur le seuillage. Les méthodes de classification non supervisée sont très nombreuses et sont caractérisées par la simplicité de leur implémentation algorithmique. Les algorithmes K-means, Fuzzy C-means et Estimation-Maximisation sont parmi ces méthodes les plus connues. Lorsqu'on considère que le niveau de gris comme caractéristique d'un pixel, la classification des pixels débouche sur le seuillage. Les méthodes de seuillage ont pour objectif de segmenter une image en plusieurs classes en les délimitant par des seuils. Pour déterminer ces seuils, on utilise l'histogramme de l'image. À chaque pic de l'histogramme est associée une classe. Dans le cas du seuillage paramétrique, les différentes classes suivent une certaine fonction de densité de probabilité. Généralement ces fonctions de densité sont supposées suivre un modèle Gaussien.

Chapitre 2

Distributions statistiques

2.1 Introduction

L'approche de seuillage paramétrique à laquelle on s'intéresse dans ce mémoire fait appel à des notions de statistique. L'objectif de ce chapitre est donc de rappeler quelques résultats probabilistes utilisés en statistique mathématique. Il présente certaines distributions des variables aléatoires continues et décrit quelques méthodes d'estimation des paramètres des distributions, ainsi quelques méthodes de test d'adéquation utilisées pour déterminer le modèle statistique d'une distribution de données.

2.2 Variable aléatoire

Une des notions fondamentales des statistiques est celle de variable aléatoire **[53].** On considère un ensemble d'individus qui sera appelé Ω. Un individu de cet ensemble sera noté w . On note $X(w)$ une caractéristique de l'individu w. La quantité $X(.)$ est appelée variable aléatoire (v.a.). Les valeurs possibles que peut prendre $X(w)$ quand $w \in \Omega$ détermine la nature de la variable aléatoire. Ainsi, si $X(w)$ prend ses valeurs dans $\Re$, on parlera de variable aléatoire continue, si $X(w)$ prend ses valeurs dans un ensemble fini ou dénombrable, $X(w)$ sera alors appelée v.a. discrète.

2.3 Lois de probabilité discrètes

Pour complètement définir une loi de probabilité d'une v.a. discrète X , il suffit de définir la probabilité d'occurrence de chaque valeur k que peut prendre cette v.a. En d'autres termes, la donnée des quantités $P(X = k)$ et ceci pour toutes les valeurs k possibles déterminent une loi de probabilité particulière. De façon équivalente, pour complètement caractériser une loi de probabilité, il suffit de définir sa fonction de répartition, définie par : $F(n) = \sum_{k \le n} P(X \le k)$

Cette fonction s'interprète comme la probabilité que la v.a. X soit au plus égale à n. C'est évidemment une fonction positive et croissante.

Exemple **:** On lance une pièce ayant la probabilité P de tomber sur « pile » ; soit X la variable valant 1 si le résultat est pile, et 0 sinon: $P(X = 0) = 1 - P$; $P(X = 1) = P$; la loi P_X est appelée loi de Bernoulli $B(P)$ de paramètre P.

2.4 Lois de probabilité continues

2.4.1 Variable aléatoire continue

a. Définition

On dit que X est une variable aléatoire continue (v.a.c) si sa fonction de répartition F est continue et dérivable à gauche et à droite de tout point x de $\Re$.

La fonction dérivée f de F est dite fonction densité de probabilité de X et vérifie les relations :

$$\forall x \in R, f(x) = F'(x) \quad et \quad F(x) = \int_{-\infty}^{x} f(t)dt \tag{2.1}$$

Le support d'une v.a.c. X est un intervalle ou une réunion d'intervalles.

Le résultat suivant résume les principales propriétés de la fonction de densité de probabilité f d'une v.a.c. X .

b. Propriétés

- $\forall x \in R, f(x) \geq 0$ (f est positive).
- f est continue sur $\Re$ sauf peut être en un nombre fini de points où elle admet une limite finie à gauche et une limite à droite.
- L'intégrale $\int_{-\infty}^{+\infty} f(x)dx$ est convergente et on a $\int_{-\infty}^{+\infty} f(x)dx = 1$

c. Loi de probabilité d'une v.a.c.

- Pour tout nombre réel on a :

$$P(X \leq a) = F(a) = \int_{-\infty}^{a} f(x)dx \quad et \quad P(X > a) = \int_{a}^{+\infty} f(x)dx = 1 - F(a) \tag{2.2}$$

- Pour tout réel $\alpha \; et \; \beta \;\; tels \; que \;\; \alpha \leq \beta \; on \; a$:

$$P(a \leq X < b) = F(b) - F(a) = \int_{a}^{b} f(x)dx \tag{2.3}$$

d. Propriétés d'une fonction de répartition : La fonction de répartition F d'une variable aléatoire vérifie les conditions suivantes :

- F est croissante.
- F est continue et dérivable sur $\Re$ sauf peut être en un nombre de points où elle est continue à gauche ou à droite.
- $\lim_{x \to -\infty} F(x) = 0 \quad et \quad \lim_{x \to +\infty} F(x) = 1$

2.4.2 Caractéristiques d'une v.a.c

2.4.2.1 Espérance mathématique

On considère une v.a.c X de fonction de densité f :

Définition : On appelle espérance mathématique de X le nombre réel noté $E(X)$ défini par l'intégrale $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$ si elle est convergente.

Propriété :

- Si j est une application définie de $\Re$ dans $\Re$ et X une v.a.c. de densité f , alors la composée $Y = j(X)$ définit aussi une v.a.c. et son espérance mathématique $E(j(X))$ existe si et seulement si l'intégrale $\int_{-\infty}^{+\infty} |j(x)| f(x)dx$ est convergente. De plus on a : $E(f(X)) = \int_{-\infty}^{+\infty} j(x) f(x)dx$
- En particulier si j désigne une fonction affine ; $j(x) = a x + b$ avec $a, b \in R$, alors on établit que : $E(aX + b) = aE(X) + b$ (linéarité de l'espérance mathématique).

2.4.2.2 Variance et écart-type

Définition de la variance : Soit X une v.a.c. de fonction de densité de probabilité f . On appelle variance de X le nombre réel, noté $V(X)$, et défini par l'intégrale : $V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx$ si elle est convergente.

Définition de l'écart- type : La variance est toujours positive ou nulle (car étant l'intégrale d'une fonction positive). La racine carrée de la variance est appelée écart-type de X et noté $s(X)$. On a donc : $s(X) = \sqrt{V(X)}$ ou $s^2(X) = V(X)$.

2.4.3 Fonction caractéristique

Définition : Soit X une v.a.c. de fonction de densité de probabilité f . On appelle fonction caractéristique de X la fonction définie de $\Re$ dans $\Re$ par $\Psi_X(t) = E(e^{itx}) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx$ si l'intégrale est convergente.

2.4.4 Fonction génératrice des moments : La fonction génératrice des moments d'une v.a. X est définie par : $M_X(t) = E(e^{tX}), \quad t \in R$, lorsque son espérance existe.

Cette fonction comme son nom l'indique, est utilisée afin de générer les moments associées à la distribution de probabilité de la variable aléatoire X .

- Si à X est associée une densité de probabilité continue $f(x)$, alors la fonction génératrice des moments est donnée par : $M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x)dx$
- Si la densité de probabilité n'est pas continue, la fonction génératrice des moments peut être obtenue par : $M_X(t) = \int_R e^{tx} dF(x)$, où $F(x)$ est la fonction de répartition de X .

D'autres notions sur les paramètres statistiques sont rappelées dans l'annexe A.

2.4.5 Familles usuelles des distributions continues

Nous donnons, dans cette section les principales distributions continues univariées à travers leurs graphes et leurs fonctions de densités de probabilité et leurs caractéristiques statistiques essentielles (moyenne et variance).

A. Distribution uniforme

Soient a et b deux réels tels que $a < b$. Une v.a.c. X suit une loi uniforme entre a et b, notée $U(a, b)$, si l'événement a une chance égale de se produire dans l'intervalle $[a, b]$.

Sa fonction de densité de probabilité est donnée par :

$$f(x) = \begin{cases} \frac{1}{b-a} & si\ x \in [a, b] \\ 0 & si\ non \end{cases} \qquad (2.4)$$

Elle est illustrée sur la figure (2.1).

Sa fonction de répartition est donnée par :

$$F(x) = \begin{cases} 0 & si \quad x < a \\ \frac{x-a}{b-a} & si \quad x \in [a, b] \\ 1 & si \quad x > b \end{cases} \qquad (2.5)$$

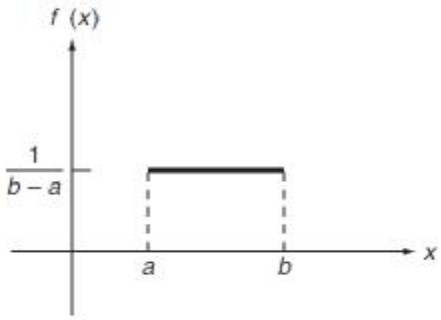


Figure 2.1 : Distribution Uniforme

B. Distribution Gamma

Une v.a.c. X suit une loi Gamma de paramètres $k \in N$ *et* θ, notée : $g(q,k)$, si sa fonction de densité de probabilité associée à X est :

$$f(x) = \frac{q^{k}}{\Gamma(k)} x^{k-1} e^{-qx}, x > 0 \tag{2.6}$$

où la fonction gamma Γ est l'intégrale récurrente telle que : $\Gamma(k) = \int_0^{+\infty} e^{-x} x^{k-1} dx$, θ est un paramètre d'échelle et k un paramètre de forme.

Cette distribution est illustrée sur la figure (2.2)

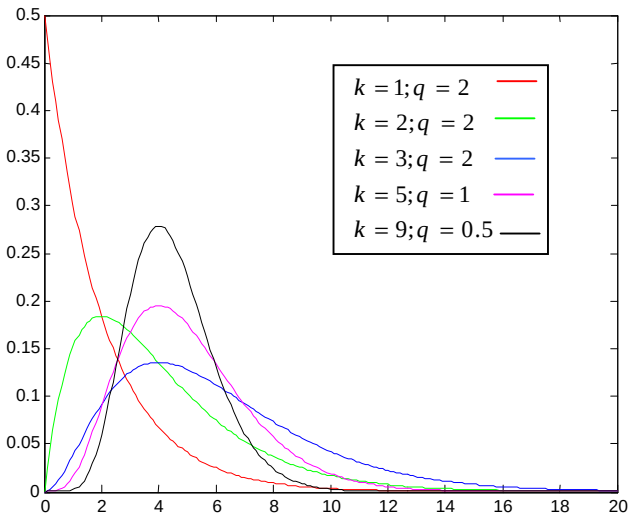


Figure 2.2 : Distribution Gamma

Sa fonction de répartition prend la forme suivante :

$$F(x) = \frac{\theta^{k}}{\Gamma(k)} \int_0^x u^{k-1} e^{-\theta k} du \tag{2.7}$$

C. **Distribution Bêta**

Une v.a.c. suit une loi bêta de paramètres $\alpha > 0 \; et \; \beta > 0$, notée $\mathrm{B}(\alpha, \beta)$, si sa fonction de densité de probabilité associéeest donnée par :

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)}, \quad x \in R$$

(2.8)

où $\mathrm{B}(a,b)$ est la fonction bêta définie par l'intégrale d'Euler:

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt; \alpha > 0, \beta > O. \tag{2.9}$$

$\alpha \; et \; \beta$ sont des paramètres de forme.

La figure (2.3) donne, pour quelques valeurs de $\alpha \; et \; \beta$, l'allure de cette fonction de densité de probabilité.

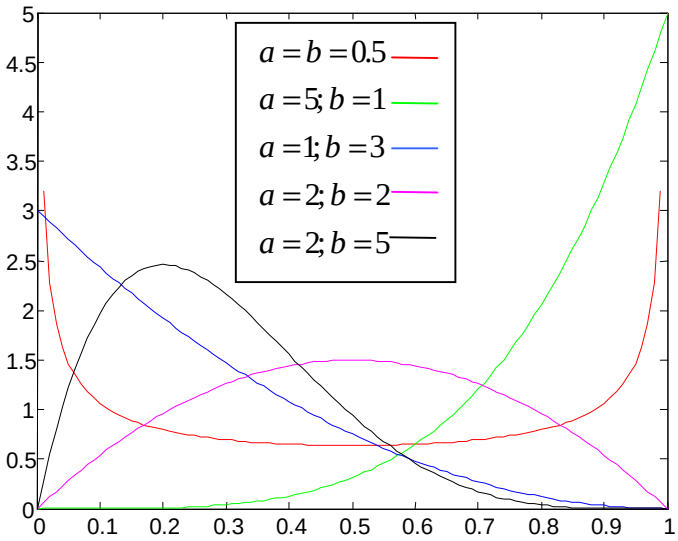


Figure 2.3 : Distribution Beta

Sa fonction de répartition est :

$$F(x) = \begin{cases} 0, & pour \; x < 0 \\ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x u^{\alpha-1}(1-u)^{\beta-1}du, & pour \; 0 \le x \le 1 \\ 1, & pour \; x > 1 \end{cases} \tag{2.10}$$

D. Distribution log-normale

Une v.a.c. suit une loi log-normale de paramètres $\mu \in R \; et \; \sigma > 0$, notée $L(\mu, \sigma)$, si sa fonction de densité de probabilité prend la forme suivante:

$$f(x) = \frac{1}{x\, s\, \sqrt{2p}} \exp\left\{ -\frac{1}{2}\left(\frac{\ln\, x - m}{s} \right)^2 \right\}, x > 0 \qquad (2.11)$$

La figure (2.4) donne, pour quelques valeurs de σ avec $\mu = 0$, l'allure de cette fonction de densité de probabilité.

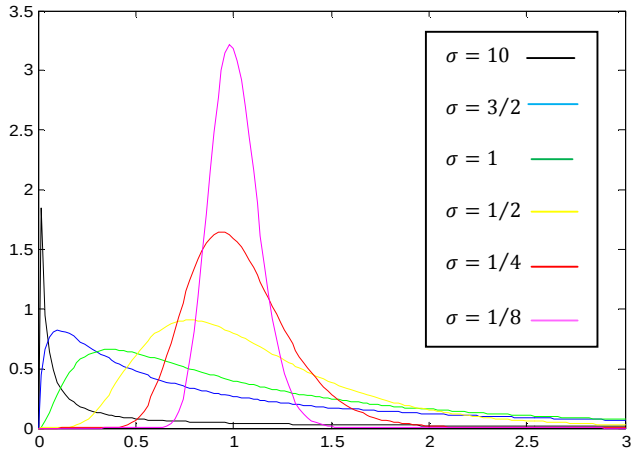


Figure 2.4 : Distribution Log-Normale

Sa fonction de répartition est donnée par :

$$F(x) = \frac{1}{2} + \frac{1}{2} erf\left[\frac{ln(x)-\mu}{\sigma\sqrt{2}}\right] \qquad (2.12)$$

où $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ est la fonction d'erreur.

E. Distribution normale ou de Gauss-Laplace

La loi normale joue un rôle particulièrement important dans la théorie des probabilités et dans les applications pratiques. La particularité fondamentale de la loi normale la distinguant des autres lois est que c'est une loi vers laquelle tendent les autres lois pour des conditions se rencontrant fréquemment en pratique.

La v.a.c. x suit une loi normale de paramètres $m \in R$ et $s > 0$, notée par $N(m, s)$ si sa fonction densité est donnée par :

$$f(x)=\frac{1}{s\sqrt{2p}}\exp\left[-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right], x\in R. \quad (2.13)$$

- Le paramètre m représente la moyenne, il détermine la position de la courbe, l'axe $x=m$ étant un axe de symétrie.
- Le paramètre σdésigne l'écart type, il détermine l'échelle (la position des valeurs autour de la moyenne).

La figure (2.5) montre l'allure de cette fonction pour différentes valeurs des paramètres $\mu\ et\ \sigma$.

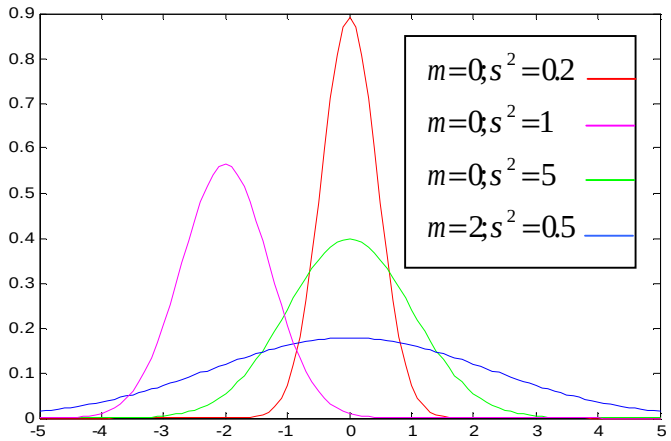


Figure 1.5 : Distribution Normale

La loi $N(\mu,\sigma)$ a pour fonction de répartition :

$$F(x)=\frac{1}{(2\pi)^{1/2}\sigma}\int_{-\infty}^{x} exp\left[-\frac{(u-m)^2}{2\sigma^2}\right]du,\ -\infty<x<+\infty \quad (2.14)$$

Dans le cas particulier $N(0,1)$, la loi normale est définie par :

$$\forall x\in R, \Phi(x)=\frac{1}{\sqrt{2p}}\int_{-\infty}^{x}\exp\left(\frac{-t^2}{2}\right)dt \quad (2.15)$$

- **Propriété de la fonction Φ**

Outre les propriétés d'une fonction de répartition, la fonction Φ vérifie les propriétés suivantes :

i) elle est indéfiniment dérivable et $\Phi'(x)=f(x)$.

ii) elle est strictement croissante de $]-\infty,+\infty[$ dans $]0,1[$. Elle est donc bijective et la réciproque est la fonction quantile Φ^{-1}.

iii) $\forall x\in R, \Phi(-x)=1-\Phi(x)$ (compte tenu de la parité de la fonction densité) et en particulier

iv) $\Phi(0) = 0.50$.

- **Approximations et valeurs de Φ**

Il n'existe pas d'expression explicite pour la fonction Φ, mais on fait appel à des méthodes numériques pour faire un calcul approché de l'intégrale. Par exemple un développement de série de Taylor à l'ordre 5 autour de 0 permet d'établir que : $\Phi(x) \simeq \frac{1}{2} + 0.3989423\left(x - \frac{x^3}{6} - \frac{x^5}{40}\right)$.

Cette approximation est performante pour $|x| < 2$.

F. Distribution exponentielle

Une v.a.c. X suit une loi exponentielle $E(\mu)$ de paramètre $m \in R$, si sa fonction densité s'exprime par : $f(x) = \frac{1}{m} e^{-\frac{x}{m}}$ (2.16)

On dit aussi que X suit une loi exponentielle de paramètre $l = \frac{1}{m}$ telle que la fonction de densité $E(\lambda)$ est:

$$f(x) = \lambda e^{-\lambda x} \quad (2.17)$$

La figure (2. 6) donne, pour quelques valeurs de l, l'allure de la fonction densité de probabilité.

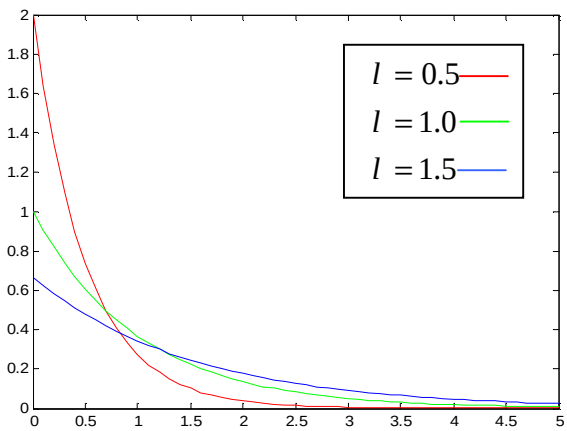


Figure 2.6 : Distribution Exponentielle

Sa fonction de répartition est :

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & pour\ x \geq 0 \\ 0 & sinon \end{cases} \quad (2.18)$$

G. Distribution de Khi2

On considère n variables aléatoires $X_1,..., X_n$ indépendantes suivant toutes la loi normale $N(0,1)$. La variable $Q = \sum_{i=1}^{v} X_i^2$ suit une loi du Khi2 à v degrés de liberté, notée $c^2(v)$. Sa fonction de densité de probabilité est:

$$f(x) = \frac{1}{2^{v/2}\,\Gamma\left(\frac{v}{2}\right)} . x^{\frac{v}{2}-1} . e^{\frac{x}{2}} \text{ avec } x \geq 0 \quad (2.19)$$

La figure (2.7) donne, pour quelques valeurs de v , l'allure de la fonction de densité de probabilité.

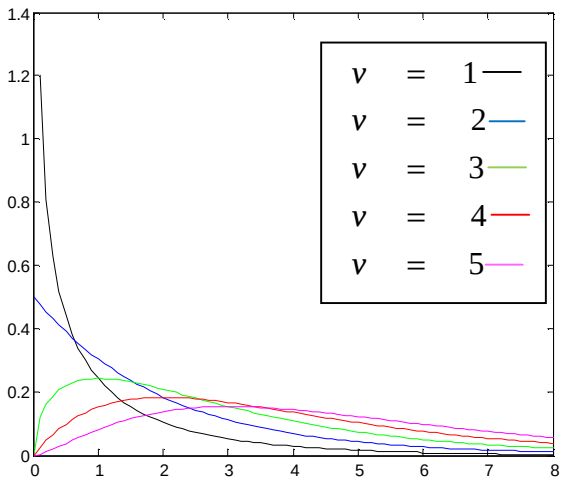


Figure 2.7 : Distribution Khi2

Sa fonction de répartition est :

$$\boldsymbol{F(x) = \frac{\gamma(v/2, x/2)}{\Gamma(v/2)}} \quad (2.20)$$

où γ est la fonction Gamma incomplète ($\gamma(a,x) = \int_0^x e^{-t} \cdot t^{a-1} dt$).

H. Distribution de Weibull

Une v.a.c. suit une loi de Weibull, notée $W(k,\lambda)$ de paramètres $k>0$ et $l>0$, si sa fonction densité de probabilité est:

$$f(x)=(k/l)(x/l)^{(k-1)}e^{-(x/l)^k} \tag{2.21}$$

où k est le paramètre de forme et $l>0$ le paramètre d'échelle.

La figure (2.8) donne, pour quelques valeurs de k , l'allure de la fonction de densité de probabilité :

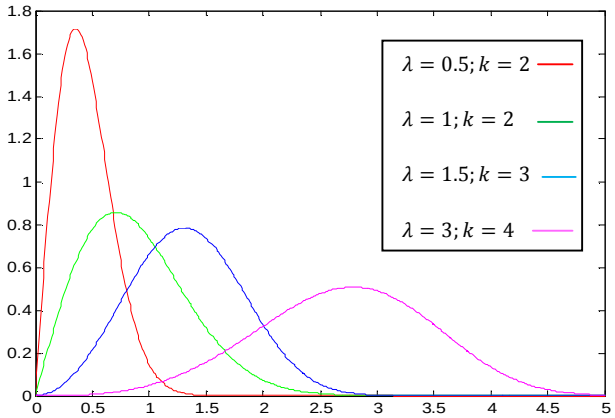


Figure 2.8 : Distribution Weibull

Remarque : on peut introduire un troisième paramètre θ dit de localisation. Dans ce cas, la fonction de densité de probabilité prend la forme suivante:

$$f(x)=\frac{k}{l}\left(\frac{x-q}{l}\right)^{k-1}e^{-\left(\frac{x-q}{l}\right)^k} \tag{2.22}$$

La fonction de répartition pour la loi de Weibull à 3-paramètres est définie par :

$$F(x)=1-e^{-\left(\frac{x-q}{l}\right)^k} \tag{2.23}$$

I. Distribution de Rayleigh

Une v.a.c. suit une loi de Rayleigh $R(\sigma)$ de paramètre $s>0$, si sa fonction densité prend la forme suivante:

$$f(x) = \frac{x}{s^2} \exp\left(\frac{-x^2}{2s^2}\right), \quad x \in [0 ; +\infty [\tag{2.24}$$

La figure (2.9) illustre, pour quelques valeurs de s , l'allure de la fonction de densité de probabilité.

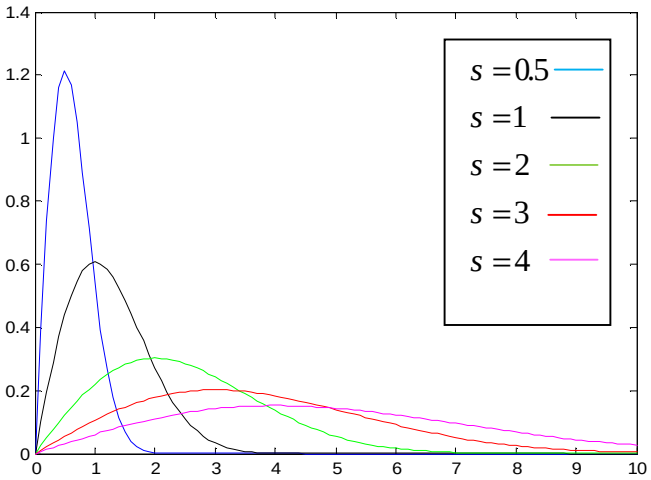


Figure 2.9 : Distribution Rayleigh

Sa fonction de répartition est :

$$F(x) = 1 - \left(\frac{x^2}{2\sigma^2}\right) \tag{2.25}$$

Les fonctions de densité de probabilité ainsi que les différentes caractéristiques statistiques principales (moyenne et variance) de chaque distribution citée précédemment sont résumés dans le tableau (2.1).

Remarque : Il existe bien d'autres distributions. On peut citer par exemple, les distributions de Student, Pareto, Cauchy, etc. Pour plus de détails sur toutes ces distributions, on peut se référer aux ouvrages **[54], [55].**

Distribution	*Fonction de densité de probabilité*	*Moyenne* $E(X)$	*Variance* $V(X)$
Uniforme $U(a,b)$	$f(x)=\begin{cases}\frac{1}{b-a} & si \quad x\in[a,b]\\ 0 & si \quad non\end{cases}$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
Normale $N(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2p}}\exp\left[-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right], x\in R.$	m	s^2
Gamma $g(q,k)$	$f(x)=\frac{q^k}{\Gamma(k)}x^{k-1}e^{-qx}, x>0$ Où $\Gamma(k)=\int_0^{+\infty}e^{-x}x^{k-1}dx$	$\frac{k}{q}$	$\frac{k}{q^2}$
Beta $B(\alpha,\beta)$	$f(x)=\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$ Où $B(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}dt; x>0, y>$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
Lognormal $L(\mu,\sigma)$	$f(x)=\frac{1}{xs\sqrt{2p}}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-m}{s}\right)^2\right\}, x>0$	$\left(m+\frac{s^2}{2}\right)$	$\left[e^{s^2}-1\right]e^{2m+s^2}$
Exponentielle $E(\lambda)$	$f(x)=l\,e^{-lx}$	$\frac{1}{l}$	$\frac{1}{l^2}$
Chi2 $\chi^2(v)$	$f(x;v)=\frac{1}{2^{v/2}\Gamma\left(\frac{v}{2}\right)}.x^{\frac{v}{2}-1}.e^{\frac{x}{2}} \quad x\geq 0$	v	$2v$
Weibull $W(k,\lambda)$	$f(x;k,l)=(k/l)(x/l)^{(k-1)}e^{-(x/l)^k}$	$l\Gamma\left(1+\frac{1}{k}\right)$	$l^2\Gamma\left(1+\frac{2}{k}\right)-E^2(X)$
Rayleigh $R(\sigma)$	$f(x;s)=\frac{x}{s^2}\exp\left(\frac{-x^2}{2s^2}\right) \quad x\geq 0$	$s\sqrt{\frac{p}{2}}$	$\frac{4-p}{2}s^2$

Tableau 2.1 : Tableau récapitulatif des lois de probabilité continues usuelles et leurs caractéristiques statistiques

2.5 Relations entre distributions

Une notion très importante des distributions statistiques et qui joue un rôle majeur en reconnaissance des lois de probabilité est les liens existants entre elles. En effet, sous certaines conditions, deux distributions peuvent être similaires. Quelques relations entre certaines distributions citées dans la table (2.1) sont données comme suit :

a) La distribution normale $N(\mu,\sigma)$pour les paramètres $\mu = 0$ et $\sigma = 1$ $(N(0,1))$ est reliée aux distributions suivantes :

- $\mathrm{B}(\alpha,\beta)$ quand les paramètres $\alpha\ et\beta$ tendent vers ∞ .
- $\chi^2(v)$ quand le paramètre v tend vers ∞ et la somme des carrés de v distributions normales unitaires indépendantes$N(0,1)$:$\sum_{i=1}\left(N(0,1)\right)_i^2 \approx \chi^2(v)$
- $\gamma(\theta,k)$ quand le paramètre σ tend vers ∞
- $L(\mu,\sigma)$ quand le paramètre σ tend vers 0

b) La distribution Log-Normale $L(\mu,\sigma)$ est liée à la distribution normale $N(\mu,\sigma)$ telle que :

$$L(\mu,\sigma) = e^{N(\mu,\sigma)}$$

- Pour des petites valeurs de σ, la distribution normale $N(\log(\mu),\sigma)$ donne une approximation de la distribution Log-Normale $L(\mu,\sigma)$

c) La distribution Beta $\mathrm{B}(\alpha,\beta)$ devient uniforme $U(a,b)$ avec $a = 0$ et $b = 1$ pour $\alpha = \beta = 1$ c.à.d. $\mathrm{B}(1,1) = U(0,1)$

- Elle est reliée à la distribution Gamma telle que :$\mathrm{B}(\alpha,\beta) = \frac{\gamma(1,\alpha)}{\gamma(1,\alpha)+\gamma(1,\beta)}$
- Elle aussi reliée à la distribution Chi2 $\chi^2(v)$ telle que :$\mathrm{B}\left(\frac{v_1}{2},\frac{v_2}{2}\right) = \frac{\chi^2(v)}{\chi^2(v_1)+\chi^2(v_2)}$

d) La distribution Chi2 $\chi^2(v)$ avec $v = 2$ est équivalente à la distribution exponentielle $E(\lambda)$ pour $\lambda = 2$, c.à.d., $\chi^2(2) = E(2)$ et à la distribution Rayleigh $R(\sigma)$ pour $\sigma = 1$, c.à.d. , $\chi^2(2) = R(1)$.

- Elle est liée à la distribution Gamma comme suit : $\chi^2(v) = 2\gamma\left(1,\frac{v}{2}\right) = \gamma\left(2,\frac{v}{2}\right)$
- Elle est liée à la distribution Beta $\mathrm{B}(\alpha,\beta)$ comme suit :$\mathrm{B}\left(\frac{v_1}{2},\frac{v_2}{2}\right) = \frac{\chi^2(v_1)}{\chi^2(v_1)+\chi^2(v_2)}$

- Elle est équivalente à la somme des carrés de v distributions normales unitaires indépendantes $N(0,1)$: $\chi^2(v) = \sum_{i=1}^{v}\big(N(0,1)\big)$ *et* $\sum_{i=1}^{v}\big(N(0,1)\big)_i^2 = \sum_{i=1}^{v}\left(\frac{N(\mu_i,\sigma_i)-\mu_i}{\sigma_i}\right)^2$
- Pour des grandes valeurs de v, la distribution Chi2 peut être approximée par des transformations de la distribution normale :

$$\chi^2(v) = \frac{1}{2}\left[(2v-1)^{\frac{1}{2}} + N(0,1)\right]^2 \tag{2.26}$$

$$\chi^2(v) = \frac{1}{2}\left[1 - \frac{2}{9v} + \left(\frac{2}{9v}\right)^{\frac{1}{2}} N(0,1)\right]^3 \tag{2.27}$$

La première approximation dite de Fisher est moins significative que la deuxième dite approximation de Wilson-Hilferty.

e) La distribution exponentielle $E(\lambda)$ est un cas particulier de la distribution Gamma $\gamma(\theta, \mathrm{k})$, $(E(\lambda) = \gamma(\lambda, 1))$ et de la distribution Weibull $W(k,\lambda)$, c.à.d., $E(\lambda) = W(k,1)$. Elle est aussi reliée à la distribution uniforme : $E(\lambda) = -\lambda log U(0,1)$

- La somme de θ distributions exponentielles indépendantes $E(\lambda)$ donne la distribution Gamma $\gamma(\theta,k)$ pour le paramètre entier k : $\gamma(\theta,k) = \sum_{i=1}^{k}\big(E(\lambda)\big)_i$

f) La distribution Gamma est liée à :

- La distribution exponentielle : $\gamma(\theta,1) = E(\lambda)$et $\gamma(\theta,k) = \sum_{i=1}^{k}\big(E(\lambda)\big)_i$
- La distribution Chi2 : $\gamma(1,k) = \frac{1}{2}\chi^2(2k)$ avec k est un entier.
- La distribution Beta : $\mathrm{B}(\alpha,\beta) = \frac{\gamma(1,\alpha)}{\gamma(1,\alpha)+\gamma(1,\beta)}$
- $\sum_{i=1}^{N}\gamma(\theta,k_i) = \gamma(\theta,k)$ avec $k = \sum_{i=1}^{N} k_i$

g) La distribution Rayleigh $R(\sigma)$ est liée à la distribution Weibull $W(\alpha,\beta)$ pour $\alpha = \beta = 2$, c.à.d. $R(\sigma) = W(2,2)$ et aussi à la distribution Chi2 :$R(\sigma = 1) = \chi^2(v = 2)$

- Le carré de la distribution Rayleigh est relié à la distribution exponentielle $E(\lambda)$ pour $\lambda = \frac{1}{2\sigma^2}$: $[R(\lambda)]^2 = E\left(\frac{1}{2\sigma^2}\right)$

- La distribution Rayleigh est aussi liée aux distributions normales indépendantes : $R(\sigma) \approx \left[\left(N(0,\sigma)\right)_1^2 + \left(N(0,\sigma)\right)_2^2\right]^{\frac{1}{2}}$

2.6 Estimation des paramètres des distributions

Toute étude statistique repose sur un ensemble d'observations, c'est-à-dire sur les valeurs empiriques x_i (obtenues dans le cadre d'une expérience statistique) d'une variable aléatoire X, issu d'une loi P de paramètres $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ entièrement ou partiellement inconnue. Considérons n répétitions indépendantes de cette expérience statistique et désignons par $x_1, x_2, ..., x_n$ l'ensemble de valeurs observées et par $(X_1, \dots, X_n)$ un échantillon.

Le problème de l'estimation des paramètres inconnus se pose lorsqu'on cherche, à partir de l'échantillon X_n, le vecteur de paramètre $\hat{\theta}$ de la loi P.

$\hat{q}$ est une approximation de q dépendant de l'échantillon $(X_1, X_2, ..., X_n)$.

2.6.1 Méthodes d'estimation

2.6.1.1 Méthode du maximum de vraisemblance

L'estimation du maximum de vraisemblance est une méthode statistique courante utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon donné.

La fonction de vraisemblance, notée $L(X_1, ..., X_n; q)$, est fonction des probabilités conditionnelles qui décrit le paramètre q d'une loi statistique en fonction des valeurs x_i supposées connues. Elle s'exprime à partir de la fonction de densité de probabilité conditionnelle $f(x/q)$ par :

$$L(X_1, ..., X_n; q) = \prod_{i=1}^{n} f(X_i; q) \qquad (2.28)$$

cette formule n'est valable que si on suppose que les X_i sont indépendants entre eux.

L'estimation du vecteur paramètre $\hat{\theta}$ revient à maximiser la fonction de vraisemblance pour que les probabilités des réalisations observées soient aussi maximales. Ceci constitue un problème d'optimisation dont la solution est celle du système suivant :

$$(I) \quad \begin{cases} \dfrac{\partial L}{\partial q} = 0 \\ \left(\dfrac{\partial^2 L}{\partial q^2} \right)_{q=\hat{q}} < 0 \end{cases} \tag{2.29}$$

A cause de la forme particulière des densités de probabilité des distributions usuelles de probabilité, il est plus aisé d'utiliser le logarithme de la vraisemblance, $\log L(X_1,\ldots, X_n, q)$, si $f(x,q) > 0, \forall x, \forall q$.

Le système (I) est donc équivalent au système (II) suivant :

$$(II) \quad \begin{cases} \dfrac{\partial \log L}{\partial q} = 0 \\ \left(\dfrac{\partial^2 \log L}{\partial q^2} \right)_{q=\hat{q}} < 0 \end{cases} \tag{2.30}$$

La première équation du système (II) $\dfrac{\partial \log L}{\partial q} = 0$ s'appelle équation de vraisemblance.

Pour un échantillon indépendant X_n , l'équation de vraisemblance s'écrit:

$$\sum_{i=1}^{n} \frac{\partial \log f(x_i, q)}{\partial q} = 0 \tag{2.31}$$

La résolution de cette équation par rapport à chaque paramètre $\theta_j (j = 1, \ldots, M)$ permet d'aboutir à la solution $\hat{\theta}$.

2.6.1.2 Méthode des moments

Cette procédure d'estimation repose sur la propriété de convergence presque sûre des moments empiriques d'un échantillon $X_n = (X_1, \ldots, X_n)$, extrait de X, vers les moments théoriques correspondants de X **[56]**.

a. Moments

Le moment d'ordre $k \in N^*$, s'il existe, d'une variable aléatoire X est défini par :

$$m_k = E[X^k] = \int x^k f(x) dx \tag{2.32}$$

avec $f(x)$ la fonction de densité de probabilité.

Le moment centré d'ordre $k>1$ est défini par :

$$m_k^c = [\ (X - E[X])^k\ ] = \int_R (x - m_1)^k\ f(x)dx \qquad (2.33)$$

b. Moments empiriques

On appelle moment empirique, non centré, d'ordre k la quantité suivante :

$$m_k(e,n) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

(2.34)

et le moment empirique, centré, d'ordre k est défini par :

$$m_k^c(e,n) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k$$

(2.35)

où $\bar{X} = m_1(e,n)$ est la moyenne empirique de l'échantillon.

Ø Estimation par la méthode des moments

Soit le vecteur paramètre $q = (q_1,...q_m)$ à estimer : On note par $m_k(q)$ le moment théorique d'ordre k de X , qu'il soit centré ou non, et par $m_k(e,n)$ le moment empirique d'ordre k.

Définition : On appelle estimateur $\hat{\theta}$ de q , obtenu par la méthode des moments (EMM), la solution du système d'équations suivant :

$$\begin{cases} m_1(q) = m_1(e,n) \\ \quad\vdots \\ m_m(q) = m_m(e,n) \end{cases} \qquad (2.36)$$

Remarque : Le choix des moments est guidé par la facilité de résolution du système. On peut prendre des moments tous centrés, ou tous non centrés, ou un mélange de moments centrés et

non centrés. En outre, il n'y a aucune raison de choisir les m premiers moments, sinon la simplicité des calculs.

Le tableau (2.2) donne les expressions et la méthode d'estimation des paramètres estimés des différentes distributions présentées dans le tableau (2.1).

Distribution	Paramètres estimés	Méthode d'estimation des paramètres
Uniforme $U(a,b)$	$a=\bar{x}-3^{1/2}s$ et $b=\bar{x}+3^{1/2}s$	Moments
Normale $N(\mu,\sigma^2)$	$m=\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i \quad s^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2$	Maximum de vraisemblance
Gamma $g(q,k)$	$q=\frac{\bar{x}}{s^2}$ and $k=\left(\frac{\bar{x}}{s}\right)^2$	Moments
Beta $\mathrm{B}(a,b)$	$a=\bar{x}\left(\frac{\bar{x}(1-\bar{x})}{s^2}-1\right) \quad b=(1-\bar{x})\left(\frac{\bar{x}(1-\bar{x})}{s^2}-1\right)$	Moments
Lognormal $L(\mu,\sigma)$	$m=\frac{1}{N}\sum_{i=1}^{N}\log(x_i) \quad s^2=\frac{1}{(N-1)}\sum_{i=1}^{N}[\log(x_i-m)]^2$	Moments
Exponentielle $E(\lambda)$	$l=\frac{1}{\bar{x}}$	Maximum de vraisemblance
Chi2 $\chi^2(v)$	$n=\bar{x}$	Moments
Weibull $W(k,\lambda)$	$k=\left(\frac{1}{N}\sum_{i=1}^{N}x_i^l\right)^{1/l}$ $l=\frac{N}{\left(\frac{1}{k}\right)\sum_{i=1}^{N}x_i^l\log(x_i)-\sum_{i=1}^{N}\log(x_i)}$	Maximum de vraisemblance
Rayleigh $R(\sigma)$	$s=\left(\frac{1}{2N}\sum_{i=1}^{N}x_i^2\right)^{1/2}$	Maximum de vraisemblance

Tableau 2.2 : Tableau récapitulatif des expressions des paramètres estimés des différentes distributions ainsi que le type de la méthode utilisée

2.7 Distributions généralisées

2.7.1 Distribution Gaussienne généralisée

Il s'agit d'une famille de lois qui constitue une extension de la loi Gaussienne. Elle est caractérisée par trois paramètres : la moyenne m, la variance s^2 et le paramètre de forme g. C'est ce dernier paramètre qui permet de couvrir en plus de la loi Gaussienne $(g = 2)$, des lois pointues dites sur-Gaussiennes $(g < 2)$ et des lois aplaties dites sous-Gaussiennes $(g > 2)$. comme la montre la figure (2.10).

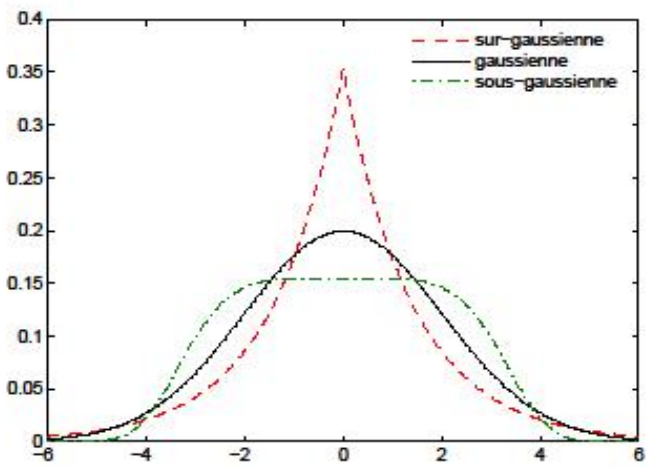


Figure 2.10 : Allure de la fonction de densité de probabilité gaussienne généralisée de moyenne nulle et de variance 2 pour différentes valeurs du paramètre de forme : sur-gaussienne $(g = 1)$, gaussienne $(g = 2)$ et sous-gaussienne $(g = 5)$

Une variable aléatoire Gaussienne généralisée $x \in \Re$ a pour la fonction de densité **[57] :**

$$f(x) = \frac{h(g)g}{[2\Gamma(1/g)]} \exp\left[-(h(g)|x-m|)^g\right] \tag{2.37}$$

avec $h(g) = \left[\frac{\Gamma(3/g)}{s^2\Gamma(1/g)}\right]^{\frac{1}{2}}$ où $\Gamma(g)$ est la fonction Gamma telle que : $\Gamma(g) = \int_0^\infty z^{g-1}\exp(-z)dz$

L'estimation des paramètres de la Gaussienne généralisée n'est pas aussi simple que dans le cas Gaussien. L'utilisation de la méthode des moments donne les mêmes expressions que celles dans le cas gaussien pour la moyenne et la variance, c'est-à-dire :

$$\hat{m} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \hat{s}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \hat{m}\right)^2 \tag{2.38}$$

Par contre, le paramètre de forme peut être retrouvé en résolvant numériquement l'équation ci-après obtenue avec le moment d'ordre 4 **[58]** :

$$\hat{m}_4 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{m}^4\right) = s^4 \frac{\Gamma\left(\frac{5}{g}\right)\Gamma\left(\frac{1}{g}\right)}{\Gamma^2\left(\frac{3}{g}\right)} \tag{2.39}$$

Provost et al. ont également proposé dans **[59]** une méthode d'estimation hybride consistant à estimer le paramètre de forme et la variance par maximum de vraisemblance et la moyenne par la méthode des moments. Les estimateurs correspondants sont les suivants :

Moyenne : $\hat{m} = \frac{1}{N}\sum_{i=1}^{N} x_i$ (2.40)

Variance : $\hat{s}^2 = \frac{\Gamma(3/\hat{g})}{\Gamma(1/\hat{g})}\left[\left(\frac{\hat{g}}{N}\right)^{\frac{1}{\hat{g}}} G_{\hat{g}}\right]^2$ (2.41)

Paramètre de forme : solution de l'équation :

$$g + \Psi(1/g) + \log(g/N) + \log G_g - g\frac{G_g'}{G_g} = 0 \tag{2.42}$$

Notons que Ψ est la fonction digamma telle que : $\Psi(x) = \frac{\partial}{\partial l}\int_0^{\infty} z^{x-1}\exp(-z)dz$ et G_g est la $g-$*norme* à la puissance N : $G_g = \sum_{i=1}^{N}\left|x_i - \hat{m}\right|^g$ et G_g' sa dérivée par rapport à g : $G_g' = \frac{\partial G_g}{\partial g}$

Cette technique d'estimation a été, notamment, utilisée pour la classification non supervisée des images de télédétection SPOT **[60]**.

2.7.2 Distribution Gamma généralisée

Une version plus flexible de la distribution Gamma est obtenue en lui ajoutant un troisième paramètre. C'est la distribution gamma généralisée.

Une variable aléatoire suit une loi Gamma généralisée, si pour $x \in \Re$ on a**: [61]**

$$f(x;a,b,c) = ac(ax)^{bc-1} e^{-(ax)^c} / \Gamma(b) \qquad (2.43)$$

a et b sont les mêmes paramètres employés pour la distribution Gamma. c est le paramètre qui caractérise cette distribution généralisée (pour $c = 1$ on obtient la loi gamma ordinaire).

La figure (2.11) donne, pour quelques valeurs de c dans le cas où $a = 1$ et $b = 2$, l'allure de la fonction de densité de probabilité.

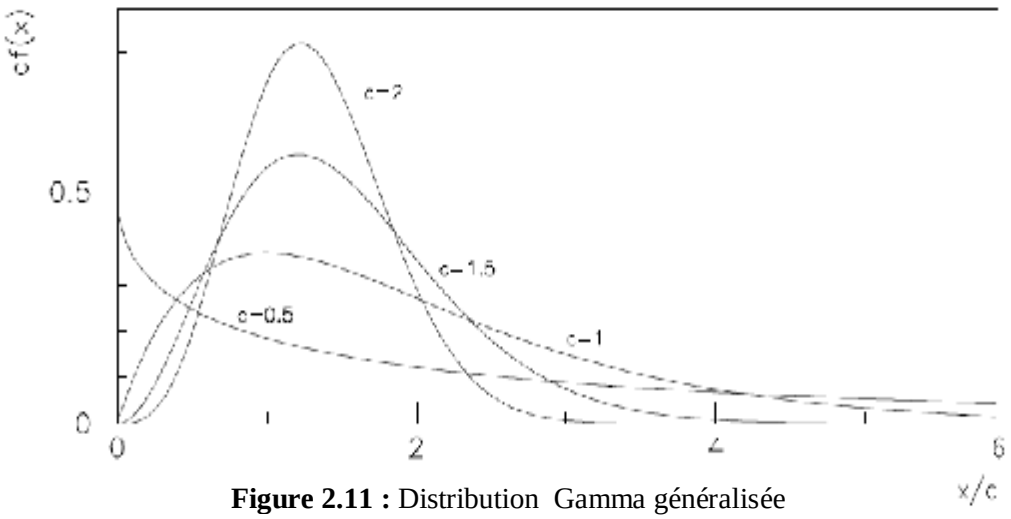


Figure 2.11 : Distribution Gamma généralisée

La distribution gamma généralisée est une forme générale qui pour certaines combinaisons de paramètres permet de décrire d'autres distributions (tableau 2.3).

Distribution	a	b	c
Gamma	a	b	1
Chi2	½	n/2	1
Exponentielle	1/α	1	1
Weibull	1/σ	1	η
Rayleigh	$1/\alpha\sqrt{2}$	1	2
Normale	$1/\sqrt{2}$	½	2

Tableau 2.3 : Distribution Gamma généralisée et ses relations avec d'autres distributions

2.7.3 Système des distributions de Pearson

L'intérêt du système de Pearson est qu'il couvre une marge gamme de formes des distributions (huit familles de lois) avec un nombre très limité de paramètres puisque les 4 premiers moments suffisent.

2.7.3.1 Définition

Une fonction de densité f, sur $\Re$, appartient au système des distributions de Pearson, si elle est solution de l'équation différentielle suivante :

$$\frac{1}{f(x)}\frac{df(x)}{dx} = -\frac{x+a}{c_0 + c_1 x + c_2 x^2} \tag{2.44}$$

où les paramètres a, c_0, c_1, c_2 varient suivant la forme de la fonction f. Leur valeurs caractérisent complètement le système de distributions de Pearson. Pour les estimer, on utilise souvent la méthode des moments.

Les solutions de cette équation différentielle sont fortement liées à l'existence et au type des solutions du polynôme $P(x) = c_0 + c_1 x + c_2 x^2$. On ne peut, par conséquent, pas donner une forme générique des solutions. On ne s'intéresse, de plus, qu'aux solutions qui sont des densités.

2.7.3.2 Distributions appartenant au système de Pearson

On peut, répartir les fonctions de densité de probabilité $f(x)$ solutions de l'équation différentielle précédente en huit familles distinctes $\{F_1, \ldots\ldots, F_8\}$ selon le polynôme $p(x)$.

- La famille F_4 correspond au cas où $P(x)$n'a pas de racines réelles. Ses éléments ont des distributions de cette famille ont une fonction de densité de probabilité:

$$f_4(x) = K_4\left[C_0 + c_2(x + C_1)^2\right]^{-\left(\frac{1}{2c_2}\right)} \exp\left(-\frac{a - C_1}{\sqrt{C_0 c_2}}\arctan\left(\sqrt{\frac{c_2}{C_0}}(x + C_1)\right)\right) \tag{2.45}$$

avec $C_0 = c_0 - \frac{c_1^2}{4c_2}, C_1 = \frac{c_1}{2c_2}$.

- La famille F_7 englobe les distributions de densités :

$$f_7(x) = K_7 \left[c_0 + c_2 x^2\right]^{\left(\frac{1}{2c_2}\right)} \exp\left(-\frac{a}{\sqrt{c_0 c_2}} \arctan\left(\sqrt{\frac{c_2}{c_0}} x\right)\right) \tag{2.46}$$

Elle découle de la famille F_4 dans laquelle $c_1 = 0$.

K_4 et K_7 sont les constantes de normalisation $K_i = \int_{IR} f_i(x)dx, \qquad i = 4,7$

Les autres distributions sont très utilisées en statistiques, et ont un réel intérêt pratique.

- La famille $F_1 \cup F_2$ est la famille des lois Bêtas, de densité :

$$f_{1,2}(x) = \frac{1}{b(p,q)} \frac{(x - b_1)^{p-1}(b_2 - x)^{q-1}}{(b_2 - b_1)^{p+q-1}} \tag{2.47}$$

Avec $b_1 = -\frac{1}{2c_2}\left(c_1 + \sqrt{\Delta}\right)$, $b_2 = -\frac{1}{2c_2}\left(c_1 - \sqrt{\Delta}\right)$, $p = \frac{a+b_1}{c_2(b_2 - b_1)} + 1$, $q = -\frac{a+b_2}{c_2(b_2 - b_1)} + 1$,

Où $\Delta = c_1^2 - 4c_0c_2$

Ces deux types correspondent au cas où $P(x)$ a deux racines réelles de signes opposés. F_2 étant le cas où $p = q$.

- La famille F_6, pour laquelle $P(x)$ a deux racines réelles distinctes et de même signe, est la famille des lois Bêta du second type. Leur fonction de densité de probabilité est de la forme :

$$f_6(x) = \frac{s^q}{B(p,q)} \frac{(x - r)^{p-1}}{(x - (r - s))^{p+q}} \tag{2.48}$$

Avec $s = \frac{\sqrt{\Delta}}{c_2}$, $r = -\frac{1}{2c_2}\left(c_1 - \sqrt{\Delta}\right)$, $p = \frac{a+r}{c_2 s} + 1$, $q = \frac{1}{c_2} - 1$

- La famille F_5 est la famille des lois inverse-Gammas, obtenue dans le cas où $P(x)$ est un carré parfait :

$$f_5(x) = \frac{1}{\Gamma(q)} \frac{(x - r)^{-q-1}}{p^q} \exp\left(-\frac{2}{p(x - r)}\right) \tag{2.49}$$

Avec $p = \frac{c_2}{a - \frac{c_1}{2c_2}}$, $q = \frac{1}{c_2} - 1$, $r = -\frac{c_1}{2c_2}$

- Dans le cas où $P(x)$ n'est pas de second degré $(c_2 = 0)$ et que $c_1 \neq 0$, on obtient la famille F_3 des lois Gamma dont la fonction de densité est :

$$f_3(x) = \frac{1}{\Gamma(q)} \frac{(x-r)^{q-1}}{p^q} \exp\left(-\frac{(x-r)}{p}\right) \tag{2.50}$$

avec $p = c_1$, $q = \frac{1}{c_1}\left(\frac{c_0}{c_1} - a\right) + 1$, $r = -\frac{c_0}{2c_1}$.

- Si $c_2 = c_1 = 0$, on obtient la famille F_8 est la famille de lois normales dont la fonction de densité est :

$$f_8(x) = \frac{1}{\sqrt{2ps^2}} \exp\left(-\frac{(x-m)^2}{2s^2}\right) \text{Avec } m = -a, s^2 = c_0. \tag{2.51}$$

Remarque : La définition des distributions du système de Pearson par la même équation différentielle permet de déterminer ces distributions uniquement à partir de leurs quatre premiers moments.

2.7.3.3 Graphe de Pearson

Le système de Pearson permet de décrire huit familles de distributions en se limitant au calcul de quelques paramètres qui sont la moyenne m_1 et les moments centrés m_q.

Soit X une variable aléatoire dont la densité appartient au système de Pearson, son moment d'ordre 1 est donné par:

$$m_1 = E(X) \tag{2.52}$$

et pour $q = 2,3,4$, les moments centrés d'ordre q :

$$\mu_q = E[(X - E[X])^q] \tag{2.53}$$

grâce auxquels on définit les coefficients :

$$g_1 = \frac{(m_3)^2}{(m_2)^3} \quad \text{et} \quad g_2 = \frac{m_4}{(m_2)^2} \tag{2.54}$$

$\sqrt{g_1}$ est appelée « skewness » et g_2 « kurtosis ». Ces paramètres décrivent respectivement l'asymétrie et l'aplatissement d'une distribution. Toute distribution de ce système est identifiée dans le graphe de Pearson (figure 2.12), exprimant g_2 en fonction de g_1 .

Les quatre coefficients de l'équation différentielle sont alors reliés à m_1 et m_2 par les relations **[62]**:

$$a = \frac{(g_2+3)\sqrt{g_1 m_2}}{10g_2 - 12g_1 - 18} - m_1, \tag{2.55}$$

$$c_0 = \frac{m_2(4g_2 - 3g_1) - m_1(g_2+3)\sqrt{g_1 m_2} + (m_1)^2(2g_2 - 3g_1 - 6)}{10g_2 - 12g_1 - 18} \tag{2.56}$$

$$c_1 = \frac{(g_2+3)\sqrt{g_1 m_2} - 2m_1(2g_2 - 3g_1 - 6)}{10g_2 - 12g_1 - 18} \tag{2.57}$$

$$c_2 = \frac{(2g_2 - 3g_1 - 6)}{10g_2 - 12g_1 - 18} \tag{2.58}$$

Les solutions de l'équation différentielle suivant les valeurs de g_1, g_2 et du paramètre l , permettant de discuter les différentes solutions de $c_0 + c_1 x + c_2 x^2$, où :

$$l = \frac{c_1^2}{4c_0 c_2} = \frac{g_1(g_2+3)^2}{4(4g_2 - 3g_1)(2g_2 - 3g_1)(2g_2 - 3g_1 - 6)} \tag{2.59}$$

Une fonction de densité de probabilité f appartient affectée à l'une des huit familles $\{F_1, ..., F_8\}$ selon les règles suivantes **[62], [63]** :

$$\{f \in F_1\} \Leftrightarrow \{l < 0\},$$

$$\{f \in F_2\} \Leftrightarrow \{l = 0 \, avec \, \{g_1 = 0 \, et \, g_2 < 3\}\},$$

$$\{f \in F_3\} \Leftrightarrow \{l \, est \, \infty \, car \, \{2g_2 - 3g_1 - 6 = 0\}\},$$

$$\{f \in F_4\} \Leftrightarrow \{0 < l < 1\},$$

$$\{f \in F_5\} \Leftrightarrow \{l = 1\},$$

$$\{f \in F_6\} \Leftrightarrow \{l > 1\},$$

$$\{f \in F_7\} \Leftrightarrow \{l = 0 \, avec \, \{g_1 = 0 \, et \, g_2 > 3\}\},$$

$$\{f \in F_8\} \Leftrightarrow \{l = 0 \, avec \, \{g_1 = 0 \, et \, g_2 3\}\}. \tag{2.60}$$

La connaissance des moments $m_1, m_2, m_3 et m_4$ permet, donc, de déterminer la famille F_i à laquelle appartient une distribution, ainsi que les paramètres de cette distribution.

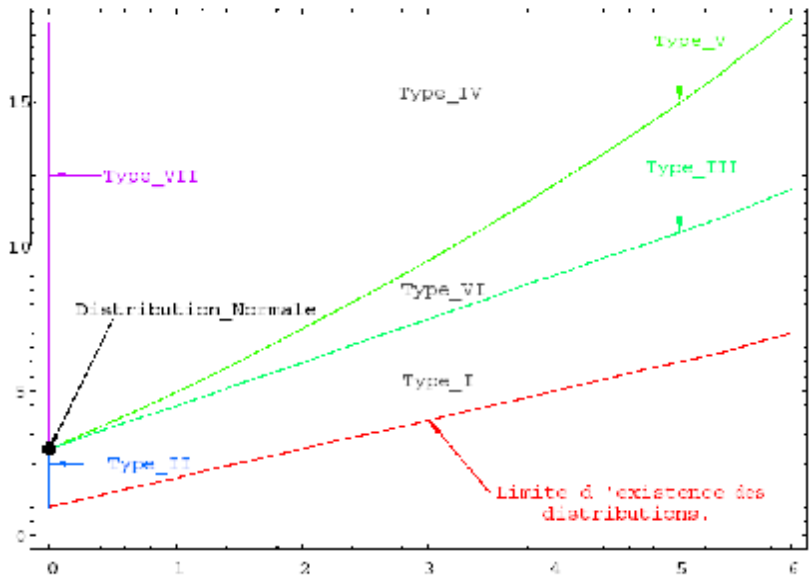


Figure 2.12 : Graphe de Pearson

2.8 Tests d'adéquation

Les tests d'adéquation permettent de vérifier si une variable aléatoire donnée suit une loi de distribution connue à priori (loi binomiale, loi normale etc..) dite théorique et notée $f(x)$.

Les tests d'adéquation ne permettent pas de trouver la loi d'une variable aléatoire, mais seulement d'accepter ou de rejeter une hypothèse simple mise à priori, généralement, notée $H_0 \, ou \, H_1$ telle que :

H_0: la distribution de probabilité théorique qui a engendrée l'échantillon.

H_1: la distribution de probabilité théorique qui n'a pas engendré l'échantillon.

Ils consistent à comparer la distribution observée (distribution empirique) dans l'échantillon à une distribution théorique.

2.8.1 Le test de Khi deux (c^2)

Il est sans doute le plus utilisé. Il consiste en premier lieu de répartir les valeurs de l'échantillon de taille N en m classes ou catégories $c_1, c_2, \dots, c_m$. Soient $n_1, n_2, \dots, n_m$ les

effectifs observés (fréquences empiriques) où n_i représentent le nombre d'éléments de l'échantillon ayant une valeur dans la classe c_i. n_i représente l'histogramme empirique.

Soient $p_1,..., p_m$ les probabilités des m classes calculées à partir d'une loi théorique donnée, parfaitement spécifiée, de fonction de répartition connue F . La répartition de l'échantillon de taille N sur les m classes suit une loi multinomiale $M(N, p_1,..., p_m)$. Soit F_n la loi empirique observée à partir de l'échantillon.

On définit « distance de Khi deux » entre la loi théorique F et la loi empirique observée F_n la quantité :

$$D(F_n, F) = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i} \tag{2.61}$$

où np_i représentent les effectifs théoriques. np_i doit être supérieur à 5 sinon, on doit procéder au regroupement des classes.

En notant $\hat{p}_i = \frac{n_i}{n}$, la proportion empirique, on peut écrire :

$$D(F_n, F) = D(\hat{p}, p) = n\sum_{i=1}^{m} \frac{(\hat{p}_i - p_i)^2}{p_i} \tag{2.62}$$

$$D(\hat{p}, p) = \sum_{i=1}^{m} \frac{n_i^2}{np_i} - n \tag{2.63}$$

$D(\hat{p}, p)$ s'appelle aussi distance entre les distributions p et $\hat{p}$ centrée en p .

Pour savoir si les observations proviennent bien de la loi théorique F, on pose le problème du test d'adéquation comme suit : $\begin{cases} H_0 : \ la\ loi\ de\ X\ est\ F \\ H_1 : \ la\ loi\ de\ X\ n'estpas\ F \end{cases}$

Intuitivement, si X suit approximativement la loi F , $D(\hat{p}, p)$ doit être « petit ».

La distance $D(\hat{P}, P)$ suit en générale une loi de χ^2 de degrés de liberté $m - 1$:

$$D(F_n, F) = D(\hat{P}, P) \rightarrow \chi^2(m-1) \tag{2.64}$$

d'où le nom de ce test.

Pour accepter l'hypothèse H_0, on cherche une valeur critique χ^2_α dans la loi de χ^2 à $m-1$ degrés de liberté. Si $D(\hat{P}, P) > \chi^2_\alpha$ alors l'hypothèse est accepter sinon elle est rejetée.

2.8.2 Le test de Kolmogorov, Kuiper, Cramer-Von Mises

Soit X une v.a. de loi inconnue P, de fonction de répartition $F(X)$ supposée continue. Soit X_n un échantillon $(X_1, \dots, X_n)$ de X. On désire ajuster la loi P inconnue à une loi donnée P_0 de fonction de répartition continue F_0 connue à priori. On définit une fonction de répartition empirique $F_n(X)$ à partir de l'échantillon X_n de sorte que $F_n(X)$ converge presque sûrement vers $F(X)$. On dira que F_n est un estimateur sans biais de F .

$F_n(x)$ est un histogramme cumulé qui peut être estimé en ordonnant les valeurs observées $(x_1, x_2, \dots, x_n)$ pour ensuite déterminer $F_n(x_1) = \frac{1}{n}, F_n(x_2) = \frac{2}{n}, \dots, F_n(x_n) = 1$.

Les tests d'adéquation qui vont suivre sont, comme le test du c^2, fondés sur des statistiques fonctions de F_n et F assimilables à des distances ou des « pseudo-distance » entre lois de probabilités. On définit les distances suivantes :

$$K_n^+ = \sup_x \left(F_n(X) - F_0(X)\right) \tag{2.65}$$

$$K_n^- = \sup_x \left(F_0(X) - F_n(X)\right) \tag{2.66}$$

a. Test de Kolmogorov

Dans ce cas, la distance entre les distributions F_n et F est :

$$K_n = \sup_x \left|F_n(X) - F_0(X)\right| = \max\left(K_n^+, K_n^-\right) \tag{2.67}$$

La réponse au problème du test d'adéquation est assez intuitive ; on acceptera (H_0) si la statistique K_n prend des valeurs « faibles » c.à.d. si $K_n > D_n$, où D_n est une valeur critique qu'on peut lire dans la table de Kolmogorov ou estimée d'une manière numérique.

b. Test de Kuiper

La distance entre les deux distributions est:

$$V_n = K_n^+ + K_n^- \tag{2.68}$$

Comme pour le test de Kolmogorov, l'hypothèse H_0 que l'échantillon X_n soit issu de la loi F est acceptée si $V_n > d$ où d est une valeur critique.

c. Test de Cramer-Von mises

La distance entre les deux distributions empiriques et théoriques est :

$$CVM = n\int_{\Re}\left(F_n(x) - F_0(x)\right)^2 dF_0(x) \quad (2.69)$$

Cette quantité peut être estimée par :

$$CVM = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F(x_i)\right]^2 \quad (2.70)$$

Notons que les x_i doivent être ordonnés dans l'ordre croissant.

On accepte l'hypothèse H_0 si CVM supérieur à une valeur critique.

d. Test d'Anderson-Darling

Il est défini par la distance suivante :

$$AD = n\int_{-\infty}^{+\infty}\left(F_n(x) - F_0(x)\right)^2\left[F_0(x)\left(1 - F_0(x)\right)\right]^{-1}\mathrm{dx} \quad (2.71)$$

Qu'on peut exprimer par :

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[logF(x_i) + log\left(1 - F(x_{n+1-i})\right)\right] \quad (2.72)$$

Comme pour les autres tests, on accepte l'hypothèse H_0 si AD est supérieur à une valeur critique.

2.8.3 Test de Kullback-Leibler

La distance de Kullback-Leibler est souvent utilisée pour comparer deux distributions. Soient $F\ et\ G$ ces deux distributions. Supposons que la distribution F admet la fonction de densité de probabilité f et la distribution G admet une fonction de densité de probabilité g. On appelle distance de Kullback-Leibler entre les deux distributions $F\ et\ G$ la quantité :

$$d(F,G) = \int ln\frac{f(x)}{g(x)}f(x)dx \quad (2.73)$$

Qu'on peut exprimer par : $d(F,G) = \sum_{i=1}^{n} f(x_i)log\frac{f(x_i)}{g(x_i)}$ (2.74)

2.9 Conclusion

Dans ce chapitre, nous avons présenté une étude détaillée des familles usuelles de distributions continues. Après avoir rappelé leurs propriétés essentielles, nous avons déterminé les différentes méthodes d'estimation de leurs paramètres. Nous avons également cité quelques résultats de la théorie de décision (tests d'adéquation) en vue de déterminer le type de la distribution statistique qui a engendré un échantillon de données.

Les notions statistiques présentées dans ce chapitre vont nous servir pour appréhender la méthode de seuillage que nous présentons dans le prochain chapitre.

Chapitre 3

Seuillage paramétrique basé sur l'approximation de l'histogramme par un mélange de différentes distributions

3.1 Introduction

On s'intéresse dans ce chapitre à la segmentation d'image en niveaux de gris par l'approche paramétrique de seuillage. Nous présentons, en premier lieu, le principe général ainsi qu'un état de l'art de cette approche. Puis nous proposons une méthode de seuillage basée sur l'approximation de l'histogramme par un mélange de distributions dont le modèle statistique de chaque distribution est déterminé automatiquement parmi les modèles statistique d'une famille de huit distribuions disponibles. Cette méthode permet de déterminer des seuils qui séparent les modes de l'histogramme une fois que les paramètres de chaque distribution soient calculés. La méthode de seuillage proposée s'articule essentiellement autour des trois étapes : l'identification du modèle statistique des distributions composant l'histogramme de l'image à segmenter, l'estimation de leurs paramètres et la recherche des seuils en fonction de ces paramètres.

3.2 Principe du seuillage paramétrique

La méthode de seuillage que nous proposons appartient à l'approche paramétrique et considère l'histogramme de l'image comme un mélange de distributions. Ainsi, on suppose que l'histogramme peut être approximé par une somme pondérée de K distributions où chaque distribution correspond à une classe ou à un mode de l'histogramme tel que :

$$f(i/\Theta) = \sum_{k=1}^{K} P_k f_k(i/\theta_k) \qquad i = 0,1,\dots,L-1 \tag{3.1}$$

Où $\Theta = (P_k, \theta_k); k = 1,\dots,K$ est le vecteur de paramètres à estimer. P_k est la probabilité a priori de la $k^{ème}$ composante qui satisfait : $P_k \geq 0 \;\; et \;\; \sum_{k=1}^{K} P_k = 1$. On suppose que la probabilité conditionnelle $f_k(i/\theta_k)$ d'un niveau de gris i à la classe C_k appartient à une famille de distributions disponibles.

Le problème du seuillage paramétrique consiste alors à chercher soit les paramètres $\Theta = (P_k, \theta_k)$ de chaque distribution en minimisant le critère suivant (Fig. 3.1) :

$$J = \sum_{i=0}^{L-1} \big(h(i) - f(i/\Theta)\big)^2 \tag{3.2}$$

soit les seuils $T_1, T_2, \dots, T_{k-1}$ correspondent aux niveaux de gris qui égalisent deux distributions successives (Fig. 3.1).

$$P_k f_k(i/\theta_k) = P_{k+1} f_{k+1}(i/\theta_{k+1}) \qquad k = 1,\dots,K-1 \tag{3.3}$$

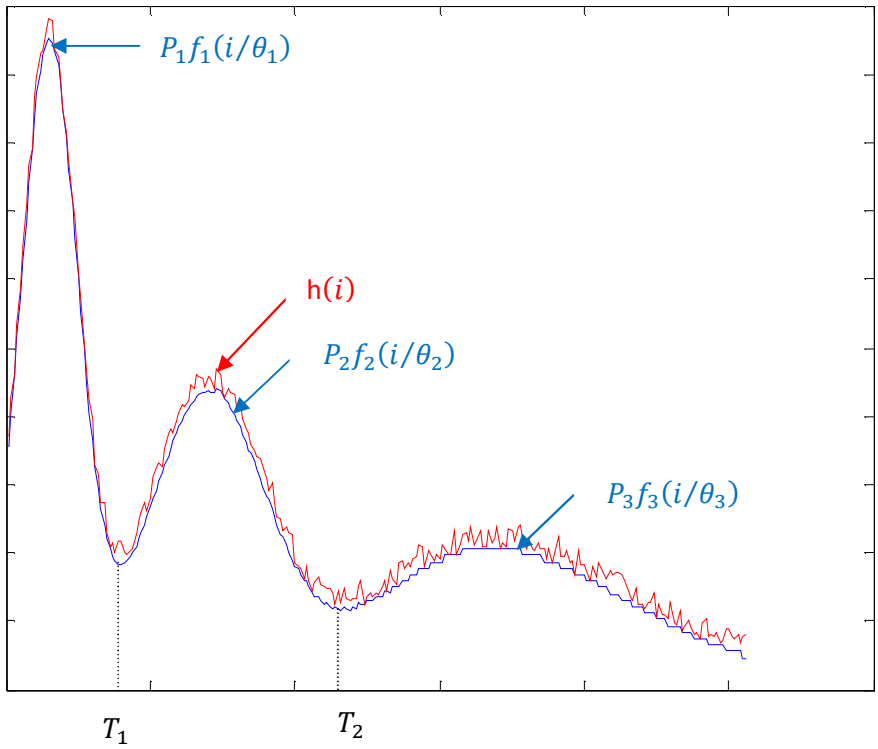


Figure (3.1) : Principe du seuillage paramétrique

Ce sujet a réellement fait l'objet d'une attention particulière. Plusieurs techniques ont été proposées dans la littérature afin de résoudre le problème du seuillage basé sur l'approche paramétrique. La plus part de ces méthodes considère que les modèles statistiques du mélange sont de même type Gaussien **[46, 64-68]**, Gamma **[69-72]**, Beta **[73]**, Rayleigh **[74]**, etc. D'autres méthodes font appel à des modèles de distributions généralisées (Gauss généralisée **[75-77]** et Rayleigh généralisée **[78]**). Une autre catégorie de méthodes considère que le mélange est constitué de modèles de distributions différentes appartenant au système de Pearson **[79], [80]** ou à une famille de distributions englobant les distributions les plus connues, Gauss, Gamma, Beta et Lognormal **[81]** et Lognormal, Nakagami, Gauss, Rayligh, Weibull, K-distribution,… **[82]**.

Dans les méthodes de la première et la deuxième catégorie, les modèles de distributions sont fixés et connus à priori, tandis que pour la troisième catégorie, les modèles de chaque distribution doit être identifié au préalable. Cette identification est réalisée grâce aux paramètres skewness et kurtosis (voir paragraphe 2.7.3.3) pour le système de Pearson, aux moments d'ordre 3 et 4 dans **[81]** et log-cumulant dans **[82]**.

Les méthodes basées sur l'identification des modèles des distributions pouvant approximer un histogramme sont très intéressantes car elles offrent plus de flexibilité.

3.3 Méthode du seuillage proposée

La méthode de seuillage que nous proposons consiste à déterminer les seuils comme étant les niveaux de gris égalisant deux distributions successives qui approximent l'histogramme (Fig.3.1).

Pour ce faire, nous supposons que les distributions peuvent avoir des modèles différents et qu'elles appartiennent à une famille de huit distributions à savoir la distribution Gaussienne $N(\mu,\sigma^2)$, Beta $\mathrm{B}(a,b)$, Gamma $g(q,k)$, Exponentielle $E(\lambda)$, Log-normale $L(\mu,\sigma)$, Chi2 $\chi^2(v)$, Weibull $W(k,\lambda)$ et Rayleigh $R(\sigma)$. Les paramètres statistiques de chaque distribution sont donnés dans le tableau (2.1) du chapitre précédent.

Pour déterminer le modèle de chaque distribution, nous avons utilisé les tests d'adéquations présentés également dans le chapitre précédent.

3.3.1 Identification du modèle d'une distribution

Afin d'identifier le modèle d'une distribution parmi la famille des lois données, nous devons choisir l'un des trois tests d'adéquations : le test de Kolmogorov-Smirnov(KS), le test de Chi2 $\left(c^2\right)$ et le test de Kullback-Leibler (KL). Rappelons que le but de ces tests est de calculer une distance entre l'histogramme réel $h(i)$ et l'histogramme estimé $h_e(i)$. Ces distances sont définies comme suit :

- Distance de KS : $D(KS) = \sup_i |h_e(i) - h(i)|$, $i \in \{0,1,\dots,L-1\}$
- Distance de Chi2 : $D(\chi^2) = \sum_i \frac{\left(h_e(i)-h(i)\right)^2}{h(i)}$
- Distance de KL : $D(KL) = \sum_i h(i) * log\left(\frac{h(i)}{h_e(i)}\right)$

Pour identifier la nature d'une distribution, nous avons comparé les distances entre $h(i)$ et $h_e(i)$. A chaque fois nous évaluons $h_e(i)$ en choisissant un modèle de distribution parmi les huit distributions disponibles, puis nous calculons la distance minimale par l'une des trois distances citées dessus.

Notons que l'histogramme estimé est obtenu, pour un modèle de distribution donnée, en estimant en premier lieu les paramètres du modèle à partir des valeurs de l'histogramme par la méthode du maximum de vraisemblance ou la méthode des moments puis en déterminant pour chaque niveau de gris la valeur de la fonction de densité de probabilité du modèle de distribution choisi.

Afin d'évaluer les performances de ces tests, nous les avons appliqué à des histogrammes unimodales générées artificiellement. Notons qu'en pratique, les histogrammes unimodales correspondent aux images qui contiennent beaucoup de pixels appartenant au fond et peu de pixels de l'objet, c'est le cas par exemple des images gradient ou des images obtenues par différance entre deux images successives d'une séquence d'images **[83]**.

Les tableaux (3.1) à (3.8) regroupent les valeurs des distances de KS, χ^2 et KL entre l'histogramme généré respectivement par l'un des modèles de distributions et les histogrammes estimés par chaque modèle de distribution.

Tous ces tableaux montrent que les trois tests d'adéquation nous ont permis d'identifier correctement le modèle de la distribution à partir duquel l'histogramme original a été généré. En effet, si on prend l'exemple de l'histogramme généré à partir du modèle Normal (Gauss), les valeurs minimales (représentées en gras) des trois distances correspondent à celles pour les quelles l'histogramme estimés suit une distribution Normale.

	Normal	Log-normal	Beta	Gamma	Exponentiel	Chi2	Weibull	Rayleigh
$D(KS)$	**39.2**	360.5	209.4	265.1	2148	151 .6	284.1	1684 .7
$D(\chi^2)$	**80.5052**	2530	770	1220	2170	1500	1410	12600
$D(KL)$	**186.0877**	2302	723	1106	7590	1482	975	7358

Tableau (3.1) : Distances de KS, χ^2 et de KL entre l'histogramme unimodal (Normal) et les histogrammes estimés par chaque modèle de distribution

	Normal	Log-normal	Beta	Gamma	Exponentiel	Chi2	Weibull	Rayleigh
$D(KS)$	1018.7	**122.7989**	385 .4	329	4096.7	1159	1438.2	2886
$D(\chi^2)$	5540	**77.3163**	810	590	133140	10830	13100	132370
$D(KL)$	3671	**141.3430**	919	743	68077	5496	6238	29992

Tableau (3.2) : Distances de KS, χ^2 et de KL entre l'histogramme unimodal (Log-normal) et les histogrammes estimés par chaque modèle de distribution

	Normal	Log-normal	Beta	Gamma	Exponentiel	Chi2	Weibull	Rayleigh
$D(KS)$	1143	2396	**504 .1674**	1533	1649	68628	1377	1425
$D(\chi^2)$	901000	933000	**1232**	870000	845000	11964000	914000	1290000
$D(KL)$	0.0032	10100	**2519 .7**	6200	6400	3685900	5500	3800

Tableau (3.3) : Distances de KS, χ^2 et de KL entre l'histogramme unimodal (Beta) et les histogrammes estimés par chaque modèle de distribution

	Normal	Log-normal	Beta	Gamma	Exponentiel	Chi2	Weibull	Rayleigh
$D(KS)$	367.1	213.5	1972.2	**155.3**	1602.2	1213 .6	345.5	705
$D(\chi^2)$	3360	810	65170	**525.4047**	104360	17580	3660	37070
$D(KL)$	2700	1000	167810	**520.3359**	44700	20300	2200	10800

Tableau (3.4) : Distances de KS, χ^2et de KL entre l'histogramme unimodal (Gamma) et les histogrammes estimés par chaque modèle de distribution

	Normal	Log-normal	Beta	Gamma	Exponentiel	Chi2	Weibull	Rayleigh
$D(KS)$	3719	2223	1430	1483	**302.1266**	10781	1411	4636
$D(\chi^2)$	17902	2825	1124	1035	**243.6223**	81668	1134	30619
$D(KL)$	2863	392	503	66008	**138.3705**	3375	4888	26600

Tableau (3.5) : Distances de KS, χ^2et de KL entre l'histogramme unimodal (Exponentiel) et les histogrammes estimés par chaque modèle de distribution

	Normal	Log-normal	Beta	Gamma	Exponentiel	Chi2	Weibull	Rayleigh
$D(KS)$	17.8	86	4960 .2	0.7	74.7	**0.54**	16.8	32.7
$D(\chi^2)$	100	17.9652	3455200	0.7328	5800	**0.6907**	100	1200
$D(KL)$	6.8997	28.4724	3.0675	3.0875	837.4330	**2.3231**	57.9953	237.9663

Tableau (3.6) : Distances de KS, χ^2et de KL entre l'histogramme unimodal (Chi2) et les histogrammes estimés par chaque modèle de distribution

	Normal	Log-normal	Beta	Gamma	Exponentiel	Chi2	Weibull	Rayleigh
$D(KS)$	540.6	1700.4	129.4	1378.2	5397.7	2375.2	**146.4244**	4047.4
$D(\chi^2)$	890	10300	5520	6000	333000	73450	**54.8603**	170890
$D(KL)$	1953	7142	5132	5367	73850	13644	**60.0406**	35627

Tableau (3.7) : Distances de KS, χ^2et de KL entre l'histogramme unimodal (Weibull) et les histogrammes estimés par chaque modèle de distribution

	Normal	Log-normal	Beta	Gamma	Exponentiel	Chi2	Weibull	Rayleigh
$D(KS)$	347.4	523.6	175	236.3	1925.8	2337	42.5	**33.7025**
$D(\chi^2)$	2693	5440	761	1353	80339	53671	119	**101.6659**
$D(KL)$	3524	6358	848	1464	19178	82458	**242.5303**	245

Tableau (3.8) : Distances de KS, χ^2et de KL entre l'histogramme unimodal (Rayleigh) et les histogrammes estimés par chaque modèle de distribution

Le tableau (3.9) donne les valeurs des paramètres originaux à partir desquels les histogrammes sont générés et les valeurs des paramètres estimés à partir des lois des distributions précédemment identifiées. Dans tous les cas les valeurs estimées restent très proches des valeurs réelles.

	Normal		Log-normal		Beta		Gamma		Exponentiel	Chi2	Weibull		Rayleigh
	μ	σ	μ	σ	α	β	θ	k	λ	v	k	λ	σ
$h(i)$	50	10	3	0 .25	0 .5	0 .5	8	5	12.5	15	18	5	25
$h_e(i)$	50	9.8957	3.0045	0 .247	0.6324	0 .6324	8.2406	4.8407	12.6866	15	18.1	5.0279	24.8367

Tableau (3.9) : Estimation des paramètres des de l'histogramme généré $h(i)$ et l'histogramme estimé $h_e(i)$.

Les figures (3.2) à (3.9) montrent l'allure des histogrammes générés et des histogrammes estimés.

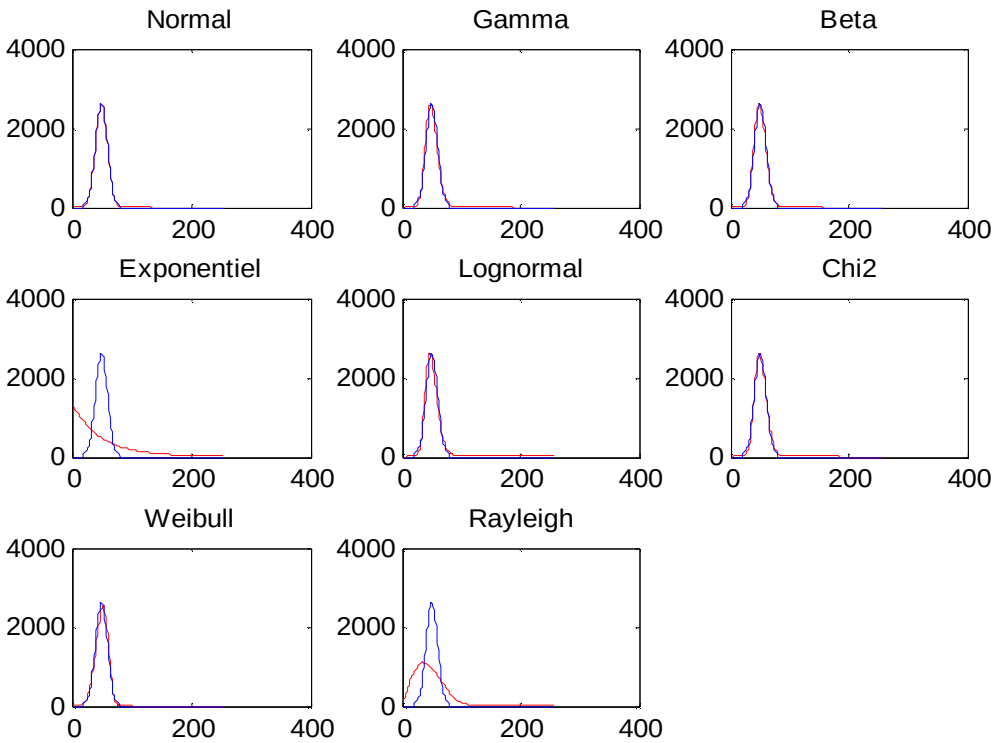


Figure (3.2) : Comparaison entre l'histogramme généré à partir du modèle Normal et les histogrammes estimés

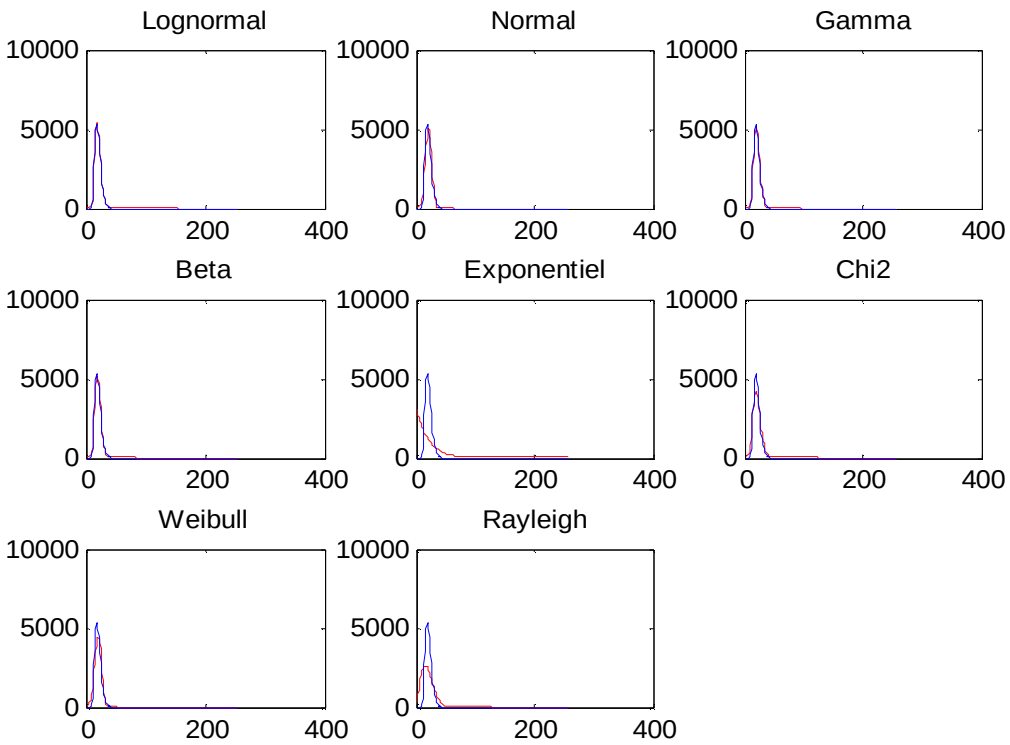


Figure (3.3) : Comparaison entre l'histogramme généré à partir du modèle Log-normal et les histogrammes estimés

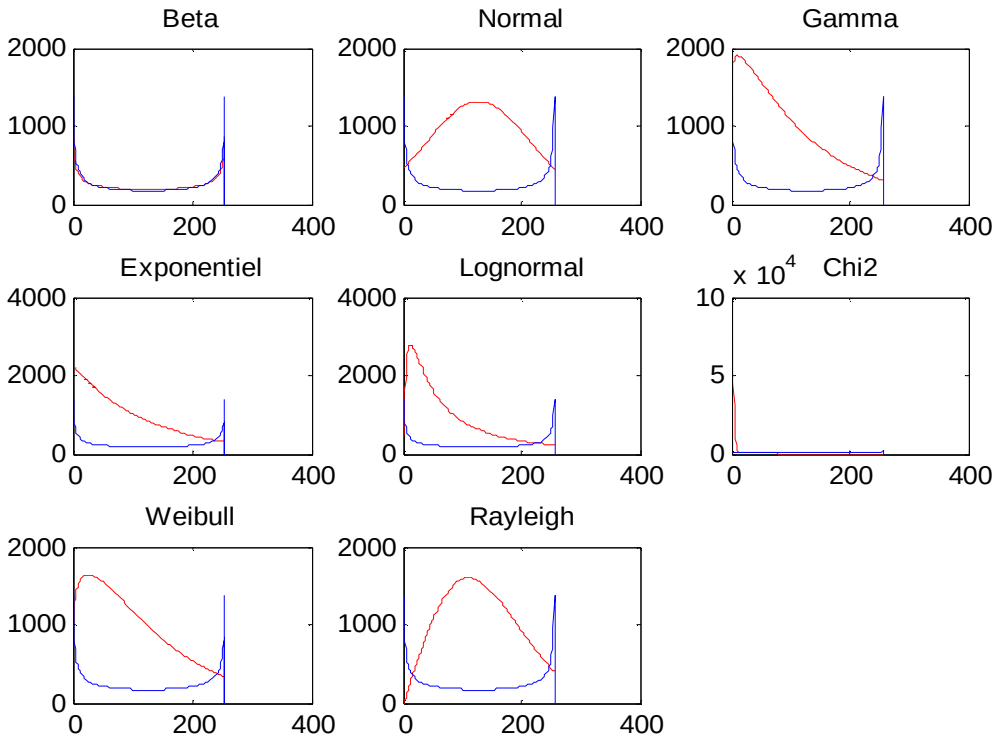


Figure (3.4) : Comparaison entre l'histogramme généré à partir du modèle Beta et les histogrammes estimés

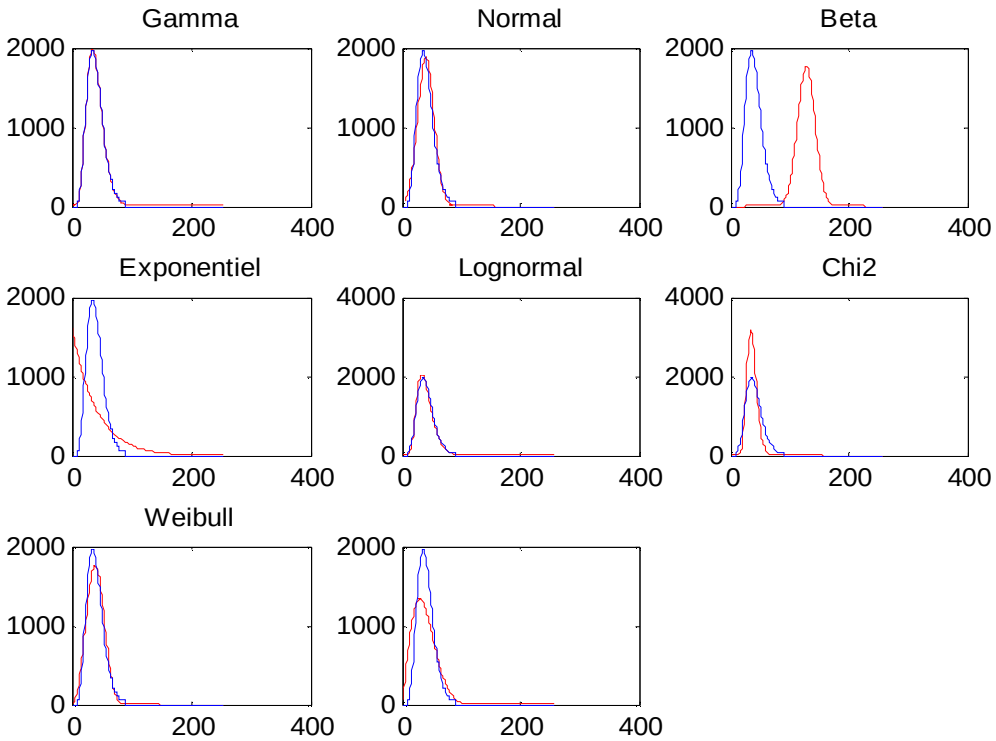


Figure (3.5) : Comparaison entre l'histogramme généré à partir du modèle Gamma et les histogrammes estimés

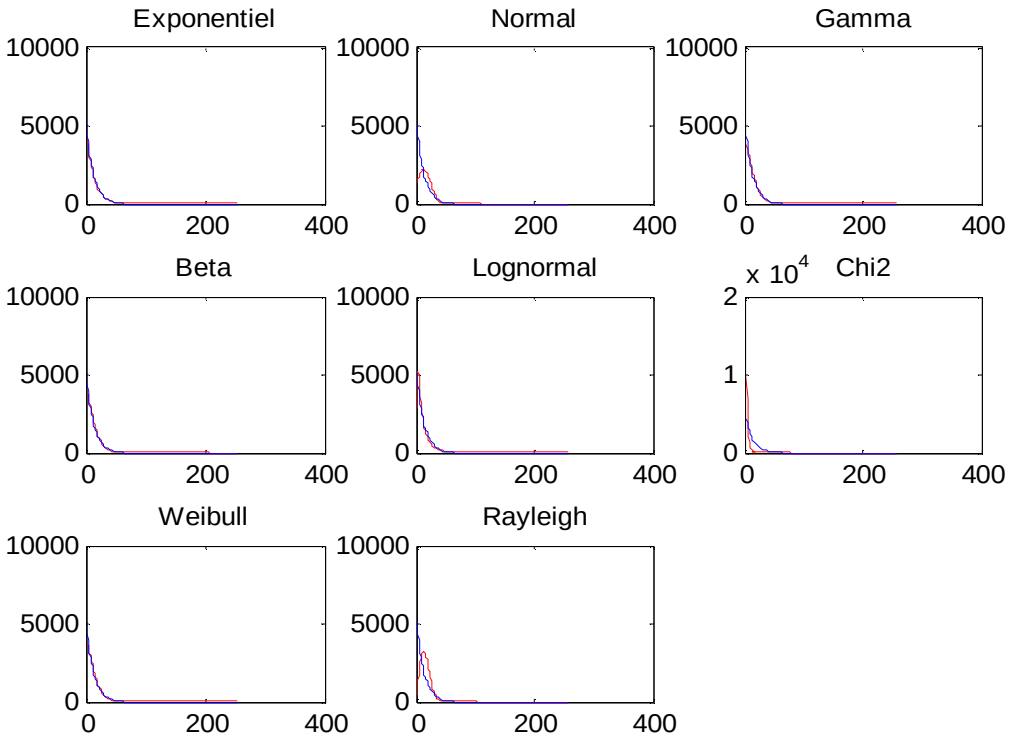


Figure (3.6) : Comparaison entre l'histogramme généré à partir du modèle Exponentiel et les histogrammes estimés

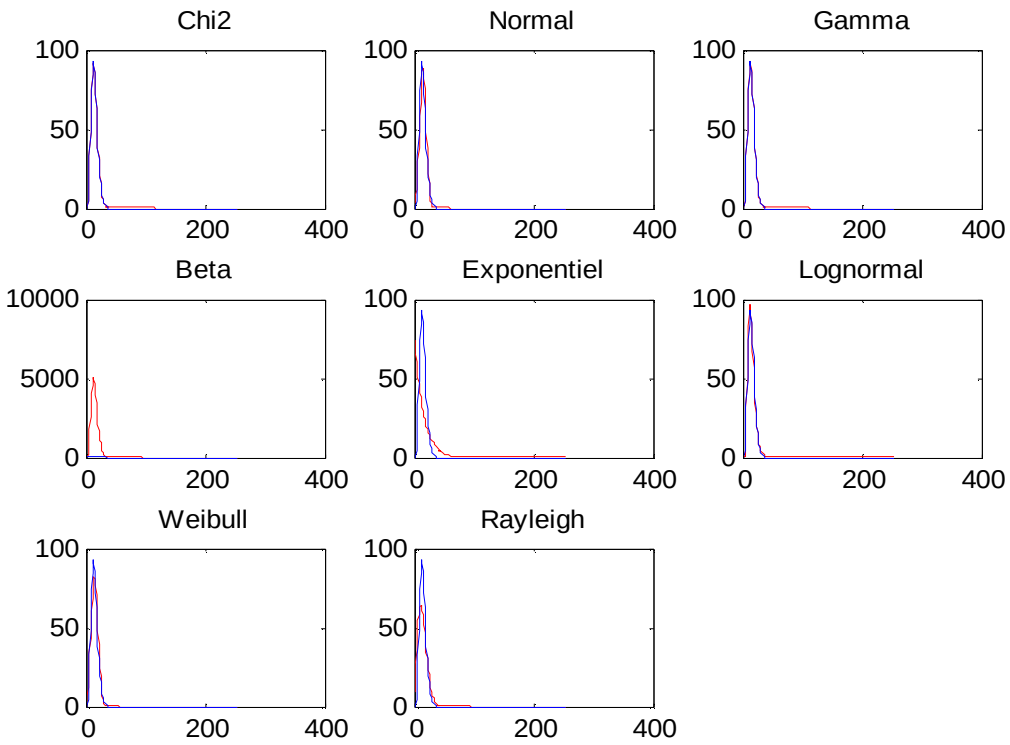


Figure (3.7) : Comparaison entre l'histogramme généré à partir du modèle Chi2 et les histogrammes estimés

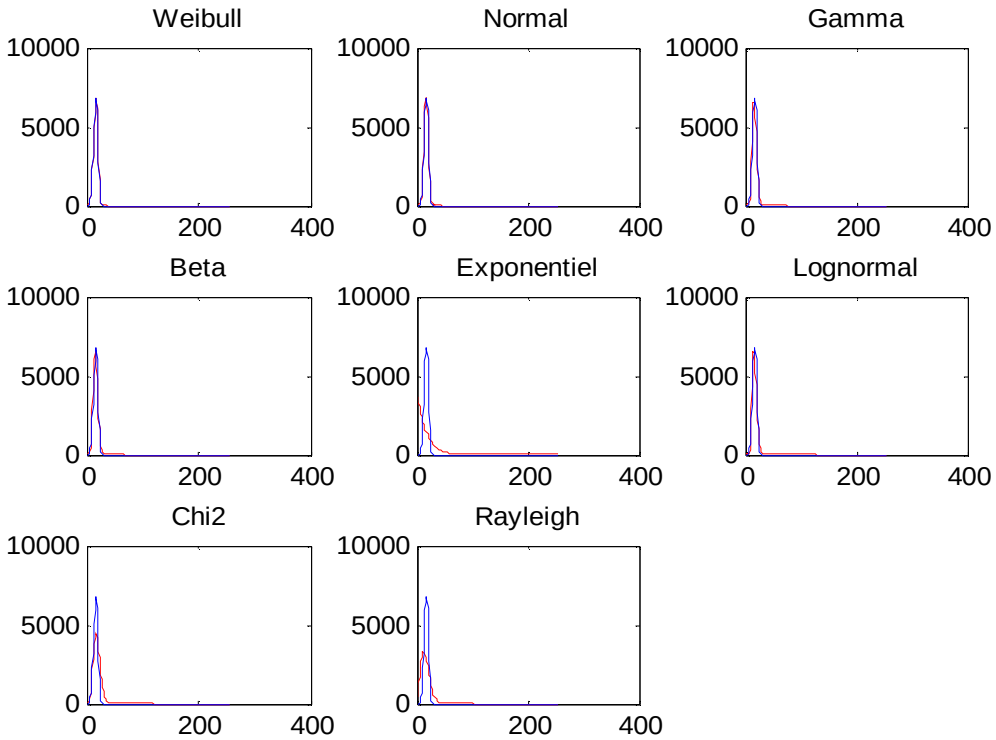


Figure (3.8) : Comparaison entre l'histogramme généré à partir du modèle Weibull et les histogrammes estimés

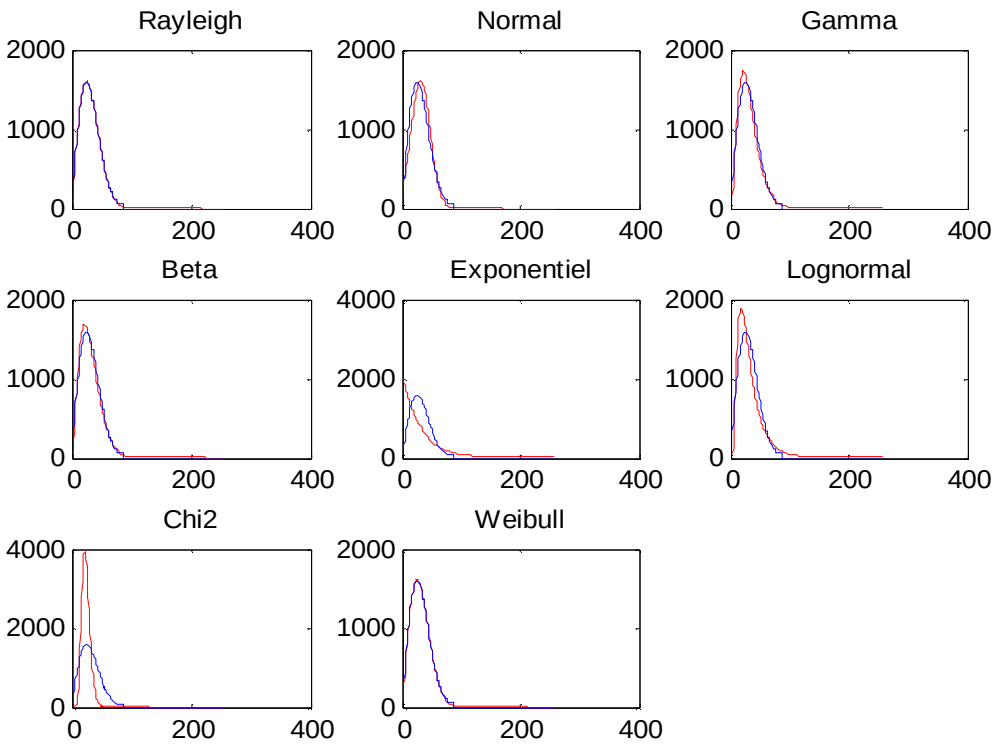


Figure (3.9) : Comparaison entre l'histogramme généré à partir du modèle Rayleigh et les histogrammes estimés

En résumé, les trois tests d'adéquation conduisent aux mêmes résultats puisqu'ils ont permis d'identifier à chaque fois le modèle statistique de la distribution unimodale sauf dans le cas de la distribution Rayleigh qui a été identifiée comme une distribution de Weibull par le test de Kullback-Leibler.

3.3.2 Seuillage d'un histogramme bimodal

Comme nous l'avons évoqué au début du paragraphe 3.3, la méthode de seuillage que nous proposons consiste à identifier les modèles statistiques qui approximent les distributions qui composent l'histogramme puis à déterminer la valeur des seuils égalisant deux à deux les distributions ainsi identifiées.

Pour être plus précis, cette méthode consiste à initialiser aléatoirement la valeur de seuil $T^{(0)}$. Ce seuil initial nous permet de séparer l'histogramme $h(i)$ en deux classes. La première contient les niveaux de gris $C_1 = \{0,1,\dots,T^{(0)}\}$ et la deuxième les niveaux de gris $C_2 = \{T^{(0)}+1,\dots,L-1\}$. La première classe est décrite par l'histogramme $h_1(i)$ $(i \in C_1)$ et la deuxième par l'histogramme $h_2(i)$ $(i \in C_2)$ tel que $h(i) = h_1(i) + h_2(i)$. L'histogramme de

chaque classe est alors approximé par une distribution dont le modèle statistique est choisi parmi les 8 modèles des distributions disponibles. Pour cela, nous avons utilisé le test d'adéquation de Kolmogorov Smirnov (KS).

A l'issue de cette étape, le modèle statistique $h_{1e}(i)$ et $h_{2e}(i)$ de chaque partie de l'histogramme ainsi que leurs paramètres sont connus tels que :

$$h_{1e}(i) = P_1 f_1(i/\theta_1) \;\; et \;\; h_{2e}(i) = P_2 f_2(i/\theta_2) \tag{3.4}$$

et l'histogramme bimodal de l'image est approximé par :

$$h_e(i) = \sum_{k=1}^{2} P_k f_k(i/\theta_k) \tag{3.5}$$

avec θ_k le vecteur des paramètres de la $i^{ème}$ distribution et P_k la probabilité a priori de la classe C_k telles que : $P_k \geq 0 \;\; et \;\; \sum_{k=1}^{2} P_k = 1$.

Notons que les probabilités à priori des deux classes sont estimées par :

$$P_1 = \frac{n_1}{n_1+n_2} \;\; et \;\; P_2 = \frac{n_2}{n_1+n_2} \tag{3.6}$$

où $n_1 \;\; et \; n_2$ représentent le nombre de pixels dans chaque classe $C_1 \;\; et \;\; C_2$ respectivement tel que :

$$n_1 = \sum_{i=0}^{T} h(i) \;\; et \; n_2 = \sum_{i=T+1}^{L-1} h(i) \tag{3.7}$$

On procède alors au calcul du seuil. Pour cela, on a utilisé la règle de décision Bayesienne.

3.3.3 Estimation du seuil

La règle de décision Bayesienne est une méthode de classification supervisée qui permet d'affecter d'une manière optimale un pixel de niveau de gris i à une classe donnée. La classification bayesienne présente l'avantage d'être optimale dans le sens de la maximisation des probabilités à priori et donc de la minimisation de la probabilité de l'erreur de la classification.

D'une manière générale, on suppose que l'histogramme est approximé par un mélange de K distributions ($K = 2$ dans le cas du seuillage simple). Les distributions sont délimitées par $K-1$ seuils $T_0, T_1, \ldots, T_K \;\; avec \; T_0 = 0 \;\; et \; T_K = L-1$ le niveau de gris maximal. T_j est le seuil qui sépare les pixels de l'image en deux classes $C_j \; et \; C_{j+1}$ ou l'histogramme en deux modes. On

note par P_j la probabilité à priori de la classe C_j. La probabilité de l'erreur de classification de la classe $C_j\ pour\ j \in [1, K]$ est donnée par **[81]**:

$$P_{erreur}(C_j) = P_j\left(\int_0^{T_{j-1}} f_j(x/\theta_j)dx + \int_{T_j}^{+\infty} f_j(x/\theta_j)dx\right) \quad (3.8)$$

avec $f_j(x/\theta_j)$ est la fonction de densité de probabilité de la $j^{ème}$ distribution correspondant au $j^{ème}$ mode de l'histogramme. Ainsi, la probabilité d'erreur de la classification des classes $C_1,\ldots,C_j,\ldots,C_K$ est donnée par :

$$P_{erreur}(C_1, \ldots, C_K) = \sum_{j=1}^{K} P_j \int_0^{T_{j-1}} f_j(x/\theta_j)dx + \int_{T_j}^{+\infty} f_j(x/\theta_j)\,dx \quad (3.9)$$

Le but de la décision Bayesienne dans ce cas est de trouver les seuils $T_1, \ldots, T_j, \ldots, T_{K-1}$ qui minimisent la fonction $P_{erreur}(C_1, \ldots, C_K)$. La solution de ce problème est donnée par :

$$\frac{\partial}{\partial T_j} P_{erreur}(C_1, .., C_K) = 0 \quad (3.10)$$

ou par la règle suivante :

$$P_j f_j(x/\theta_j)(T_J) = P_{j+1} f_{j+1}(x/\theta_{j+1})(T_j) \qquad \forall j = 1, \ldots, K-1 \quad (3.11)$$

En remplaçant les fonctions de densité de probabilité $f_j(T_j) et\ f_{j+1}(T_{j+1})$ par leurs valeurs et en prenant le logarithme des deux membres, nous pouvons obtenir le seuil optimal T_j qui sépare les deux modes $j\ et\ j+1\ pour\ j = 1, \ldots, K-1$.

A titre d'exemple, pour un histogramme bimodal composé par un mélange de deux distributions Gaussiennes $\aleph_1(\mu_1, \sigma_1^2)$ et $\aleph_2(\mu_2, \sigma_2^2)$ avec des probabilités à priori $P_1\ et\ P_2$, l'équation (3.11) s'écrit :

$$\frac{P_1}{s_1\sqrt{2p}}\exp\left[-\frac{1}{2s_1^2}(x-m_1)^2\right] = \frac{P_2}{s_2\sqrt{2p}}\exp\left[-\frac{1}{2s_2^2}(x-m_2)^2\right] \quad (3.12)$$

Le logarithme des deux membres de cette équation nous donne :

$$\log\left(\frac{P_1}{s_1\sqrt{2p}}\right) - \frac{1}{2s_1^2}(x-m_1)^2 = \log\left(\frac{P_2}{s_2\sqrt{2p}}\right) - \frac{1}{2s_2^2}(x-m_2)^2 \quad (3.13)$$

$$\frac{1}{2}\left(\frac{1}{s_2^2} - \frac{1}{s_1^2}\right)x^2 + \left(\frac{m_1}{s_1^2} - \frac{m_2}{s_2^2}\right)x + \left(\log\frac{P_1 s_2}{P_2 s_1} - \frac{m_1^2}{2s_1^2} + \frac{m_2^2}{2s_2^2}\right) = 0 \quad (3.14)$$

si on pose

$$A=\frac{1}{2}\left(\frac{1}{s_2^2}-\frac{1}{s_1^2}\right) \quad B=\left(\frac{m_1}{s_1^2}-\frac{m_2}{s_2^2}\right) \quad C=\left(\log\frac{P_1 s_2}{P_2 s_1}-\frac{m_1^2}{2s_1^2}+\frac{m_2^2}{2s_2^2}\right) \tag{3.15}$$

on aboutit à l'équation de deuxième ordre suivante :

$$f(x)=Ax^2+Bx+C=0 \tag{3.16}$$

dont la solution est :

$$x^*=T_j^*=\frac{-B\pm\sqrt{B^2-4AC}}{2A} \tag{3.17}$$

Pour le calcul des seuils dans le cas des autres mélanges, nous avons suivi les mêmes étapes que la procédure précédente. Cependant, dans la plus part des cas, nous avons obtenu des équations non linéaires que nous avons résolu par la méthode itérative de Newton -Raphson. Les équations finales de chaque mélange ainsi que les seuils obtenus sont donnés en annexe B.

Rappelons que le principe de la méthode de Newton-Raphson consiste à rechercher la racine approchée de la solution exacte en suivant un processus itératif. Si $f(x)\ avec\ x\in\Re$, est la fonction non linéaire à résoudre, alors la solution x^* de $f(x)=0$ est donnée par :

$$x^*=\lim_{n\to\infty} x^{(n+1)} \qquad avec: \qquad x^{(n+1)}=x^{(n)}-\frac{f(x^{(n)})}{f'(x^{(n)})} \tag{3.18}$$

où $f'(x)$ est la fonction dérivée.

En pratique, ce processus est itéré à partir d'une solution initiale $x^{(0)}$ jusqu'à convergence $\left(x^{(n+1)}\approx x^{(n)}\ ou\ f\left(x^{(n+1)}\approx 0\right)\right)$.

Le nouveau seuil obtenu divise l'histogramme en deux classes (cas bimodal). L'histogramme de chaque classe est à nouveau approximé par une distribution dont le modèle est choisi parmi les 8 modèles disponibles. Les paramètres de chaque distribution ainsi identifiée sont estimés et le seuil est recalculé. Les procédures d'approximation des histogrammes, d'estimations des paramètres et de calcul du seuil sont réitérées jusqu'à ce que le seuil ne change pas.

L'algorithme de la méthode de seuillage proposée est résumé dans les étapes suivantes :

1. Déterminer l'histogramme de l'image.
2. Fixer le seuil initial $T^{(i)}$(i=0).
3. Décomposer l'histogramme en deux classes C_1 et C_2 selon le seuil initial $T^{(i)}$. La classe C_1 contient les pixels ayant les niveaux de gris entre 0 et $T^{(i)}$ et la classe C_2 entre $T^{(i)}+1$ et L-1.
4. Identifier les modèles des distributions des deux classes en utilisant le test de KS.
5. Estimer les paramètres des distributions identifiées en utilisant la méthode de maximum de vraisemblance ou la méthode des moments selon le type de la distribution.
6. Calculer les probabilités à priori de chaque classe.
7. Calculer le nouveau seuil T en résolvant l'équation obtenue selon les distributions identifiées.
8. Si $T^{(i)} \neq T$

   i=i+1

   $T^{(i)} = T$

   Aller à 3
9. Segmenter l'image avec le seuil T .

Algorithme 3.1 : Méthode de seuillage proposée

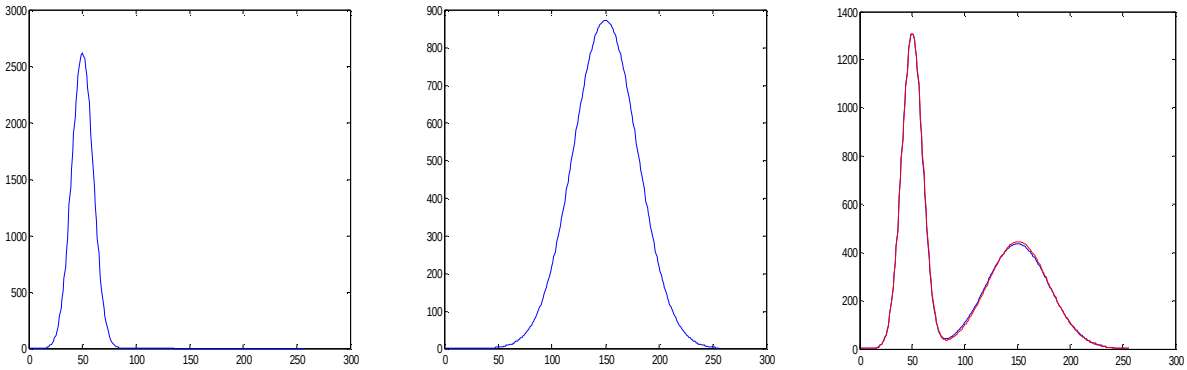
Remarque : Le modèle de distribution de chaque classe peut changer d'une itération à une autre.

3.4 Evaluation du seuillage d'un histogramme artificiel par la méthode proposée

Afin d'évaluer les performances de la méthode de seuillage proposée, nous l'avons appliqué à un ensemble d'histogrammes générés artificiellement. Notons que l'utilisation de ces histogrammes permet d'évaluer objectivement les performances de la méthode de seuillage du fait que les modèles statistiques sont parfaitement connus et contrôlables.

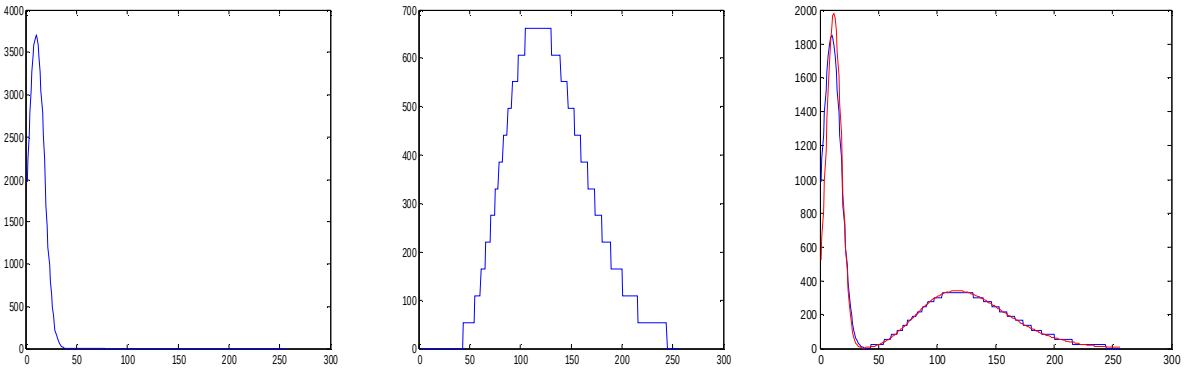
Les figures (3.10) à (3.17) montrent quelques exemples d'histogrammes bimodaux générés artificiellement à partir des modèles statistiques différents et leurs approximations par un mélange de deux distributions identifiées par le test de KS et délimité par le seuil optimal obtenu par notre algorithme.

Sur ces figures les courbes, titrées par (a) et (b), représentent les distributions générées respectivement pour le premier mode et le deuxième mode. Les figures titrées (c) regroupent l'histogramme original et le mélange des distributions identifiées.



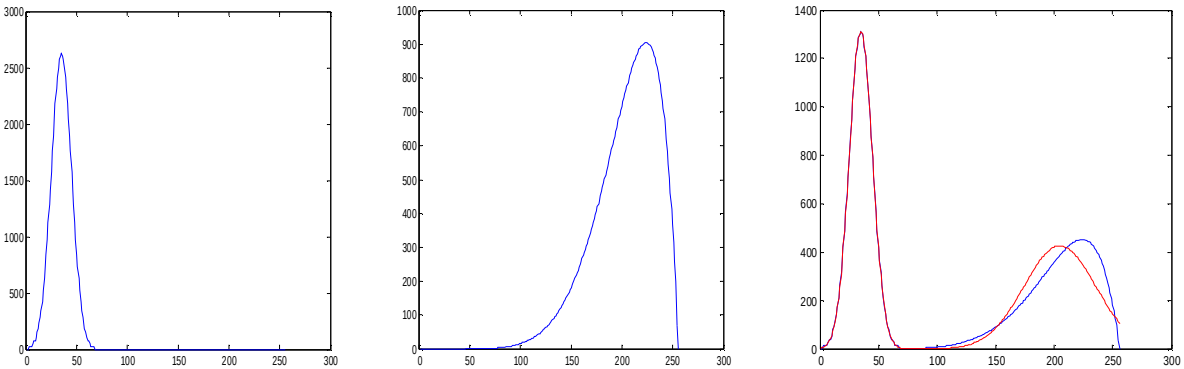
(a) Gaussienne **(b) Gaussienne** **(c) Histogramme généré VS. mélange estimé**

Figure (3.10) : Histogramme généré et estimé par un mélange de deux distributions Gaussiennes.



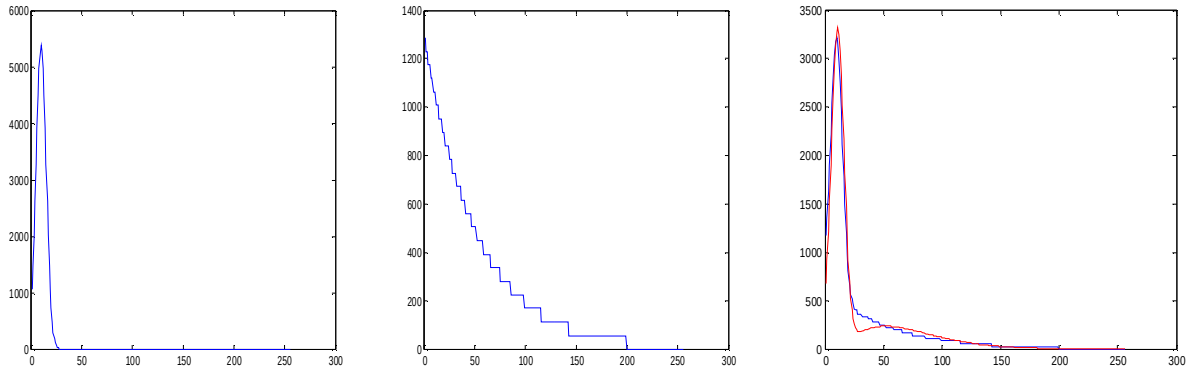
(a) Gaussienne **(b) Gamma** **(c) Histogramme généré VS. mélange estimé**

Figure (3.11) : Histogramme généré et estimé par un mélange de deux distributions Gauss et Gamma.



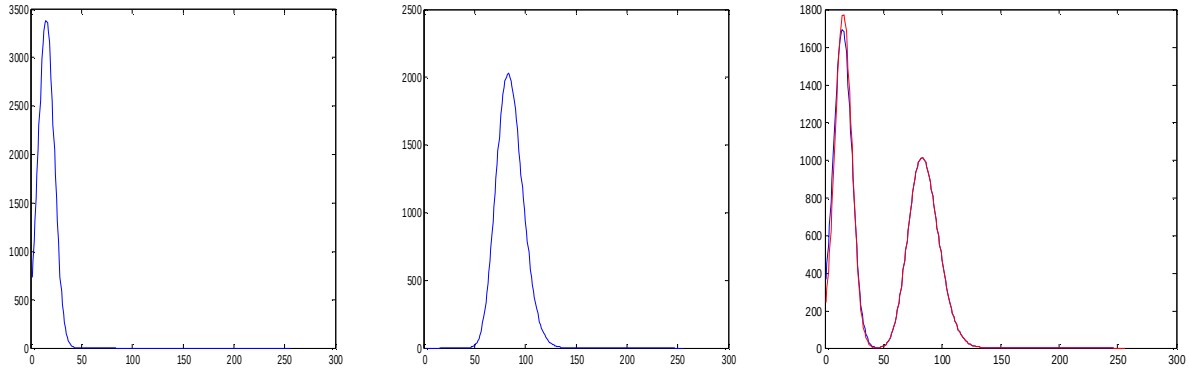
(a) Gaussienne **(b) Beta** **(c) Histogramme généré VS. mélange estimé**

Figure (3.12) : Histogramme généré et estimé par un mélange de deux distributions Gauss et Beta.



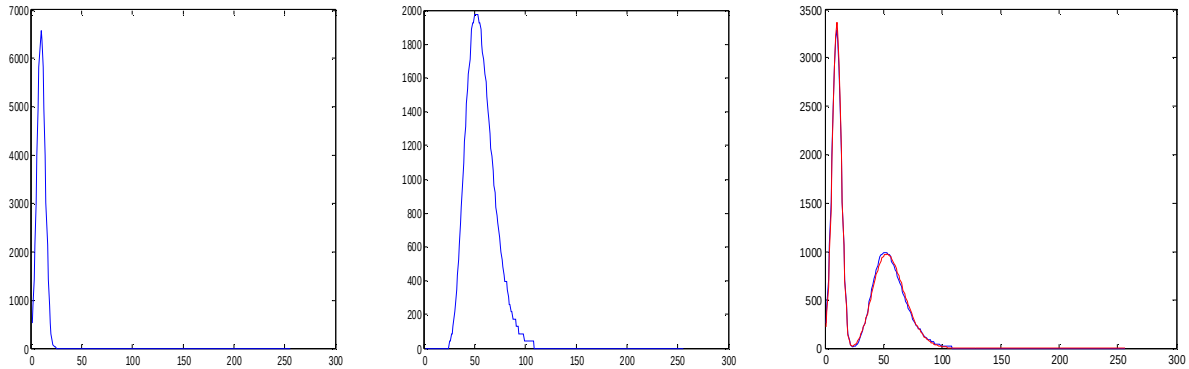
(a) Gaussienne **(b) Exponentielle** **(c) Histogramme généré VS. mélange estimé**

Figure (3.13) : Histogramme généré et estimé par un mélange de deux distributions Gauss et Exponentielle.



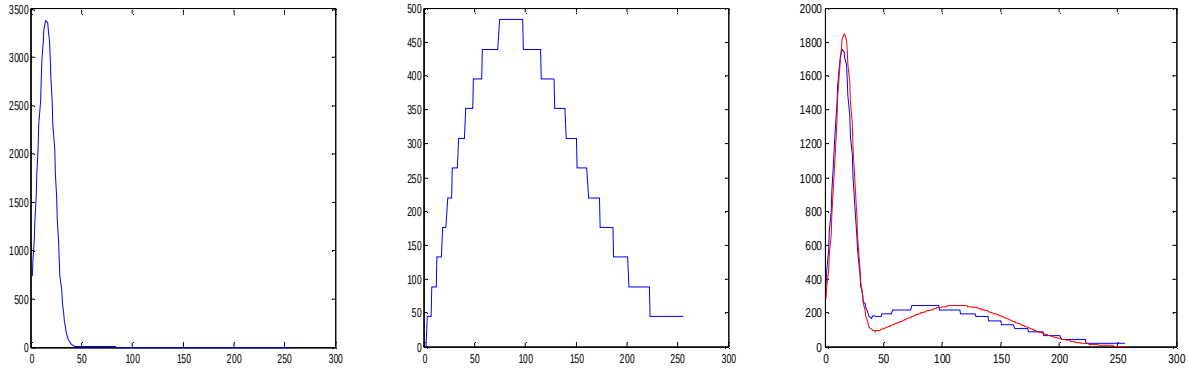
(a) Gaussienne **(b) Chi2** **(c) Histogramme généré VS. mélange estimé**

Figure (3.14) : Histogramme généré et estimé par un mélange de deux distributions Gauss et Chi2.



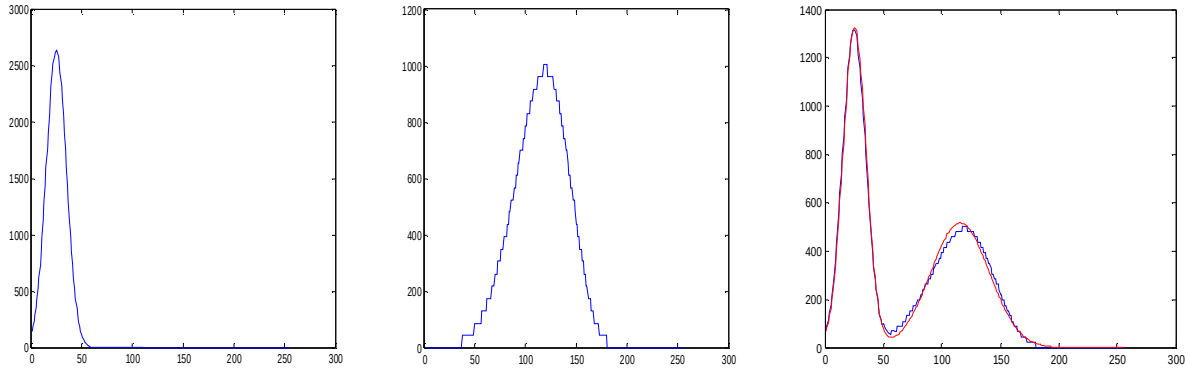
(a) Gaussienne **(b) Lognormale** **(c) Histogramme généré VS. mélange estimé**

Figure (3.15) : Histogramme généré et estimé par un mélange de deux distributions Gauss et Lognormal.



(a) Gaussienne **(b) Rayleigh** **(c) Histogramme généré VS. mélange estimé**

Figure (3.16) : Histogramme généré et estimé par un mélange de deux distributions Gauss et Rayleigh.



(a) Gaussienne **(b) Weibull** **(c) Histogramme généré VS. mélange estimé**

Figure (3.17) : Histogramme généré et estimé par un mélange de deux distributions Gauss et Weibull.

En observant ces figures, nous pouvons constater une bonne identification des distributions de tous les mélanges proposés du fait que les histogrammes estimés sont très proches des histogrammes générés (originaux).

Le tableau (3.10) résume les résultats obtenus pour les huit cas de mélanges en spécifiant les paramètres utilisés pour la génération des mélanges des histogrammes de départ, ainsi que leurs valeurs estimées une fois que les modèles des distributions sont identifiés. Dans la plus part des cas, la distribution générée est reconnue par le test de KS. En effet, les distributions Gaussienne, Gamma et Chi2 sont identifiées correctement dans tout les cas. Par contre, les distributions Beta, Rayleigh et weibull sont reconnues comme étant des Gaussiennes et les distributions Exponentielle et Log-normale sont confondues aux modèles Gamma. Cependant, sur les figures précédentes, nous remarquons que ces distributions sont proches.

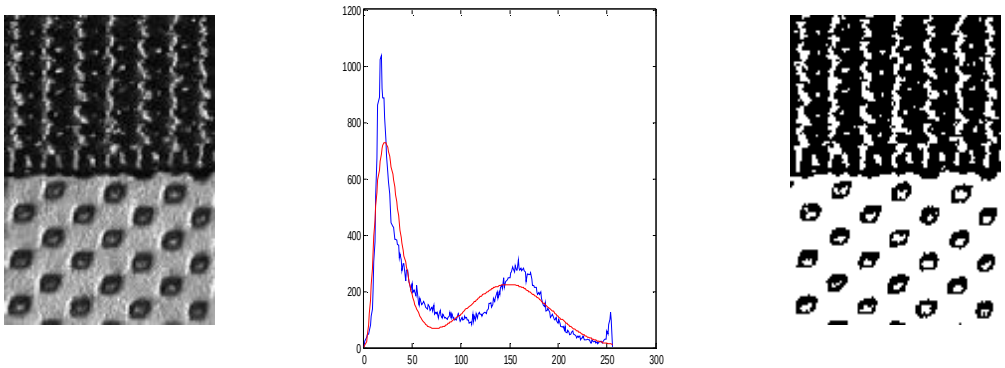
Classe	Distributions générées	Paramètres	Distributions identifiées	Paramètres estimés	seuil
1^{er} *mélange*					80
Premier mode	Gauss	μ =50 ; σ = 10	Gauss	μ_{est}=50,1293 ; σ_{est}=10.0909	
Deuxième mode	Gauss	μ = 150;σ = 30	Gauss	μ_{est}=150.6356 ; σ_{est}=29.1451	
$2^{ème}$ *mélang*					70
Premier mode	Gauss	μ =10 ; σ =8	Gauss	μ_{est}=11.7716 ; σ_{est}=6.5925	
Deuxième mode	Gamma	θ =10 ; k = 13	Gamma	θ_{est} =10.5603; k_{est} =12.2203	
$3^{ème}$ *mélang*					81
Premier mode	Gauss	μ =35;σ = 10	Gauss	μ_{est}=35.1810 ; σ_{est}=10.4060	
Deuxième mode	Beta	α = 8 ; β = 2	Gauss	μ_{est}=205.1782 ; σ_{est}=30.1849	
$4^{ème}$*mélang*					27
Premier mode	Gauss	μ =10;σ = 5	Gauss	μ_{est}=10.7899 ; σ_{est}=5.5101	
Deuxième mode	Exponentielle	λ =52	Gamma	θ_{est} = 3.5354;k_{est} = 20.5271	
$5^{ème}$ *mélang*					50
Premier mode	Gauss	μ =15;σ = 8	Gauss	μ_{est}=15.3648; σ_{est}=7.3657	
Deuxième mode	Chi2	v =85	Chi2	v_{est} =85	
$6^{ème}$ *mélang*					23
Premier mode	Gauss	μ =10; σ = 4	Gauss	μ_{est}=10.0903 ; σ_{est}=3.8785	
Deuxième mode	Lognormal	μ =4;σ = 1/4	Gamma	θ_{est} = 5.4159; k_{est} = 1.8632	
$7^{ème}$*mélang*					37
Premier mode	Gauss	μ =15;σ = 8	Gauss	μ_{est}=16.3786; σ_{est}=7.8661	
Deuxième mode	Rayleigh	r =85	Gauss	μ_{est}=112.9682 ; σ_{est}=48.9040	
$8^{ème}$*mélang*					54
Premier mode	Gauss	μ =25;σ = 10	Gauss	μ_{est}=25 .4696; σ_{est}=9.9801	
Deuxième mode	Weibull	K = 125 ; λ =5	Gauss	μ_{est}=115.6889; σ_{est}=24.9877	

Tableau (3.10) : Paramètres des distributions générées et estimées

3.5 Tests et résultats expérimentaux

Nous allons présenter maintenant les résultats de l'application de la méthode de seuillage proposée sur un ensemble de douze images réelles. Les figures (3.18) à (3.29) montrent ces images, les résultats de l'approximation de l'histogramme par un mélange de deux distributions

et les images segmentées correspondantes. Les modèles des distributions identifiées, leurs paramètres ainsi que la valeur du seuil sont donnés pour chaque image dans les tableaux présentés en dessous de la figure correspondante.

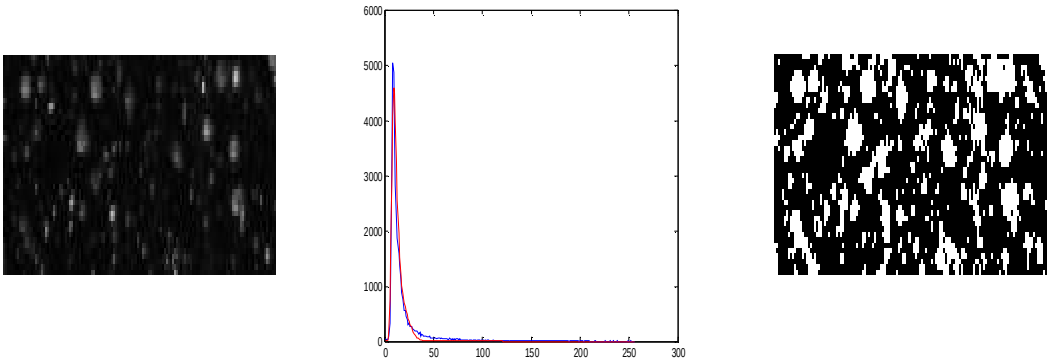


(a) (b) (c)

Figure (3.18) : (a) : Image 1 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Gamma	$\theta_1 = 4.1714\,; k_1 = 6.9337$	$P_1 = 0.4883$	65	1
2	Gauss	$\mu = 148.9281;\ \sigma = 42.9730$	$P_2 = 0.5117$		

Tableau (3.11) : Résultats obtenus sur l'image 1

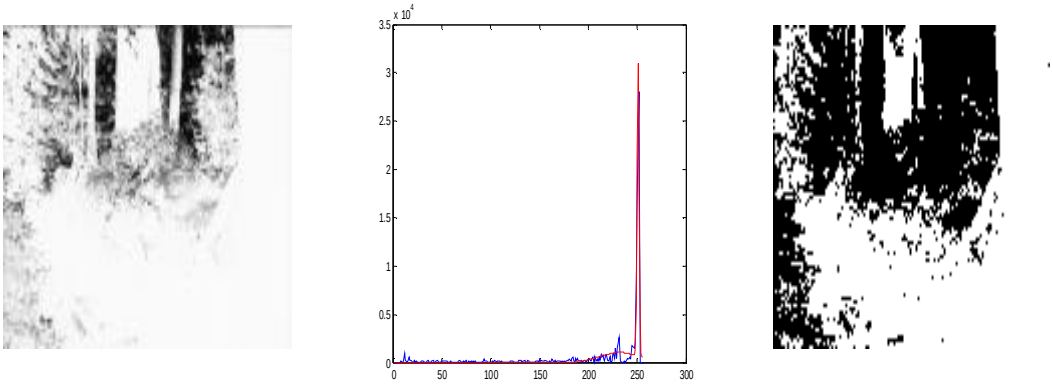


(a) (b) (c)

Figure (3.19) : (a) : image 2 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Gamma	$\theta = 2.4294; k = 15.7053$	P_1 =0.7072	16	3
2	Chi2	$v = 18$	P_2 =0.2928		

Tableau (3.12) : Résultats obtenus sur l'image 2

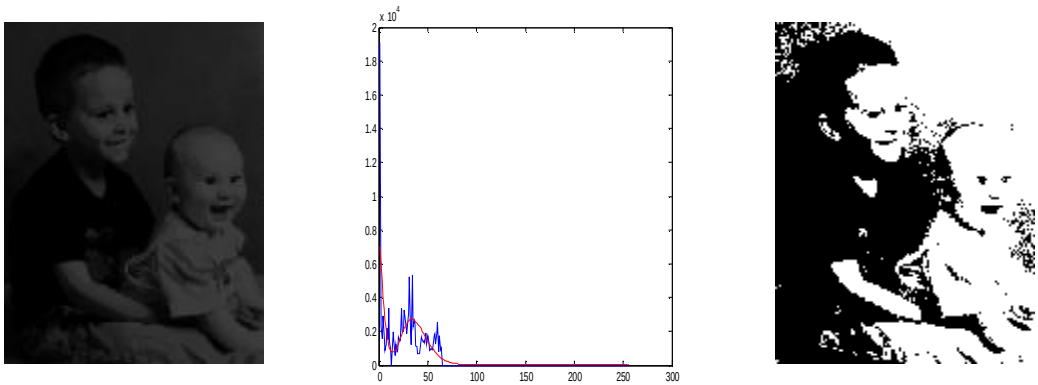


(a) (b) (c)

Figure (3.20) : (a) : Image3 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Chi2	$v = 234$	P_1 =0.5809	248	6
2	Gauss	$\mu = 250.9019;\ \sigma = 1.0267$	P_2 =0.4191		

Tableau (3.13) : Résultats obtenus sur l'image 3

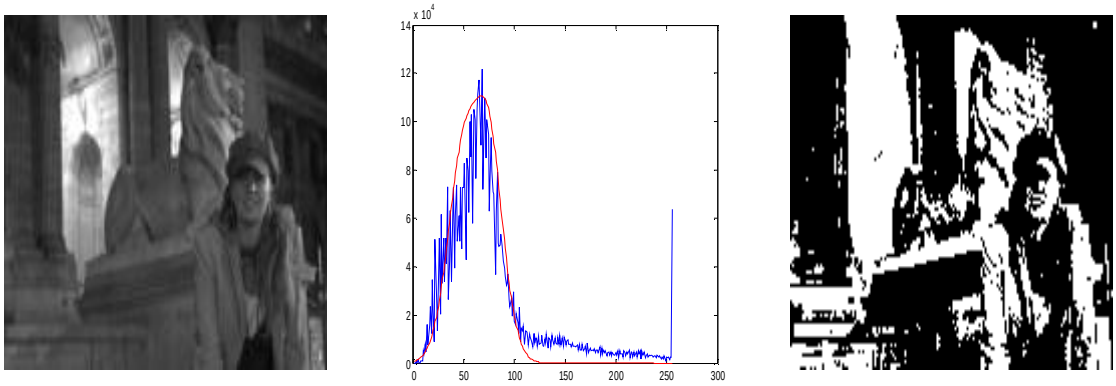


(a) (b) (c)

Figure (3.21) : (a) : Image 4 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Beta	$\alpha = 1.0851; \beta = 57.8234$	P_1 =0.3302	27	1
2	Gamma	$\theta = 8.3761; k = 4.5580$	P_2 =0.6698		

Tableau (3.14) : Résultats obtenus sur l'image 4

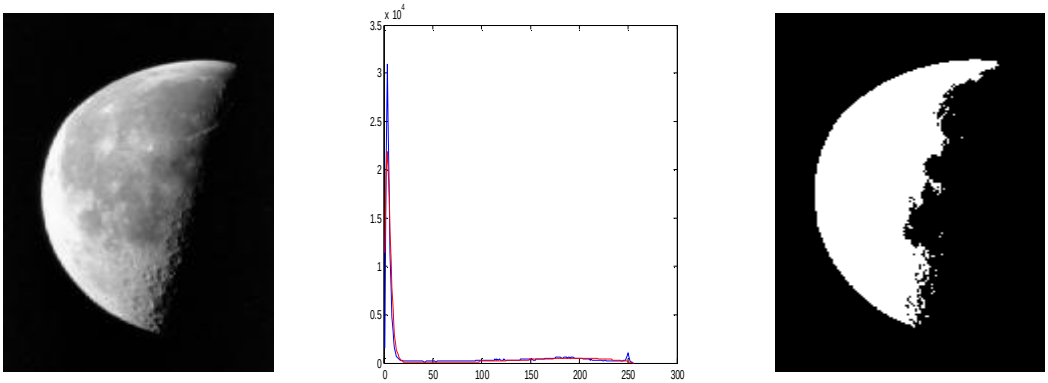


(a) (b) (c)

Figure (3.22) : (a) : Image 5 ;(b) ; Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Gauss	$\mu = 125.1642; \sigma = 50.6451$	$P_1 = 0.6607$	77	1
2	Chi2	$v_2 = 80$	$P_2 = 0.3393$		

Tableau (3.15) : Résultats obtenus sur l'image 5

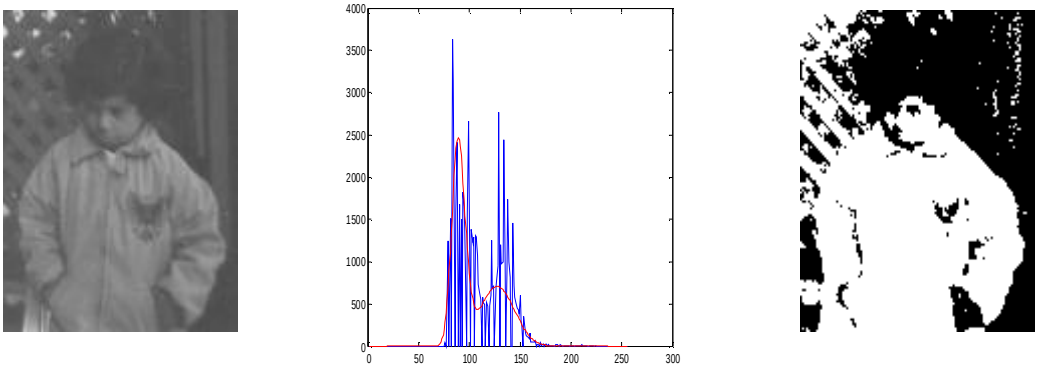


(a) (b) (c)

Figure (3.23) : (a) : Image 6 ;(b) : Histogramme réel et estimé; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Chi2	$v = 5$	$P_1 = 0.7237$	90	1
2	Beta	$\alpha = 0.9559;\ \beta = 26.9901$	$P_2 = 0.2763$		

Table au (3.16) : Résultats obtenus sur l'image 6

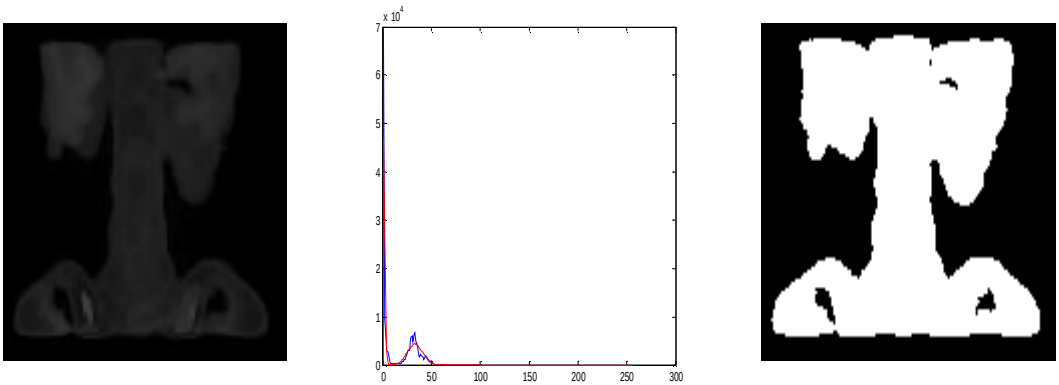


(a) (b) (c)

Figure (3.24) : (a) : Image 7;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Gamma	$\theta = 60.0552; k = 2.1281$	P_1 =0.4306	100	1
2	Beta	$\alpha = 28.4221; \beta = 28.6984$	P_2 =0.5694		

Tableau (3.17) : Résultats obtenus sur l'image 7

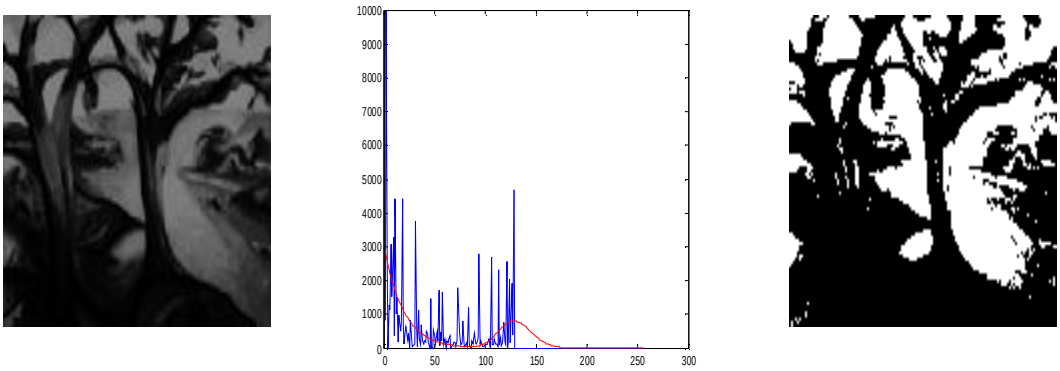


(a) (b) (c)

Figure (3.25) : (a) : image 8 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Gamma	$\theta = 2.9158; k = 0.5445$	P_1 =0.5162	7	5
2	Gauss	$\mu = 32.9187; \sigma = 8.0074$	P_2 =0.4838		

Tableau (3.18) : Résultats obtenus sur l'image 8

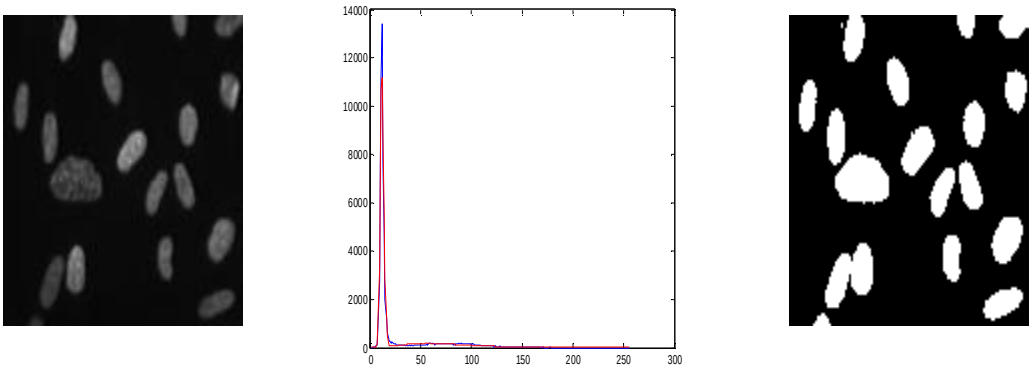


(a) (b) (c)

Figure (3.26) : (a) : Image 9 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Exponentiel	$\lambda = 13.4810$	P_1 =0.6404	91	2
2	Beta	$v = 130$	P_2 =0.3596		

Tableau (3.19) : Résultats obtenus sur l'image 9

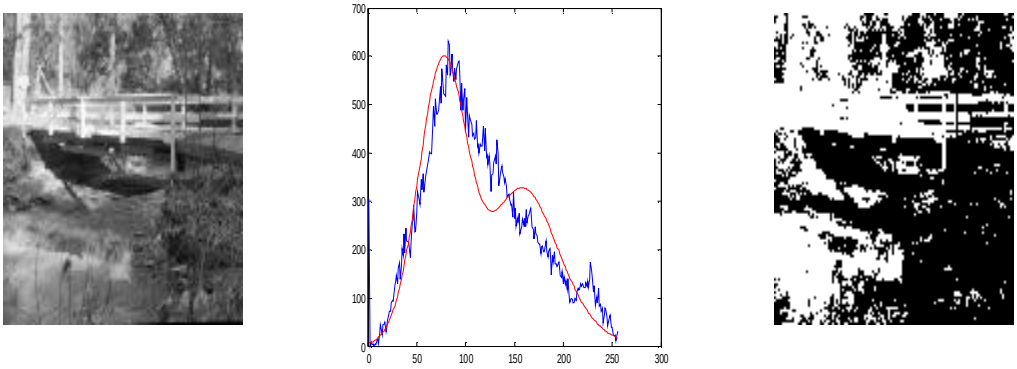


(a) (b) (c)

Figure (3.27) : (a) : Image 10 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Gamma	$\theta_1 = 41.1357; k_1 = 0.2902$	$P_1 = 0.8028$	19	2
2	Gamma	$\theta_2 = 4.3903; k_2 = 16.6000$	P_2 =0.1972		

Tableau (3.20) : Résultats obtenus sur l'image 10

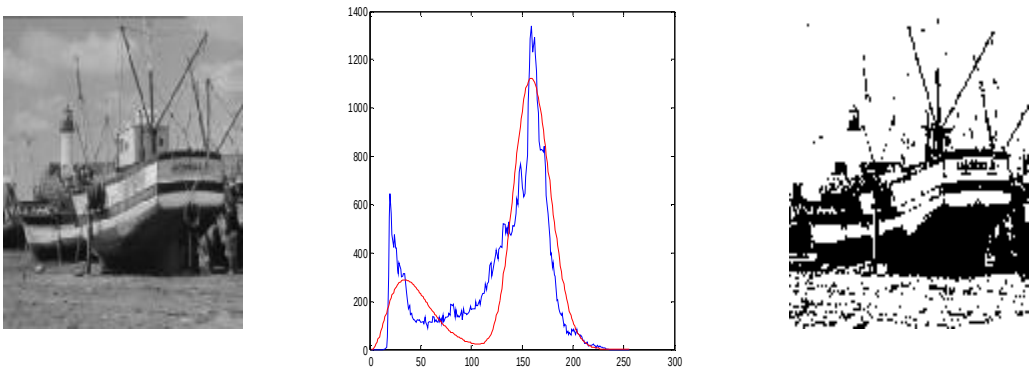


(a) (b) (c)

Figure (3.28) : (a) : Image 11 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Gauss	$\mu = 80.6135;\ \sigma = 27.9718$	$P_1 = 0.6234$	125	6
2	Gamma	$\theta = 27.1848; k = 6.3100$	$P_2 = 0.3766$		

Tableau (3.21) : Résultats obtenus sur l'image 11



(a) (b) (c)

Figure (3.29) : (a) : Image 12 ;(b) : Histogramme réel et estimé ; (c) : Image segmentée

Classe	Distribution	Paramètres	Probabilités à priori	Seuil	Nombre d'itérations
1	Gamma	$\theta = 2.9248; k = 21.6851$	$P_1 = 0.3180$	94	4
2	Chi2	$v = 161$	$P_2 = 0.6820$		

Tableau (3.22) : Résultats obtenus sur l'image 12

Ces résultats montrent clairement que l'hypothèse qu'un histogramme peut être approximer par un mélange de distributions de même type n'est pas toujours vérifiée. En effet, sur la plus part des images testées ayant des histogrammes bimodales, l'approximation de ces histogrammes par deux distributions ayant des modèles différents permettent de mieux localiser le seuil séparant ces deux distributions et ce en quelques itérations.

Afin de mieux évaluer les performances de la méthode de seuillage proposée, nous avons comparé ses résultats avec ceux obtenus par les méthodes qui considèrent que l'histogramme est composé d'un mélange de deux distributions de même type, Gaussien **[46]** ou Beta **[73]**. Notons que dans le premier cas (mélange Gaussien), les paramètres statistiques sont déterminés par l'algorithme EM alors que pour le deuxième cas, nous avons appliqué la méthode proposée mais sans la phase d'identification du modèle de distribution puisque le modèle de chaque distribution est fixé par la loi Beta. Comme critère de comparaison, nous avons utilisé la distance de Kolmogorov-Smirnov entre l'histogramme original de l'image est l'histogramme estimé. Plus la distance de KS est faible, meilleur est la méthode de seuillage.

Le tableau (3.23) regroupe les résultats comparatifs, avec en gras la méthode de seuillage qui fournit la plus faible distance.

	Méthode proposée			**Beta_Beta**			**Gauss_Gauss**		
	Seuil	Temps	KS	Seuil	Temps	KS	Seuil	Temps	KS
Image1	65	0.663920	**360.6061**	63	3.663192	370.1570	80	0.622052	1.0370e+003
Image2	16	1.778193	**1.1459e+003**	18	1.666348	1.2783e+003	27	0.433979	5.0459e+003
Image3	248	3.450874	**1.0235e+004**	231	2.360758	1.9131e+004	210	1.215252	2.8070e+004
Image4	27	1.61658	**1.1052e+004**	14	3.716729	1.1897e+004	45	0.370025	1.9065e+004
Image5	77	31.497500	**63725**	77	91.025462	63725	108	83.565687	1.2167e+005
Image6	100	7.770496	9.9496e+003	15	2.163436	**5.9031e+003**	74	1.569134	3.0924e+004
Image7	100	1.150459	**2.4628e+003**	111	0.849109	2.6553e+003	110	0.846205	3.6280e+003
Image8	7	4.187398	**2.2437e+004**	9	2.518959	2.4065e+004	38	0.870093	6.9426e+004
Image9	60	4.849608	**7.2487e+003**	50	0.826003	7.3999e+003	43	2.747342	9.9800e+003
Image10	19	1.049883	**2.1920e+003**	19	1.075353	2.2006e+003	39	0.507008	1.3394e+004
Image 11	125	4.330324	**297.2499**	215	6.731621	304.0945	129	4.852870	631.9917
Image 12	94	3.649833	**441.9282**	109	3.879729	465.5268	99	1.418775	1.3420e+003

Tableau (3.23) : Comparaison des résultats de la méthode proposée et celles du cas de mélange Gaussien et le mélange de deux Beta.

Ces résultats montrent clairement la performance de la méthode proposée par rapport aux cas de mélange Gaussien ou mélange de distribution Beta. En effet, sur toutes les images testées, la distance de KS est faible par rapport aux deux autres cas, ce qui dénote d'une meilleure segmentation. Ce tableau renferme également le temps de calcul (t en secondes) consommé par chaque méthode. Il montre que ces temps sont très proches et que la procédure qui détermine le type du modèle de chaque distribution ne consomme pas beaucoup de temps.

3.6 Conclusion

Dans ce chapitre, nous avons décrit en détails une méthode de seuillage basée sur la modélisation statistique de l'histogramme. Son principe consiste à approximer l'histogramme par un mélange de distributions dont les modèles sont à priori inconnus. L'algorithme proposé permet de déterminer automatiquement le modèle de chaque distribution pour ensuite déduire les seuils à partir des paramètres des distributions identifiées. Les résultats expérimentaux portant sur différents types d'images ont montré l'efficacité de cette méthode de segmentation.

Conclusion générale

Conclusion générale

Dans ce mémoire, nous nous somme intéressé à la segmentation d'images par seuillage paramétrique qui considère que l'histogramme est composé d'un mélange de distributions où chaque distribution correspond à une classe.

Dans ce cas, le problème de seuillage devient un problème du choix des lois qui composent le mélange et d'estimation des paramètres de ces lois et du calcul du seuil à partir de ces lois. Généralement, ces lois sont considérées de même type (Gaussienne) et les paramètres sont estimés par des méthodes générales de type EM. Cependant, dans plusieurs cas, cette hypothèse s'avère inadéquate.

Nous avons ainsi proposé une méthode qui consiste à déterminer automatiquement le type de chaque distribution d'un mélange pour ensuite déduire les seuils à partir des paramètres des distributions. Pour cela, nous avons considéré une famille de huit lois admissibles pour le mélange : Gaussien, Gamma, Beta, Exponentielle, Chi2, Lognormal, Rayleigh et Weibull. Les paramètres statistiques de chaque distribution sont déterminés par la méthode du maximum de vraisemblance ou la méthode des moments, après avoir identifié le modèle de chaque distribution parmi ces huit modèles donnés. Cette identification est effectuée en utilisant le test d'adéquation de Kolmogorov-Smirnov qui se base sur une distance minimale entre la distribution empirique et la distribution candidate.

Afin d'évaluer l'algorithme proposé, nous l'avons testé sur un ensemble de douze images en niveaux de gris. Le critère d'évaluations a été définit pour comparer les résultats obtenus avec les autres méthodes de seuillage (cas de mélange Gaussien et mélange de distributions Beta) afin de conclure sur son validité.

Les résultats expérimentaux portant sur différents types d'images ont montré l'efficacité de l'algorithme de segmentation développé. Ceci est du à la capacité de l'algorithme de s'adapter au contexte de l'image.

Les perspectives de ce travail concernent principalement l'extension de notre algorithme de segmentation sur les images ayant des histogrammes multimodales. Il s'agira d'approximer l'histogramme par un mélange de plusieurs distributions pour ensuite déduire plusieurs seuils.

Références

Références

[1] Y. Delignon, R. Garello et A. Hillion. *Etude statistique d'images SAR de la surface de la mer*. Colloque GRETSI 1991, Juan-Les-pins, pp. 573-576, 1991.

[2] S. W. Zucker. *Region growing: childhood and adolescence.* Computer Vision Graphics and Image Processings, vol. 5, pp. 382-399, 1976.

[3] S. Philipp et J. P. Cocquerez. *Analyse d'images: filtrage et segmentation*. Masson, 1995.

[4] L. G. Roberts. *Machine perception of three dimensional solids.* In J. T. and al. Tippet, editor, Optical and Electro-optical Information Processing, pp. 159-197. MIT Press, Cambridge, 1965.

[5] J. M. S. Prewitt. *Object enhancement and extraction*. In PPP70, pp. 75-149, 1970.

[6] I. Sobel. *Neighbourhood coding of binary images for fast contour following and general array binary processing*. Computer Graphics and Image Processing, vol. 8, pp. 127-135, 1978.

[7] R. Kirsch. *Computer determination of the constituent structures of biomedical images*. Computer and Biomedical Research, vol. 4, No. 3, pp. 315-328. USA, 1971.

[8] J. Canny. *A computational approach to edge detection*. IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 8, No. 6, pp. 679-698, 1986.

[9] R. Deriche. *Using Canny's criteria to derive a recursively implemented optimal edge detector*. International Journal of Computer Vision, pp. 167-187, 1987.

[10] J. Shen and S. Castan. *An optimal linear operator for edge detection*. In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition , pp. 109-114, Miami Beach, Florida, USA, 1986.

[11] S. Castan, J. Zhao et J. Shen. *Une famille de détecteurs de contours basée sur filtre exponentiel optimal*. In AFCET-RFIA, 1989.

[12] M. Kass, A. Wikin and D. Terzopoulos. *Snakes: active contour models.* Computer Vision, Graphics and Image Processing, pp.321-331, 1988.

[13] R. Haralick and L. Shapiro. *Image segmentation techniques.* Computer Vision Graphics Image Processing, vol. 29, pp. 100-132, 1985.

[14] R. Gonzalez and R. Woods. *Digital image processing*. Addison-Wesley Publishing Company, Reading, MA 1993.

[15] A. Tremeau and N. Borel. *A region growing and merging algorithm to color segmentation.* Pattern Recognition, vol. 30, No. 7, pp. 1191-1204, 1997.

[16] S. L. Horowitz and S. Pavlidis. *Picture segmentation by a directed split and merge procedure*. In *2nd Int*. Joint Conf. on Pattern Recognition, pp. 424-433, 1974.

[17] M. Fontaine. *Segmentation non supervisée d'images couleur par analyse de la connexité des pixels*. Thèse de doctorat, Université de Lille 1. 2001.

[18] R. Schettini. *A segmentation algorithm for color images.* Pattern Recognition Letters, vol. 14, pp. 499–506, 1993.

[19] K. Saarinen. *Color image segmentation by a watershed algorithm and region adjacency graph processing.* In *ICIP'94: Int. Conf. on Image Processing*, pp. 1021–1024, 1994.

[20] A. Trémeau and P. Colantoni. *Region adjacency graph applied to color image segmentation*. IEEE Trans. In Image Processing, pp.735–744, 2000.

[21] A. Nakib. *Conception de métaheuristiques d'optimisation pour la segmentation d'images. Application à des images biomédicales,* Thèse de doctorat, Université de Paris 12-Val de Marne. 2007.

[22] N. Monmarché. *Algorithmes des fourmis artificielles: applications à la classification et à l'optimisation*, Thèse de doctorat, Université de Tours. 2000.

[23] K. Takahashi, H. Nakatani and K. Abe. *Color image segmentation using ISODATA clustering method*. 2nd Asian Conf. On Computer Vision, vol. 1, pp 523–527, 1995.

[24] A. P. Dempster, N. M. Laird and D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society, vol. 39, pp. 1-38, 1977.

[25] R. Horaud et O. Monga. *Vision par ordinateur : outils fondamentaux.* Editions Hermès, 1993.

[26] A. S. Abutaleb. *Automatic thresholding of grey-level pictures using two-dimensional entropy*. Comput. Using Graphics Image Process. 47 22-32, 1989.

[27] A. D. Brink. Thresholding *of digital images using two-dimensional entropies.* Pattern Recognition, vol. 25, pp. 803-808, 1992.

[28] J. Z. Li and N. Q. Li. T*he automatic thresholding of grey-level pictures via two-dimensional Otsu method*. Acta Automat. Sinica (In Chinese), vol. 19, pp. 101-105, 1993.

[29] J. Gong, L. Li and W. Chen. *Fast recursive algorithms for two-dimensional thresholding.* Pattern Recognition, vol. 31, No. 3, pp. 295-300, 1998.

[30] J. Bernsen. *Dynamic thresholding of grey-level images proc*. Internat. Conf. Pattern Recognition. pp. 1251-1255, 1986.

[31] M. Kamel and A. Zhao. *Extraction of binary character/ Graphics images from greyscale document images.* Graphical Models and Image Processing, vol. 55, No. 3, pp. 203-217, 1993.

[32] C. K. Chow and T.Kaneko. *Automatic boundary detection of left ventricle from cineangiograms.* Comput. Biomed. Res, vol. 5, pp. 388-410, 1972.

[33] Q. D. Trier and T. Taxt. *Improvement of integrated function algorithm of binarization of document images.* Pattern Recognition Letters, vol. 16, No. 3, pp. 277-283, 1995.

[34] K. Chehdi et D. Coquin. *Binarisation d'images par seuillage local optimal maximisant un critère d'homogénéité.* Troisième colloque GRETSI-Juan-les-pins, pp. 1096-1072, 1991.

[35] N. B. Venkateswarlu and R.D.Boyle. *New segmentation thechnic for document image analysis.* Image and Vision Computing to Appear, 1994.

[36] N. B. Venkateswarlu. *Implimentation of some image thresholding algorithms on a connection machine*, Pattern Recognition Letters, vol. 16, pp. 759-768, 1995.

[37] J. Bernsen. *Dynamic thresholding of grey-level images proc.* Internat. Conf. Pattern Recognition, pp. 1251-1255, 1986.

[38] Q. D. Trier and A. K. Jain. *Goal directed evaluation of binarization methods*, IEEE Transactions on Pattern Analysis ans Machine Intellegence, vol. 17, No. 12, 1995

[39] K. V. Mardia and T. J. Hainsworth. *A special thresholding method for image segmentation*, IEEE Trans. Pattern Analysis and Mach. Intellegence, vol. 10, No. 6, pp. 919-927, 1988.

[40] N. Otsu. *A threshold selection method for grey level histograms*, IEEE Trans. On System, Man and Cybernetics, vol. SMC-9, No. 1, 1979.

[41] T. Pun. A *new method for gray-level picture thresholding using the entropy of histogram.* Signal Proc., vol. 2, pp. 223-237, 1980.

[42] E. J. N. Kapur, P. K. Sahoo and A. K. C. Wong. *A new methode for gray-level picture thresholding using the entropy of the histogram.* Comput. Vision Graphics Image Process, vol. 29, pp. 273-285, 1985.

[43] G. Johannnsen and J. Bille. *A threshold selection method using information measures*. In Proc. V. I. Int. Conf. On Pattern Recognition, Munich, 1983.

[44] C. H. Li and C. K. Lee. *Minimum cross entropy thresholding.* Pattern Recognition, vol. 26 No. 4, pp. 617-625, 1993.

[45] P. Sahoo, C. Wilkins and J. Yeager. *Threshold selection using Renyi's entropy.* Pattern Recognition, vol. 30, No. 1, pp. 71-84, 1997.

[46] Z.-K. Huang and K.-W. Chau. *A new thresholding method based on Gaussian mixture model.* Model, Appl. Ms th. Comput, doi: 10.1016/j.amc.2008.05.130, 2008.

[47] D. E Goldberg. *Genitic algorithms in search, optimization and machine learning*. Massachusetts: Addison-Wesly, 1989.

[48] S. Kirkpatrick, C. Gelatt and M. P.Vecchi. *Optimisation by simulated annealing*. Science, vol. 220, pp. 671-680, 1983.

[49] W. Billbro, G. Logenthiran and S. Rajala. *Optimal thresholding: a new approach*. Pattern Recognition Letters, vol. 11, pp. 803-810, 1990.

[50] Y. Collette et P. Siarry. *Optimisation multiobjectif*. Eyrolles, 2002.

[51] J. Dréo, A. Pétrowski, P. Siarry et E. Taillard. *Métaheuristiques pour l'optimisation difficile*. Eyrolles, 2003.

[52] J. Kittler and J. Illingworth. *Minimum error thresholding*. Pattern Recognition, vol. 19, pp. 41-47, 1986.

[53] M. Evans. *Statistical distributions.* A Wiley-Interscience Publication, 1993.

[54] N. L. Johnson, S. Kotz and N. Balakrishnan. *Continuous univariate distributions.* Volume1**,** Wiley-Interscience Publication, 1970.

[55] T. T. Soong. *Fundamentals probability and statistics for engineers*. A Wiley-Interscience Publication, 2004.

[56] P. Tassi. *Méthodes statistiques. Methodes statistiques*. Collection : Economie et Statistiques Avancées. Economica, Paris, 1985.

[57] F. Flitti. *Techniques de réduction de données et analyse d'images multispectrales astronomiques par arbres de Markov*. Thèse de doctorat, Université Louis Pasteur-Strasbourg I, 2005.

[58] M. K. Varanasi and B. Aazhang. *Parametric generalized gaussian density estimation. J. Acoust. Soc. Amer*, 86, pp. 1404-1415, 1989.

[59] J. N. Provost, C. Collet, P. Rostaing, P. Pérez and P. Bouthemy. *Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps*. Computer Vision and Image Understanding, pp. 155-174, 2004.

[60] J. N. Provost. *Classification bathymétrique en imagerie multispectrale SPOT*. Thèse de doctorat, Université de Bretagne Occidentale - Ecole Navale (Laboratoire GTS), 2001.

[61] C. Walck. *Hand-book on statistical distributions for experimentalists*. Particle Physics Group Fysikum University of Stockholm , 2007.

[62] E. Monfrini. *Identifiabilité et méthode des moments dans les mélanges généralisés de distributions du système de Pearson*. Thèse de doctorat, Université Paris, 2002.

[63] S. Derrode. S*ystème de Pearson : description mathématique du système de Pearson implémentation en C++ à l'aide de la librairie GSL*. Ecole Centrale Marseille (ECM), 2009.

[64] M. Sezgin and B. Sankur. *Survey over image thresholding techniques and quantitative performance evaluation*. J. of Electronic Imaging, vol. 13, pp. 146-165, 2004.

[65] P. K. Sahoo and al. *A survey of thresholding techniques*. Computer Vision, Graphics, and Image Processing, vol. 41, pp. 233-260, 1988.

[66] R. C. Gonzalez and R. E. Woods. *Digital image processing using Matlab*. Pearson Prentice Hall, 2004.

[67] H. Maitre. *Le traitement des images*. Paris : Hermes, 2003.

[68] J. P. Cocquerez et S. Philipp. *Analyse d'images et segmentation*. Masson, 1995.

[69] A. El Zaart, D. Ziou, S. Wang and Q. Jiang. *Segmentation of SAR image.* Pattern Recognition, vol. 35, pp. 713-724, 2002.

[70] A. Al-Haussain and A. El-Zaart. *Moment-Preserving thresholding using Gamma distribution.* Second International Conference Engineering and Technology, vol. 6-325, 2010.

[71] R. Al-Attas and A. El-Zaart. *Thresholding of medical images using minimum cross entropy*. IFMBE Proceedings 15, pp. 296-299, 2007.

[72] A. El-Zaart. *Images thresholding using ISODATA technique with Gamma distribution*. Pattern Recognition and Image Analysis, vol. 20, pp. 29-41, 2010.

[73] A. Al-Saleh and A. El-Zaart. *Unsupervised learning technique for skin images segmentation using a mixture of beta distributions*. IFMBE Proceedings 15, pp. 304-307, 2007.

[74] X. Jinghao, Z. Yujin and L. Xinggang. *Rayleigh-distribution based minimum error thresholding for SAR images*. Journal of Electronics, vol. 16, No. 4, 1999.

[75] Y. Bazi, L. Bruzzone and F. Melgani. *Image thresholding based on the EM algorithm and the generalized Gaussian distribution*. Pattern Recognition, pp. 619-634, 2006.

[76] M. S. Allili, N. Bouguila and D. Ziou. *Finite generalized Gaussian mixture modeling and applications to image and video foreground segmentation*. TIn Proc. of Fourth Canadian Conference on Computer and Robot Vision, pp. 183-190, Montreal, Canada, 2007.

[77] Y. Bazi, L. Bruzzone and F. Melgrani. *Image thresholding based on the EM algorithm and the generalized Gaussian distribution*. Pattern Recognition, pp. 619-634, 2007.

[78] E. Kuruoglu and J. Zerubia. *Modeling SAR images with a generalization of Rayleigh distribution.* IEEE Transactions on Image Processing, vol. 13, No. 4, 2004.

[79] A. Bougarradh, S. Mhiri et F. Ghorbel. *Segmentation non supervisée d'images angiographiques de la rétine par le système de Pearson et le reéchantillonnage bootstrap***.** Tunisie 2010.

[80] A. Marzouki,Y. Delignon, H. C. Quelle et W. Pieczynski. S*egmentation non supervisée d'images satellite utilisant un modèle hiérarchique généralisé.* Quatorzième Colloque RETSI-JUAN-LES-PINS. 1993.

[81] A. El Zaart and D. Ziou. *Statistical modeling of heterogenuous multimodel image histogram using parametric distribution.* Int. Journal of Remote Sensing, pp. 2277-2294, 2007.

[82] G. Moser, J. Zerubia and S. B. Serpico. *Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation.* IEEE Trans. on Geoscience and Remote Sensing, vol. 44, No. 1, 2006.

[83] R. Horaud et O. Monga. *Vision par ordinateur outils fondamentaux*. Hermès, 1993.

Annexes

Annexe A : Notions générales sur les paramètres statistiques

A.1. Paramètres de position

Les paramètres de position, aussi appelés valeurs centrales, servent à caractériser l'ordre de grandeur des données.

a. Moyenne arithmétique : Elle est plus souvent appelée moyenne, et est en général notée $\bar{x}$, elle est calculée en utilisant la formule:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

b. Moyenne géométrique : La moyenne géométrique $(\bar{x}_g)$ est toujours inférieure (ou égale) à la moyenne arithmétique. Elle est donnée par :

$$\bar{x}_g = \left[\prod_{i=1}^{n} x_i\right]^{1/n}$$

On peut remarquer que : $\log(\bar{x}_g) = \frac{1}{n}\sum_{i=1}^{n} \log(x_i)$

En d'autres termes, le log de la moyenne géométrique est la moyenne arithmétique du log des données. Elle est très souvent utilisée pour les données distribuées suivant une loi normale.

c. Moyenne harmonique : La moyenne harmonique $(\bar{x}_h)$ est toujours inférieure (ou égale) à la moyenne géométrique, elle est en général utilisée pour calculer les moyennes sur des intervalles de temps qui séparent des événements. Elle est donnée par :

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

On peut remarquer :

$$\frac{1}{\bar{x}_h} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}$$

d. Médiane : La médiane, notée M_e, est la valeur (observée ou possible) de la variable statistique, dans la série d'observations ordonnée en ordre croissant ou décroissant, qui partage cette série en deux parties, chacune comprenant le même nombre d'observations de part et d'autre de M_e.

e. **Les quartiles :** Les quartiles sont au nombre de trois. La médiane est le deuxième. Le premier quartile q_1 est la valeur telle que 75% des observations lui sont supérieures (ou égales) et 25% inférieures (ou égales). Le troisième quartile q_3 est la valeur que 25% des observations lui sont supérieurs (ou égales) et 75% inférieures (ou égales).

f. **Mode :** Le mode, noté M_0, (ou valeur dominante) est la valeur de la variable statistique la plus fréquente que l'on observe dans une série.

A.2. Paramètres de dispersion

Ces paramètres (comme son nom l'indique) mesurent la dispersion des données.

a. **La variance :** Elle est définie comme la moyenne des carrés des écarts à la moyenne, soit :

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

La variance s'exprime dans l'unité au carré des données.

b. **L'écart-type :** est la racine carrée de la variance : $s = \sqrt{S^2}$

c. **L'étendue ou amplitude :** est définie comme la différence entre le maximum et le minimum.

d. **Le coefficient de variation :** est définie comme le rapport entre l'écart-type et la moyenne :

$$CV = \frac{s}{\bar{x}}$$

A.3. Paramètres de forme

Les paramètres Skewness et Kurtosis sont construits à partir des moments centrés d'ordre 2, 3 et 4 qui mesurent respectivement la symétrie et l'aplatissement de la distribution dont l'échantillon est issu.

Pour une loi centrée réduite, ces coefficients sont nuls.

Les moments centrés d'ordre 3 et 4 sont définis par :

$$m_3 = \frac{1}{n} = \sum_{i=1}^{n}(x_i - \bar{x})^3 \; ; \qquad m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$$

A partir de ces définitions, les paramètres Skewness et Kurtosis sont respectivement définis par :

$$g_1 = \frac{m_3}{S^3} \quad ; \quad g_2 = \frac{m_4}{S^4} - 3$$

Annexe B : Estimation de seuil

Cette partie de l'annexe expose les équations égalisant deux distributions définies par leurs paramètres dans le but de déterminer les seuils.

· **Mélange des distributions Gauss et Gamma**

$$Ax^2 + Bx + C\log x + D = 0$$

$$A = -\frac{1}{2s^2} \quad B = \frac{m}{s^2} + \frac{1}{q} \quad C = 1 - k \quad D = \left(\log \frac{P_1}{s\sqrt{2p}} - \frac{m^2}{2s^2} - \log \frac{P_2}{q^k \Gamma(k)} \right)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^2 + B * T_{old} + C * \log(T_{old}) + D}{2A * T_{old} + B + (C / T_{old})}$$

· **Mélange des distributions Gauss et Beta**

$$Ax^2 + Bx + C\log x + D\log(1-x) + E = 0$$

$$A = -\frac{1}{2s^2} \quad B = \frac{m}{s^2} \quad C = (1-a) \quad D = (1-b) \quad E = \left(\log \frac{P_1}{s\sqrt{2p}} - \frac{m^2}{2s^2} - \log \frac{P_2}{\mathrm{B}(a,b)} \right)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^2 + B * T_{old} + C * \log(T_{old}) + D * \log(1 - T_{old}) + E}{2A * T_{old} + B + (C / T_{old}) - (D / (1 - T_{old}))}$$

· **Mélange des distributions Gauss et Log-normal**

$$Ax^2 + Bx + C\log x + (D\log x)^2 + E = 0$$

$$A = -\frac{1}{2s_1^2} \quad B = \frac{m_1}{s_1^2} \quad C = \left(1 - \frac{m_2}{s_2^2}\right) \quad D = \frac{1}{2s_2^2} \quad E = \left(\log \frac{P_1 s_2}{P_2 s_1} - \frac{m_1^2}{2s_1^2} + \frac{m_2^2}{2s_2^2} \right)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^2 + B * T_{old} + C * \log(T_{old}) + (D * \log(T_{old}))^2 + E}{2A * T_{old} + B + (C / T_{old}) + (2D * \log(T_{old}) / (T_{old}))}$$

· **Mélange des distributions Gauss et Exponentielle**

$$Ax^2 + Bx + C = 0$$

$$A = -\frac{1}{2s^2} \quad B = \left(\frac{m}{s^2} + l \right) \quad C = \left(\log \frac{P_1}{s\sqrt{2p}} - \frac{m^2}{2s^2} - \log(P_2 l) \right)$$

$$T^* = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

- **Mélange des distributions Gauss et Weibull**

$$Ax^k + Bx^2 + Cx + D\log x + E = 0$$

$$A = \frac{1}{l^k} \quad B = -\frac{1}{2s^{2}} \quad C = \frac{m}{s^{2}} \quad D(1-k) \quad E = \left(\log \frac{P_1}{s\sqrt{2p}} - \frac{m^2}{2s^{2}} - \log(P_2 k) + k\log(l) \right)$$

$$T_{new} = T_{old} - \frac{A*T_{old}^k + B*T_{old}^2 + C*T_{old} + D*\log(T_{old}) + E}{k*A*T_{old}^{k-1} + 2B*T_{old} + C + (D/T_{old})}$$

- **Mélange des distributions Gauss et Rayleigh**

$$Ax^2 + Bx - \log x + C = 0$$

$$A = \frac{1}{2}\left(\frac{1}{s_2^2} - \frac{1}{s_1^2} \right) \quad B = \frac{m}{s_1^2} \quad C = \left(\log \frac{P_1 s_2}{P_2 s_1} - \frac{m_1^2}{2s_1^2} - \frac{P_2}{s_2^2} \right)$$

$$T_{new} = T_{old} - \frac{A*T_{old}^2 + B*T_{old} - \log(T_{old}) + C}{2A*T_{old} + B - (1/T_{old})}$$

- **Mélange des distributions Gauss et Chi2**

$$Ax^2 + Bx + C\log x + D = 0$$

$$A = -\frac{1}{2s^{2}} \quad B = \left(\frac{m}{s^{2}} - \frac{1}{2} \right) \quad C = \left(1 - \frac{k}{1} \right) \quad D = \left(\log \frac{P_1}{s\sqrt{2p}} - \frac{m^2}{2s^{2}} - \log\left(\frac{P_2}{2^{k/2}\Gamma[k/2]} \right) \right)$$

$$T_{new} = T_{old} - \frac{A*T_{old}^2 + B*T_{old} + C*\log(T_{old}) + D}{2A*T_{old} + B + (C/T_{old})}$$

- **Mélange de deux distributions Gamma**

$$Ax + B\log x + C = 0$$

$$A = \left(\frac{1}{q_2} - \frac{1}{q_1} \right) \quad B = (k_1 - k_2) \quad C = \log\left(\frac{P_1 q_2^{k_2} \Gamma(k_2)}{P_1 q_1^{k_1} \Gamma(k_1)} \right)$$

$$T_{new} = T_{old} - \frac{A*T_{old} + B*\log(T_{old}) + C}{A + (B/T_{old})}$$

· **Mélange des distributions Gamma et Beta**

$$Ax + B\log x + C\log(1-x) + D = 0$$

$$A = \frac{1}{q} \quad B = (a-k) \quad C = (b-1) \quad D = \log\left(\frac{P_1}{q^k \Gamma(k)}\right) - \log\left(\frac{P_2}{b(a,b)}\right)$$

$$T_{new} = T_{old} - \frac{A*T_{old} + B*\log(T_{old}) + C*\log(1-T_{old}) + D}{A + (B/T_{old}) - (C/(1-T_{old}))}$$

· **Mélange des distributions Gamma et Log-normal**

$$Ax + B\log x + C*(\log(x))^2 + D = 0$$

$$A = -\frac{1}{q} \quad B = \left(k - \frac{m}{s^2}\right) \quad C = \frac{1}{2s^2} \quad D = \log\left(\frac{P_1}{q^k \Gamma(k)}\right) - \log\left(\frac{P_2}{s\sqrt{2p}}\right) + \left(\frac{m^2}{2s^2}\right)$$

$$T_{new} = T_{old} - \frac{A*T_{old} + B*\log(T_{old}) + C*(\log(T_{old}))^2 + D}{A + (B/T_{old}) + (2C*\log(T_{old})/(T_{old}))}$$

· **Mélange des distributions Gamma et Exponentielle**

$$Ax + B\log x + C = 0$$

$$A = \left(l - \frac{1}{q}\right) \quad B = (k-1) \quad D = \log\left(\frac{P_1}{q^k \Gamma(k)}\right) - \log(P_2\, l)$$

$$T_{new} = T_{old} - \frac{A*T_{old} + B*\log(T_{old}) + C}{A + (B/T_{old})}$$

· **Mélange des distributions Gamma et Weibull**

$$Ax^k + Bx + C\log(x) + D = 0$$

$$A = \frac{1}{l^{k_2}} \quad B = -\frac{1}{q} \quad C = (k_1 - k_2) \quad D = \left(\log\frac{P_1}{q^{k_1}\Gamma(k_1)} - \log\left(\frac{P_2 k_2}{l}\right) + (k_2 - 1)\log(l)\right)$$

$$T_{new} = T_{old} - \frac{A*T_{old}^k + B*T_{old} + C*\log(T_{old}) + D}{k*A*T_{old}^{k-1} + B + (C/T_{old})}$$

· **Mélange des distributions Gamma et Rayleigh**

$$Ax^2 + Bx + C\log x + D = 0$$

$$A = \frac{1}{2s^2} \quad B = -\frac{1}{q} \quad C = (k-2) \quad D = \log\left(\frac{P_1}{q^k \Gamma(k)}\right) - \log\left(\frac{P_2}{s^2}\right)$$

$$T_{new} = T_{old} - \frac{A*T_{old}^2 + B*T_{old} + C*\log(T_{old}) + D}{2A*T_{old} + B + (C/T_{old})}$$

- **Mélange des distributions Gamma et Chi2**

$$Ax + B\log x + C = 0$$

$$A = \left(\frac{1}{2} - \frac{1}{q}\right) \quad B = \left(k_1 - \frac{k_2}{2}\right) \quad C = \log\left(\frac{P_1}{q^{k_1}\Gamma(k_1)}\right) - \log\left(\frac{P_2}{2^{k_2/2}\Gamma[k_2/2]}\right)$$

$$T_{new} = T_{old} - \frac{A*T_{old} + B*\log(T_{old}) + C}{A + (B/T_{old})}$$

- **Mélange de deux distributions Beta**

$$A*\log(x) + B*\log(1-x) + C = 0$$

$$A = (a_1 - a_2) \quad B = (b_1 - b_2) \quad C = \log\left(\frac{P_1 \mathrm{B}(a_2, b_2)}{P_2 \mathrm{B}(a_1, b_1)}\right) \quad T_{new} = T_{old} - \frac{A*\log(T_{old}) + B*\log(1-T_{old}) + C}{(A/T_{old}) - (B/(1-T_{old}))}$$

- **Mélange des distributions Beta et Log-normal**

$$A*\log(x) + B*\log(1-x) + C*(\log(x))^2 + D = 0$$

$$A = \left(a - \frac{m}{s^2}\right) \quad B = (b-1) \quad C = \frac{1}{2s^2} \quad D = \log\left(\frac{P_1}{\mathrm{B}(a,b)}\right) - \log\left(\frac{P_2}{s\sqrt{2p}}\right) + \frac{m^2}{2s^2}$$

$$T_{new} = T_{old} - \frac{A*\log(T_{old}) + B*\log(1-T_{old}) + C*(\log(1-T_{old})) + D}{(A/T_{old}) - (B/(1-T_{old})) + (2C*\log(T_{old})/T_{old})}$$

- **Mélange des distributions Beta et Exponentielle**

$$Ax + B*\log(x) + C*\log(1-x) + D = 0$$

$$A = l \qquad B = (a - 1) \qquad C = (b - 1) \qquad D = \log\left(\frac{P_1}{\mathrm{B}(a,b)}\right) - \log(P_2\, l)$$

$$T_{new} = T_{old} - \frac{A * x + B * \log(T_{old}) + C * \log(1 - T_{old}) + D}{A + (B / T_{old}) - (C/(1 - T_{old}))}$$

- **Mélange des distributions Beta et Weibull**

$$Ax^k + B\log(x) + C\log(1-x) + D = 0$$

$$A = \frac{1}{l^k} \quad B = (a - k) \quad C = (b - 1) \quad D = \log\left(\frac{P_1}{\mathrm{B}(a,b)}\right) - \log\left(\frac{P_2 k}{l}\right) + (k-1)\log(l)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^k + B * \log(T_{old}) + C * \log(1 - T_{old}) + D}{k * A * T_{old}^{k-1} + (B / T_{old}) - (C/(1 - T_{old}))}$$

- **Mélange des distributions Beta et Rayleigh**

$$Ax^2 + B\log(x) + C\log(1-x) + D = 0$$

$$A = \frac{1}{2s^2} \quad B = (a - 2) \quad C = (b - 1) \quad D = \log\left(\frac{P_1}{\mathrm{B}(a,b)}\right) - \log\left(\frac{P_2}{s^2}\right)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^2 + B * \log(T_{old}) + C * \log(1 - T_{old}) + D}{2A * T_{old} + (B/T_{old}) - (C/(1 - T_{old}))}$$

- **Mélange des distributions Beta et Chi2**

$$Ax + B\log(x) + C\log(1-x) + D = 0$$

$$A = \frac{1}{2} \quad B = \left(a - \frac{k}{2}\right) \quad C = (b - 1) \quad D = \log\left(\frac{P_1}{\mathrm{B}(a,b)}\right) - \log\left(\frac{P_2}{2^{k/2}\Gamma[k/2]}\right)$$

$$T_{new} = T_{old} - \frac{A * T_{old} + B * \log(T_{old}) + C * \log(1 - T_{old}) + D}{A + (B/T_{old}) - (C/(1 - T_{old}))}$$

- **Mélange de deux distributions Log-normal**

$$Ax^2 + Bx + D = 0$$

$$A = \frac{1}{2}\left(\frac{1}{s_1^2} - \frac{1}{s_2^2}\right) \quad B = \left(\frac{m_2}{s_2^2} - \frac{m_1}{s_1^2}\right) \quad C = \left(\log\frac{P_1 s_2}{P_2 s_1} - \frac{m_1^2}{2s_1^2} + \frac{m_2^2}{2s_2^2}\right)$$

$$T^* = \exp\left(\frac{-B \pm \sqrt{B^2 - 4AC}}{2A}\right)$$

· **Mélange des distributions Log-normal et Exponentiel**

$$Ax + B\log(x) + C(\log(x))^2 + D = 0$$

$$A = l \quad B = \left(\frac{m}{s^2} - 1\right) \quad C = -\frac{1}{2s^2} \quad D = \log\left(\frac{P_1}{s\sqrt{2p}}\right) - \frac{m^2}{2s^2} - \log(P_2\, l)$$

$$T_{new} = T_{old} - \frac{A * T_{old} + B * \log(T_{old}) + C * (\log(T_{old}))^2 + D}{A + (B/(T_{old})) + (2C * \log(T_{old})/T_{old})}$$

· **Mélange des distributions Log-normal et Weibull**

$$Ax^k + B\log(x) + C(\log(x))^2 + D = 0$$

$$A = \frac{1}{l^k} \quad B = \left(\frac{m}{s^2} - k\right) \quad C = -\frac{1}{2s^2} \quad D = \log\left(\frac{P_1}{s\sqrt{2p}}\right) - \log\left(\frac{P_2 k}{l}\right) + (k-1)\log(l) - \frac{m^2}{2s^2}$$

$$T_{new} = T_{old} - \frac{A * T_{old}^k + B * \log(T_{old}) + C * (\log(T_{old}))^2 + D}{A * k * T_{old}^{k-1} + (B/(T_{old})) + (2C * \log(T_{old})/T_{old})}$$

· **Mélange des distributions Log-normal et Rayleigh**

$$Ax^2 + B\log(x) + C(\log(x))^2 + D = 0$$

$$A = \frac{1}{2s_2^2} \quad B = \left(\frac{m}{s_1^2} - 2\right) \quad C = -\frac{1}{2s_1^2} \quad D = \log\left(\frac{P_1}{s_1\sqrt{2p}}\right) - \log\left(\frac{P_2}{s_2^2}\right) - \frac{m^2}{2s_1^2}$$

$$T_{new} = T_{old} - \frac{A * T_{old}^2 + B * \log(T_{old}) + C * (\log(T_{old}))^2 + D}{2A * T_{old} + (B/(T_{old})) + (2C * \log(T_{old})/T_{old})}$$

· **Mélange des distributions Log-normal et Chi2**

$$Ax + B\log(x) + C(\log(x))^2 + D = 0$$

$$A = -\frac{1}{2} \quad B = \left(\frac{m}{s^2} - \frac{k}{2}\right) \quad C = -\frac{1}{2s^2} \quad D = \log\left(\frac{P_1}{s\sqrt{2p}}\right) - \log\left(\frac{P_2}{2^{k/2}\Gamma(k/2)}\right) - \frac{m^2}{2s^2}$$

$$T_{new} = T_{old} - \frac{A * T_{old} + B * \log(T_{old}) + C * (\log(T_{old}))^2 + D}{A + (B/(T_{old})) + (2C * \log(T_{old})/T_{old})}$$

- **Mélange de deux distributions Exponentielle**

$$\log\left(\frac{P_1 l_1}{P_2 l_2}\right)+(l_2-l_1)x=0$$

$$T_{new}=\frac{\log\left(\frac{P_1 l_1}{P_2 l_2}\right)}{l_1-l_2}$$

- **Mélange des distributions Exponentielle et Weibull**

$$Ax^k+Bx+C\log(x)+D=0$$

$$A=\frac{1}{l_2^k}\quad B=l_1\quad C=(1-k)\quad D=\log(P_1 l_1)-\log\left(\frac{P_2 k}{l_2}\right)+(k-1)\log(l_2)$$

$$T_{new}=T_{old}-\frac{A*T_{old}^k+B*T_{old}+C*\log(T_{old})+D}{A*k*T_{old}^{k-1}+B+(C/(T_{old}))}$$

- **Mélange des distributions Exponentiel et Rayleigh**

$$Ax^2+Bx-\log(x)+C=0$$

$$A=\frac{1}{2s^2}\quad B=-l\quad C=\log(P_1 l)-\log\left(\frac{P_2}{s_2^2}\right)$$

$$T_{new}=T_{old}-\frac{A*T_{old}^2+B*T_{old}-\log(T_{old})+C}{2A*T_{old}+B-(1/(T_{old}))}$$

- **Mélange des distributions Exponentiel et Chi2**

$$Ax+B\log(x)+C=0$$

$$A=\left(\frac{1}{2}-l\right)\quad B=\left(1-\frac{k}{2}\right)\quad C=\log(P_1 l)-\log\left(\frac{P_2}{2^{k/2}\Gamma(k/2)}\right)$$

$$T_{new}=T_{old}-\frac{A*T_{old}+B*\log(T_{old})+C}{A+(B/(T_{old}))}$$

- **Mélange de deux distributions Weibull**

$$Ax^{k_1}+Bx^{k_2}+C\log(x)+D=0$$

$$A = -\frac{1}{l_1^{k_1}} \quad B = \frac{1}{l_2^{k_2}} \quad C = (k_1 - k_2) \quad D = \log\left(\frac{P_1 k_1}{P_2 k_2}\right) - k_1 \log(l_1) + k_2 \log(l_2)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^{k_1} + B * T_{old}^{k_2} + C * \log(T_{old}) + D}{A * k_1 * T_{old}^{k_1 - 1} + B * k_2 * T_{old}^{k_2 - 1} + (C/(T_{old}))}$$

· **Mélange des distributions Weibull et Rayleigh**

$$Ax^{k_1} + Bx^2 + C\log(x) + D = 0$$

$$A = -\frac{1}{l^k} \quad B = \frac{1}{2s^2} \quad C = (k-2) \quad D = \log(P_1 k) - k\log(l) - \log\left(\frac{P_1}{s^2}\right)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^{k} + B * T_{old}^{2} + C * \log(T_{old}) + D}{A * k * T_{old}^{k-1} + 2B * T_{old} + (C/(T_{old}))}$$

· **Mélange des distributions Weibull et Chi2**

$$Ax^{k_1} + Bx + C\log(x) + D = 0$$

$$A = -\frac{1}{l^{k_1}} \quad B = \frac{1}{2} \quad C = \left(k_1 - \frac{k_2}{2}\right) \quad D = \log(P_1 k_1) - k_1 \log(l) - \log\left(\frac{P_2}{2^{k/2}\Gamma(k/2)}\right)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^{k} + B * T_{old} + C * \log(T_{old}) + D}{A * k * T_{old}^{k-1} + B + (C/(T_{old}))}$$

· **Mélange de deux distributions Rayleigh**

$$Ax^2 + B = 0$$

$$A = \frac{1}{2}\left(\frac{1}{s_2^2} - \frac{1}{s_1^2}\right) \quad B = \log\left(\frac{P_1 s_2^2}{P_2 s_1^2}\right)$$

$$T_{new} = T_{old} - \frac{A * T_{old}^2 + B}{2A * T_{old}}$$

· **Mélange des distributions Rayleigh et Chi2**

$$Ax^2 + Bx + C\log(x) + D = 0$$

$$A = -\frac{1}{2s^2} \quad B = \frac{1}{2} \quad C = \left(2 - \frac{k}{2}\right) \quad D = \log\left(\frac{P_1}{s^2}\right) - \log\left(\frac{P_2}{2^{k/2}\Gamma(k/2)}\right)$$

$$T_{new} = T_{old} - \frac{A*T_{old}^2 + B*T_{old} + C\log(T_{old}) + D}{2A*T_{old} + B + (C/(T_{old}))}$$

· **Mélange de deux distributions Chi2**

$$A*\log(x) + B = 0$$

$$A = \frac{1}{2}(k_1 - k_2) \qquad B = \log\left(\frac{P_1}{2^{k_1/2}\Gamma(k_1/2)}\right) - \log\left(\frac{P_2}{2^{k_2/2}\Gamma(k_2/2)}\right)$$

$$T_{new} = \exp\left(-\frac{B}{A}\right)$$