**II. Règle bayésienne avec modèle gaussien**

On suppose dans cette section que, conditionnellement à ,𝑥= (,......,) est l’observation d’un vecteur aléatoire gaussien ℵ (, ) ; est un vecteur de ℝ𝑝 et une

matrice (𝑝, 𝑝) symétrique et définie positive. La densité de la loi, au sein de la classe k, s'écrit donc :

=

L'affectation de x à une classe se fait en maximisant 𝑃(𝐺𝑘). par rapport à k soit encore

la quantité :

ln (𝑃 () − ln𝑑𝑒𝑡()− (𝑦−)′ (𝑦−)

Dans les applications, on distingue en fait entre deux modèles d'analyse discriminante selon que l’on suppose que les sont différentes d’un groupe à un autre (modèle hétéroscédastique) ou que ces matrices sont identiques (modèle homoscédastique).

1. **Hétéroscédasticité**

Dans le cas général, il n'y a pas d'hypothèse supplémentaire sur la loi de x et donc les matrices sont en fonction de k. Le critère d’affectation est alors quadratique en x. Les probabilités 𝑃(𝐺𝑘) sont supposées connues mais il est nécessaire d'estimer les moyennes 𝜇𝑘 ainsi que les covariances en maximisant, compte tenu de l'hypothèse de normalité, la vraisemblance. Ceci conduit à estimer la moyenne.

=

Par la moyenne empirique de x dans la classe k pour l'échantillon d'apprentissage et Σ𝑘 par la matrice de covariance empirique :

=

Pour ce même échantillon.

**2. Homoscédasticité**

On suppose dans ce cas que les lois de chaque classe partagent la même structure de covariance =Σ. Supprimant les termes indépendants de k, le critère à maximiser devient :

ln(𝑃 ())− ′𝜇+′𝑥

qui est cette fois linéaire en x. Les moyennes sont estimées comme précédemment tandis que Σ est estimée par la matrice de covariance intra empirique :

=

Si, de plus les probabilités 𝑃(𝐺𝑘) sont égales, après estimation le critère s'écrit :

**−𝑥̅**

**Règle bayésienne avec estimation non paramétrique**

En statistique, une estimation est dite non paramétrable ou fonctionnelle lorsque le nombre de paramètres à considérer est infini. Autrement dit, au lieu d'estimer un ensemble fini de paramètres comme dans les modèles paramétriques (par exemple, la moyenne et l'écart-type d'une loi normale), on cherche ici à estimer une fonction complète. Cette fonction peut être, par exemple, une fonction de régression y=f(x) ou encore la densité de probabilité.

Dans ce cadre, on ne fait pas l'hypothèse que la densité de probabilité correspond à une forme spécifique (comme une loi normale). Au contraire, on cherche à construire une estimation de la fonction de densité f.

L'intérêt principal de cette approche est qu'elle ne repose sur aucune hypothèse forte concernant la distribution des données, à l'exception d'une certaine régularité de f.

Dans le cadre de l'analyse discriminante, ces méthodes permettent d'estimer directement les densités (𝑦). On considère ici deux approches : la méthode du noyau et celle des k plus proches voisins.

**Méthode de noyau**

**Estimation de densité**

Soit ,……..., n observation équipondérées d'une v.a.r. continue x de densité f inconnue.

Soit K(x) (le noyau) une densité de probabilité unidimensionnelle (sans f) et h un réel strictement positif. On appelle estimation de f par la méthode du noyau la fonction

(𝑥)=

Il est immédiat de vérifier que :

∀ k ϵℝ, (𝑥)≥0 𝑒𝑡

h est appelé largeur de fenêtre ou paramètre de lissage ; plus h est grand, l'estimation est régulière. Le noyau K est choisi centré en 0, unimodal et symétrique. Les cas les plus usuels sont la densité gaussienne, celle uniforme sur [-1 ; 1] ou triangulaire : 𝐾(𝑥)=[1−|𝑥|]𝕝[−1 ;1](𝑥). La forme de noyau n'est pas très déterminante sur la qualité de l'estimation contrairement à la valeur de h.

K plus proches voisins

On cherche les k points plus proche de nouvel individu x au sens d'une métrique à préciser, et on classe x dans le groupe (classe) plus représenté : la probabilité à posteriori s'obtient comme par la discrimination par boule mais n'a pas un grand sens si k est faible.

Cette méthode consiste à enchainer les étapes suivantes :

i. Choix d'un entier 𝑘 : 𝑛≥𝑘≥1

ii. calculer les distances (𝑥, ),i= 1,2,3,......,n où M est la métrique de Mahalanobis c'est-à-dire la matrice inverse de la matrice de covariance (ou de variance-intra)

retenir les k observations (,......,) pour lesquelles ces distances sont les plus petites

iv. compter les nombres de fois (,......,) que ces k observations apparaissent dans des classes

v. estimer les densités par

(𝑥)=

Où (𝑥) est le volume de l'ellipsoïde {𝑧;(𝑧−𝑦) ′𝑀(𝑧−𝑦) = (𝑥, 𝑥(𝑘)}.

Remarque : Pour k=1, x est affecté à la classe du plus proche élément.