

MÉTHODE NON PARAMÉTRIQUE (B à N)

(13 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **méthode non paramétrique** est une méthode statistique adaptée à des modèles non explicitement paramétrés, ou dépendant d'un nombre infini de « paramètres ».

(i) La **Statistique non paramétrique** a pour objet l'étude de tout **modèle statistique** qui n'est pas un **modèle paramétrique** (cf aussi **paramétrique**) : le modèle en question est alors appelé **modèle non paramétrique**, et toute méthode adaptée à son étude est appelée « méthode non paramétrique ».

Un modèle non paramétrique se note généralement $(\Omega, \mathcal{T}, \mathcal{P})$, où $\mathcal{P}$ désigne une **famille** de **mesures de probabilité** définies sur $\mathcal{T}$ mais ne dépendant pas d'un nombre fini de paramètres scalaires.

On ne peut donc pas indexer $\mathcal{P}$ à l'aide d'un **paramètre** tq $\theta \in \Theta$, avec $\Theta \subset V$ (**espace vectoriel** de dimension finie Q) ou encore $\Theta \triangleleft V$ (sous-espace de V).

$\mathcal{P}$ est toujours indexable par elle-même, ie $\mathcal{P} = (P)_{P \in \mathcal{P}}$. De même, le modèle fondamental initial, celui associé à l'**espace fondamental** initial $(\Omega, \mathcal{T})$, est indexable selon une famille d'**espaces probabilisés** $(\Omega, \mathcal{T}, P)_{P \in \mathcal{P}}$ (cf infra).

(ii) Néanmoins, $\mathcal{P}$ doit généralement faire l'objet d'autres hypothèses. En effet, il est toujours nécessaire d'admettre des **conditions de régularité** (à caractère mathématique), souvent assez larges : eg absolue continuité des **lois de probabilité** (cf **loi absolument continue**), **dérivabilité** des **densités de probabilité** ou des **fonctions de répartition**, etc.

Cette nécessité conduit souvent à assimiler **méthode affranchie** et **méthode non paramétrique**.

(iii) Si $(\mathcal{X}, \mathcal{B})$ est un **espace d'observation** et $X : \Omega \mapsto \mathcal{X}$ une **va** donnée (**échantillon** ou **statistique**), on définit (comme pour un modèle paramétrique) un **modèle image** de $(\Omega, \mathcal{T}, \mathcal{P})$ par X . Ce modèle peut se noter $(\mathcal{X}, \mathcal{B}, \mathcal{P}^X)$, avec :

$$(1) \quad \mathcal{P}^X = \{P^X = X(P), \forall P \in \mathcal{P}\}$$

(famille des **lp**, ie des **mesures images** de P par X).

Ce modèle (1) est aussi un modèle non paramétrique, donc relève encore des méthodes non paramétriques (cf **décision statistique**).

(ii) Les méthodes non paramétriques se sont développées à partir des outils de base suivants :

(a) **analyse combinatoire** : celle-ci, associée aux notions de **statistique d'ordre** et de **statistique de rang**, permet de définir de nombreux **tests non paramétriques** spécifiques ;

(b) **robustesse** statistique. Ce concept sert à étudier les « réactions » d'une **procédure statistique** (paramétrique aussi bien que non paramétrique) à des défauts d'**hypothèses statistiques** courantes, ie lorsque ces hypothèses ne sont pas vérifiées : eg **normalité** ou **symétrie** ou unimodalité des lois (cf **loi unimodale**), **indépendance** des **observations**, existence de **moments**, etc ;

(c) **influence statistique**, ie étude de l'effet des variations de certaines **observations** sur les statistiques utilisées, donc sur l'**optimalité** des procédures qui en font usage (cf **aberration**, **courbe d'influence**) ;

(d) **estimation** de **densités** ou de **fr** (cf aussi **reconnaissance des formes**) ;

(e) test d'**hypothèses statistiques**. Ainsi, le **test d'adéquation** d'une **loi** à des observations a pour objectif de tester l'hypothèse selon laquelle un jeu d'observations donné a été généré par la loi en question.

(iii) En pratique, la **normalité asymptotique** des **statistiques** utilisées est fréquente, même lorsque la normalité du modèle n'est pas vérifiée ni même supposée a priori, notamment à distance finie (ie dans des échantillons de « petite » taille $N \ll +\infty$).

Ceci explique l'importance de la **loi des grands nombres** et du **théorème de la limite centrale** dans l'étude des propriétés statistiques des méthodes non paramétriques.

(iv) Ce sont des raisons historiques qui expliquent que les méthodes non paramétriques sont distinguées des méthodes paramétriques :

(a) d'abord leur développement respectif s'est déroulé de façon relativement indépendante ;

(b) d'un point de vue « technique », l'**optimisation** de fonctions dépendant d'un nombre fini de « variables » (eg une **vraisemblance** comportant un nombre fini de **paramètres**), avec ou sans **contraintes**, a longtemps été la seule méthode mathématique connue ou pratiquée (cf **programmation mathématique**).

L'**optimisation** de fonctions dépendant d'un nombre « infini » de « variables » (ou dépendant d'autres fonctions) (calcul « fonctionnel ») s'est développé relativement plus tard. La Statistique non paramétrique a dû, entre-temps, mettre en oeuvre des procédures spécifiques.

Le développement de la théorie générale de l'optimisation (**calcul des variations**, etc) a permis de progresser dans le traitement de problèmes non paramétriques : ainsi, la **méthode du maximum de vraisemblance**, longtemps présentée dans le seul cadre paramétrique, a été étendue au cadre non paramétrique (cf **méthode des cribles**).

(v) Les développements de la **théorie de la décision** statistique ont tendu à rapprocher, voire unifier, les deux types de méthodes. Ainsi, la **théorie de la robustesse** a bénéficié des apports du calcul fonctionnel (**dérivée de FRÉCHET**, étude de **fonctionnelles** diverses, etc).