

EXTRAPOLATION (H3)

(12 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

Un **ensemble** de **données** constitué d'**observations** (numériques) peut être considéré comme insuffisant et nécessiter d'être complété en dehors du **champ des valeurs observées** : eg en dehors de l'**enveloppe convexe** du **nuage de points** décrivant ces observations.

Le but d'une **extrapolation** est d'anticiper le « comportement » du **jeu de données** disponible au-delà du champ des valeurs « couvertes » par ces données.

Moyennant des hypothèses adaptées, une extrapolation vise à permettre cette complétion (cf aussi **prévision**, **problème de prévision**) (comparer avec l'**interpolation**).

Ce problème intervient dans de nombreuses **situations statistiques**. Il peut être associé à des approches de type individuel, spatial ou temporel.

(i) Données individuelles (en **coupe instantanée**) ou spatiales. La méthode d'extrapolation se fonde généralement :

(a) sur la mise en évidence de relations entre variables observées (cf eg **relation fonctionnelle**), ce qui nécessite généralement de distinguer parmi elles des **variables endogènes** et des **variables exogènes** ;

(b) sur des méthodes d'**estimation** de ces relations ;

(c) enfin sur des hypothèses relatives aux valeurs prises par les exogènes afin d'en déduire des valeurs probables pour les endogènes (**variables d'intérêt**).

(ii) Données temporelles (cf **processus stochastique**, **série temporelle**). Le problème peut se fonder sur deux types d'approches :

(a) l'**approche « autogène »** ou **approche « endogène »** : la liste constituée de l'ensemble des variables est supposée dotée d'une **structure** interne résultant de son « passé » (phénomène de mémoire). Ceci permet la **modélisation** d'un processus dont on peut, après estimation des paramètres, poursuivre la **trajectoire** au-delà de la période d'observation. En particulier, lorsque toutes les valeurs sont numériques, on définit un **processus vectoriel** dont on peut estimer les paramètres (cf eg **modèle autorégressif**, **processus ar**, **processus à retards échelonnés**, **modèle à retards échelonnés**) ;

(b) l'**approche exogène** ou **approche mixte** : elle se fonde, comme dans le cas spatial précédent, sur le partitionnement des variables entre endogènes et exogènes, et sur la **modélisation** d'un modèle structurant leurs relations (eg **processus autorégressif de moyenne mobile**, **modèle autorégressif de moyenne mobile**, etc). Des hypothèses relatives au « cheminement » des exogènes en dehors de la période observée sont alors nécessaires : elles permettent alors d'en déduire la **trajectoire** ultérieure des endogènes (variables d'intérêt). Chaque hypothèse de cheminement est souvent appelée « **variante** » de l'extrapolation. Cette dernière est donc conditionnelle à ces hypothèses.