

PROBLÈME DE PRÉVISION (G10, H6, J9, N6)

(24 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) La **prévision** constitue souvent l'une des étapes ultimes du travail statistique portant sur un **phénomène** étudié :

(a) cette étape suit généralement d'autres étapes plus « classiques », relatives à un phénomène donné, eg :

(a)₁ discernement et délimitation du phénomène : contenu et frontières (**espace, temps**), grandeurs qualificatives ou descriptives (**variables**) ;

(a)₂ conception du mode d'**observation** du phénomène : mise en oeuvre d'informations préexistantes, « opérations pilotes » ;

(a)₃ collecte des données (**production statistique**), par **expérimentation** ou par **sondage** ;

(a)₄ description du phénomène (**Statistique descriptive**) et examen des divers problèmes associés : eg **classifications** (nomenclatures, typologies, etc) ;

(a)₅ analyse des résultats, formalisation et **modélisation** du phénomène ;

(a)₆ **estimation** ou **test d'hypothèses** relatifs à un **paramètre d'intérêt** pertinent du **modèle** retenu, etc ;

(a)₇ **validation** de la théorie relative au phénomène, du fonctionnement et de l'évolution (passée, présente et future) du système étudié.

La prévision intègre donc généralement une grande partie des phases précédentes.

(b) elle est elle-même suivie d'autres étapes. En effet, la prévision précède aussi diverses situations en rapport avec le phénomène (cf aussi **niveau, répartition, évolution**) :

(b)₁ anticipation de **dysfonctionnements** possibles du système décrivant le phénomène, au vu du modèle retenu (ie de la théorie validée) ;

(b)₂ **correction** de ces inconvénients ;

(b)₃ **conseils** d'utilisation de la prévision : cette fonction de conseil porte sur le contenu des **décisions** à prendre pr à l'avenir, ainsi que des **actions** suivies par le décideur et l'anticipation des **résultats** de ces actions.

(ii) On distingue souvent entre :

(a) **problème d'estimation**, dans lequel on cherche à évaluer un **paramètre principal** donné, à partir des observations effectuées sur le phénomène ;

(b) **problème de prévision**, dans lequel on cherche à évaluer une **variable aléatoire**, ou l'une de ses **caractéristiques** de **centralité**, donnée lorsque ses valeurs sortent du « champ d'**observation** ».

Cette distinction n'est pas absolue. On peut lui préférer celle entre :

(a) **prévision « inconditionnelle »** (ou « absolue ») : c'est le cas d'une **projection** ou d'une **extrapolation** ;

(b) **prévision « conditionnelle »** (ou « relative ») : en effet, les grandeurs à anticiper dépendent, le plus souvent, d'autres grandeurs, ou grandeurs « exogènes », considérées comme « extérieures » au modèle : **variables de contrôle, variables d'action, variables de simulation, instruments d'action**.

(iii) De façon générale, la « **meilleure** » **prévision** possible relativement à une va (resp d'une **statistique**), ou relativement à l'évolution d'un système représentant un phénomène, est la connaissance (ie estimation) de sa **loi de probabilité** elle-même : en effet, celle-ci est censée régir le « comportement » aléatoire de cette va (resp statistique) et les propriétés du modèle validé. Sa connaissance suffit donc à caractériser entièrement cette va (resp statistique).

En pratique, on se contente d'estimer cette loi ou, faute d'observations suffisamment nombreuses, d'estimer seulement une caractéristique ou un paramètre d'intérêt de cette loi. Ceci est vrai aussi bien pour :

(a) une prévision inconditionnelle : eg à partir d'une **série temporelle** unique (éventuellement vectorielle) ;

(b) que pour une prévision conditionnelle : eg lorsque la caractéristique légale ou le paramètre dépendent d'un certain nombre de **variables exogènes**. Ceci conduit à la notion générale de **prévision dans un modèle statistique**.

(iv) Une prévision peut se réaliser aussi bien en utilisant :

(a) une **série temporelle** (cf **prévision d'un processus**). Forger une « opinion » sur le « futur », ou simplement « dire des choses » à son sujet, est la situation la plus courante ;

(a) une **série spatiale** ou une **coupe instantanée** (eg inter-individuelle). Dans ce cas, on parle parfois d'**intra-prévision**, d'intrapolation (ou d'**interpolation**).

(v) D'un point de vue terminologique, on distingue souvent entre :

(a) **prévision** et « **prédiction** » (au sens « ordinaire » et non pas au sens anglo-saxon). La première est censée traduire un jugement sur le « futur » fondé sur des variables « pertinentes » (ie en rapport avec le phénomène étudié) et sur une **représentation statistique** « éprouvée » à l'aide de (ou « validée » par des) **tests** ou d'études de **robustesse**. La seconde est censée traduire un jugement sur le futur (en général, celui d'un seul individu) fondé sur des « informations » présentées comme pertinentes (jeu de cartes ou, plus généralement, tirages aléatoires, ou encore « boule de cristal », etc) mais sans modèle explicite : or, dans cette situation, la relation de cause à effet n'est pas vérifiable, la **répétition** expérimentale est impossible, etc. D'autre part, une « voyance » ne peut se confondre avec une approche bayésienne. En effet, ce type de « vision », à caractère discrétionnaire, pourrait être rapprochée d'une « opinion » tq celle émise par un « bayésien pur » qui ferait abstraction totale des observations : or, un statisticien bayésien émet généralement une opinion sur la base d'études similaires, soit antérieures, soit contemporaines ;

(b) **prévision et projection**. Une **projection** consiste, en général, à définir un modèle très simple (voire fruste), décrivant le comportement de variables censées se perpétuer dans le futur : ou encore évolution reproduisant des **structures** repérées sur le passé. Deux types de projections peuvent être distingués :

(b)₁ celui dans lequel le modèle estime des valeurs endogènes en fonction de valeurs exogènes définies a priori : eg trajectoires supposées ou admises de variables exogènes : **simulations**, etc ;

(b)₂ celui du **processus autorégressif** ou du **modèle autorégressif**, qui peuvent être multivariés (eg processus autorégressif vectoriel) ;

(c) **prévision et extrapolation**. On entend généralement par **extrapolation** une technique de calcul numérique (pouvant d'ailleurs être mise en oeuvre dans un **problème statistique**), qui procède d'un esprit voisin de celui d'une **projection**.

Une extrapolation décrit, le plus souvent, une évolution « tendancielle », souvent sans variables « explicatives » particulières (autres que le « temps », variable « omnibus »). Elle n'est cependant pas toujours fondée (cf eg **plan d'expérience**).

Cette acception de la notion de prévision englobe, en fait, celles de prédiction (au sens technique), de projection ou d'extrapolation.