PRÉVISION D'UN PROCESSUS (N6)
(09 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **théorie des processus** est un cadre adapté à l'étude d'une **procédure** d'**inférence statistique** importante : la **prévision** (cf aussi **prédiction**). La **prévision d'un processus** peut ainsi être mise en oeuvre avec un **modèle dynamique**.

Soit $X = \{(\Omega, \mathscr{T}, P), (\mathscr{X}, \mathscr{B}), (X_t)_{t \in T}\}$ un **processus stochastique**. On suppose que $P \in \mathscr{P}$ (**famille** des **probabilités** susceptibles de gouverner le comportement de X) et l'on note $P^X = X(P)$ l'une des **lois** possibles de X, laquelle doit vérifier des conditions de compatibilité (cf **système projectif de probabilités**). Par suite, on note $\mathscr{P}^X$ la famille des lois P^X, ce qui définit un **modèle de processus**.

(i) La prévision de X consiste principalement en deux types d'analyses, en quelque sorte « duales » l'une de l'autre, selon que les états ou les instants sont donnés :

(a) soit étudier l'**état** (resp l'**ensemble des états**) dans le(s)quel(s) X sera à un instant donné $u \in T$ (resp pendant un ensemble d'instants donné $U \subset T$, ou **période**). Il s'agit alors d'étudier la loi de la **va** X_u (resp du processus « restreint » $X_U = (X_u)_{u \in U}$) et, notamment, d'en estimer une **caractéristique** de **centralité** et de **dispersion** (**estimateur ponctuel** ou **estimateur ensembliste**). C'est l'**approche « espace - temps »**.

Lorsqu'on cherche à prévoir X_u (resp X_U considéré comme élément de $\mathscr{B}^{\otimes U}$), on parle de **prévision ponctuelle** ; lorsqu'on cherche à prévoir X_u (resp X_U) à l'aide de parties aléatoires $B \in \mathscr{B}$ (resp $B_U \in \mathscr{B}^{\otimes U}$), on parle de **prévision ensembliste** (approche et terminologie comparables à celles de la **théorie de l'estimation** : cf **estimateur ponctuel**, **estimateur ensembliste**).

Cette situation correspond à l'étude d'un **phénomène** décrit par un **système aléatoire** (physique, biologique, écologique, psychologique ou sociologique) évoluant au cours du **temps** et représentable par un **processus** X. Il s'agit eg de prévoir en quel point (resp en quelle portion) X_u (resp X_U) de la trajectoire de X se trouvera le système considéré à une date donnée (resp pendant une période de temps donnée).

Ce type de contextes est notamment adapté au cas d'un **système « gouvernable »** (au moins en partie), ie au cas où il existe des **variables de contrôle**, ou **variables de commande**, permettant d'agir sur les trajectoires de X (cf eg **contrôle optimal**) et où le point X_u (resp la portion X_U) doit être atteint(e) ou, au contraire, évité(e).

(b) soit étudier l'**instant** $u \in T$ (resp l'**ensemble des instants** $U \subset T$) pendant le(s)quel(s) le processus X sera dans un état donné $B \in \mathcal{B}$. L'instant u (resp l'ensemble U) est donc aléatoire et c'est l'étude de sa **loi** (notamment son **espérance** ou son **mode**, ainsi que sa **dispersion**) qui est l'objet de la prévision. C'est l'**approche « temps - espace »**.

Cette seconde situation :

(a) concerne souvent le cas où l'état donné $B \in \mathcal{B}$ du système est eg une **région « absorbante »** (ou, au contraire, une **région « réfléchissante »**) (cf **barrière absorbante**, **principe de réfléchissement**) ;

(b) ou encore correspond à une **catastrophe**.

A la différence de la situation précédente, il s'agit ici, non pas de **prévoir un événement** $B \in \mathcal{B}$, mais de **prévoir une date** u à laquelle (resp **prévoir l'ensemble des dates** U auxquelles) l'événement B (resp B_U) se produira : **instant d'arrivée** dans une file d'attente (cf **théorie des files d'attente**), **temps d'entrée** dans un état B pour une **promenade aléatoire**, **temps de séjour** d'un processus dans une région donnée, dates d'une **catastrophe**, etc.

(ii) La mise en oeuvre des méthodes de prévision nécessite l'**observation** d'une (portion de) **trajectoire** $x = (x_s)_{s \in S}$ de X (avec $S \subset T$ et $x_s = X_s(\omega)$, $\forall s \in S$).

Dans les **contextes statistiques** les plus favorables, on peut observer plusieurs trajectoires $x_n = (x_{ns})_{s \in S}$ (avec $n \in N_N^*$) de X. L'**inférence statistique** (prévision) est alors fondée sur cette **information** (cf aussi **prévision dans un modèle statistique**).

Dans certains cas, il est encore possible d'obtenir une « **amélioration** » de la prévision de X grâce à l'observation d'un autre processus, eg $Z = \{(\Omega, \mathcal{T}, P), (\mathcal{Z}, \mathcal{D}), (Z_t)_{t \in T}\}$, supposé « expliquer » X (ou du moins être très lié à X) : ceci est le cas eg lorsque X est en **régression** avec Z (ie lorsqu'il existe une fonction $\phi : \mathcal{Z} \mapsto \mathcal{X}$ tq $E(X_t / Z_t) = \phi(Z_t)$, $\forall t \in T$), ou encore s'il existe des **corrélations croisées** entre X et Z, etc.

(iii) Pour résoudre un problème de prévision avec un modèle de processus, plusieurs approches, souvent combinées, sont mises en oeuvre :

(a) l'**analyse dans l'espace des « états »**, ie dans $\mathscr{X}$ ou dans $\mathscr{B}$ (cf eg **prévision des moindres carrés d'un processus**). On met alors souvent en oeuvre un **modèle dynamique** : **modèle de régression** ou **modèle d'interdépendance** (eg **modèle d'interdépendance dynamique**). Certaines méthodes (eg les **méthodes de BOX-JENKINS**) mettent aussi en oeuvre diverses classes de modèles dynamiques : modèles associés à un **processus autorégressif**, à un **processus autorégressif de moyenne mobile**, à un **processus de moyenne mobile**, etc ;

(b) l'**analyse dans l'espace des « temps »** (parfois improprement appelé **espace des « paramètres »**), ie dans T ou dans $\mathscr{B}_T$ (tribu de parties de T tq U). Cette approche correspond surtout aux prévisions du second type décrit précédemment ;

(c) l'**analyse dans l'espace des « fréquences »** (cf eg **analyse spectrale**, **analyse harmonique**, **analyse cospectrale**), dans laquelle on décompose les variables X_t en fonction de **composantes périodiques** (parfois appelées « **périodicités cachées** »). Ces composantes sont ensuite estimées et l'on ne retient, pour la prévision, que les plus « importantes » d'entre elles : eg celles qui contribuent à une part de la **dispersion** de X jugée suffisante.