7 Séries chronologiques et prévision.

7.1 Introduction et exemples.

Face à la complexité croissante des organisations et de leur environnement, les gestionnaires cherchent à améliorer la qualité de l'information et des décisions qui en résultent. C'est dans ce contexte que les méthodes d'analyse et de prévision des séries chronologiques se sont développées depuis 20 ans.
Le pêcheur de crabe des neiges de la Gaspésie (région de la province de la province du Québec située près de l'estuaire et le golfe du Saint-Laurent) prépare le bateau et les casiers pour la saison prochaine. Que sera la campagne de pêche? Quels seront les prix? Ce scénario se répète depuis un plus de quinze ans, date à laquelle débute la pêche au crabe mais aussi pour un grand nombre d'espèces capturées dans le golfe du Saint-Laurent. Les questions posées par les pêcheurs ne sont pas non plus exclusives à cette branche industrielle. Dans une structure de marché en concurrence, les prix ne sont pas fixés à l'avance . Il importe aux gestionnaires d'essayer d'en prédire le niveau futur. De même les quantités qu'on pourra vendre d'un produit quelconque dans les douze prochains mois nous sont inconnues. Mais sont-elles prévisibles ?
Le pêcheur qui veut acheter un crabier ou l'entrepreneur désireux de transformer le crabe en produits de consommation voudront planifier l'investissement sur la base de perspective de marché. Dans un sens plus large, la planification exige d'intégrer l'information provenant de l'extérieur de l'entreprise et celle de l'intérieur. Si le preneur de décision peut controler les évènements internes à l'organisation, il ne peut le faire pour les évènements externes. C'est donc à ce niveau que la prévision prend tout son sens .
(cf. M. David J.c. Michaud, "la prévision")

La prévision fait partie de l'étude des séries chronologiques. Ce sont des suites d'observations $(x_t, t \in T)$ indexées par un ensemble ordonné T (dans ce cours, on considère des données discrètes et on suppose donc que T est un sous ensemble de $\mathbb{N}$). Les méthodes statistiques usuelles ne peuvent pas s'appliquer à ce genre de données. En effet, dans des données classiques, l'ordre des observations n'a pas de signification particulière.

Quelques exemples.

1. Ventes trimestrielles en quantité d'une entreprise

	1990	1991	1992
Tr1	860	1096	1436
Tr2	794	1021	1363
Tr3	1338	1705	2319
Tr4	1148	1505	2047

Quand on regarde cette suite de nombres, on ne peut pas en tirer grand chose, par contre toute étude des séries chronologiques commence par une représentation graphique.

Représentation graphique(1).

En observant ce graphique, on peut remarquer deux choses:
1. Les ventes de cette entreprise ont connu sur l'ensemble des 3 années un mouvement général de hausse.
2.Les fluctuations présentent d'une année sur l'autre un caractère régulier, au sens ou le deuxième et quatrième trimestre marquent toujours un recul. Ce sont des fluctuations saisonnières que l'on retrouve dans la plupart des activités économiques.
3. L'amplitude des oscillations augmentent avec la tendance au cours des 3 années.

2. Collecte mensuelle en hectolitres d'une coopérative agricole de juillet 1976 à juin 1981.

	1976	1977	1978	1979	1980	1981
Janvier		1543	1465	1605	1633	1606
.Février		1469	1452	1611	1619	1567
Mars		1780	1806	1970	2109	1948
Avril		2312	2134	2446	2466	2275
Mai		2650	2673	2811	2734	2756
Juin		2502	2553	2440	2524	2566
Juillet	2523	2410	2405	1905	2336	
Aout	2398	2376	2041	1680	2146	
Septembre	2040	2037	1837	1422	1831	
Octobre	1844	1820	1771	1681	1703	
Novembre	1622	1433	1620	1645	1493	
Décembre	1486	1375	1548	1608	1516	

On peut noter les valeurs suspectes de mai et septembre 79.

Représentation graphique (2).

3. Nombre de passagers mensuels par milliers sur les lignes internationales de janvier

   1949

La première étape de l'étude d'une série chronologique est la représentation graphique. Cette visualisation donne des informations précieuses pour choisir un modèle (croissance, oscillation, rupture,...)

7.2 Eléments constitutifs d'une série chronologique

Dans toute la suite on suppose que les intervalles entre deux observations successives sont de même longueur (par exemple les mois, les trimestres, les heures etc...)
Pour modéliser une série chronologiques, on essaye d'y repérer trois types d'effets: la tendance, le mouvement saisonnier, les irrégularités et perturbations.

A. La tendance.
La tendance à long terme ou Trend notée T_t en général est le facteur représentant l'évolution à long terme de la grandeur et traduit l'aspect général de la série: croissance

de la consommation d'électricité, croissance du trafic aérien, diminution de la population rurale, etc...par exemple.

B. Le mouvement saisonnier.
Le facteur saisonnier noté S_t en général est un mouvement périodique que l'on voit apparaître sur le graphique.

C. Les irrégularités.
Cette composante, appelée aussi mouvement résiduel et notée E_t regroupe tout ce qui reste et résulte de fluctuations aléatoires et imprévisibles. Ces fluctuations sont supposées de faible amplitude et de moyenne nulle sur un petit nombre d'observations consécutives. Dans ce cours, nous ne les modéliserons pas.

D. Les perturbations.
Ce sont des fluctuations ponctuelles de forte amplitude. Elles sont dues par exemple à une grève, à des conditions météo exceptionnelles pour l'agriculture, à un krach financier... Il convient de les éliminer avant tout traitement de la série. Pour faire comme si ces évènements n'avaient pas eu lieu, on enlève les observations que l'on remplace par des interpolations.

E. Les modèles.
On distingue 3 types de modèles:
1. le modèle additif $x_t = T_t + S_t + E_t$
2. le modèle multiplicatif $x_t = T_t * S_t * E_t$ ou $x_t = T_t * S_t + E_t$

On choisit un modèle multiplicatif si le mouvement saisonnier présente des amplitudes proportionnelles à la tendance. On peut noter que la transformation logarithmique du modèle multiplicatif se ramène au modèle additif. Dans les exemples précédents, on choisira un modèle multiplicatif pour le nombre de passagers sur les lignes aériennes, mais on choisira un modèle additif pour la production de lait.

7.3 Moyenne mobile.

Dans ce cours, nous allons présenter la méthode de la moyenne mobile. C'est une méthode dite de "filtrage", qui consiste tout d'abord à éliminer les facteurs saisonniers et les perturbations aléatoires afin de faire apparaître la composante tendancielle.

Définition 7.1 *On appelle moyenne mobile centrée de longueur p de la série x_t , la filtrée de x_t notée $M_p(t)$.*

$$\text{Si } p \text{ est impair } p = 2m+1, \quad M_p(t) = \frac{1}{p}\sum_{k=-m}^{m} x_{t+k}$$

$$Si\ p\ est\ pair\ p = 2m, \quad M_p(t) = \frac{1}{p}\left(\sum_{k=-m+1}^{m-1} x_{t+k} + \frac{x_{t-m}}{2} + \frac{x_{t+m}}{2}\right)$$

Quelques propriétés sur les moyennes mobiles:

1. Si une série est périodique de période p, alors toute suite de moyenne mobile de période p' différente de p a pour période p.

2. Si $p' = p$ les moyennes mobiles sont constantes égales à la moyenne des termes sur une période.

3. La moyenne mobile centrée transforme une série alignée en elle même. Ce filtrage est donc particulièrement bien adapté aux modèles linéaires. Plus généralement une série monotone à faible courbure en une série peu différente.

4. La moyenne mobile centrée transforme des écarts dus à des irrégularités (indépendantes de moyenne nulle et de même variance) en écarts de variance plus faible ; on dit qu'elle a un effet rabot ou qu'elle lisse la série.

En résumé, si la période du mouvement saisonnier est égale à p, alors la moyenne mobile centrée de longueur p est un filtre linéaire qui élimine le mouvement saisonnier tout en réduisant l'amplitude du mouvement résiduel. Sa valeur peut-être assimilée à la valeur de la tendance.

Quelques inconvénients de la méthode.

1. Un changement de niveau ou de pente de la tendance entraine une mauvaise approximation pendant toute une période précédant et suivant cette date. c'est pourquoi on fait l'hypothèse d'une tendance monotone à faible courbure.

2. On ne dispose pas d'information sur les $\frac{p}{2}$ dernières dates.

Malgré ces inconvénients, on admettra que dans la plupart des cas, la valeur de la tendance s'approche par la moyenne mobile centrée de longueur égale à la période du mouvement saisonnier.

Reprenons l'exemple 1 des Ventes trimestrielles en quantité d'une entreprise

	1990	1991	1992
Tr1	860	1096	1436
Tr2	794	1021	1363
Tr3	1338	1705	2319
Tr4	1148	1505	2047

Calculons les moyennes mobiles associées.

	observations	calcul	moyenne mobile
Tr1	860		
Tr2	794		
Tr3	1338	$\frac{(1/2*860+794+1338+1148+1/2*1096)}{4}$	1064
Tr4	1148	$\frac{(1/2*794+1338+1148+1096+1/2*1021)}{4}$	1122
Tr5	1096	$\frac{(1/2*1338+1148+1096+1021+1/2*1705)}{4}$	1197
Tr6	1021	$\frac{(1/2*1148+1096+1021+1705+1/2*1505)}{4}$	1287
Tr7	1705	$\frac{(1/2*1096+1021+1705+1505+1/2*1436)}{4}$	1374
Tr8	1505	$\frac{(1/2*1021+1705+1505+1436+1/2*1363)}{4}$	1459
Tr9	1436	$\frac{(1/2*1705+1505+1436+1363+1/2*2319)}{4}$	1579
tr10	1363	$\frac{(1/2*1505+1436+1363+2319+1/2*2047)}{4}$	1724
Tr11	2319		
Tr12	2047		

Graphique de la moyenne mobile.

Calcul des coefficients saisonniers. Dans toute la suite, on note p la période du mouvement saisonnier et n le nombre de période observée. On a donc $n * p$ observations et donc de $(n-1)*p$ moyennes mobiles. La méthode dépend du type de modèle (additif ou multiplicatif).

1. Modèle multiplicatif
$$x_t = T_t * S_t * (1 + E_t)$$
Si on assimile la tendance à la moyenne mobile, on calculera tout d'abord les $\frac{x_t}{M_p(t)}$ pour approcher les coefficients saisonniers. C'est l'occasion de vérifier qu'il existe bien une partie périodique, auquel cas on affine l'étude.
On peut par exemple considérer les moyennes sur les $n-1$ coefficients obtenus : on obtient
$$S_r = \frac{\sum_{i=1}^{n-1} \frac{x_{i*p+r}}{M_p(i*p+r)}}{n-1} \quad \forall r \in [1 \cdots p].$$

Définition 7.2 *Dans le cas du modèle multiplicatif, on appelle série corrigée des variations saisonnières le rapport:*
$$x_t^* = \frac{x_t}{S_t}.$$

Pour déterminer la tendance, soit on fait un ajustement de la moyenne mobile si celle ci est suffisamment lisible, soit on travaille sur x_t^* qui est en général plus régulière.

Reprenons notre exemple (1):

	observations	coef	moyenne mobile
Tr1	860		
Tr2	794		
Tr3	1338	1.26	1064
Tr4	1148	1.02	1122
Tr5	1096	0.92	1197
Tr6	1021	0.79	1287
Tr7	1705	1.24	1374
Tr8	1505	1.03	1459
Tr9	1436	0.91	1579
tr10	1363	0.79	1724
Tr11	2319		
Tr12	2047		

Si pour un même trimestre, on obtient sur les différentes années des valeurs voisines pour les coefficients calculés, le caractère saisonnier des fluctuations est bien établi.

On retient alors pour valeur du coefficient saisonnier la moyenne des coefficients calculés pour ce trimestre. Soit ici:

$$\frac{1}{2}(1.26 + 1.24) = 1.25 \text{ pour le troisième trimestre;}$$

$$\frac{1}{2}(1.02 + 1.03) = 1.025 \text{ pour le quatrième trimestre;}$$

$$\frac{1}{2}(0.92 + 0.91) = 0.915 \text{ pour le premier trimestre;}$$

$$\frac{1}{2}(0.79 + 0.79) = 0.79 \text{ pour le second trimestre;}$$

Raffinement: la somme étant égale à 3.98, on corrige ces valeurs afin d'obtenir une somme égale à 4 en multipliant par exemple par le rapport $\frac{4}{3.98}$.

On obtient finalement: le coefficients saisonniers

$$S_1 = 0.92, \quad S_2 = 0.8, \quad S_3 = 1.256, \quad S_4 = 1.03$$

et la série corrigé des variations saisonnières

$$x_1^* = 934.78, \quad x_2^* = 992.5, \quad x_3^* = 1065.28, \quad x_4^* = 1114.56, \quad x_5^* = 1191.3,$$

$$x_6^* = 1276.25, \quad x_7^* = 1357.48, \quad x*_8 = 1461.16, \quad x_9^* = 1560.87,$$

$$x_{10}^* = 1703.75, \quad x_{11}^* = 1846.33, \quad x_{12}^* = 1987.37$$

Pour trouver ensuite un modèle et faire de la prévision par exemple, on utilise la série désaisonnalisée et on cherche un ajustement linéaire par des méthodes de régression,puis on applique à ce modèle les coefficients saisonniers. On obtient ainsi une fonction $f(t)$ qui nous permet d'extrapoler notre courbe dans le futur. Dans notre exemple, $x_t^* = -746564.14 + 375.59 * t$ et donc

$$f(t) = (-746564.14 + 375.59 * t) * S_t.$$

Au premier trimestre 1993, on prévoira

$$(-746564.14 + 375.59 * 1993.0) * S_1 = 1827.792$$

Le modèle multiplicatif prédit ainsi les valeurs $x_t^* * S_t$ ou $f(t) * S_t$.
En résumé, pour avoir le modèle, on doit calculer les moyennes mobiles, puis en déduire les coefficients saisonniers standardisé et enfin la série désaisonnalisé et la tendance et le modèle.

2. Modèle additif

$$x_t = T_t + S_t + E_t.$$

On commence par chercher les coefficients: $S_t = x_t - M_p(t)$. On vérifie qu'ils sont bien périodiques et on affine en faisant la moyenne sur les $n-1$ périodes considérées.

$$S_r = \frac{1}{n-1}\sum_{i}^{n-1}(x_{i*p+r} - M_p(i*p+r))$$

Raffinement: pour standardiser la décomposition, on suppose que la somme des coefficients saisonnier est nulle et on considère plutôt $S_r = S_r - \overline{S}$.

Définition 7.3 *Dans le cas du modèle additif, on appelle série corrigée des variations saisonnières, la série des différences:*

$$x_t^* = x_t - S_t$$

Le modèle additif prédit ainsi les valeurs $T(t) + S_t$, où T est l'ajustement de le série dsaisonnalisé

Une fois le modèle établi, on prédira par exemple le trimestre 15 par $T_{15} * S_3$, le trimestre 16 par $T_{16} * S_4$ et plus généralement le trimestre $n*4+r$ par $T_{n*4+r} * S_r$.