TP Séries Chronologiques n^0 3 : Prédiction par moindres carrés et filtres linéaires

Dans ce TP, on commence par utiliser la régression par moindres carrés pour prédire les valeurs futures d'une série chronologique. Ensuite, on se sert de filtres linéaires pour estimer ou annuler la tendance et la saisonnalité d'une série chronologique.

Prédiction par moindres carrés

x_1:20	On génère la variable temps.
y_(−1)^$x + x - (0.1 * x$^2)+5*rnorm(20)	On génère une série chronologique y (sous la forme d'un vecteur). Décrire théoriquement la série (tendance, saisonnalité et bruit).
y.glm_glm(y ∼ poly(x,3))	Régression polynômiale de degré 3 de y par x. La commande glm généralise la commande lm précédemment utilisée. Elle permet notamment des régressions pondérées (voir plus bas).
plot(x,y)	Visualisation de la série.
lines(x,y.glm$fit)	Graphe du polynôme obtenu par la régression.
y.pred_predict.gam(y.glm,as.data.frame(21))	Prédiction au temps $x = 21$ obtenue par le polynôme. En remplaçant 21 par $c(25, 30, 50)$, on obtient les prédictions aux temps $x = 25$, 30 et 50. Comparer (donner des écarts numériques) avec ce que l'on aurait obtenu si le modèle avait était parfaitement identifié (on ne considère pas alors le bruit). Refaire la même chose avec un polynôme de degré 1, 2 et 10. Quel est votre choix (théorique et empirique) quant au degré de la régression?
wei_1^x	Génération d'un vecteur poids (ici uniforme).
y.glm_glm(y ∼ poly(x,2),weights=wei)	Régression polynômiale de degré 2 de y par x, pondérée par wei. Quelle est la différence avec la régression précédente de degré 2? Expliquer ce résultat. Générer un poids ne portant que sur la première donnée, puis un poids ne portant que sur la dernière. Prédire pour $x = 25$. Expliquer ces résultats.
wei_0.8^(1-x)	Génération d'un vecteur poids exponentiel. Prédire pour $x = 25$. Faire varier le paramètre $\beta = 0.8$. Quelle valeur est-elle préférable? (c'est-à-dire, quelle est la valeur (à 0.1 près) permettant d'obtenir la meilleure prédiction?)

Exercice

1. Reprendre la série X_1 et prédire la valeur de janvier 2000. Quel est votre choix pour le degré du polynôme de la régression?
   (Facultatif) Utiliser alors un poids de type exponentiel accordant peu d'importance au passé ancien. Qu'obtenez-vous? Est-ce mieux sans ou avec le poids par rapport à la vraie valeur de $X1$ en janvier 2000?

Utilisation de filtres

T_1:100	On génère un vecteur temps.
E_2*rnorm(100)	On génère un bruit blanc gaussien.
X_0.01 * $T + 1 + 2 * \sin(2 * \pi * T/5) - E$	On simule une série X (sous la forme d'un vecteur). Déterminer théoriquement la tendance et la saisonnalité (et sa période). Tracer la série chronologique. La période est-elle facilement identifiable (par l'observation)? Et la saisonnalité?

	Filtrage
filt_c(1,0,0,0,0,-1)	On fabrique un filtre.
Y_filter(X,filt)	Série filtrée. Que vaut Y_n en fonction des X_i et pour quels n (regarder dans le help)?
	Tracer la série Y. Qu'observe-t-on (tendance, saisonnier)?
	Étudier théoriquement Y (déterminer la tendance, la saisonnalité et le bruit).
	Le bruit est-il un bruit blanc?
	Comment reconstruire X lorsque seule Y est connue (ainsi que $X(1), ..., X(5)$)?
	Utiliser maintenant le filtre $c(1,1,1,1,1)/5$ et répondre aux mêmes questions.

Exercice

1. Reprendre la série X_2 et utiliser un filtre linéaire adéquat permettant d'éliminer la saisonnalité. Identifier alors la tendance. Retirer alors cette tendance à X_2 puis utiliser un autre filtre linéaire permettant d'estimer la saisonnalité. Comparer cette estimation avec la vraie saisonnalité.