

Modèles état-mesure et filtre de Kalman

Plan

Introduction aux modèles état-mesure

Exemples de représentations état-mesure

Le filtre de Kalman

Le lisseur de Kalman

Estimation par maximum de vraisemblance

Prévisions

Plan

Introduction aux modèles état-mesure

Exemples de représentations état-mesure

Le filtre de Kalman

Le lisseur de Kalman

Estimation par maximum de vraisemblance

Prévisions

Motivation

- ▶ Les modèles état-mesure constituent un cadre général permettant de :
 - ▶ unifier de nombreux modèles de séries temporelles (ARMA, tendances, saisonnalité),
 - ▶ traiter les observations manquantes de manière naturelle,
 - ▶ estimer des composantes non observables (tendance, cycle),
 - ▶ effectuer des prévisions optimales via le filtre de Kalman.
- ▶ **Idée centrale** : Distinguer ce que l'on **observe** (équation de mesure) de la **dynamique latente** (équation d'état) qui gouverne le système.

Formulation générale

- Équation de mesure (observation) :

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{d}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{H}_t)$$

- Équation d'état (transition) :

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$$

- Dimensions : $\mathbf{y}_t \in \mathbb{R}^p$ (observations), $\boldsymbol{\alpha}_t \in \mathbb{R}^m$ (état), $\boldsymbol{\varepsilon}_t \in \mathbb{R}^p$, $\boldsymbol{\eta}_t \in \mathbb{R}^r$.

Notations et hypothèses

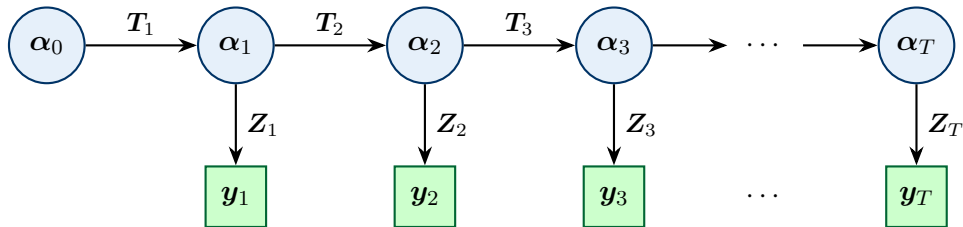
► Matrices du système :

- Z_t : matrice de mesure ($p \times m$),
- T_t : matrice de transition ($m \times m$),
- R_t : matrice de sélection des perturbations ($m \times r$),
- H_t : matrice de variance des erreurs de mesure ($p \times p$),
- Q_t : matrice de variance des perturbations d'état ($r \times r$),
- d_t, c_t : termes déterministes.

► Hypothèses fondamentales :

- $\mathbb{E}[\varepsilon_t] = \mathbf{0}, \mathbb{E}[\eta_t] = \mathbf{0}$,
- ε_t et η_s sont indépendants pour tout t, s ,
- condition initiale : $\alpha_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$,
- α_0 est indépendant de ε_t et η_t pour tout $t \geq 1$.

Représentation graphique



Les états forment une **chaîne de Markov**, les observations sont conditionnellement indépendantes sachant les états.

Plan

Introduction aux modèles état-mesure

Exemples de représentations état-mesure

Le filtre de Kalman

Le lisseur de Kalman

Estimation par maximum de vraisemblance

Prévisions

Exemple 1 : modèle de niveau local

- ▶ Marche aléatoire avec bruit de mesure :

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

- ▶ **Représentation état-mesure** : État $\alpha_t = \mu_t$ (scalaire, $m = 1$).

$$Z = 1, \quad T = 1, \quad R = 1, \quad H = \sigma_\varepsilon^2, \quad Q = \sigma_\eta^2, \quad d = c = 0$$

- ▶ Ce modèle décompose la série en une **tendance stochastique** μ_t et un **bruit de mesure** ε_t .

Exemple 2 : tendance linéaire locale

- Le modèle s'écrit :

$$y_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu_{t-1} + \nu_{t-1} + \eta_t, \quad \nu_t = \nu_{t-1} + \zeta_t$$

- **Représentation état-mesure** : Vecteur d'état $\alpha_t = (\mu_t, \nu_t)'$.

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$H = \sigma_\varepsilon^2, \quad \mathbf{Q} = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\zeta^2 \end{pmatrix}$$

Exemple 3 : processus AR(1)

- ▶ Le modèle AR(1) :

$$y_t = \phi y_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma^2), \quad |\phi| < 1$$

- ▶ **Représentation état-mesure** : État $\alpha_t = y_t$.
 - ▶ Équation de mesure : $y_t = \alpha_t$ (observation parfaite, $H = 0$),
 - ▶ Équation d'état : $\alpha_t = \phi \alpha_{t-1} + \eta_t$.

Matrices : $Z = 1$, $T = \phi$, $R = 1$, $H = 0$, $Q = \sigma^2$.

- ▶ **Remarque** : Ici, l'état est directement observable ($H = 0$). Le filtre de Kalman se réduit à l'équation de récurrence de l'AR(1).

Exemple 4 : processus AR(p)

- Le modèle AR(p) :

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \eta_t$$

- **Représentation état-mesure** (forme compagnon) : vecteur d'état

$$\alpha_t = (y_t, y_{t-1}, \dots, y_{t-p+1})'.$$

$$\mathbf{Z} = (1 \quad 0 \quad \cdots \quad 0), \quad \mathbf{T} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

- $\mathbf{R} = (1, 0, \dots, 0)'$, $\mathbf{H} = 0$, $\mathbf{Q} = \sigma^2$.

Exemple 5 : processus MA(q)

- Le modèle MA(q) :

$$y_t = \eta_t + \theta_1 \eta_{t-1} + \theta_2 \eta_{t-2} + \cdots + \theta_q \eta_{t-q}$$

- **Représentation état-mesure** : Vecteur d'état $\alpha_t = (\eta_t, \eta_{t-1}, \dots, \eta_{t-q})'$.

$$\mathbf{Z} = (1 \quad \theta_1 \quad \theta_2 \quad \cdots \quad \theta_q)$$
$$\mathbf{T} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- $H = 0, Q = \sigma^2$.

Exemple 6 : processus ARMA(1,1)

- Le modèle ARMA(1,1) :

$$y_t = \phi y_{t-1} + \eta_t + \theta \eta_{t-1}, \quad |\phi| < 1$$

- **Représentation état-mesure** (Harvey, 1989) : dimension $r = \max(1, 1 + 1) = 2$, vecteur d'état $\alpha_t \in \mathbb{R}^2$.

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 \\ \theta \end{pmatrix}$$

- $H = 0$, $Q = \sigma^2$.

Vérification : ARMA(1,1)

- ▶ Avec $\alpha_t = (\alpha_{1,t}, \alpha_{2,t})'$, l'équation d'état donne :

$$\alpha_{1,t} = \phi\alpha_{1,t-1} + \alpha_{2,t-1} + \eta_t$$

$$\alpha_{2,t} = \theta\eta_t$$

- ▶ L'observation est $y_t = \alpha_{1,t}$.
- ▶ En substituant $\alpha_{2,t-1} = \theta\eta_{t-1}$:

$$y_t = \phi y_{t-1} + \theta\eta_{t-1} + \eta_t$$

On retrouve bien le modèle ARMA(1,1). L'état $\alpha_{1,t}$ capture la dynamique AR, $\alpha_{2,t} = \theta\eta_t$ capture l'effet MA retardé.

Exemple 7 : processus ARMA(p,q) général

- Pour $y_t - \sum_{i=1}^p \phi_i y_{t-i} = \eta_t + \sum_{j=1}^q \theta_j \eta_{t-j}$, on pose $r = \max(p, q+1)$ et $\alpha_t \in \mathbb{R}^r$.
- Les matrices sont :

$$\mathbf{T} = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ \phi_{r-1} & 0 & 0 & \cdots & 1 \\ \phi_r & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{r-1} \end{pmatrix}$$

avec $\phi_j = 0$ pour $j > p$ et $\theta_j = 0$ pour $j > q$.

- $\mathbf{Z} = (1, 0, \dots, 0)$, $H = 0$, $Q = \sigma^2$.

Exemple 8 : composante saisonnière

- **Saisonnalité stochastique** (période s) :

$$\gamma_t + \gamma_{t-1} + \cdots + \gamma_{t-s+1} = \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2)$$

- **Représentation état-mesure** : Vecteur d'état $\alpha_t = (\gamma_t, \gamma_{t-1}, \dots, \gamma_{t-s+2})'$ (dimension $s - 1$).

$$\mathbf{Z} = (1 \quad 0 \quad \cdots \quad 0), \quad \mathbf{T} = \begin{pmatrix} -1 & -1 & \cdots & -1 & -1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

- $\mathbf{R} = (1, 0, \dots, 0)'$, $Q = \sigma_\omega^2$.

Plan

Introduction aux modèles état-mesure

Exemples de représentations état-mesure

Le filtre de Kalman

Le lisseur de Kalman

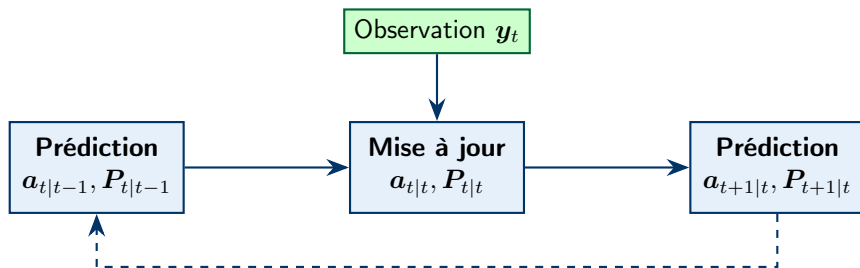
Estimation par maximum de vraisemblance

Prévisions

Objectif du filtrage

- ▶ Étant donné les observations $Y_t = \{y_1, y_2, \dots, y_t\}$, on cherche à :
 - ▶ **estimer** l'état courant α_t ,
 - ▶ **quantifier l'incertitude** de cette estimation,
 - ▶ **mettre à jour** récursivement.
- ▶ **Estimateurs optimaux :**
 - ▶ **Prédiction** : $a_{t|t-1} = \mathbb{E}[\alpha_t | Y_{t-1}]$ avec variance $P_{t|t-1}$,
 - ▶ **Filtrage** : $a_{t|t} = \mathbb{E}[\alpha_t | Y_t]$ avec variance $P_{t|t}$.
- ▶ Sous les hypothèses gaussiennes, ces estimateurs sont les **BLUE** et coïncident avec l'espérance conditionnelle.

Filtre de Kalman : vue d'ensemble



Le filtre alterne entre **prédiction** (projeter l'état via le modèle) et **correction** (incorporer la nouvelle observation).

Équations du filtre de Kalman (1/2)

► Initialisation : $\mathbf{a}_{0|0} = \mathbf{a}_0$, $\mathbf{P}_{0|0} = \mathbf{P}_0$.

► Étape de prédiction :

$$\mathbf{a}_{t|t-1} = \mathbf{T}_t \mathbf{a}_{t-1|t-1} + \mathbf{c}_t$$

$$\mathbf{P}_{t|t-1} = \mathbf{T}_t \mathbf{P}_{t-1|t-1} \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$$

Équations du filtre de Kalman (2/2)

► Étape de mise à jour :

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_{t|t-1} - \mathbf{d}_t$$

(innovation)

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t$$

(variance de l'innovation)

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1}$$

(gain de Kalman)

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{K}_t \mathbf{v}_t$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{Z}_t) \mathbf{P}_{t|t-1}$$

Preuve : lemme préliminaire

- **Lemme** (distribution conditionnelle gaussienne). Soit

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N \left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix} \right)$$

Alors :

$$\mathbf{x}|\mathbf{y} \sim N(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})$$

- On applique ce lemme avec $\mathbf{x} = \boldsymbol{\alpha}_t$ (état) et $\mathbf{y} = \mathbf{y}_t$ (observation), conditionnellement à Y_{t-1} .

Preuve du lemme (1/2)

- ▶ **Idée** : Décomposer $\mathbf{x}$ en une partie qui dépend linéairement de $\mathbf{y}$ et un résidu indépendant de $\mathbf{y}$.
- ▶ Posons $\mathbf{e} = \mathbf{x} - \boldsymbol{\mu}_x - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y)$.
- ▶ **Covariance avec $\mathbf{y}$** :

$$\begin{aligned}\text{Cov}(\mathbf{e}, \mathbf{y}) &= \text{Cov}(\mathbf{x}, \mathbf{y}) - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \text{Cov}(\mathbf{y}, \mathbf{y}) \\ &= \boldsymbol{\Sigma}_{xy} - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yy} = \mathbf{0}\end{aligned}$$

- ▶ Puisque $(\mathbf{e}, \mathbf{y})$ est gaussien (transformation linéaire de $(\mathbf{x}, \mathbf{y})$), covariance nulle implique **indépendance**.

Preuve du lemme (2/2)

► On a donc $\mathbf{x} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) + \mathbf{e}$, avec $\mathbf{e} \perp \mathbf{y}$.

► **Espérance conditionnelle :**

$$\mathbb{E}[\mathbf{x}|\mathbf{y}] = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) + \underbrace{\mathbb{E}[\mathbf{e}]}_{=\mathbf{0}} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$$

► **Variance conditionnelle.** Puisque $\mathbf{e} = (\mathbf{x} - \boldsymbol{\mu}_x) - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$ et $\text{Cov}(\mathbf{e}, \mathbf{y}) = \mathbf{0}$:

$$\begin{aligned}\mathbb{V}[\mathbf{x}|\mathbf{y}] &= \mathbb{V}[\mathbf{e}] = \text{Cov}(\mathbf{e}, (\mathbf{x} - \boldsymbol{\mu}_x) - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)) \\ &= \text{Cov}(\mathbf{e}, \mathbf{x}) + \mathbf{0} \\ &= \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} \quad \square\end{aligned}$$

Preuve : étape de prédiction

- ▶ **Hypothèse de récurrence** : $\alpha_{t-1}|Y_{t-1} \sim N(\mathbf{a}_{t-1|t-1}, \mathbf{P}_{t-1|t-1})$.
- ▶ Par l'équation d'état : $\alpha_t = \mathbf{T}_t \alpha_{t-1} + \mathbf{c}_t + \mathbf{R}_t \eta_t$.
- ▶ **Espérance conditionnelle** :

$$\mathbf{a}_{t|t-1} = \mathbb{E}[\alpha_t|Y_{t-1}] = \mathbf{T}_t \mathbf{a}_{t-1|t-1} + \mathbf{c}_t$$

- ▶ **Variance conditionnelle** (indépendance de η_t et Y_{t-1}) :

$$\mathbf{P}_{t|t-1} = \mathbb{V}[\alpha_t|Y_{t-1}] = \mathbf{T}_t \mathbf{P}_{t-1|t-1} \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$$

Preuve : distribution jointe

- Par l'équation de mesure $\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{d}_t + \boldsymbol{\varepsilon}_t$, la **distribution jointe** de $(\boldsymbol{\alpha}_t, \mathbf{y}_t)$ sachant Y_{t-1} est :

$$\begin{pmatrix} \boldsymbol{\alpha}_t \\ \mathbf{y}_t \end{pmatrix} \Big| Y_{t-1} \sim N \left(\begin{pmatrix} \mathbf{a}_{t|t-1} \\ \mathbf{Z}_t \mathbf{a}_{t|t-1} + \mathbf{d}_t \end{pmatrix}, \begin{pmatrix} \mathbf{P}_{t|t-1} & \mathbf{P}_{t|t-1} \mathbf{Z}_t' \\ \mathbf{Z}_t \mathbf{P}_{t|t-1} & \mathbf{F}_t \end{pmatrix} \right)$$

où $\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t$.

- Calcul de la covariance :

$$\begin{aligned} \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{y}_t | Y_{t-1}) &= \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t | Y_{t-1}) \\ &= \mathbf{P}_{t|t-1} \mathbf{Z}_t' + \underbrace{\text{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\varepsilon}_t)}_{=0} = \mathbf{P}_{t|t-1} \mathbf{Z}_t' \end{aligned}$$

Preuve : étape de mise à jour

- ▶ En appliquant le lemme avec $\boldsymbol{x} = \boldsymbol{\alpha}_t$ et $\boldsymbol{y} = \boldsymbol{y}_t$:
- ▶ Espérance conditionnelle :

$$\begin{aligned}\boldsymbol{a}_{t|t} &= \mathbb{E}[\boldsymbol{\alpha}_t | Y_t] = \boldsymbol{a}_{t|t-1} + \boldsymbol{P}_{t|t-1} \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} (\boldsymbol{y}_t - \boldsymbol{Z}_t \boldsymbol{a}_{t|t-1} - \boldsymbol{d}_t) \\ &= \boldsymbol{a}_{t|t-1} + \boldsymbol{K}_t \boldsymbol{v}_t\end{aligned}$$

où $\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1} \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1}$ est le gain de Kalman.

- ▶ Variance conditionnelle :

$$\begin{aligned}\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1} \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{Z}_t \boldsymbol{P}_{t|t-1} \\ &= (\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{Z}_t) \boldsymbol{P}_{t|t-1}\end{aligned}$$

Interprétation du gain de Kalman

- ▶ Le gain $\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1}$ détermine le poids accordé à l'innovation :

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{K}_t \mathbf{v}_t$$

- ▶ **Cas limites (scalaire) :**
 - ▶ $H_t \rightarrow 0$ (mesure précise) : $K_t \rightarrow 1 \Rightarrow$ confiance totale à l'observation,
 - ▶ $H_t \rightarrow \infty$ (mesure bruitée) : $K_t \rightarrow 0 \Rightarrow$ on garde la prédiction,
 - ▶ $P_{t|t-1} \rightarrow 0$ (état bien connu) : $K_t \rightarrow 0 \Rightarrow$ pas d'information nouvelle.

Formule de Joseph

- **Forme numériquement stable** pour $P_{t|t}$:

$$P_{t|t} = (I - K_t Z_t) P_{t|t-1} (I - K_t Z_t)' + K_t H_t K_t'$$

Cette forme garantit que $P_{t|t}$ reste symétrique et semi-définie positive.

- **Démonstration de l'équivalence.** En développant et utilisant $K_t F_t = P_{t|t-1} Z_t'$:

$$\begin{aligned} & (I - K_t Z_t) P_{t|t-1} (I - K_t Z_t)' + K_t H_t K_t' \\ &= P_{t|t-1} - K_t Z_t P_{t|t-1} - P_{t|t-1} Z_t' K_t' + K_t F_t K_t' \\ &= P_{t|t-1} - K_t Z_t P_{t|t-1} \end{aligned}$$

Propriétés de l'innovation

- ▶ L'innovation $\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_{t|t-1} - \mathbf{d}_t$ est la partie de $\mathbf{y}_t$ non prévisible à partir de $\mathbf{Y}_{t-1}$.
- ▶ Propriétés fondamentales :
 - ▶ $\mathbb{E}[\mathbf{v}_t | \mathbf{Y}_{t-1}] = \mathbf{0}$,
 - ▶ $\mathbb{V}[\mathbf{v}_t | \mathbf{Y}_{t-1}] = \mathbf{F}_t$,
 - ▶ $\text{Cov}(\mathbf{v}_t, \mathbf{v}_s) = \mathbf{0}$ pour $t \neq s$.
- ▶ La suite $\{\mathbf{v}_t\}_{t=1}^T$ forme un **bruit blanc** (sous hypothèses gaussiennes). C'est essentiel pour le calcul de la vraisemblance et les diagnostics du modèle.

Preuve : innovations non corrélées

- Pour $s < t$, puisque $\mathbf{v}_s$ est Y_s -mesurable et $Y_s \subset Y_{t-1}$:

$$\text{Cov}(\mathbf{v}_t, \mathbf{v}_s) = \mathbb{E}[\mathbf{v}_t \mathbf{v}_s'] \quad (\text{car } \mathbb{E}[\mathbf{v}_t] = \mathbb{E}[\mathbf{v}_s] = \mathbf{0})$$

- Par la loi des espérances itérées :

$$\begin{aligned} \mathbb{E}[\mathbf{v}_t \mathbf{v}_s'] &= \mathbb{E}[\mathbb{E}[\mathbf{v}_t \mathbf{v}_s' | Y_{t-1}]] \\ &= \mathbb{E}[\mathbb{E}[\mathbf{v}_t | Y_{t-1}] \mathbf{v}_s'] \quad (\mathbf{v}_s \text{ est } Y_{t-1}\text{-mesurable}) \\ &= \mathbb{E}[\mathbf{0} \cdot \mathbf{v}_s'] = \mathbf{0} \end{aligned}$$

Initialisation du filtre

- **Cas stationnaire.** Si les valeurs propres de $\mathbf{T}$ ont un module < 1 :

$$\mathbf{a}_{0|0} = \mathbf{0}, \quad \mathbf{P}_{0|0} \text{ solution de } \mathbf{P} = \mathbf{TPT}' + \mathbf{RQR}'$$

C'est une équation de Lyapunov, résolue par :

$$\text{vec}(\mathbf{P}) = (\mathbf{I} - \mathbf{T} \otimes \mathbf{T})^{-1} \text{vec}(\mathbf{RQR}')$$

- **Cas non stationnaire** (initialisation diffuse) :

$$\mathbf{a}_{0|0} = \mathbf{0}, \quad \mathbf{P}_{0|0} = \kappa \mathbf{I} \quad \text{avec } \kappa \rightarrow \infty$$

En pratique, on utilise le filtre de Kalman diffus exact (De Jong, 1991).

Plan

Introduction aux modèles état-mesure

Exemples de représentations état-mesure

Le filtre de Kalman

Le lisseur de Kalman

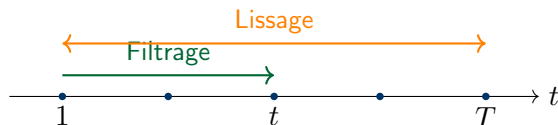
Estimation par maximum de vraisemblance

Prévisions

Motivation du lissage

- Le **filtrage** utilise Y_t pour estimer α_t . Le **lissage** utilise **toute** l'information $Y_T = \{y_1, \dots, y_T\}$:

$$\alpha_{t|T} = \mathbb{E}[\alpha_t | Y_T], \quad P_{t|T} = \mathbb{V}[\alpha_t | Y_T]$$



Le lisseur exploite l'information future : $P_{t|T} \leq P_{t|t}$.

- Initialisation : $\mathbf{a}_{T|T}$, $\mathbf{P}_{T|T}$ (issus du filtre).
- Récurrence arrière pour $t = T - 1, T - 2, \dots, 0$:

$$\mathbf{J}_t = \mathbf{P}_{t|t} \mathbf{T}'_{t+1} \mathbf{P}_{t+1|t}^{-1} \quad (\text{gain de lissage})$$

$$\mathbf{a}_{t|T} = \mathbf{a}_{t|t} + \mathbf{J}_t (\mathbf{a}_{t+1|T} - \mathbf{a}_{t+1|t})$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{J}_t'$$

- Complexité : une passe avant (filtre) + une passe arrière. Total : $O(Tm^3)$.

Preuve du lisseur (1/2)

- **Étape 1** : Le couple (α_t, α_{t+1}) sachant Y_t suit :

$$\begin{pmatrix} \alpha_t \\ \alpha_{t+1} \end{pmatrix} \Big| Y_t \sim N \left(\begin{pmatrix} \mathbf{a}_{t|t} \\ \mathbf{a}_{t+1|t} \end{pmatrix}, \begin{pmatrix} \mathbf{P}_{t|t} & \mathbf{P}_{t|t} \mathbf{T}'_{t+1} \\ \mathbf{T}_{t+1} \mathbf{P}_{t|t} & \mathbf{P}_{t+1|t} \end{pmatrix} \right)$$

car $\text{Cov}(\alpha_t, \alpha_{t+1} | Y_t) = \text{Cov}(\alpha_t, \mathbf{T}_{t+1} \alpha_t | Y_t) = \mathbf{P}_{t|t} \mathbf{T}'_{t+1}$.

- **Étape 2** : Par le lemme gaussien :

$$\alpha_t | (\alpha_{t+1}, Y_t) \sim N \left(\mathbf{a}_{t|t} + \mathbf{J}_t (\alpha_{t+1} - \mathbf{a}_{t+1|t}), \mathbf{P}_{t|t} - \mathbf{J}_t \mathbf{P}_{t+1|t} \mathbf{J}'_t \right)$$

où $\mathbf{J}_t = \mathbf{P}_{t|t} \mathbf{T}'_{t+1} \mathbf{P}_{t+1|t}^{-1}$.

Preuve du lisseur (2/2)

- **Étape 3** : Espérance lissée (loi des espérances itérées) :

$$\begin{aligned}\mathbf{a}_{t|T} &= \mathbb{E}[\boldsymbol{\alpha}_t | Y_T] = \mathbb{E}[\mathbb{E}[\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t+1}, Y_t] | Y_T] \\ &= \mathbb{E}[\mathbf{a}_{t|t} + \mathbf{J}_t(\boldsymbol{\alpha}_{t+1} - \mathbf{a}_{t+1|t}) | Y_T] \\ &= \mathbf{a}_{t|t} + \mathbf{J}_t(\mathbf{a}_{t+1|T} - \mathbf{a}_{t+1|t})\end{aligned}$$

- **Étape 4** : Variance lissée (loi de la variance totale) :

$$\begin{aligned}\mathbf{P}_{t|T} &= \mathbb{E}[\mathbb{V}[\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t+1}, Y_t] | Y_T] + \mathbb{V}[\mathbb{E}[\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t+1}, Y_t] | Y_T] \\ &= \mathbf{P}_{t|t} - \mathbf{J}_t \mathbf{P}_{t+1|t} \mathbf{J}_t' + \mathbf{J}_t \mathbf{P}_{t+1|T} \mathbf{J}_t' \\ &= \mathbf{P}_{t|t} + \mathbf{J}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{J}_t'\end{aligned}$$

Plan

Introduction aux modèles état-mesure

Exemples de représentations état-mesure

Le filtre de Kalman

Le lisseur de Kalman

Estimation par maximum de vraisemblance

Prévisions

Paramètres à estimer

- ▶ Les matrices du système dépendent de paramètres inconnus $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$.
- ▶ Exemple ARMA(1,1) : $\boldsymbol{\theta} = (\phi, \theta, \sigma^2)'$.

- ▶ Objectif :

$$\hat{\boldsymbol{\theta}}_{\text{MV}} = \arg \max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; Y_T)$$

Décomposition de la vraisemblance

- Décomposition de l'erreur de prédiction :

$$f(y_1, \dots, y_T; \boldsymbol{\theta}) = f(y_1; \boldsymbol{\theta}) \prod_{t=2}^T f(y_t | Y_{t-1}; \boldsymbol{\theta})$$

- Puisque $\mathbf{y}_t | Y_{t-1} \sim N(\mathbf{Z}_t \mathbf{a}_{t|t-1} + \mathbf{d}_t, \mathbf{F}_t)$, on a :

$$f(\mathbf{y}_t | Y_{t-1}) = (2\pi)^{-p/2} |\mathbf{F}_t|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t\right)$$

Log-vraisemblance

- La **log-vraisemblance** s'écrit :

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{pT}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t$$

- Pour évaluer $\mathcal{L}(\boldsymbol{\theta})$, il suffit d'exécuter le **filtre de Kalman** qui produit les innovations $\mathbf{v}_t$ et leurs variances $\mathbf{F}_t$.
- **Cas scalaire** ($p = 1$) :

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log F_t - \frac{1}{2} \sum_{t=1}^T \frac{v_t^2}{F_t}$$

Vraisemblance concentrée

- ▶ Si $\mathbf{Q} = \sigma^2 \mathbf{Q}^*$ et $\mathbf{H} = \sigma^2 \mathbf{H}^*$, la condition du premier ordre par rapport à σ^2 donne :

$$\hat{\sigma}^2(\boldsymbol{\theta}^*) = \frac{1}{pT} \sum_{t=1}^T v_t^{*'} F_t^{*-1} v_t^*$$

où les quantités $*$ sont calculées avec $\sigma^2 = 1$.

- ▶ On substitue $\hat{\sigma}^2$ dans la log-vraisemblance pour obtenir la **vraisemblance concentrée** $\mathcal{L}_c(\boldsymbol{\theta}^*)$ qui ne dépend plus de σ^2 .
- ▶ Cela réduit la dimension du problème d'optimisation.

- ▶ **Matrice d'information de Fisher :**

$$\mathcal{I}(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$$

Asymptotiquement : $\sqrt{T}(\hat{\boldsymbol{\theta}}_{\text{MV}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1})$.

- ▶ **Diagnostics** sur les résidus standardisés $e_t = F_t^{-1/2} v_t$:

- ▶ test de normalité : Jarque-Bera,
- ▶ test d'autocorrélation : Ljung-Box $Q(m) = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k}$,
- ▶ test ARCH sur e_t^2 .

Sélection de modèles

- Critère d'Akaike (AIC) :

$$\text{AIC} = -2\mathcal{L}(\hat{\theta}) + 2k$$

- Critère bayésien de Schwarz (BIC) :

$$\text{BIC} = -2\mathcal{L}(\hat{\theta}) + k \log(T)$$

où k est le nombre de paramètres et T le nombre d'observations.

- On sélectionne le modèle qui **minimise** le critère. Le BIC pénalise davantage la complexité et est consistant.

Plan

Introduction aux modèles état-mesure

Exemples de représentations état-mesure

Le filtre de Kalman

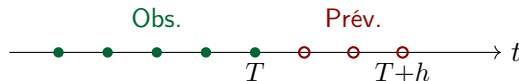
Le lisseur de Kalman

Estimation par maximum de vraisemblance

Prévisions

Types de prévisions

- ▶ **Prévision à un pas** : $\hat{y}_{T+1|T} = \mathbb{E}[y_{T+1}|Y_T]$.
- ▶ **Prévision à h pas** : $\hat{y}_{T+h|T} = \mathbb{E}[y_{T+h}|Y_T]$.
- ▶ **Stratégie** : Continuer les équations de prédiction au-delà de T , **sans** mise à jour.



Équations de prévision

- Prédiction de l'état (pour $h \geq 1$) :

$$\mathbf{a}_{T+h|T} = \mathbf{T}_{T+h} \mathbf{a}_{T+h-1|T} + \mathbf{c}_{T+h}$$

$$\mathbf{P}_{T+h|T} = \mathbf{T}_{T+h} \mathbf{P}_{T+h-1|T} \mathbf{T}_{T+h}' + \mathbf{R}_{T+h} \mathbf{Q}_{T+h} \mathbf{R}_{T+h}'$$

Initialisation : $\mathbf{a}_{T|T}$, $\mathbf{P}_{T|T}$ (issus du filtre).

- Prédiction de l'observation :

$$\hat{\mathbf{y}}_{T+h|T} = \mathbf{Z}_{T+h} \mathbf{a}_{T+h|T} + \mathbf{d}_{T+h}$$

$$\mathbf{F}_{T+h|T} = \mathbf{Z}_{T+h} \mathbf{P}_{T+h|T} \mathbf{Z}_{T+h}' + \mathbf{H}_{T+h}$$

Intervalles de prévision

- **Intervalle à $(1 - \alpha)\%$ (cas scalaire) :**

$$\hat{y}_{T+h|T} \pm z_{\alpha/2} \sqrt{F_{T+h|T}}$$

Pour $\alpha = 0,05$: $z_{0,025} \approx 1,96$.

- L'intervalle s'élargit avec h car :
 - l'incertitude sur l'état se propage,
 - les chocs futurs s'accumulent.
- Pour un modèle stationnaire : $F_{T+h|T} \rightarrow$ variance inconditionnelle quand $h \rightarrow \infty$.

Formules explicites (coefficients constants)

- Si $\mathbf{T}_t = \mathbf{T}$, etc., l'état prévu est :

$$\mathbf{a}_{T+h|T} = \mathbf{T}^h \mathbf{a}_{T|T} + \sum_{j=0}^{h-1} \mathbf{T}^j \mathbf{c}$$

- La variance de prévision est :

$$\mathbf{P}_{T+h|T} = \mathbf{T}^h \mathbf{P}_{T|T} (\mathbf{T}')^h + \sum_{j=0}^{h-1} \mathbf{T}^j \mathbf{R} \mathbf{Q} \mathbf{R}' (\mathbf{T}')^j$$

Exemple : prévision AR(1)

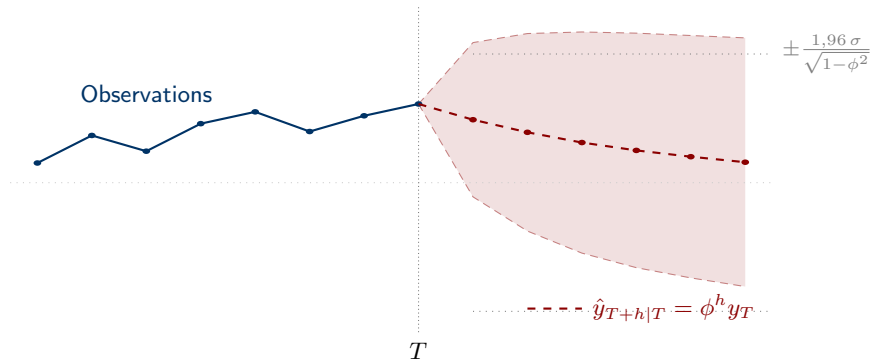
- **Rappel** : $T = \phi$, $Z = 1$, $R = 1$, $H = 0$, $Q = \sigma^2$, état $\alpha_t = y_t$.
- Puisque $H = 0$: $a_{T|T} = y_T$ et $P_{T|T} = 0$. Donc :

$$\hat{y}_{T+h|T} = \phi^h y_T$$

- **Variance de prévision** : $P_{T+h|T} = \sigma^2 \sum_{j=0}^{h-1} \phi^{2j}$, soit :

$$F_{T+h|T} = \sigma^2 \frac{1 - \phi^{2h}}{1 - \phi^2} \xrightarrow{h \rightarrow \infty} \frac{\sigma^2}{1 - \phi^2} = \mathbb{V}[y_t]$$

Exemple : prévision AR(1) – illustration



Paramètres : $\phi = 0,8$, $\sigma^2 = 1$, $y_T = 2$. La prévision décroît géométriquement et l'IC converge vers l'IC inconditionnel $\pm 1,96 \sigma / \sqrt{1 - \phi^2}$.

Exemple : prévision MA(1)

► **Rappel** : $\mathbf{T} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{Z} = (1 \quad \theta)$, $\mathbf{R} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, état $\boldsymbol{\alpha}_t = (\eta_t, \eta_{t-1})'$.

► La matrice $\mathbf{T}$ est **nilpotente** : $\mathbf{T}^2 = \mathbf{0}$. Pour $h \geq 1$:

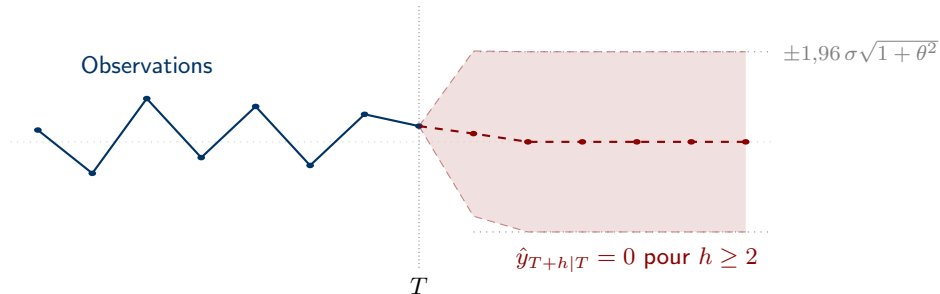
$$\mathbf{Z}\mathbf{T}^h\mathbf{a}_{T|T} = \begin{cases} \theta a_{1,T|T} & \text{si } h = 1 \\ 0 & \text{si } h \geq 2 \end{cases}$$

où $a_{1,T|T} = \mathbb{E}[\eta_T | Y_T]$ est l'innovation filtrée.

► **Variance de prévision** ($h \geq 2$) : $\mathbf{T}^h = \mathbf{0}$ et les seuls termes non nuls sont $j = 0, 1$:

$$F_{T+h|T} = \mathbf{Z}(\sigma^2\mathbf{R}\mathbf{R}' + \sigma^2\mathbf{T}\mathbf{R}\mathbf{R}'\mathbf{T}')\mathbf{Z}' = \sigma^2(1 + \theta^2) = \mathbb{V}[y_t]$$

Exemple : prévision MA(1) – illustration



Paramètres : $\theta = 0,6$, $\sigma^2 = 1$. La prévision tombe à zéro dès $h = 2$ et l'IC atteint immédiatement la variance inconditionnelle $\sigma^2(1 + \theta^2)$.

Prévision ARMA(p,q)

- Pour un ARMA(p,q), l'état filtré $\mathbf{a}_{T|T}$ contient toute l'information nécessaire à la prévision :

$$\hat{y}_{T+h|T} = \mathbf{Z}\mathbf{T}^h\mathbf{a}_{T|T}$$

- Exemple ARMA(1,1). Avec $\mathbf{T} = \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix}$:

$$\mathbf{T}^h = \begin{pmatrix} \phi^h & \phi^{h-1} \\ 0 & 0 \end{pmatrix} \quad (h \geq 1)$$

Donc : $\hat{y}_{T+h|T} = \phi^h a_{1,T|T} + \phi^{h-1} a_{2,T|T}$.

Résumé

- ▶ Les **modèles état-mesure** unifient ARMA, tendances et saisonnalité.
- ▶ Le **filtre de Kalman** fournit une estimation récursive optimale de l'état.
- ▶ Le **lisseur de Kalman** exploite toute l'information disponible.
- ▶ Le **maximum de vraisemblance** s'obtient via la décomposition de l'innovation.
- ▶ Les **prévisions** sont calculées par prolongation des équations de prédiction.