



# cours SOD333

## Filtrage Bayésien et Approximation Particulaire

### Filtre Bayésien

François Le Gland  
INRIA Rennes et IRMAR

[people.rennes.inria.fr/Francois.Le\\_Gland/ensta/](http://people.rennes.inria.fr/Francois.Le_Gland/ensta/)

24 septembre 2021



## Modèles

notations

systèmes non–linéaires / non–gaussiens

modèles de Markov cachés

chaînes de Markov partiellement observées

Filtre (et lisseur) bayésien

Distributions de Feynman–Kac



## Noyaux : rappels et notations

l'intégrale par rapport à la mesure positive  $\mu$  d'une fonction mesurable  $\phi$  est notée indifféremment

$$\mu(\phi) = \langle \mu, \phi \rangle = \int_E \phi(x) \mu(dx)$$

un noyau positif est une collection  $K(x, dx')$  indexée par  $x \in E$  de mesures positives

si  $K(x, E) = 1$  pour tout  $x \in E$ , c'est-à-dire si  $K(x, dx')$  est une distribution de probabilité pour tout  $x \in E$ , on parlera de noyau de probabilité, ou de noyau de transition, ou encore de noyau markovien



un noyau positif agit sur les fonctions, c'est-à-dire transforme une fonction  $\phi$  en une autre fonction  $K\phi$  définie par

$$K\phi(x) = \int_E K(x, dx') \phi(x')$$

pour tout  $x \in E$

il agit aussi sur les mesures positives, c'est-à-dire transforme une mesure positive  $\mu$  en une autre mesure positive  $\mu K$  définie par

$$\mu K(A) = \int_E \mu(dx) K(x, A)$$

pour tout  $A \subset E$ , également notée

$$\mu K(dx') = \int_E \mu(dx) K(x, dx')$$

où l'intégrale porte sur la variable  $x$ , vue comme mélange des mesures positives  $K(x, dx')$  par rapport à la distribution de mélange  $\mu(dx)$



on note l'identité

$$\langle \mu K, \phi \rangle = \langle \mu, K \phi \rangle$$

qui résulte du théorème de Fubini : en effet

$$\begin{aligned}
 \langle \mu K, \phi \rangle &= \int_E \mu K(dx') \phi(x') \\
 &= \int_E \left[ \int_E \mu(dx) K(x, dx') \right] \phi(x') \\
 &= \int_E \int_E \mu(dx) K(x, dx') \phi(x') \\
 &= \int_E \mu(dx) \left[ \int_E K(x, dx') \phi(x') \right] \\
 &= \int_E \mu(dx) K \phi(x) \\
 &= \langle \mu, K \phi \rangle
 \end{aligned}$$



## Systèmes non–linéaires / non–gaussiens

modèle a priori pour l'état caché à valeurs dans  $E$

$$X_k = f_k(X_{k-1}, W_k) \quad \text{avec} \quad W_k \sim p_k^W(dw)$$

condition initiale  $X_0 \sim \eta_0(dx)$

observation à valeurs dans  $\mathbb{R}^d$  dans un bruit additif possédant une densité

$$Y_k = h_k(X_k) + V_k \quad \text{avec} \quad V_k \sim q_k^V(v) dv$$

les v.a.  $X_0, W_1, \dots, W_k, \dots$  et  $V_0, V_1, \dots, V_k, \dots$  sont

*indépendantes mais pas nécessairement gaussiennes*

pour la suite, il suffit de savoir

- ▶ *simuler* une v.a. selon  $\eta_0(dx)$  ou selon  $p_k^W(dw)$
- ▶ *évaluer* pour tout  $v \in \mathbb{R}^d$ , la fonction  $q_k^V(v)$



**Proposition** la suite  $\{X_k\}$  est une chaîne de Markov à valeurs dans  $E$ , c'est-à-dire que conditionnellement aux états passés  $X_{0:k-1}$ , l'état présent  $X_k$  ne dépend que de  $X_{k-1}$  (statistique exhaustive)

$$\mathbb{P}[X_k \in dx \mid X_{0:k-1}] = \mathbb{P}[X_k \in dx \mid X_{k-1}]$$

et la loi de  $X$  est complètement déterminée par

- ▶ la loi initiale

$$\mathbb{P}[X_0 \in dx] = \eta_0(dx)$$

- ▶ et le *noyau de transition*

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = Q_k(x, dx')$$



**Preuve** compte tenu que  $X_{0:k-1}$  et  $W_k$  sont indépendants

$$\begin{aligned}\mathbb{E}[\phi(X_k) \mid X_{0:k-1}] &= \mathbb{E}[\phi(f_k(X_{k-1}, W_k)) \mid X_{0:k-1}] \\ &= \int_{\mathbb{R}^p} \phi(f_k(X_{k-1}, w)) p_k^W(dw)\end{aligned}$$

ne dépend que de  $X_{k-1}$ , c'est-à-dire que

$$\mathbb{E}[\phi(X_k) \mid X_{0:k-1}] = \mathbb{E}[\phi(X_k) \mid X_{k-1}]$$

d'où la **propriété de Markov**, et

$$\int_E Q_k(x, dx') \phi(x') = \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] = \int_{\mathbb{R}^p} \phi(f_k(x, w)) p_k^W(dw)$$

ce qui définit **implicitement** le noyau de transition □





**Remarque** si  $f_k(x, w) = b_k(x) + w$ , et si la distribution de probabilité  $p_k^W(dw)$  admet une densité encore notée  $p_k^W(w)$ , c'est-à-dire si  $p_k^W(dw) = p_k^W(w) dw$ , alors

$$Q_k(x, dx') = p_k^W(x' - b_k(x)) dx'$$

c'est-à-dire que le noyau  $Q_k(x, dx')$  admet une densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}^m$

en effet, le changement de variable  $x' = b_k(x) + w$  donne immédiatement

$$Q_k \phi(x) = \int_{\mathbb{R}^m} \phi(b_k(x) + w) p_k^W(w) dw = \int_{\mathbb{R}^m} \phi(x') p_k^W(x' - b_k(x)) dx'$$



**Remarque** le noyau  $Q_k(x, dx')$  n'admet pas nécessairement de densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}^m$   
 en effet, conditionnellement à  $X_{k-1} = x$ , le vecteur aléatoire  $X_k$  appartient nécessairement au sous-ensemble

$$\mathcal{M}(x) = \{x' \in \mathbb{R}^m : \text{il existe } w \in \mathbb{R}^p \text{ tel que } x' = f_k(x, w)\}$$

et dans le cas où  $p < m$  ce sous-ensemble  $\mathcal{M}(x)$  est généralement, sous certaines hypothèses de régularité, une sous-variété différentielle de dimension  $p$  dans l'espace  $\mathbb{R}^m$ , c'est-à-dire un sous-ensemble de mesure de Lebesgue nulle

la distribution de probabilité conditionnelle du vecteur aléatoire  $X_k$  sachant  $X_{k-1} = x$  ne peut donc pas avoir de densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}^m$



**Proposition** la suite  $\{Y_k\}$  vérifie l'hypothèse de *canal sans mémoire*, c'est-à-dire que pour tout instant  $n$

- ▶ conditionnellement aux états cachés  $X_{0:n}$  les observations  $Y_{0:n}$  sont mutuellement indépendantes, ce qui se traduit par

$$\mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n}] = \prod_{k=0}^n \mathbb{P}[Y_k \in dy_k \mid X_{0:n}]$$

- ▶ pour tout  $k = 0 \dots n$ , la loi conditionnelle de  $Y_k$  sachant  $X_{0:n}$  ne dépend que de  $X_k$ , ce qui se traduit par

$$\mathbb{P}[Y_k \in dy_k \mid X_{0:n}] = \mathbb{P}[Y_k \in dy_k \mid X_k]$$

avec la *densité d'émission*

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = q_k^V(y - h_k(x)) dy$$



on définit la *fonction de vraisemblance*

$$g_k(x) = q_k^V(Y_k - h_k(x))$$

qui mesure l'adéquation de  $x \in \mathbb{R}^m$  avec l'observation  $Y_k$   
 la loi conditionnelle jointe des observations  $Y_{0:n}$  sachant les états  $X_{0:n}$   
 vérifie

$$\mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n} = x_{0:n}] = \prod_{k=0}^n q_k^V(y_k - h_k(x_k)) dy_0 \cdots dy_n$$



**Preuve** compte tenu que  $V_{0:n}$  et  $X_{0:n}$  sont mutuellement indépendants

$$\begin{aligned}
 & \mathbb{E}[\phi_0(Y_0) \cdots \phi_n(Y_n) \mid X_{0:n}] \\
 &= \mathbb{E}[\phi_0(h_0(X_0) + V_0) \cdots \phi_n(h_n(X_n) + V_n) \mid X_{0:n}] \\
 &= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \phi_0(h_0(X_0) + v_0) \cdots \phi_n(h_n(X_n) + v_n) \mathbb{P}[V_{0:n} \in dv_{0:n}] \\
 &= \prod_{k=0}^n \int_{\mathbb{R}^d} \phi_k(h_k(X_k) + v_k) \mathbb{P}[V_k \in dv_k] = \prod_{k=0}^n \mathbb{E}[\phi_k(Y_k) \mid X_k]
 \end{aligned}$$

et le changement de variable  $y = h_k(x) + v$  donne immédiatement

$$\begin{aligned}
 \mathbb{E}[\phi_k(Y_k) \mid X_k = x] &= \int_{\mathbb{R}^d} \phi_k(h_k(x) + v) q_k^V(v) dv \\
 &= \int_{\mathbb{R}^d} \phi_k(y) q_k^V(y - h_k(x)) dy \quad \square
 \end{aligned}$$



## Modèles de Markov cachés

plus généralement, on peut supposer que les états cachés  $\{X_k\}$  forment une chaîne de Markov à valeurs dans un espace  $E$  qui peut être très général, par exemple

- ▶ un espace hybride continu / discret
- ▶ un sous-ensemble défini par des contraintes
- ▶ une variété différentielle
- ▶ un graphe, etc.

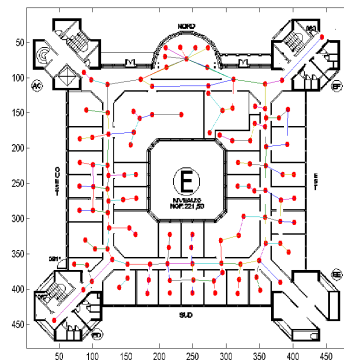
y compris trajectorienl, de loi initiale

$$\mathbb{P}[X_0 \in dx] = \eta_0(dx)$$

et de *noyau de transition*

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = Q_k(x, dx')$$





**Figure:** modèle simplifié : l'utilisateur se déplace sur un graphe de Voronoï, qui prend en compte automatiquement les contraintes de déplacement





on suppose également que les observations  $\{Y_k\}$  vérifient l'hypothèse de *canal sans mémoire*, c'est-à-dire que pour tout instant  $n$

- ▶ conditionnellement aux états cachés  $X_{0:n}$  les observations  $Y_{0:n}$  sont mutuellement indépendantes, ce qui se traduit par

$$\mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n}] = \prod_{k=0}^n \mathbb{P}[Y_k \in dy_k \mid X_{0:n}]$$

- ▶ pour tout  $k = 0 \cdots n$ , la loi conditionnelle de  $Y_k$  sachant  $X_{0:n}$  ne dépend que de  $X_k$ , ce qui se traduit par

$$\mathbb{P}[Y_k \in dy \mid X_{0:n}] = \mathbb{P}[Y_k \in dy \mid X_k]$$

avec la *densité d'émission*

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = g_k(x, y) \lambda_k^F(dy)$$

où la mesure positive  $\lambda_k^F(dy)$  définie sur  $F$  ne dépend pas de  $x \in E$



par abus de notation on définit la *fonction de vraisemblance*

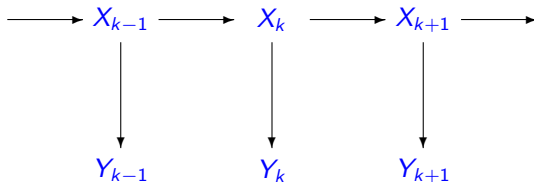
$$g_k(x) = g_k(x, Y_k)$$

qui mesure l'adéquation de  $x \in E$  avec l'observation  $Y_k$

la loi conditionnelle jointe des observations  $Y_{0:n}$  sachant les états  $X_{0:n}$  vérifie

$$\mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n} = x_{0:n}] = \prod_{k=0}^n g_k(x_k, y_k) \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n)$$

la situation est complètement décrite par



où les flèches représentent la dépendance entre variables aléatoires



pour la suite il suffit de savoir

- ▶ *simuler* pour tout  $x \in E$ , une v.a. selon le noyau de transition  $Q_k(x, dx')$
- ▶ *évaluer* pour tout  $x' \in E$ , la fonction de vraisemblance  $g_k(x')$

pour tout instant  $k = 1 \cdots n$



## Chaînes de Markov partiellement observées

encore plus généralement, on peut considérer un modèle où les états cachés  $\{X_k\}$  ne forment plus nécessairement une chaîne de Markov, mais où *conjointement* états cachés et observations  $\{X_k, Y_k\}$  forment une chaîne de Markov à valeurs dans  $E \times F$

- ▶ de loi initiale

$$\mathbb{P}[X_0 \in dx, Y_0 \in dy] = \gamma_0(y, dx) \lambda_0^F(dy)$$

où la mesure positive  $\lambda_0^F(dy)$  définie sur  $F$ , ne dépend pas de  $x \in E$

- ▶ et de *noyau de transition*

$$\begin{aligned} \mathbb{P}[X_k \in dx', Y_k \in dy' \mid X_{k-1} = x, Y_{k-1} = y] \\ = R_k(y, y', x, dx') \lambda_k^F(y, dy') \end{aligned}$$

où la mesure positive  $\lambda_k^F(y, dy')$  définie sur  $F$ , dépend de l'observation précédente  $y \in F$  mais ne dépend pas de la transition  $(x, x') \in E$



la loi jointe des états cachés  $X_{0:n}$  et des observations  $Y_{0:n}$  vérifie

$$\begin{aligned} \mathbb{P}[X_{0:n} \in dx_{0:n}, Y_{0:n} \in dy_{0:n}] &= \gamma_0(y_0, dx_0) \lambda_0^F(dy_0) \prod_{k=1}^n R_k(y_{k-1}, y_k, x_{k-1}, dx_k) \lambda_k^F(y_{k-1}, dy_k) \\ &= [\gamma_0(y_0, dx_0) \prod_{k=1}^n R_k(y_{k-1}, y_k, x_{k-1}, dx_k)] \lambda_0^F(dy_0) \prod_{k=1}^n \lambda_k^F(y_{k-1}, dy_k) \end{aligned}$$

ce modèle très général inclut comme cas particulier les modèles de Markov cachés, avec

$$\gamma_0(y, dx) = \eta_0(dx) g_0(x, y) \quad \text{et} \quad R_k(y', x, dx') = Q_k(x, dx') g_k(x', y')$$



## Modèles

### Filtre (et lisseur) bayésien

modèles de Markov cachés : représentation probabiliste

modèles de Markov cachés : équation récurrente

chaînes de Markov partiellement observées : représentation intégrale

chaînes de Markov partiellement observées : équation récurrente

## Distributions de Feynman–Kac



# Modèles de Markov cachés : représentation probabiliste

**Théorème** le filtre bayésien  $\mu_n$  défini par

$$\langle \mu_n, \phi \rangle = \mathbb{E}[\phi(X_n) \mid Y_{0:n}]$$

admet la représentation probabiliste

$$\langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, \mathbf{1} \rangle} \quad \text{avec} \quad \langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)]$$

**Remarque** l'espérance porte seulement sur les états cachés successifs  $X_{0:n}$  : les fonctions de vraisemblance  $g_k(x)$  sont définies par abus de notation comme

$$g_k(x) = g_k(x, Y_k)$$

pour tout  $k = 0 \cdots n$ , et dépendent implicitement des observations, mais celles-ci sont considérées comme *fixées*



**Preuve** d'après la formule de Bayes, et d'après la propriété de canal sans mémoire, la loi jointe des états cachés  $X_{0:n}$  et des observations  $Y_{0:n}$  vérifie

$$\begin{aligned} \mathbb{P}[X_{0:n} \in dx_{0:n}, Y_{0:n} \in dy_{0:n}] &= \mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n} = x_{0:n}] \mathbb{P}[X_{0:n} \in dx_{0:n}] \\ &= \mathbb{P}[X_{0:n} \in dx_{0:n}] \prod_{k=0}^n g_k(x_k, y_k) \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n) \end{aligned}$$

en intégrant par rapport aux variables  $x_{0:n}$ , on obtient la loi jointe des observations  $Y_{0:n}$ , c'est-à-dire

$$\begin{aligned} \mathbb{P}[Y_{0:n} \in dy_{0:n}] &= \int_E \cdots \int_E \prod_{k=0}^n g_k(x_k, y_k) \mathbb{P}[X_{0:n} \in dx_{0:n}] \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n) \\ &= \mathbb{E} \left[ \prod_{k=0}^n g_k(X_k, y_k) \right] \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n) \end{aligned}$$





d'après la formule de Bayes, il vient

$$\begin{aligned}
 & \mathbb{P}[X_{0:n} \in dx_{0:n}, Y_{0:n} \in dy_{0:n}] \\
 &= \mathbb{P}[X_{0:n} \in dx_{0:n}] \prod_{k=0}^n g_k(x_k, y_k) \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n) \\
 &= \mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] \mathbb{P}[Y_{0:n} \in dy_{0:n}] \\
 &= \mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] \mathbb{E}\left[\prod_{k=0}^n g_k(X_k, y_k)\right] \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n)
 \end{aligned}$$

et on obtient

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] = \frac{\prod_{k=0}^n g_k(x_k, y_k) \mathbb{P}[X_{0:n} \in dx_{0:n}]}{\mathbb{E}\left[\prod_{k=0}^n g_k(X_k, y_k)\right]}$$



comme cette identité est vérifiée pour toute suite  $y_{0:n}$  d'observations, on obtient

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n}] = \frac{\prod_{k=0}^n g_k(x_k) \mathbb{P}[X_{0:n} \in dx_{0:n}]}{\mathbb{E}[\prod_{k=0}^n g_k(X_k)]}$$

soit

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n}] \propto \underbrace{\prod_{k=0}^n g_k(x_k)}_{g_{0:n}(x_{0:n})} \underbrace{\eta_0(dx_0) \prod_{k=1}^n Q_k(x_{k-1}, dx_k)}_{\eta_{0:n}(dx_{0:n})}$$

avec

$$g_{0:n}(x_{0:n}) = \prod_{k=0}^n g_k(x_k)$$

et avec la loi jointe des états cachés  $X_{0:n}$

$$\eta_{0:n}(dx_{0:n}) = \mathbb{P}[X_{0:n} \in dx_{0:n}] = \eta_0(dx_0) \prod_{k=1}^n Q_k(x_{k-1}, dx_k)$$



pour toute fonction  $f$  définie sur l'espace des trajectoires

$$\begin{aligned} \mathbb{E}[f(X_{0:n}) \mid Y_{0:n}] &= \frac{\int_E \cdots \int_E f(x_{0:n}) \prod_{k=0}^n g_k(x_k) \mathbb{P}[X_{0:n} \in dx_{0:n}]}{\mathbb{E}\left[\prod_{k=0}^n g_k(X_k)\right]} \\ &= \frac{\mathbb{E}\left[f(X_{0:n}) \prod_{k=0}^n g_k(X_k)\right]}{\mathbb{E}\left[\prod_{k=0}^n g_k(X_k)\right]} \end{aligned}$$

en particulier pour toute fonction ne dépendant que de l'état final et pas de la trajectoire entière

$$\mathbb{E}[\phi(X_n) \mid Y_{0:n}] = \frac{\mathbb{E}\left[\phi(X_n) \prod_{k=0}^n g_k(X_k)\right]}{\mathbb{E}\left[\prod_{k=0}^n g_k(X_k)\right]} = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, \mathbf{1} \rangle} \quad \square$$



de la même manière, on montre que le prédicteur bayésien  $\mu_n^-$  défini par

$$\langle \mu_n^-, \phi \rangle = \mathbb{E}[\phi(X_n) \mid Y_{0:n-1}]$$

admet la représentation probabiliste

$$\langle \mu_n^-, \phi \rangle = \frac{\langle \gamma_n^-, \phi \rangle}{\langle \gamma_n^-, \mathbf{1} \rangle} \quad \text{avec} \quad \langle \gamma_n^-, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^{n-1} g_k(X_k)]$$

et on montre que le lisseur bayésien  $\mu_{k|n}$  défini par

$$\langle \mu_{k|n}, \phi \rangle = \mathbb{E}[\phi(X_k) \mid Y_{0:n}]$$

admet la représentation probabiliste

$$\langle \mu_{k|n}, \phi \rangle = \frac{\langle \gamma_{k|n}, \phi \rangle}{\langle \gamma_{k|n}, \mathbf{1} \rangle} \quad \text{avec} \quad \langle \gamma_{k|n}, \phi \rangle = \mathbb{E}[\phi(X_k) \prod_{p=0}^n g_p(X_p)]$$



## Modèles de Markov cachés : équation récurrente

**Théorème** la suite  $\{\mu_k\}$  vérifie l'équation récurrente suivante

$$\mu_{k-1} \xrightarrow{\text{prédiction}} \mu_k^- = \mu_{k-1} Q_k \xrightarrow{\text{correction}} \mu_k = g_k \cdot \mu_k^-$$

avec la condition initiale  $\mu_0^- = \eta_0$

**Remarque** dans l'énoncé du théorème, la notation

$$\mu_{k-1} Q_k(dx') = \int_E \mu_{k-1}(dx) Q_k(x, dx')$$

désigne l'action du noyau markovien  $Q_k(x, dx')$  sur la distribution de probabilité  $\mu_{k-1}(dx)$ , et la notation

$$g_k \cdot \mu_k^- = \frac{g_k \mu_k^-}{\langle \mu_k^-, g_k \rangle}$$

désigne le produit projectif de la distribution de probabilité a priori  $\mu_k^-(dx')$  et de la fonction de vraisemblance  $g_k(x')$

**Proposition** de manière équivalente, en une seule étape

$$\mu_k = \frac{\mu_{k-1} R_k}{\langle \mu_{k-1} R_k, 1 \rangle}$$

avec le noyau positif (non-normalisé)  $R_k(x, dx') = Q_k(x, dx') g_k(x')$

**Preuve** pour toute fonction  $\phi$

$$\begin{aligned} \langle \mu_{k-1} R_k, \phi \rangle &= \int_E \left[ \int_E \mu_{k-1}(dx) R_k(x, dx') \right] \phi(x') \\ &= \int_E \int_E \mu_{k-1}(dx) Q_k(x, dx') g_k(x') \phi(x') \\ &= \int_E \left[ \int_E \mu_{k-1}(dx) Q_k(x, dx') \right] g_k(x') \phi(x') \\ &= \int_E \mu_k^-(dx') g_k(x') \phi(x') = \langle \mu_k^-, g_k \phi \rangle \end{aligned}$$



en particulier pour  $\phi \equiv 1$

$$\langle \mu_{k-1} R_k, 1 \rangle = \langle \mu_k^-, g_k \rangle$$

et en normalisant, on vérifie que

$$\frac{\langle \mu_{k-1} R_k, \phi \rangle}{\langle \mu_{k-1} R_k, 1 \rangle} = \frac{\langle \mu_k^-, g_k \phi \rangle}{\langle \mu_k^-, g_k \rangle} = \langle \mu_k, \phi \rangle \quad \square$$



**Preuve du Théorème** on procède en deux étapes, correspondant respectivement aux étapes de prédiction et de correction, et en raisonnant d'abord sur les versions non normalisées, puis en normalisant

► **étape de prédiction** expression de  $\mu_n^-$  en fonction de  $\mu_{n-1}$   
on remarque immédiatement que

$$\langle \gamma_n^-, 1 \rangle = \mathbb{E} \left[ \prod_{k=0}^{n-1} g_k(X_k) \right] = \langle \gamma_{n-1}, 1 \rangle$$

c'est-à-dire que la constante de normalisation est conservée





en utilisant la propriété de Markov, on a

$$\begin{aligned}
 \langle \gamma_n^-, \phi \rangle &= \mathbb{E}[\phi(X_n) \prod_{k=0}^{n-1} g_k(X_k)] \\
 &= \mathbb{E}[\mathbb{E}[\phi(X_n) \mid X_{0:n-1}] \prod_{k=0}^{n-1} g_k(X_k)] \\
 &= \mathbb{E}[\mathbb{E}[\phi(X_n) \mid X_{n-1}] \prod_{k=0}^{n-1} g_k(X_k)] \\
 &= \mathbb{E}[Q_n \phi(X_{n-1}) \prod_{k=0}^{n-1} g_k(X_k)] = \langle \gamma_{n-1}, Q_n \phi \rangle = \langle \gamma_{n-1} Q_n, \phi \rangle
 \end{aligned}$$

pour toute fonction  $\phi$  définie sur  $E$



la dernière égalité exprime simplement que

$$\begin{aligned}
 \langle \gamma_{n-1}, Q_n \phi \rangle &= \int_E \gamma_{n-1}(dx) \left[ \int_E Q_n(x, dx') \phi(x') \right] \\
 &= \int_E \left[ \int_E \gamma_{n-1}(dx) Q_n(x, dx') \right] \phi(x') \\
 &= \int_E \gamma_{n-1} Q_n(dx') \phi(x') = \langle \gamma_{n-1} Q_n, \phi \rangle
 \end{aligned}$$

comme la fonction  $\phi$  est quelconque, on en déduit que

$$\gamma_n^- = \gamma_{n-1} Q_n$$

et en normalisant, on obtient

$$\mu_n^- = \frac{\gamma_n^-}{\langle \gamma_n^-, 1 \rangle} = \frac{\gamma_{n-1} Q_n}{\langle \gamma_{n-1}, 1 \rangle} = \mu_{n-1} Q_n$$



► **étape de correction** expression de  $\mu_n$  en fonction de  $\mu_n^-$   
on a simplement

$$\begin{aligned} \langle \gamma_n, \phi \rangle &= \mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)] \\ &= \mathbb{E}[\phi(X_n) g_n(X_n) \prod_{k=0}^{n-1} g_k(X_k)] = \langle \gamma_n^-, g_n \phi \rangle = \langle g_n \gamma_n^-, \phi \rangle \end{aligned}$$

pour toute fonction  $\phi$  définie sur  $E$ , où la dernière égalité exprime simplement que

$$\langle \gamma_n^-, g_n \phi \rangle = \int_E [g_n(x) \phi(x)] \gamma_n^-(dx) = \int_E \phi(x) [g_n(x) \gamma_n^-(dx)] = \langle g_n \gamma_n^-, \phi \rangle$$

comme la fonction  $\phi$  est quelconque, on en déduit que

$$\gamma_n = g_n \gamma_n^-$$

et en normalisant, on obtient

$$\mu_n = \frac{\gamma_n}{\langle \gamma_n, 1 \rangle} = \frac{g_n \gamma_n^-}{\langle \gamma_n^-, g_n \rangle} = \frac{g_n \mu_n^-}{\langle \mu_n^-, g_n \rangle} \quad \square$$



**Preuve alternative du Théorème** on rappelle la représentation probabiliste de la loi conditionnelle des états cachés  $X_{0:k}$  sachant les observations  $Y_{0:k}$

$$\begin{aligned} \mathbb{P}[X_{0:k} \in dx_{0:k} \mid Y_{0:k}] &\propto \eta_0(dx_0) \prod_{p=1}^k Q_p(x_{p-1}, dx_p) \prod_{p=0}^k g_p(x_p) \\ &\propto g_k(x_k) Q_k(x_{k-1}, dx_k) \mathbb{P}[X_{0:k-1} \in dx_{0:k-1} \mid Y_{0:k-1}] \end{aligned}$$

en intégrant par rapport aux variables  $x_{0:k-1}$  (et dans le membre de droite, d'abord par rapport aux variables  $x_{0:k-2}$  puis par rapport à la variable  $x_{k-1}$ ), on obtient la loi conditionnelle de l'état caché  $X_k$  sachant les observations  $Y_{0:k}$ , c'est-à-dire le filtre bayésien

$$\begin{aligned} \mu_k(dx_k) &= \mathbb{P}[X_k \in dx_k \mid Y_{0:k}] \\ &\propto g_k(x_k) \int_E Q_k(x_{k-1}, dx_k) \mathbb{P}[X_{k-1} \in dx_{k-1} \mid Y_{0:k-1}] \\ &\propto g_k(x_k) \int_E \mu_{k-1}(dx_{k-1}) Q_k(x_{k-1}, dx_k) \end{aligned}$$



on introduit la fonction

$$v_k(x) = \mathbb{E} \left[ \prod_{q=k+1}^n g_q(X_q) \mid X_k = x \right]$$

pour tout  $k = 0, 1, \dots, n-1$ , avec la convention  $v_n \equiv 1$  pour  $k = n$

**Proposition**

$$\gamma_{k|n} = v_k \gamma_k \quad \text{et} \quad \mu_{k|n} = \frac{v_k \mu_k}{\langle \mu_k, v_k \rangle} = v_k \cdot \mu_k$$

pour tout  $k = 0, 1, \dots, n$ , c'est-à-dire que le lisseur bayésien est absolument continu par rapport au filtre bayésien et la fonction  $v_k$  peut s'interpréter à une constante de normalisation près comme la densité (ou dérivée de Radon–Nikodym)



**Preuve** en séparant explicitement les termes qui dépendent des états passés  $X_{0:k}$  et les termes qui dépendent des états futurs  $X_{k+1:n}$ , puis en conditionnant par rapport aux états passés  $X_{0:k}$  et en utilisant la propriété de Markov, on obtient

$$\begin{aligned}
 \langle \gamma_{k|n}, \phi \rangle &= \mathbb{E} \left[ \phi(X_k) \prod_{p=0}^k g_p(X_p) \prod_{q=k+1}^n g_q(X_q) \right] \\
 &= \mathbb{E} \left[ \phi(X_k) \prod_{p=0}^k g_p(X_p) \mathbb{E} \left[ \prod_{q=k+1}^n g_q(X_q) \mid X_{0:k} \right] \right] \\
 &= \mathbb{E} \left[ \phi(X_k) \prod_{p=0}^k g_p(X_p) \mathbb{E} \left[ \prod_{q=k+1}^n g_q(X_q) \mid X_k \right] \right] \\
 &= \mathbb{E} \left[ \phi(X_k) v_k(X_k) \prod_{p=0}^k g_p(X_p) \right] = \langle \gamma_k, v_k \phi \rangle
 \end{aligned}$$



comme la fonction  $\phi$  est quelconque, on en déduit que

$$\gamma_{k|n} = v_k \gamma_k$$

en particulier, la constante de normalisation vérifie

$$\langle \gamma_{k|n}, \mathbf{1} \rangle = \langle \gamma_k, v_k \rangle$$

et en normalisant on obtient

$$\mu_{k|n} = \frac{\gamma_{k|n}}{\langle \gamma_{k|n}, \mathbf{1} \rangle} = \frac{v_k \gamma_k}{\langle \gamma_k, v_k \rangle} = \frac{v_k \mu_k}{\langle \mu_k, v_k \rangle} = v_k \cdot \mu_k \quad \square$$



plus généralement, on introduit la fonction

$$f_k(x) = \mathbb{E}[\phi(X_n) \prod_{q=k+1}^n g_q(X_q) \mid X_k = x]$$

pour tout  $k = 0, 1 \dots n-1$ , avec la convention  $f_n = \phi$  pour  $k = n$

**Proposition** la suite de fonctions  $\{f_k\}$  vérifie l'équation récurrente dans le sens rétrograde

$$f_{k-1} = Q_k(g_k f_k)$$

pour tout  $k = n, n-1 \dots 1$ , avec la condition initiale  $f_n = \phi$  pour  $k = n$   
 dans le cas particulier  $\phi \equiv 1$ , la suite de fonctions  $\{v_k\}$  vérifie l'équation récurrente dans le sens rétrograde

$$v_{k-1} = Q_k(g_k v_k)$$

pour tout  $k = n, n-1 \dots 1$ , avec la condition initiale  $v_n \equiv 1$  pour  $k = n$





## Preuve par définition

$$\begin{aligned}
 f_k(x_k) &= \mathbb{E}[\phi(X_n) \prod_{q=k+1}^n g_q(X_q) \mid X_k = x_k] \\
 &= \int_E \cdots \int_E \phi(x_n) \prod_{q=k+1}^n g_q(x_q) \mathbb{P}[X_{k+1:n} \in dx_{k+1:n} \mid X_k = x_k] \\
 &= \int_E \cdots \int_E \prod_{q=k+1}^n g_q(x_q) \prod_{q=k+1}^n Q_q(x_{q-1}, dx_q) \phi(x_n)
 \end{aligned}$$



il en résulte que

$$\begin{aligned}
 f_{k-1}(x_{k-1}) &= \int_E \cdots \int_E \prod_{q=k}^n g_q(x_q) \prod_{q=k}^n Q_q(x_{q-1}, dx_q) \phi(x_n) \\
 &= \int_E \cdots \int_E g_k(x_k) Q_k(x_{k-1}, dx_k) \prod_{q=k+1}^n g_q(x_q) \prod_{q=k+1}^n Q_q(x_{q-1}, dx_q) \phi(x_n) \\
 &= \int_E g_k(x_k) Q_k(x_{k-1}, dx_k) \\
 &\quad \int_E \cdots \int_E \prod_{q=k+1}^n g_q(x_q) \prod_{q=k+1}^n Q_q(x_{q-1}, dx_q) \phi(x_n) \\
 &= \int_E f_k(x_k) g_k(x_k) Q_k(x_{k-1}, dx_k) \\
 &= Q_k(g_k f_k)(x_{k-1}) \quad \square
 \end{aligned}$$



dans cette approche, on résoud séparément

- ▶ une équation récurrente dans le sens direct pour le filtre bayésien  $\mu_k$
- ▶ une équation récurrente dans le sens rétrograde pour la fonction  $v_k$

et on obtient le lisseur bayésien  $\mu_{k|n}$  grâce à la relation

$$\mu_{k|n} = \frac{v_k \mu_k}{\langle \mu_k, v_k \rangle} = v_k \cdot \mu_k$$



il existe des approches alternatives pour organiser les calculs

par exemple dans une deuxième approche, on résoud

- ▶ d'abord une équation récurrente dans le sens direct pour le filtre bayésien  $\mu_k$
- ▶ puis une équation récurrente dans le sens rétrograde pour le lisseur bayésien  $\mu_{k|n}$

et dans une troisième approche, on résoud

- ▶ d'abord une équation récurrente dans le sens rétrograde pour la fonction  $v_k$
- ▶ puis une équation récurrente dans le sens direct pour le lisseur bayésien  $\mu_{k|n}$

examen du 23 octobre 2020



## Chaînes de Markov partiellement observées : représentation intégrale

**Théorème** le filtre bayésien  $\mu_n$  défini par

$$\langle \mu_n, \phi \rangle = \mathbb{E}[\phi(X_n) \mid Y_{0:n}]$$

admet la représentation intégrale

$$\langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, \mathbf{1} \rangle}$$

avec

$$\langle \gamma_n, \phi \rangle = \int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k) \phi(x_n)$$

**Remarque** la mesure positive  $\gamma_0(dx)$  et les noyaux positifs (non-normalisés)  $R_k(x, dx')$  sont définis par abus de notation comme

$$\gamma_0(dx) = \gamma_0(Y_0, dx) \quad \text{et} \quad R_k(x, dx') = R_k(Y_{k-1}, Y_k, x, dx')$$

pour tout  $k = 1 \cdots n$ , et dépendent implicitement des observations



**Preuve** d'après la propriété de Markov, la loi jointe des états cachés  $X_{0:n}$  et des observations  $Y_{0:n}$  vérifie

$$\begin{aligned} \mathbb{P}[X_{0:n} \in dx_{0:n}, Y_{0:n} \in dy_{0:n}] \\ = [\gamma_0(y_0, dx_0) \prod_{k=1}^n R_k(y_{k-1}, y_k, x_{k-1}, dx_k)] \lambda_0^F(dy_0) \prod_{k=1}^n \lambda_k^F(y_{k-1}, dy_k) \end{aligned}$$

en intégrant par rapport aux variables  $x_{0:n}$ , on obtient la loi jointe des observations  $Y_{0:n}$ , c'est-à-dire

$$\begin{aligned} \mathbb{P}[Y_{0:n} \in dy_{0:n}] = \left[ \int_E \cdots \int_E \gamma_0(y_0, dx_0) \prod_{k=1}^n R_k(y_{k-1}, y_k, x_{k-1}, dx_k) \right] \\ \lambda_0^F(dy_0) \prod_{k=1}^n \lambda_k^F(y_{k-1}, dy_k) \end{aligned}$$



d'après la formule de Bayes, il vient

$$\begin{aligned}
 & \mathbb{P}[X_{0:n} \in dx_{0:n}, Y_{0:n} \in dy_{0:n}] \\
 &= [\gamma_0(y_0, dx_0) \prod_{k=1}^n R_k(y_{k-1}, y_k, x_{k-1}, dx_k)] \lambda_0^F(dy_0) \prod_{k=1}^n \lambda_k^F(y_{k-1}, dy_k) \\
 &= \mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] \mathbb{P}[Y_{0:n} \in dy_{0:n}] \\
 &= \mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] \\
 & \quad \left[ \int_E \cdots \int_E \gamma_0(y_0, dx_0) \prod_{k=1}^n R_k(y_{k-1}, y_k, x_{k-1}, dx_k) \right] \\
 & \quad \lambda_0^F(dy_0) \prod_{k=1}^n \lambda_k^F(y_{k-1}, dy_k)
 \end{aligned}$$



et on obtient

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] = \frac{\gamma_0(y_0, dx_0) \prod_{k=1}^n R_k(y_{k-1}, y_k, x_{k-1}, dx_k)}{\int_E \cdots \int_E \gamma_0(y_0, dx_0) \prod_{k=1}^n R_k(y_{k-1}, y_k, x_{k-1}, dx_k)}$$

comme cette identité est vérifiée pour toute suite  $y_{0:n}$  d'observations, on obtient

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n}] = \frac{\gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k)}{\int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k)}$$





pour toute fonction  $f$  définie sur l'espace des trajectoires

$$\mathbb{E}[f(X_{0:n}) | Y_{0:n}] = \frac{\int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k) f(x_{0:n})}{\int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k)}$$

et en particulier pour toute fonction ne dépendant que de l'état final et pas de la trajectoire entière

$$\mathbb{E}[\phi(X_n) | Y_{0:n}] = \frac{\int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k) \phi(x_n)}{\int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k)} = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, \mathbf{1} \rangle}$$





de la même manière, on montre que le lisseur bayésien  $\mu_{k|n}$  défini par

$$\langle \mu_{k|n}, \phi \rangle = \mathbb{E}[\phi(X_k) \mid Y_{0:n}]$$

admet la représentation intégrale

$$\langle \mu_{k|n}, \phi \rangle = \frac{\langle \gamma_{k|n}, \phi \rangle}{\langle \gamma_{k|n}, \mathbf{1} \rangle}$$

avec

$$\langle \gamma_{k|n}, \phi \rangle = \int_E \cdots \int_E \gamma_0(dx_0) \prod_{p=1}^n R_p(x_{p-1}, dx_p) \phi(x_k)$$



## Chaînes de Markov partiellement observées : équation récurrente

**Théorème** la suite  $\{\mu_k\}$  vérifie l'équation récurrente suivante

$$\mu_k = \frac{\mu_{k-1} R_k}{\langle \mu_{k-1} R_k, 1 \rangle} \quad \text{avec} \quad \mu_0 = \frac{\gamma_0}{\langle \gamma_0, 1 \rangle}$$



Preuve on a

$$\begin{aligned}
 \langle \gamma_n, \phi \rangle &= \int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k) \phi(x_n) \\
 &= \int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^{n-1} R_k(x_{k-1}, dx_k) \int_E R_n(x_{n-1}, dx_n) \phi(x_n) \\
 &= \int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^{n-1} R_k(x_{k-1}, dx_k) R_n \phi(x_{n-1}) \\
 &= \langle \gamma_{n-1}, R_n \phi \rangle = \langle \gamma_{n-1} R_n, \phi \rangle
 \end{aligned}$$

comme la fonction  $\phi$  est quelconque, on en déduit que

$$\gamma_n = \gamma_{n-1} R_n$$

et en normalisant, on obtient

$$\mu_n = \frac{\gamma_n}{\langle \gamma_n, 1 \rangle} = \frac{\gamma_{n-1} R_n}{\langle \gamma_{n-1} R_n, 1 \rangle} = \frac{\mu_{n-1} R_n}{\langle \mu_{n-1} R_n, 1 \rangle}$$





on introduit la fonction

$$v_k(x_k) = \int_E \cdots \int_E \prod_{q=k+1}^n R_q(x_{q-1}, dx_q)$$

pour tout  $k = 0, 1, \dots, n-1$ , avec la convention  $v_n \equiv 1$  pour  $k = n$

### Proposition

$$\gamma_{k|n} = v_k \gamma_k \quad \text{et} \quad \mu_{k|n} = \frac{v_k \mu_k}{\langle \mu_k, v_k \rangle} = v_k \cdot \mu_k$$

pour tout  $k = 0, 1, \dots, n$ , c'est-à-dire que le lisseur bayésien est absolument continu par rapport au filtre bayésien et la fonction  $v_k$  peut s'interpréter à une constante de normalisation près comme la densité (ou dérivée de Radon–Nikodym)



**Preuve** en séparant explicitement les termes qui dépendent des variables  $x_{0:k}$  et les termes qui dépendent des variables  $x_{k+1:n}$ , on obtient

$$\begin{aligned}
 \langle \gamma_k | n, \phi \rangle &= \int_E \cdots \int_E \gamma_0(dx_0) \prod_{p=1}^n R_p(x_{p-1}, dx_p) \phi(x_k) \\
 &= \int_E \cdots \int_E \gamma_0(dx_0) \prod_{p=1}^k R_p(x_{p-1}, dx_p) \phi(x_k) \prod_{q=k+1}^n R_q(x_{q-1}, dx_q) \\
 &= \int_E \cdots \int_E \gamma_0(dx_0) \prod_{p=1}^k R_p(x_{p-1}, dx_p) \phi(x_k) \\
 &\quad \int_E \cdots \int_E \prod_{q=k+1}^n R_q(x_{q-1}, dx_q) \\
 &= \int_E \cdots \int_E \gamma_0(dx_0) \prod_{p=1}^k R_p(x_{p-1}, dx_p) \phi(x_k) v_k(x_k) = \langle \gamma_k, v_k \phi \rangle
 \end{aligned}$$



comme la fonction  $\phi$  est quelconque, on en déduit que

$$\gamma_{k|n} = v_k \gamma_k$$

en particulier, la constante de normalisation vérifie

$$\langle \gamma_{k|n}, \mathbf{1} \rangle = \langle \gamma_k, v_k \rangle$$

et en normalisant on obtient

$$\mu_{k|n} = \frac{\gamma_{k|n}}{\langle \gamma_{k|n}, \mathbf{1} \rangle} = \frac{v_k \gamma_k}{\langle \gamma_k, v_k \rangle} = \frac{v_k \mu_k}{\langle \mu_k, v_k \rangle} = v_k \cdot \mu_k \quad \square$$



plus généralement, on introduit la fonction

$$f_k(x) = \int_E \cdots \int_E \prod_{q=k+1}^n R_q(x_{q-1}, dx_q) \phi(x_n)$$

pour tout  $k = 0, 1, \dots, n-1$ , avec la convention  $f_n = \phi$  pour  $k = n$

**Proposition** la suite de fonctions  $\{f_k\}$  vérifie l'équation récurrente dans le sens rétrograde

$$f_{k-1} = R_k f_k$$

pour tout  $k = n, n-1, \dots, 1$ , avec la condition initiale  $f_n = \phi$  pour  $k = n$   
 dans le cas particulier  $\phi \equiv 1$ , la suite de fonctions  $\{v_k\}$  vérifie l'équation récurrente dans le sens rétrograde

$$v_{k-1} = R_k v_k$$

pour tout  $k = n, n-1, \dots, 1$ , avec la condition initiale  $v_n \equiv 1$  pour  $k = n$





## Preuve par définition

$$\begin{aligned}
 f_{k-1}(x_{k-1}) &= \int_E \cdots \int_E \prod_{q=k}^n R_q(x_{q-1}, dx_q) \phi(x_n) \\
 &= \int_E \cdots \int_E R_k(x_{k-1}, dx_k) \prod_{q=k+1}^n R_q(x_{q-1}, dx_q) \phi(x_n) \\
 &= \int_E R_k(x_{k-1}, dx_k) \int_E \cdots \int_E \prod_{q=k+1}^n R_q(x_{q-1}, dx_q) \phi(x_n) \\
 &= \int_E R_k(x_{k-1}, dx_k) f_k(x_k) \\
 &= R_k f_k(x_{k-1}) \quad \square
 \end{aligned}$$



## Modèles

## Filtre (et lisseur) bayésien

## Distributions de Feynman–Kac

équation récurrente

décomposition d'importance et représentation probabiliste



## Distributions de Feynman–Kac : équation récurrente

distribution non-normalisée et distribution normalisée associée

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)] \quad \text{et} \quad \langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, \mathbf{1} \rangle}$$

avec  $\{X_k\}$  chaîne de Markov à valeurs dans  $E$ , caractérisée par

- ▶ la loi initiale  $\eta_0(dx)$
- ▶ et le *noyau de transition*  $Q_k(x, dx')$

et avec la *fonction de pondération / sélection*  $g_k(x)$



plus généralement

$$\langle \gamma_n, \phi \rangle = \int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k) \phi(x_n)$$

avec

- ▶ la mesure positive  $\gamma_0(dx)$  non-normalisée
- ▶ et le noyau positif  $R_k(x, dx')$  non-normalisé

équation récurrente (déjà vue)

$$\gamma_k = \gamma_{k-1} R_k \quad \text{et} \quad \mu_k = \frac{\mu_{k-1} R_k}{\langle \mu_{k-1} R_k, 1 \rangle}$$



## Distributions de Feynman–Kac : représentation probabiliste

décomposition d'importance (non unique)

$$\gamma_0(dx) = g_0^{\text{imp}}(x) \eta_0^{\text{imp}}(dx) \quad \text{et} \quad R_k(x, dx') = g_k^{\text{imp}}(x, x') Q_k^{\text{imp}}(x, dx')$$

comme le produit de

- ▶ la *fonction de pondération* positive  $g_0^{\text{imp}}(x)$  ou  $g_k^{\text{imp}}(x, x')$
- ▶ et la distribution de probabilité  $\eta_0^{\text{imp}}(dx)$  ou le *noyau markovien*  $Q_k^{\text{imp}}(x, dx')$

en pratique, la décomposition doit être telle qu'il est facile

- ▶ de *simuler* pour tout  $x \in E$ , une v.a. selon le noyau markovien  $Q_k^{\text{imp}}(x, dx')$
- ▶ d'*évaluer* pour tout  $x, x' \in E$ , la fonction de pondération  $g_k^{\text{imp}}(x, x')$

avec la convention  $g_0^{\text{imp}}(x, x') = g_0^{\text{imp}}(x')$

attention : la fonction de pondération  $g_k^{\text{imp}}(x, x')$  dépend de la transition  $(x, x')$  et pas seulement de l'état d'arrivée  $x'$



cas particulier des modèles de Markov cachés : premier exemple de décomposition

$$\gamma_0(dx) = g_0(x) \eta_0(dx) \quad \text{et} \quad R_k(x, dx') = Q_k(x, dx') g_k(x')$$

autre exemple de décomposition

$$\gamma_0(dx) = \langle \eta_0, g_0 \rangle \underbrace{\frac{g_0(x) \eta_0(dx)}{\langle \eta_0, g_0 \rangle}}_{\hat{\eta}_0(dx)} \quad \text{et} \quad R_k(x, dx') = \underbrace{R_k(x, E)}_{\hat{g}_k(x)} \underbrace{\frac{R_k(x, dx')}{R_k(x, E)}}_{\hat{Q}_k(x, dx')}$$

où

$$\hat{g}_k(x) = R_k(x, E) = \int_E Q_k(x, dx') g_k(x')$$

interprétation : mesure quantitative du recouvrement entre l'application  $g_k(x')$  et la distribution de probabilité  $Q_k(x, dx')$

et où

$$\hat{Q}_k(x, dx') = \frac{R_k(x, dx')}{R_k(x, E)}$$



pour une décomposition d'importance donnée

$$\begin{aligned}
 \langle \gamma_n, \phi \rangle &= \int_E \cdots \int_E \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k) \phi(x_n) \\
 &= \int_E \cdots \int_E \eta_0^{\text{imp}}(dx_0) \prod_{k=1}^n Q_k^{\text{imp}}(x_{k-1}, dx_k) \phi(x_n) \prod_{k=0}^n g_k^{\text{imp}}(x_{k-1}, x_k) \\
 &= \mathbb{E}[\phi(X_n^{\text{imp}}) \prod_{k=0}^n g_k^{\text{imp}}(X_{k-1}^{\text{imp}}, X_k^{\text{imp}})]
 \end{aligned}$$

où la suite  $\{X_k^{\text{imp}}\}$  est une chaîne de Markov, caractérisée par

- ▶ la loi initiale  $\eta_0^{\text{imp}}(dx)$
- ▶ et le *noyau de transition*  $Q_k^{\text{imp}}(x, dx')$

et avec la *fonction de pondération / sélection*  $g_k^{\text{imp}}(x, x')$



## Rationale : chaîne de Markov à valeurs transitions

on définit la v.a.  $X_k^{\text{tr}} = (X_{k-1}^{\text{imp}}, X_k^{\text{imp}})$  à valeurs dans l'ensemble produit  $E^{\text{tr}} = E \times E$ , et la v.a.  $X_0^{\text{tr}} = X_0^{\text{imp}}$  à valeurs dans  $E$

**Proposition** la suite  $\{X_k^{\text{tr}}\}$  est une chaîne de Markov, caractérisée par

- ▶ la loi initiale  $\eta_0^{\text{tr}}(dx) = \eta_0^{\text{imp}}(dx)$
- ▶ et le noyau de transition  $Q_k^{\text{tr}}(x_1, x_2, dx'_1, dx'_2)$  défini par

$$Q_k^{\text{tr}}(x_1, x_2, dx'_1, dx'_2) = \delta_{x_2}(dx'_1) Q_k^{\text{imp}}(x'_1, dx'_2)$$

c'est-à-dire que le point de départ de la nouvelle transition est juste le point d'arrivée de la transition précédente et le point d'arrivée de la nouvelle transition est distribué à partir du point de départ selon le noyau de transition  $Q_k^{\text{imp}}(x'_1, dx'_2)$

**Remarque** pour simuler une v.a.  $X^{\text{tr}}$  distribuée selon  $Q_k^{\text{tr}}(x_1, x_2, dx'_1, dx'_2)$ , où  $(x_1, x_2)$  est fixé, il suffit de simuler une v.a.  $X'$  distribuée selon  $Q_k^{\text{imp}}(x_2, dx'_2)$  et de poser  $X^{\text{tr}} = (x_2, X')$





on considère la distribution non-normalisée et la distribution normalisée associée

$$\langle \gamma_n^{\text{tr}}, f \rangle = \mathbb{E}[f(X_n^{\text{tr}}) \prod_{k=0}^n g_k(X_k^{\text{tr}})] \quad \text{et} \quad \langle \mu_n^{\text{tr}}, f \rangle = \frac{\langle \gamma_n^{\text{tr}}, f \rangle}{\langle \gamma_n^{\text{tr}}, 1 \rangle}$$

avec les fonctions de sélection

$$g_0(x, x') = g_0^{\text{imp}}(x') \quad \text{et} \quad g_k(x, x') = g_k^{\text{imp}}(x, x')$$

on vérifie que

$$\begin{aligned} \langle \gamma_n, \phi \rangle &= \mathbb{E}[\phi(X_n^{\text{imp}}) \prod_{k=1}^n g_k^{\text{imp}}(X_{k-1}^{\text{imp}}, X_k^{\text{imp}})] \\ &= \mathbb{E}[\phi \circ \pi(X_n^{\text{tr}}) \prod_{k=0}^n g_k(X_k^{\text{tr}})] = \langle \gamma_n^{\text{tr}}, \phi \circ \pi \rangle \end{aligned}$$

où l'application

$$\pi : (x, x') \in E^{\text{tr}} = E \times E \longmapsto x' \in E$$

pointe sur l'état d'arrivée de la transition