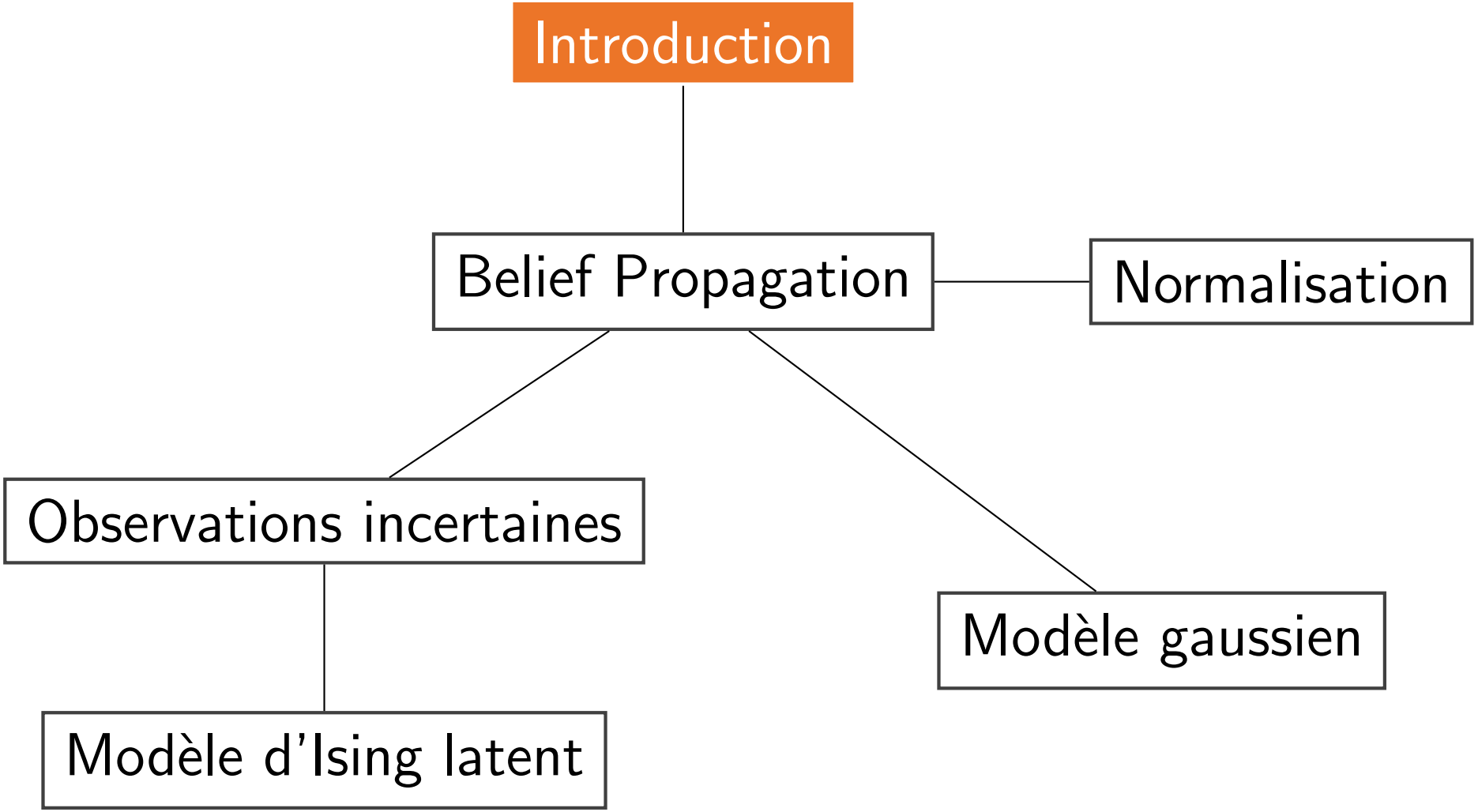
Modélisation probabiliste et inférence par l'algorithme Belief Propagation

Victorin MARTIN

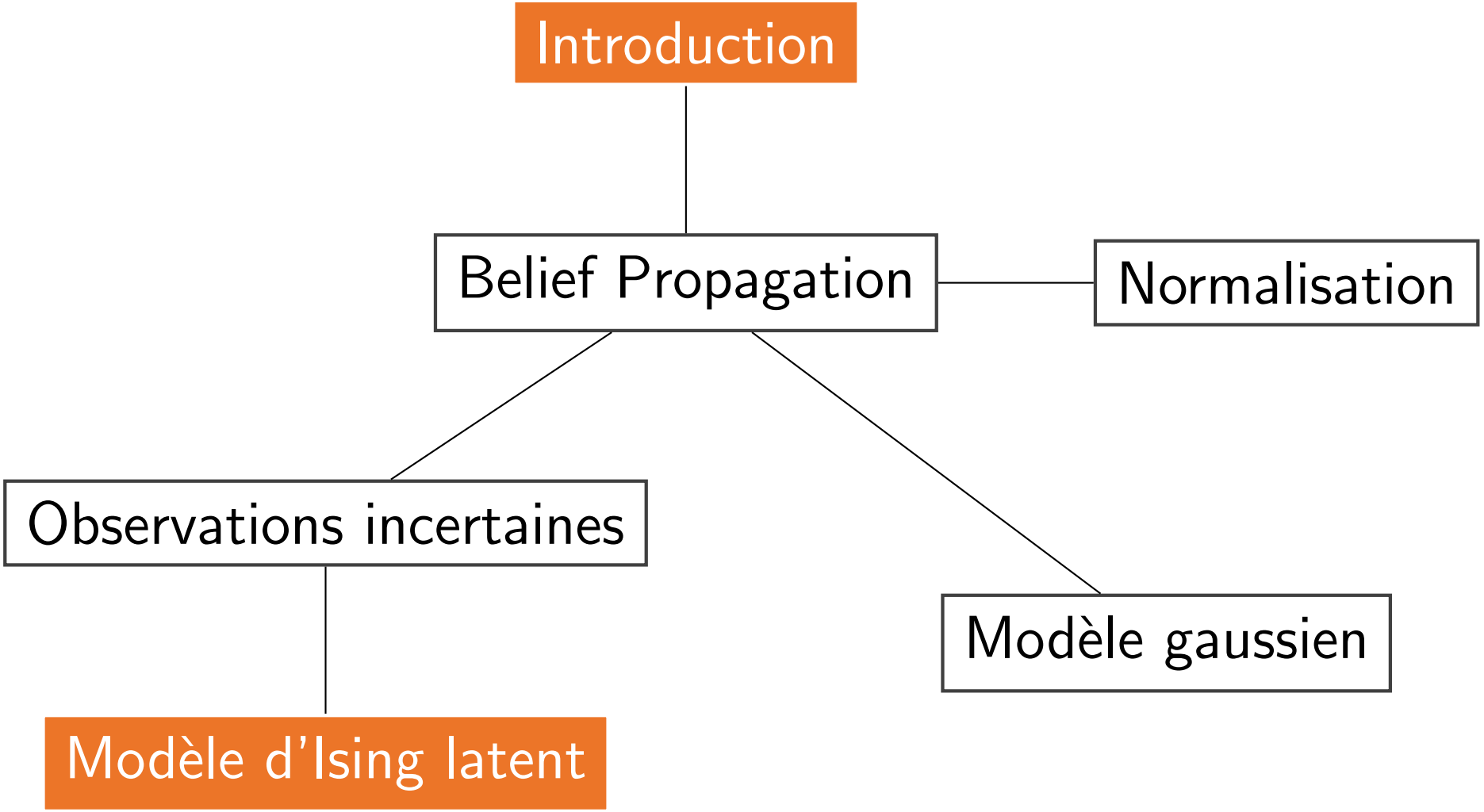
Directeur de thèse: Arnaud de LA FORTELLE.
Encadrants: Cyril FURTLEHNER & Jean-Marc LASGOUTTES.

INRIA Rocquencourt
Équipes-projets IMARA et TAO

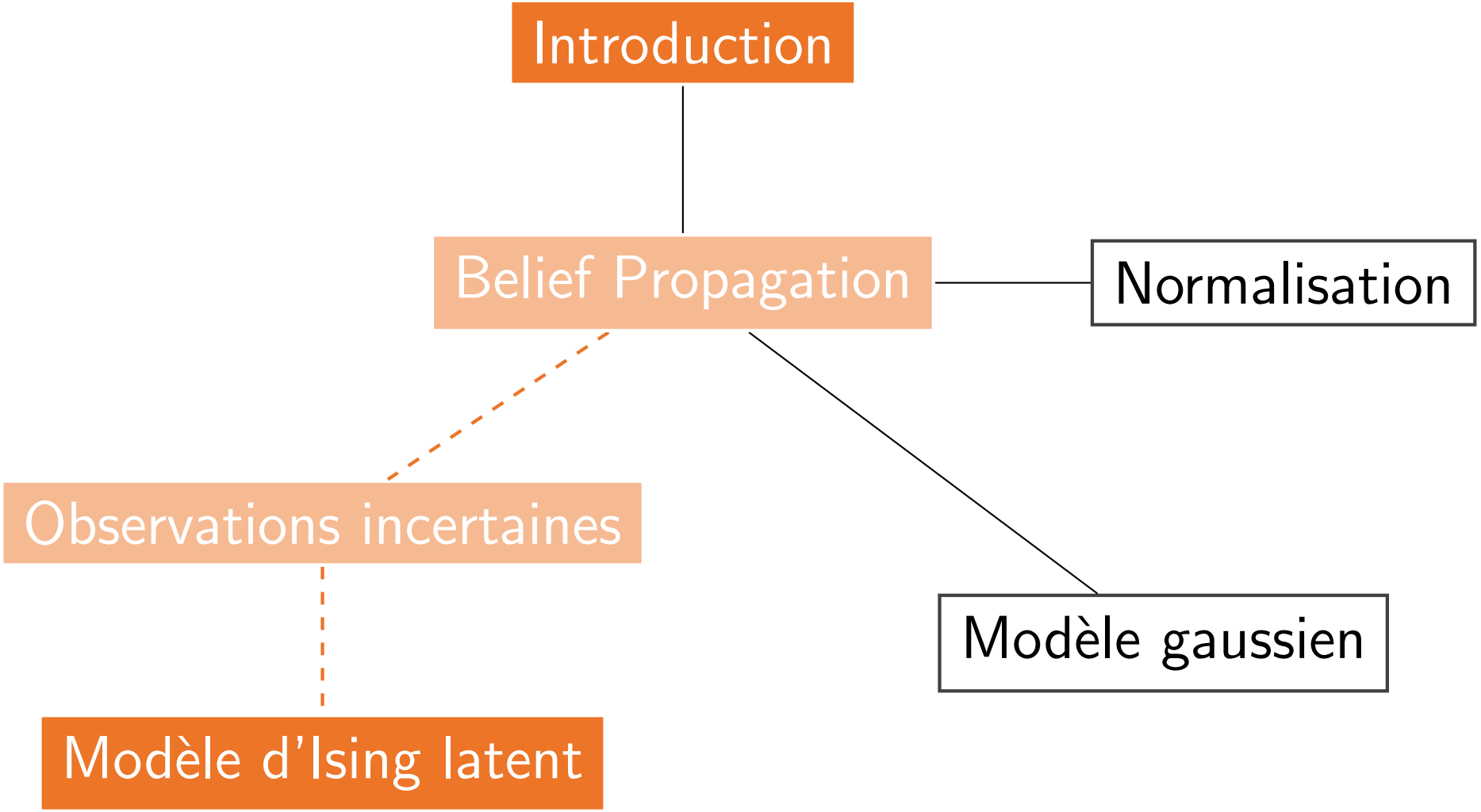
Présentation vs. mémoire



Présentation vs. mémoire



Présentation vs. mémoire



Contexte : données issues de véhicules sondes

Projet TRAVESTI

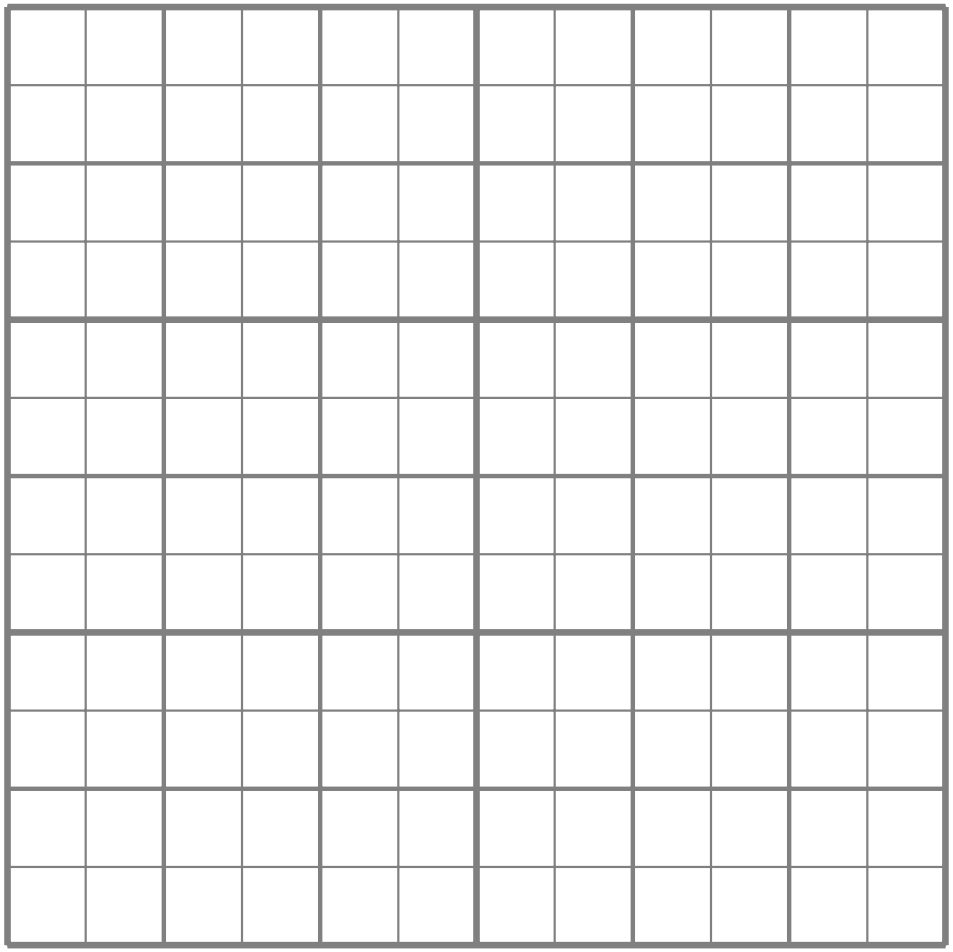
- *TRAffic Volume Estimation by Spatio-Temporal Inference*,
- Modélisation de systèmes complexes de grande taille à partir d'interactions locales dans but de prédire leur comportement macroscopique.

Projet PUMAS

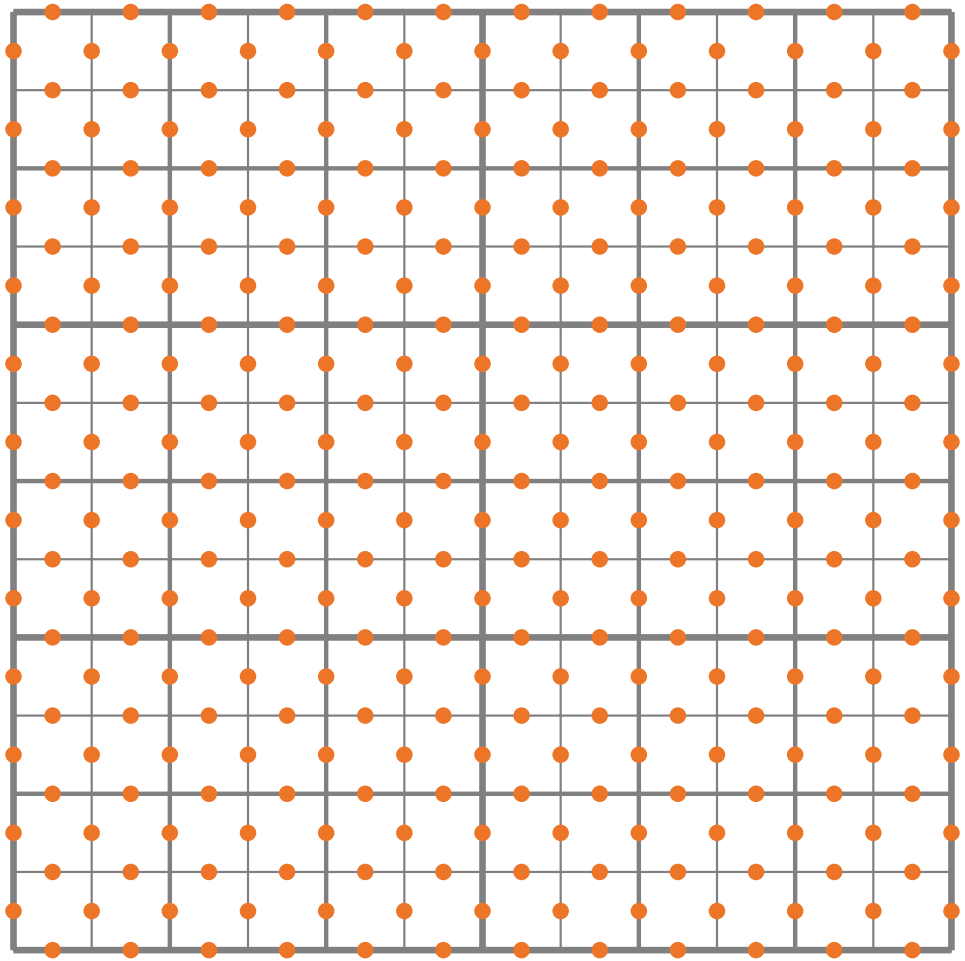
- *Plateforme Urbaine de Mobilité Avancée et Soutenable*.
- Objectif : prédiction de temps de trajet sur un réseau routier urbain et péri-urbain.
- Mise en place sur l'agglomération de Rouen.

Approche basée sur les véhicules plutôt que sur l'infrastructure

Approche classique : capteurs fixes

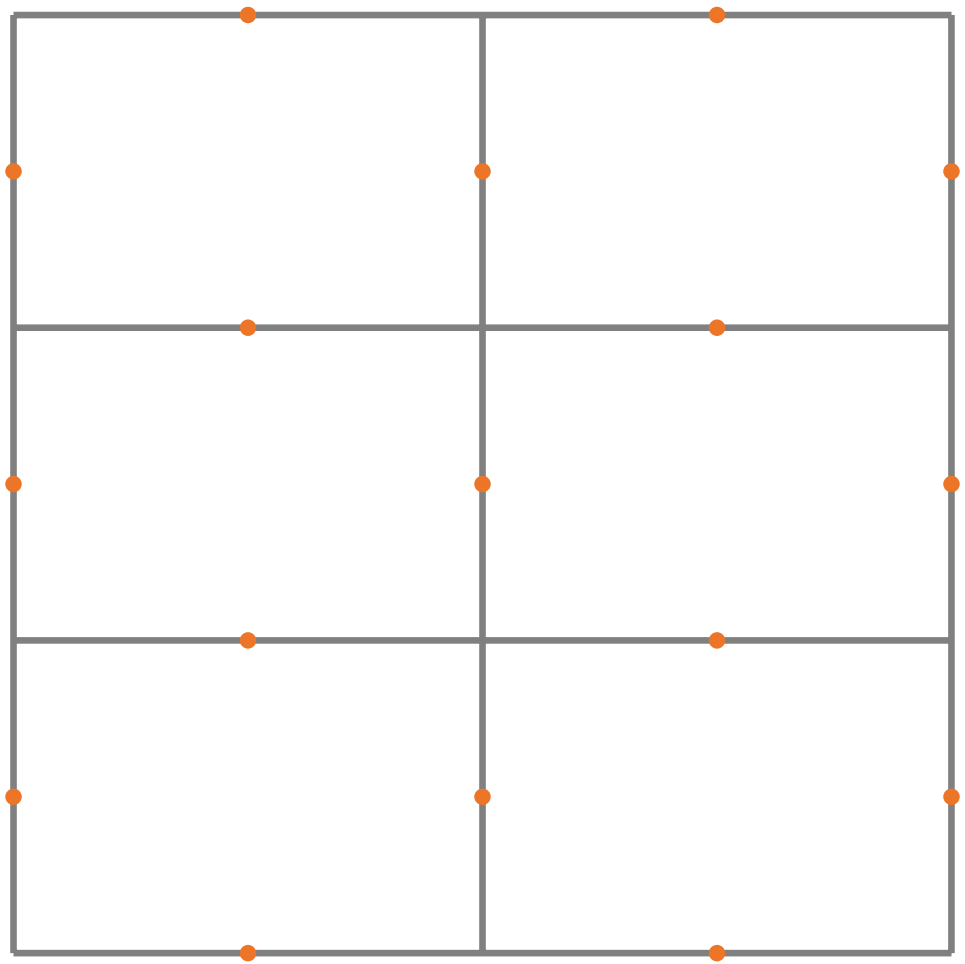


Approche classique : capteurs fixes



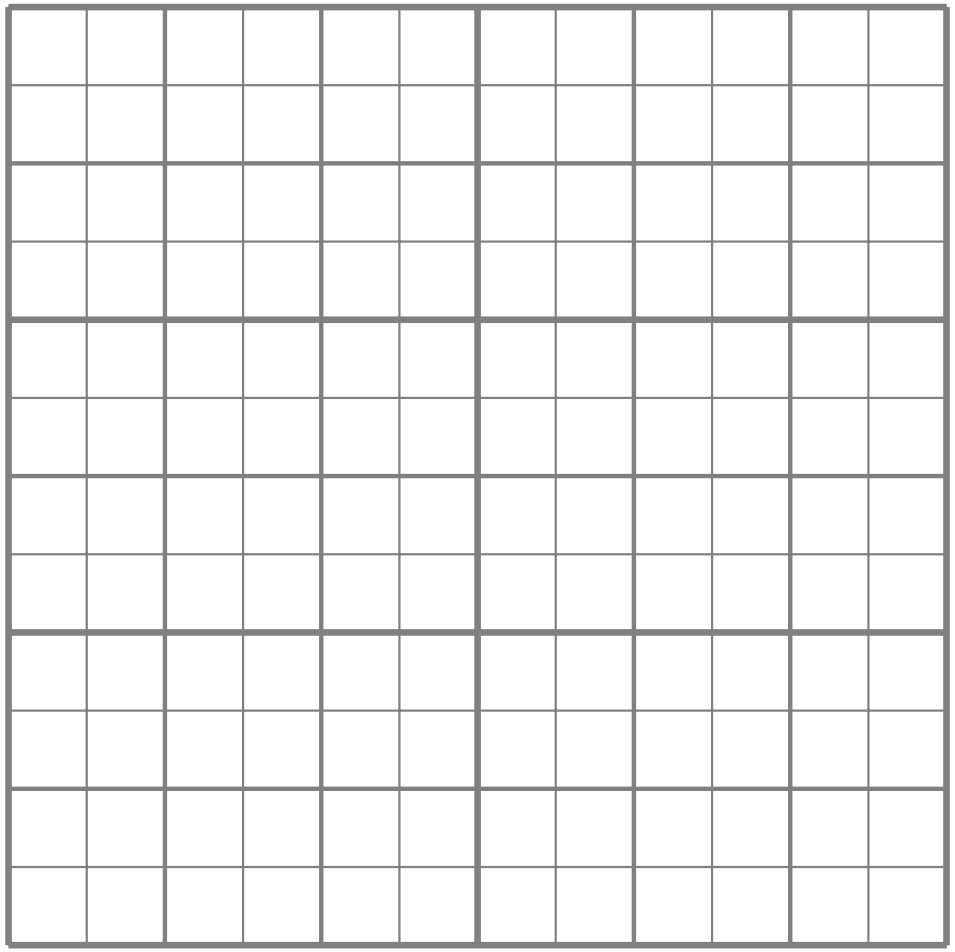
624 capteurs

Approche classique : capteurs fixes



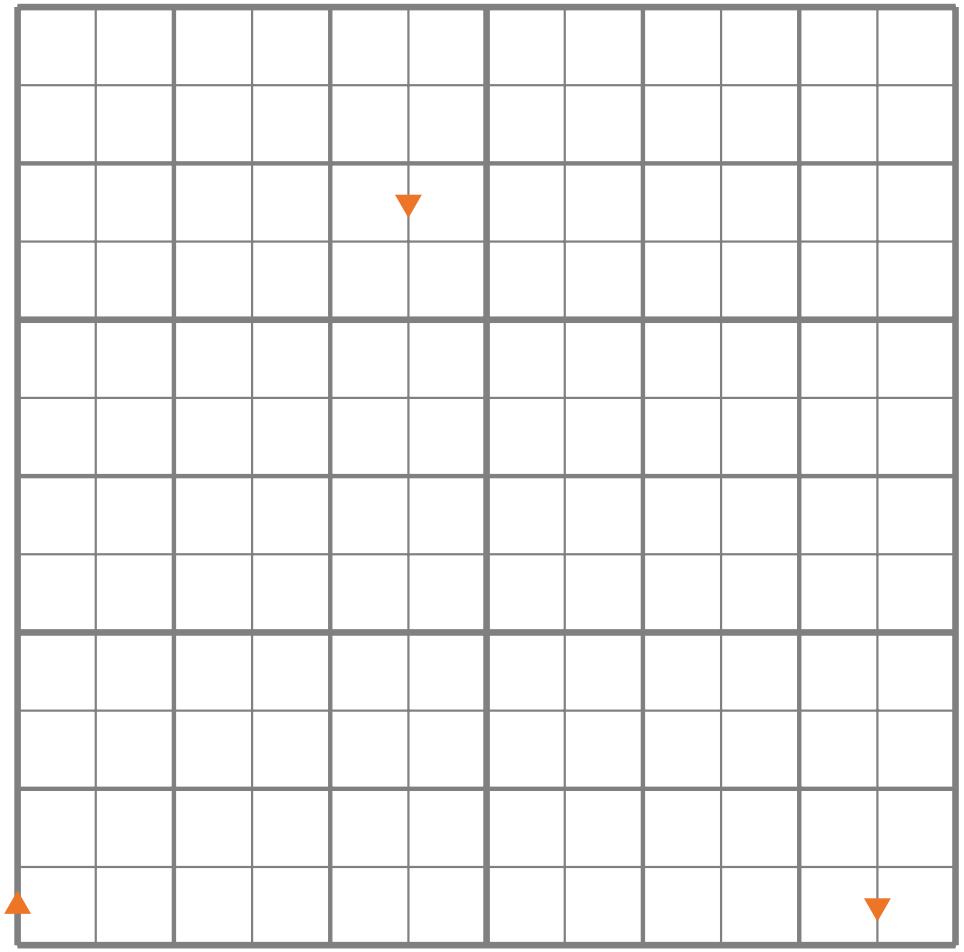
34 Capteurs

Approche basée sur des véhicules sondes



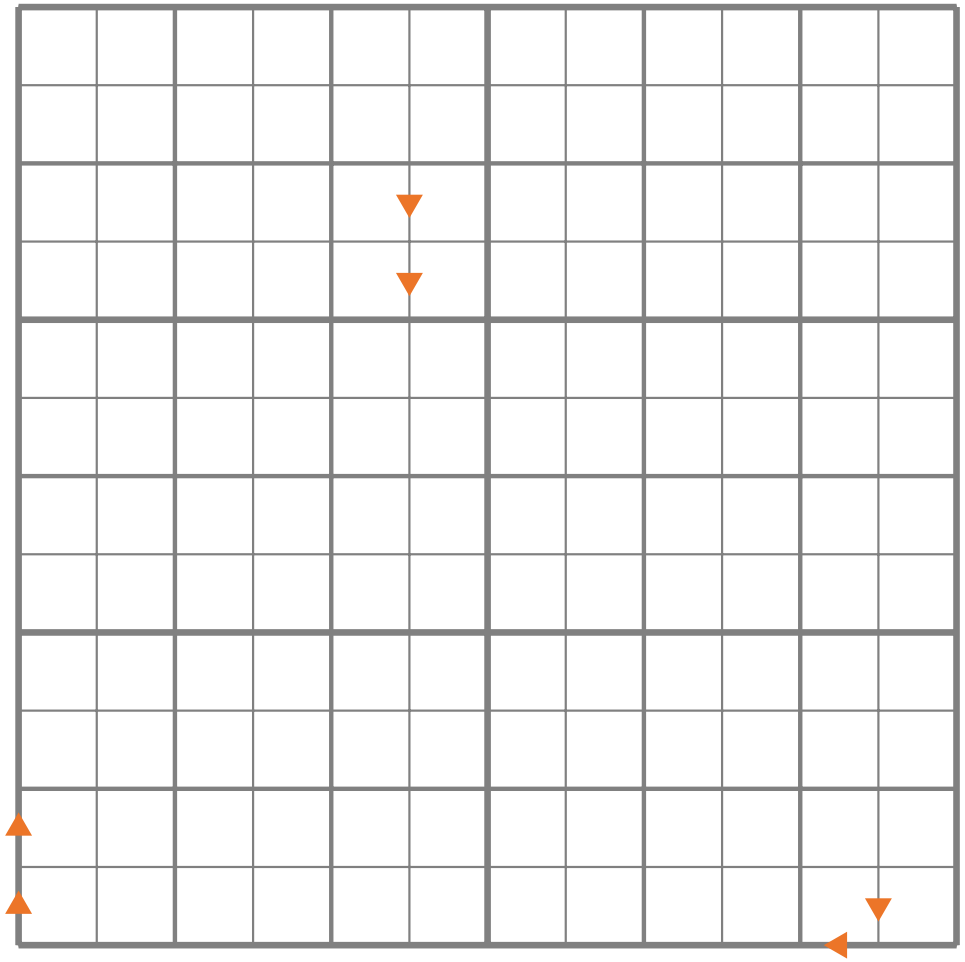
Capteurs mobiles

Approche basée sur des véhicules sondes



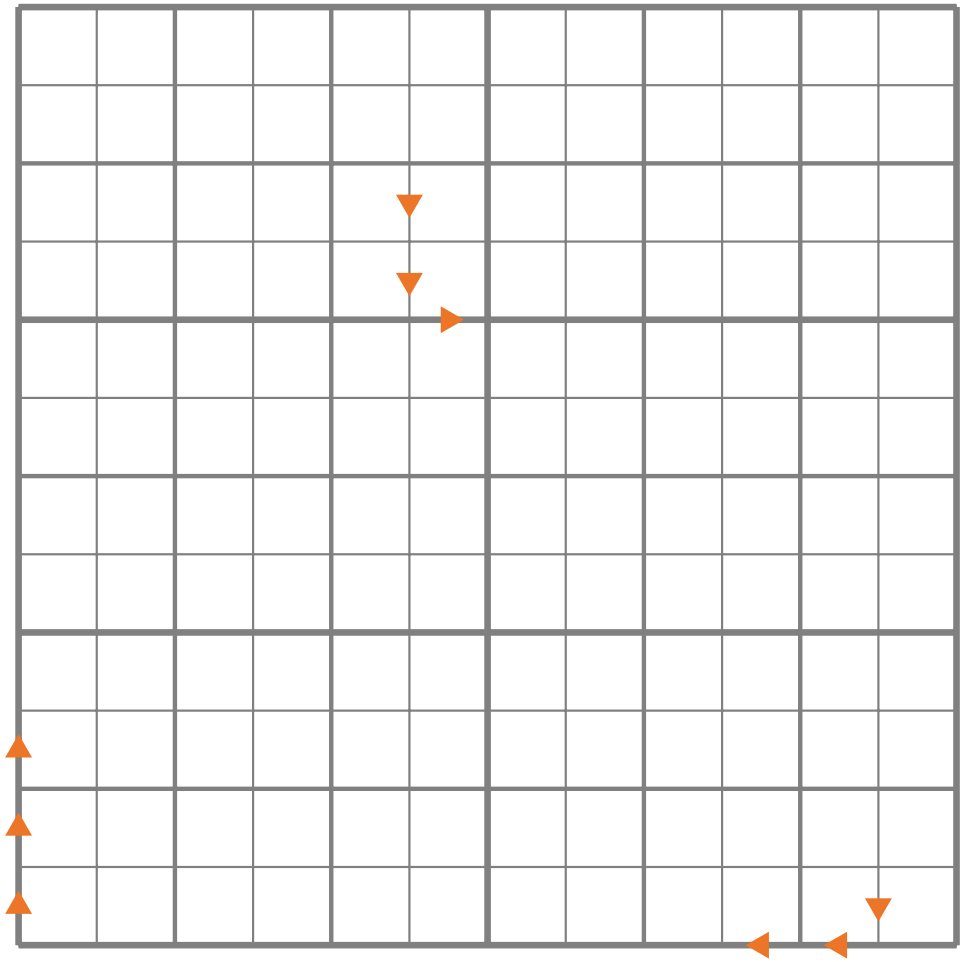
Capteurs mobiles

Approche basée sur des véhicules sondes



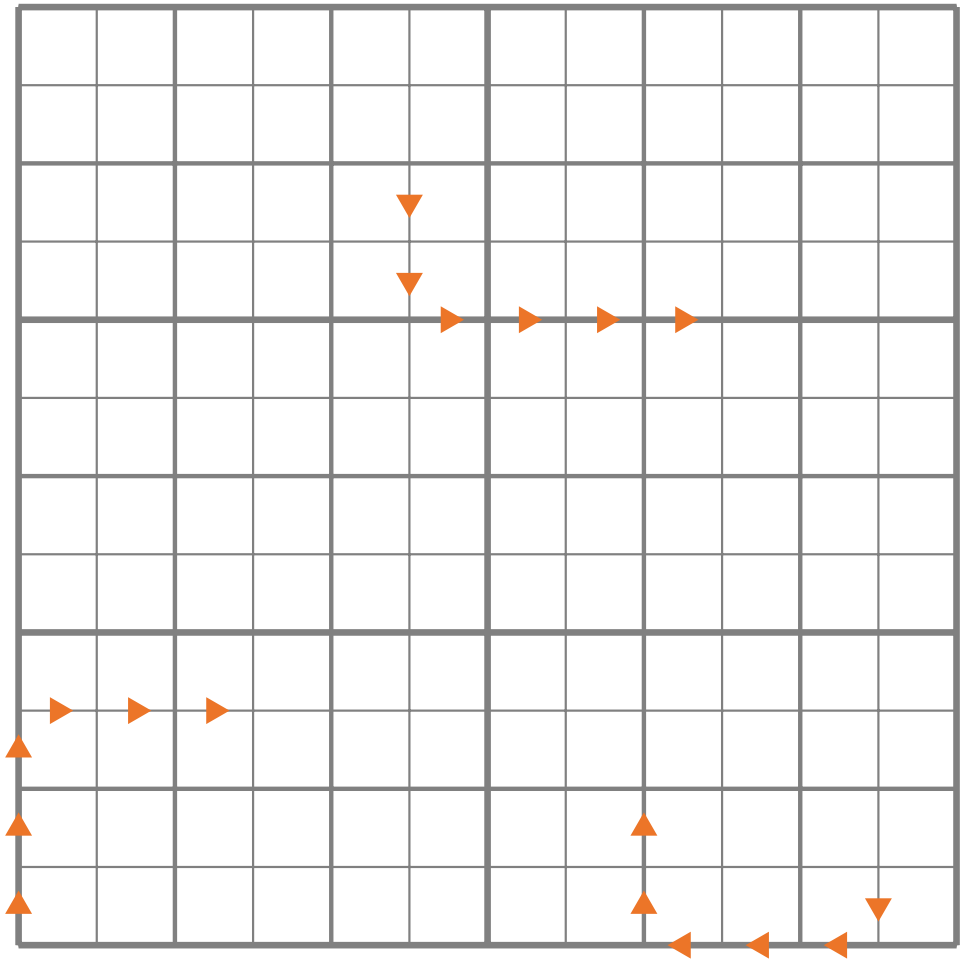
Capteurs mobiles

Approche basée sur des véhicules sondes



Capteurs mobiles

Approche basée sur des véhicules sondes



Capteurs mobiles

Un modèle statistique du trafic

- Nécessité d'un modèle du fait :
  - d'observations très partielles de l'état du réseau,
  - d'observations locales,

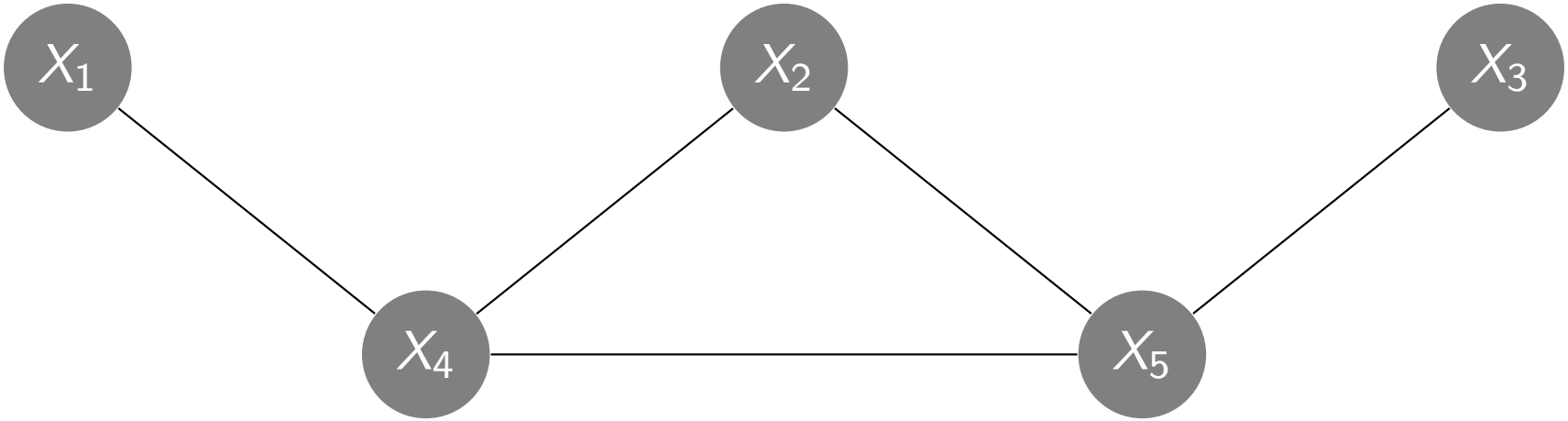
Un modèle statistique du trafic

- Nécessité d'un modèle du fait :
  - d'observations très partielles de l'état du réseau,
  - d'observations locales,
- Pourquoi statistique ?
  - complexité du phénomène (feux rouges, nombreux axes, etc.),
  - connaissance imparfaite du réseau,

Un modèle statistique du trafic

- Nécessité d'un modèle du fait :
  - d'observations très partielles de l'état du réseau,
  - d'observations locales,
- Pourquoi statistique ?
  - complexité du phénomène (feux rouges, nombreux axes, etc.),
  - connaissance imparfaite du réseau,
- Choix d'un champ markovien aléatoire (MRF)
  - modèle d'interactions locales : la congestion se propage de proche en proche,
  - choix le moins discriminant (maximum d'entropie).

Deux tâches à accomplir

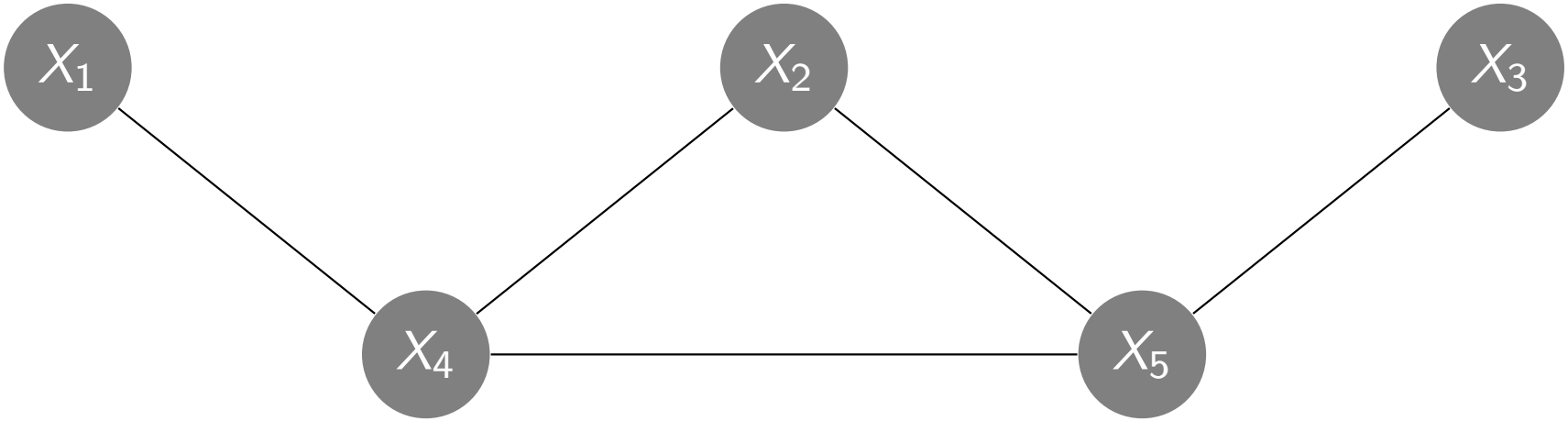


Deux tâches à accomplir

Estimation du modèle

Loi jointe de forme produit (théorème de Hammersley-Clifford).

$$\mathbb{P}(\mathbf{X}) = \prod_{(ij)} \Theta_{ij}(X_i, X_j)$$

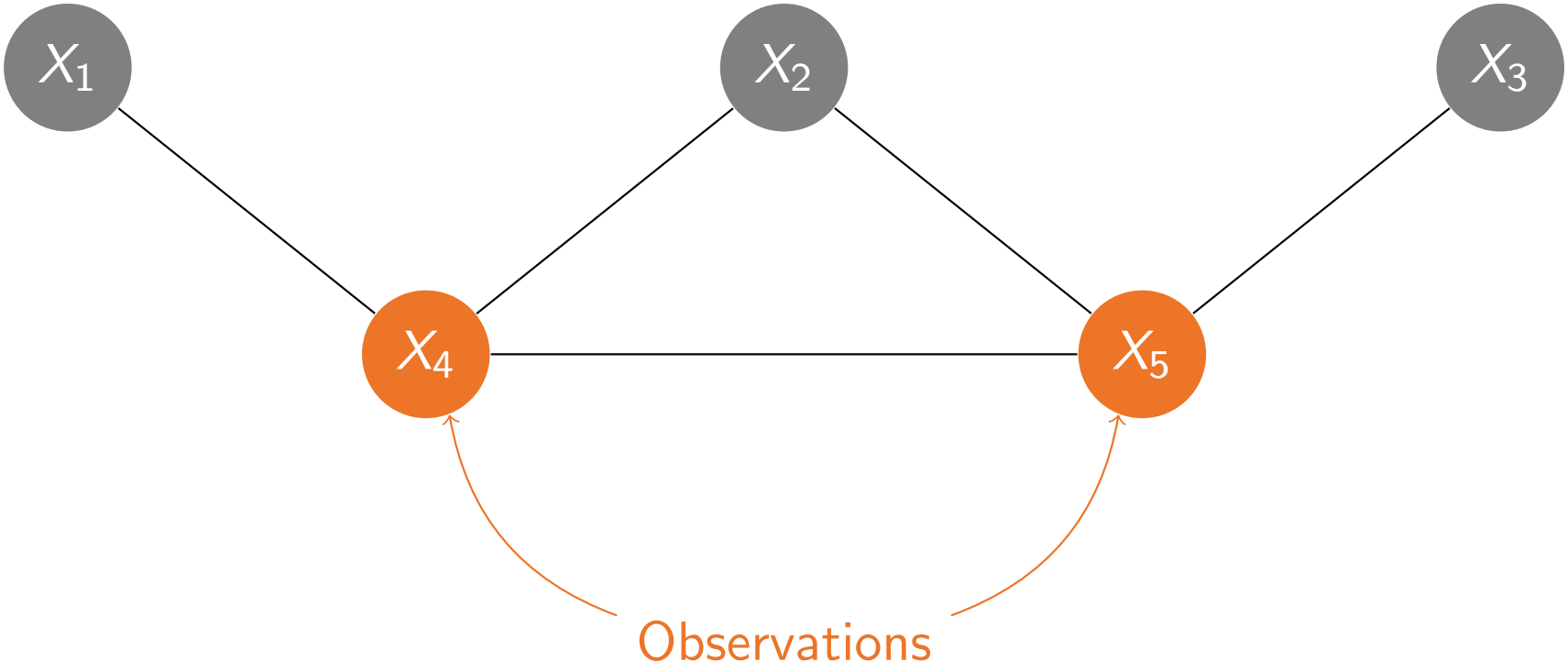


Deux tâches à accomplir

Estimation du modèle

Loi jointe de forme produit (théorème de Hammersley-Clifford).

$$\mathbb{P}(\mathbf{X}) = \prod_{(ij)} \Theta_{ij}(X_i, X_j)$$

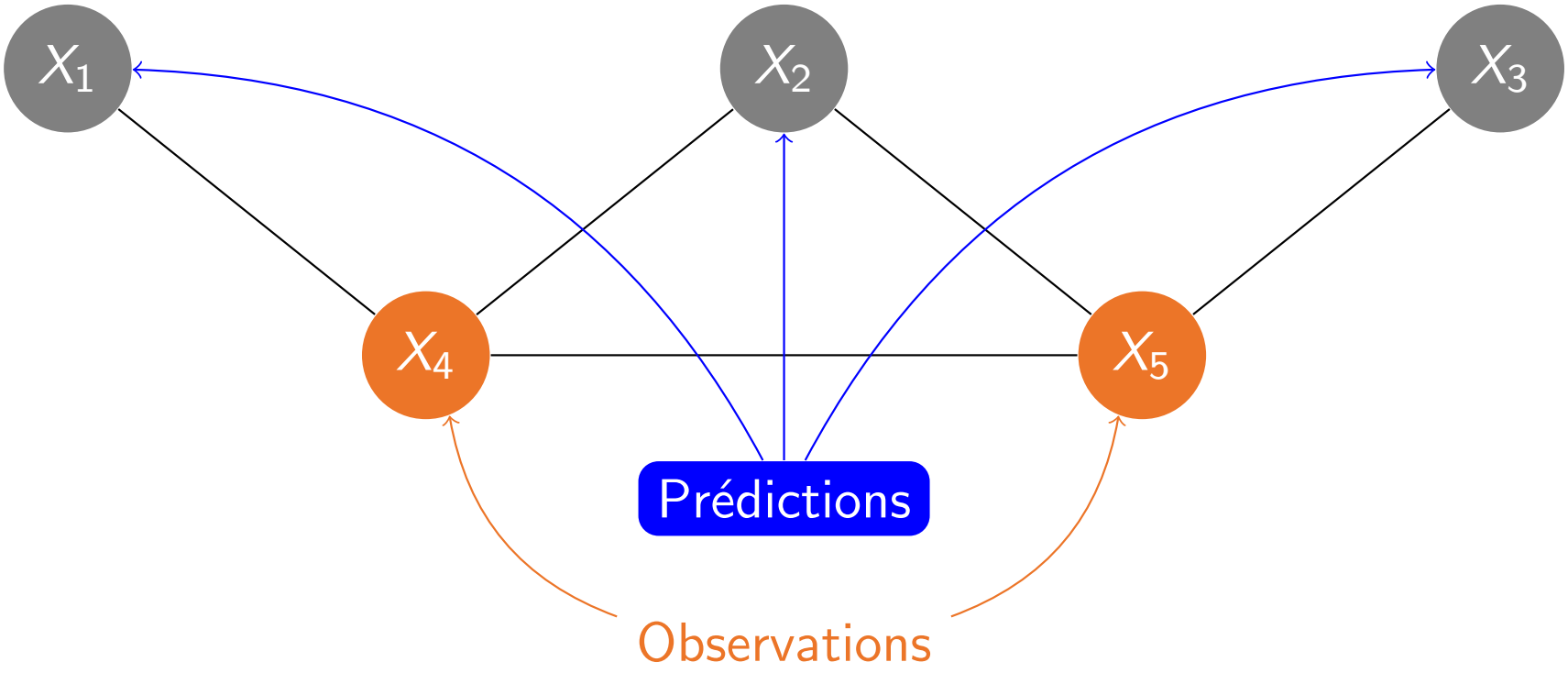


Deux tâches à accomplir

Estimation du modèle

Loi jointe de forme produit (théorème de Hammersley-Clifford).

$$\mathbb{P}(\mathbf{X}) = \prod_{(ij)} \Theta_{ij}(X_i, X_j)$$



Prédiction d'une variable aléatoire

Prédiction optimale $\hat{\theta}_\ell(X)$

- Associée à une fonction de perte ℓ : $\hat{\theta}_\ell(X) = \text{argmin}_z \, \mathbb{E}[\ell(X, z)]$,
- Exemples : $\hat{\theta}_{L^2}(X) = \mathbb{E}[X]$, $\hat{\theta}_{L^1}(X) = q_X^{0.5}$.

Prédiction d'une variable aléatoire

Prédiction optimale $\hat{\theta}_\ell(X)$

- Associée à une fonction de perte ℓ : $\hat{\theta}_\ell(X) = \text{argmin}_z \, \mathbb{E}[\ell(X, z)]$,
- Exemples : $\hat{\theta}_{L^2}(X) = \mathbb{E}[X]$, $\hat{\theta}_{L^1}(X) = q_X^{0.5}$.

Un problème théoriquement trivial

- $\hat{\theta}_\ell(X_k)$ entièrement déterminé par la distribution marginale de X_k.

$$\mathbb{P}(X_k) = \sum_{\mathbf{X} \setminus X_k} \mathbb{P}(\mathbf{X}) d\mathbf{X}$$

- Supposons que $\mathbf{X}$ contient n variables binaires $\rightarrow 2^{n-1}$ termes dans la somme

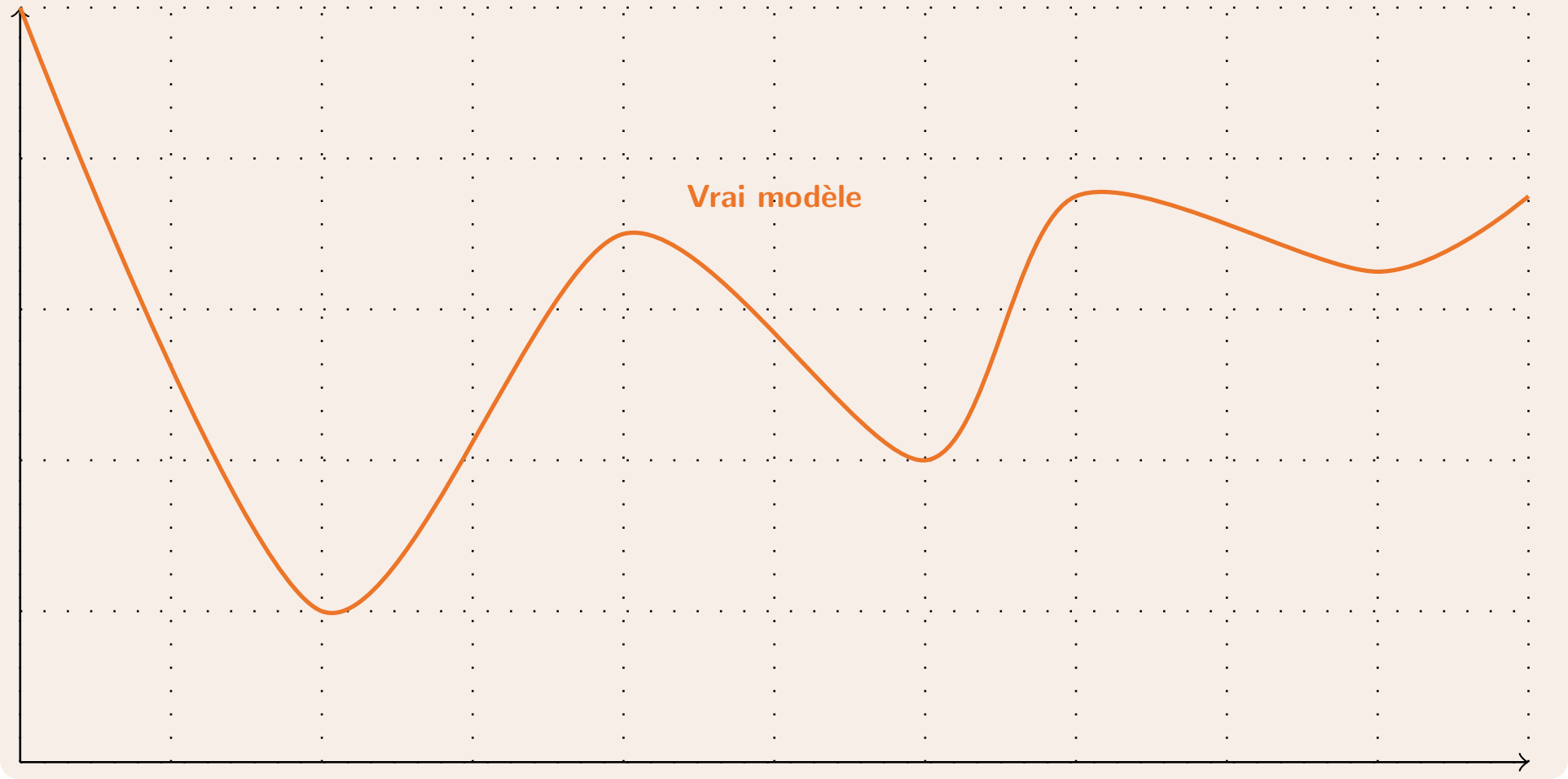
Nécessité de méthodes approchées.

Estimation du modèle

Compatibilité avec la méthode de prédiction

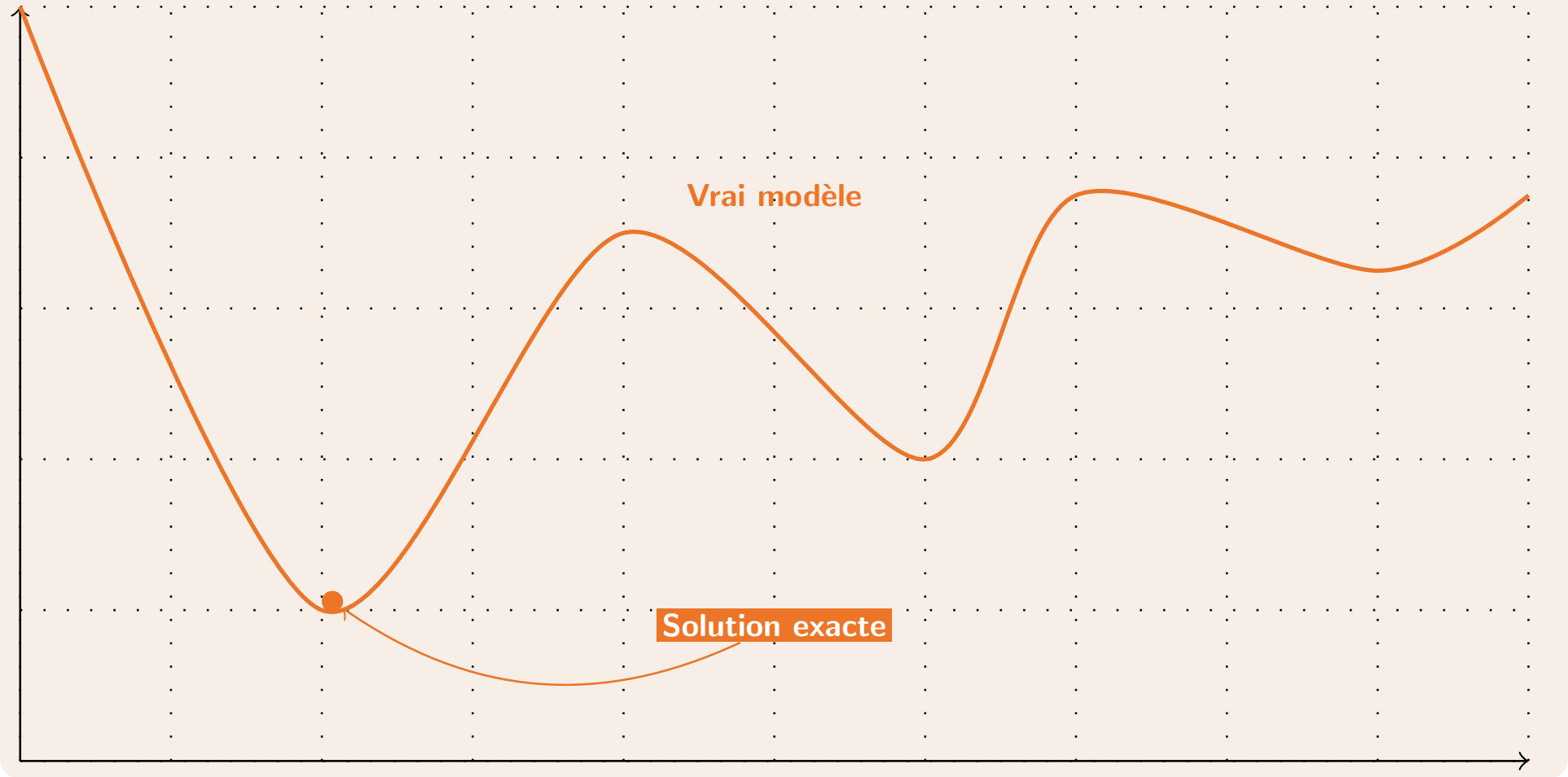
Estimation du modèle

Compatibilité avec la méthode de prédiction



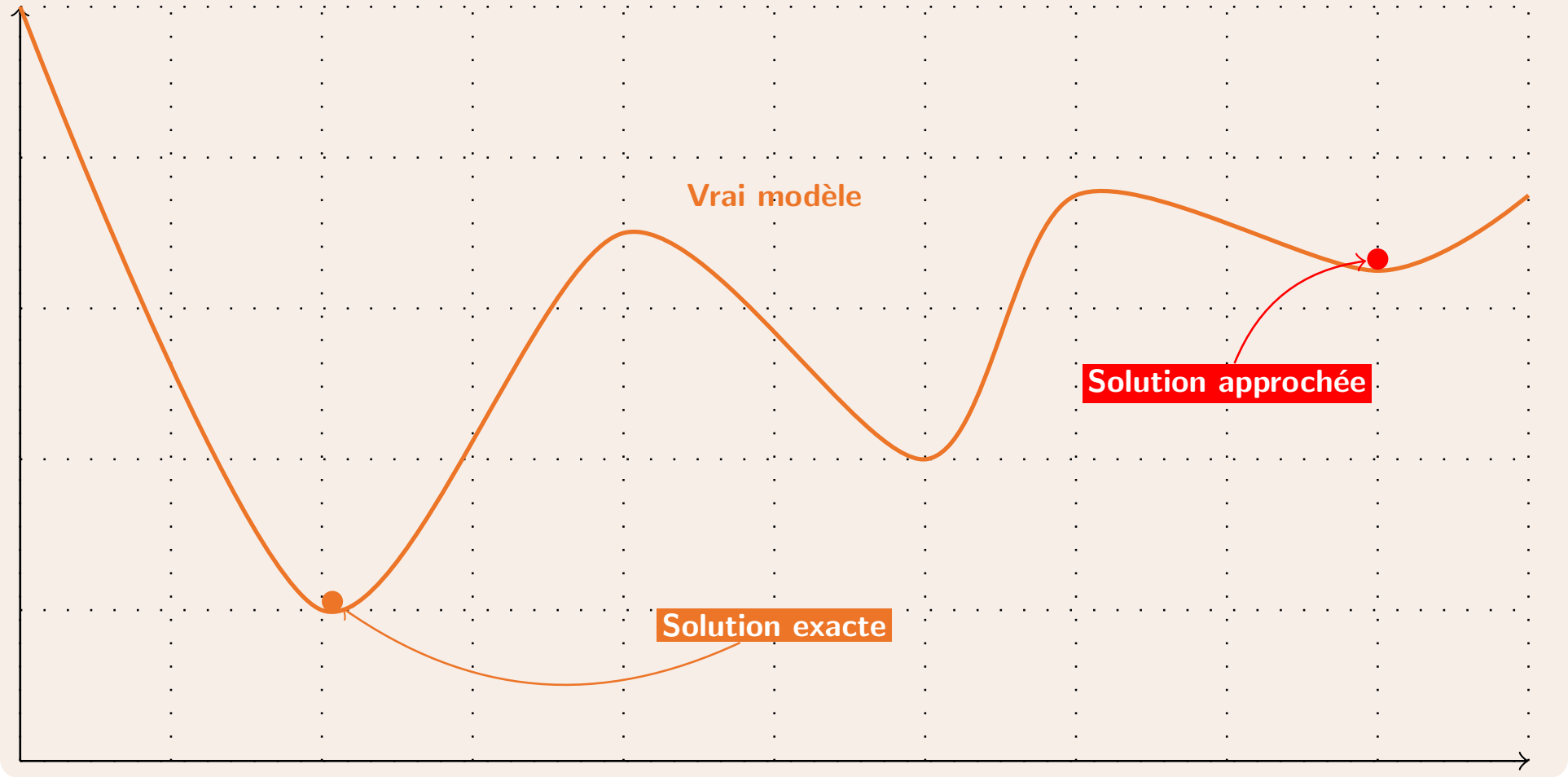
Estimation du modèle

Compatibilité avec la méthode de prédiction



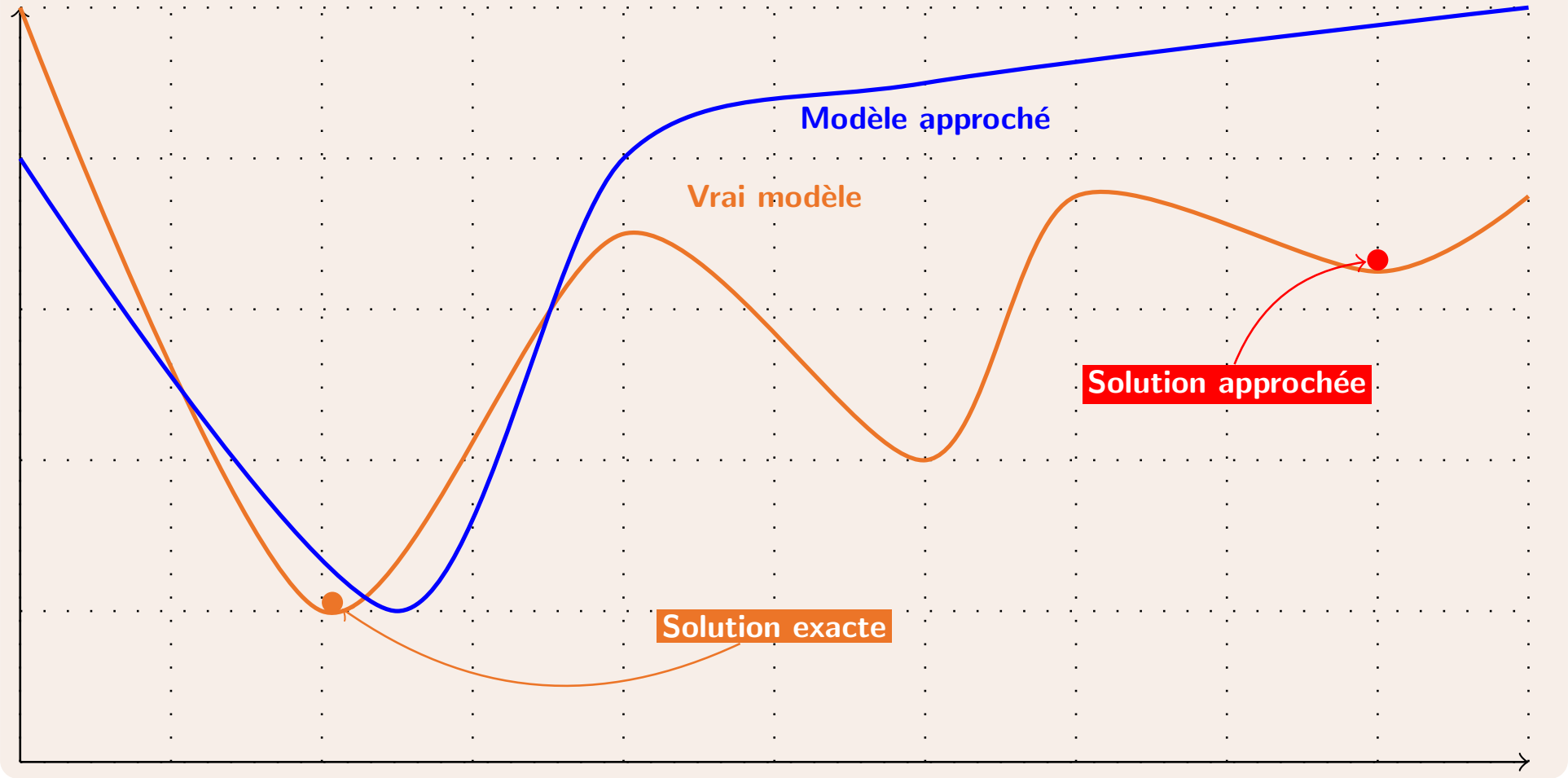
Estimation du modèle

Compatibilité avec la méthode de prédiction



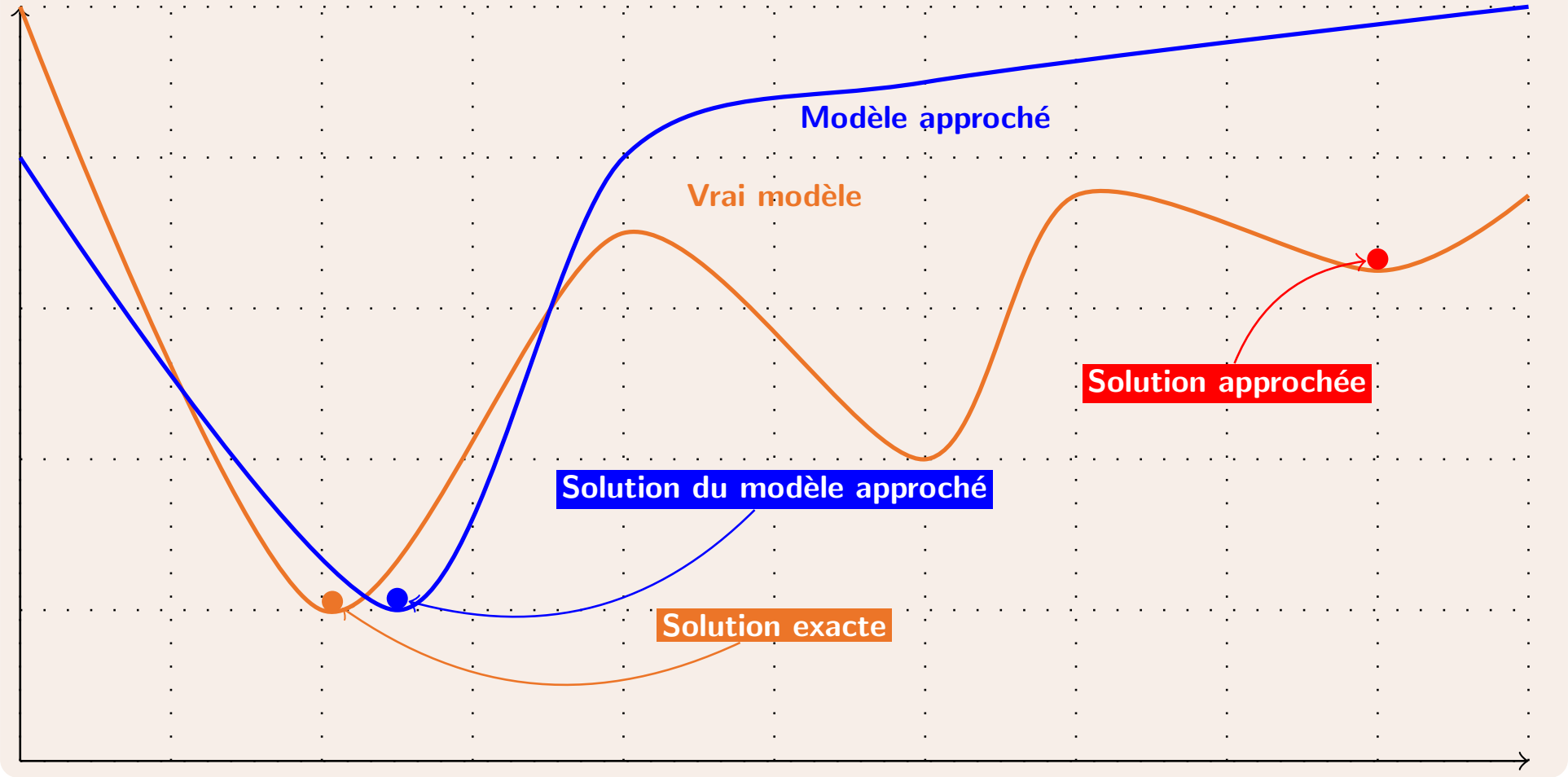
Estimation du modèle

Compatibilité avec la méthode de prédiction



Estimation du modèle

Compatibilité avec la méthode de prédiction



Deux tâches à accomplir conjointement

Estimation du modèle

- À partir d'observations historiques incomplètes, choisir les fonctions Θ_{ij} telles que

$$\mathbb{P}(\mathbf{X}) = \prod_{(ij)} \Theta_{ij}(X_i, X_j)$$

Prédiction en temps réel - Inférence

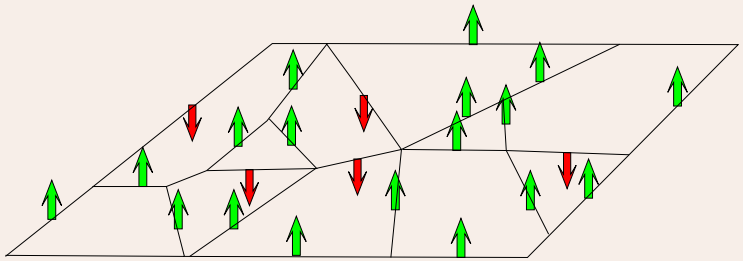
- L'identité des variables observées $\mathbf{X}_{\mathcal{O}}$ n'est pas contrôlée,
- Calculer (une statistique) de $\mathbb{P}(X_k|\mathbf{X}_{\mathcal{O}})$.

Envisagées comme un ensemble.

Pourquoi un modèle d'Ising ?

Champ markovien aléatoire de variables binaires

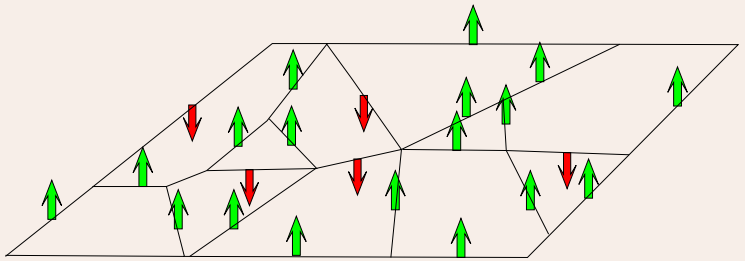
- variables ne pouvant prendre que deux valeurs ± 1,
- $\mathbb{P}(\mathbf{s}) = \frac{1}{Z} \exp\left(\sum_{i,j} J_{ij} s_i s_j + \sum_i h_i s_i\right)$.



Pourquoi un modèle d'Ising ?

Champ markovien aléatoire de variables binaires

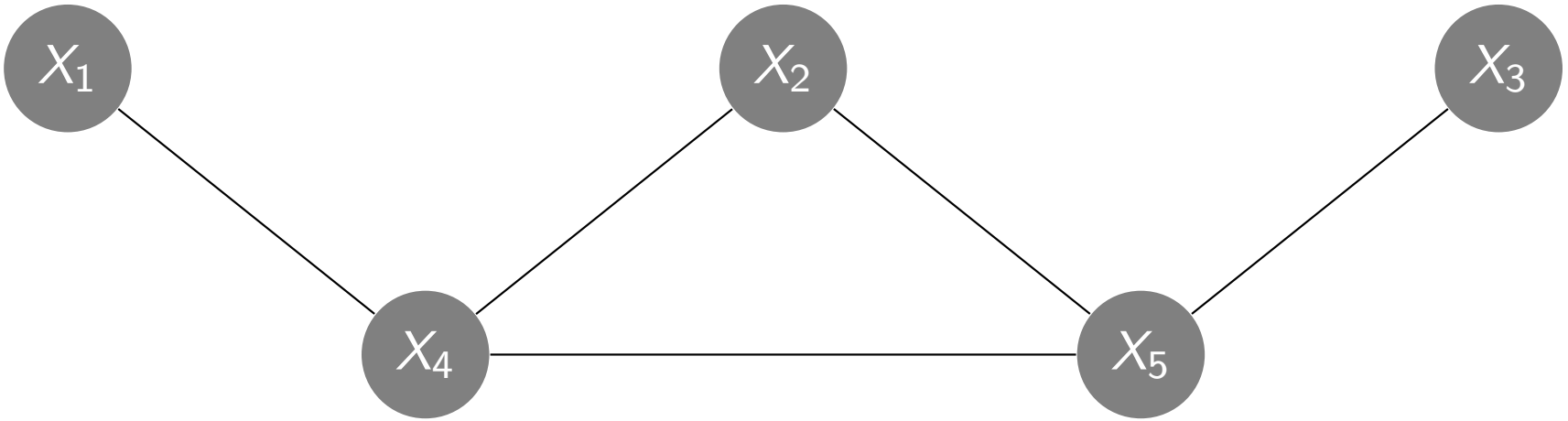
- variables ne pouvant prendre que deux valeurs ± 1,
- $\mathbb{P}(\mathbf{s}) = \frac{1}{Z} \exp\left(\sum_{i,j} J_{ij} s_i s_j + \sum_i h_i s_i\right)$.



- Modélisation binaire intuitive de l'état du trafic
  - chaque segment est soit fluide soit congestionné.
- Modèle d'interactions locales
  - Couplage est assez fort ▶ zones où les spins sont uniformes.
  - Équivalent de parties saturées du réseau.

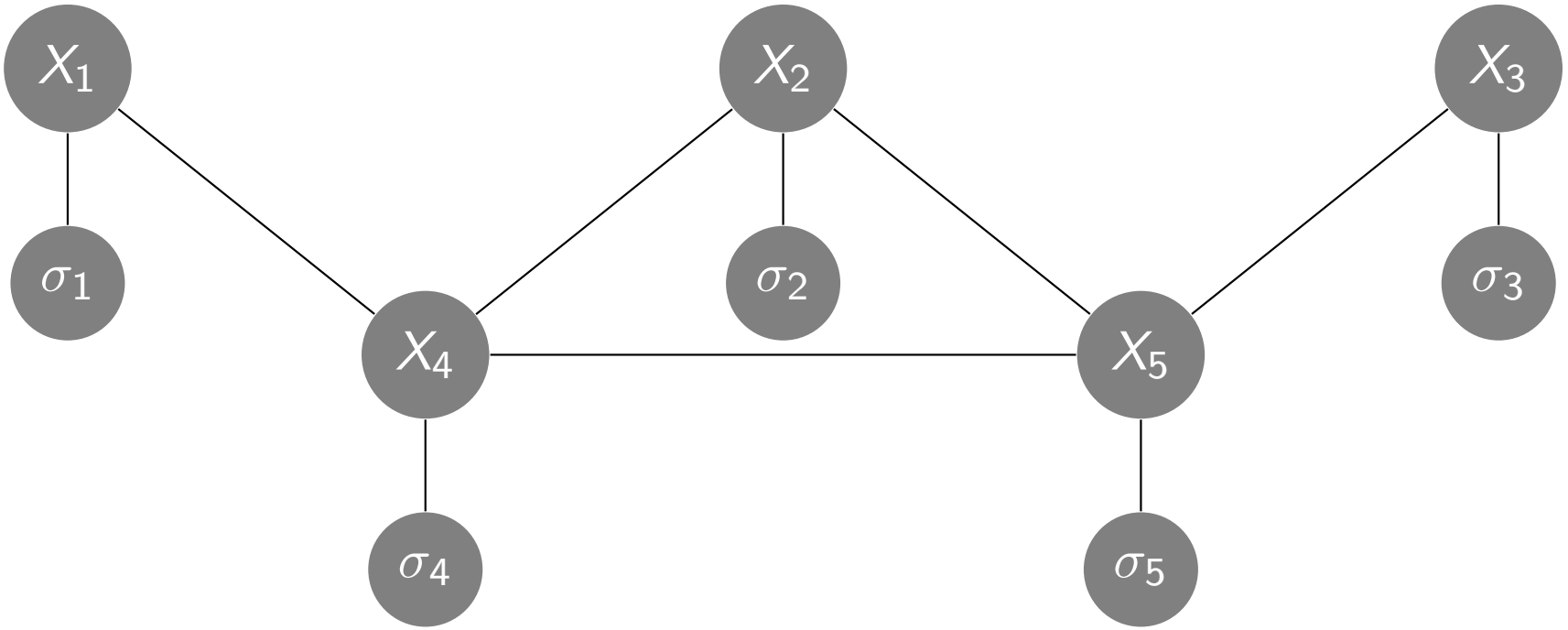
Notre approximation

MRF de variables réelles

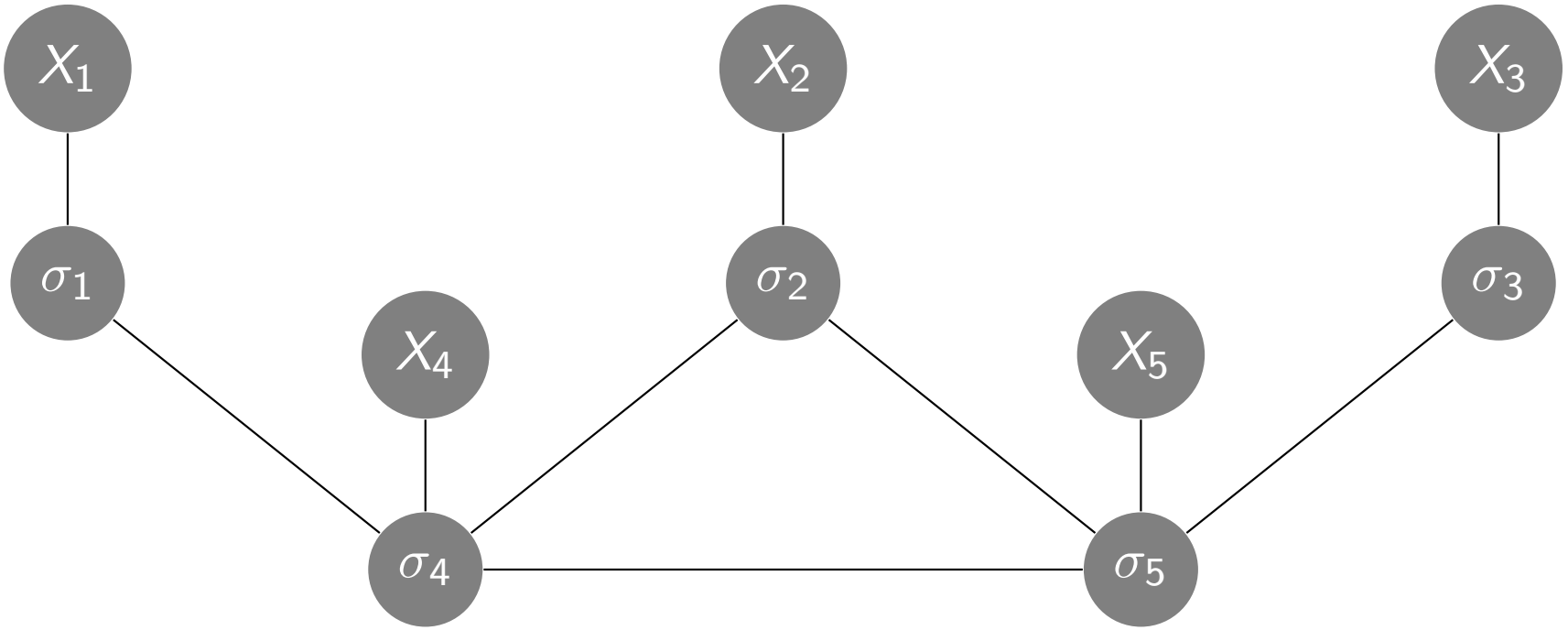


Notre approximation

MRF de variables réelles

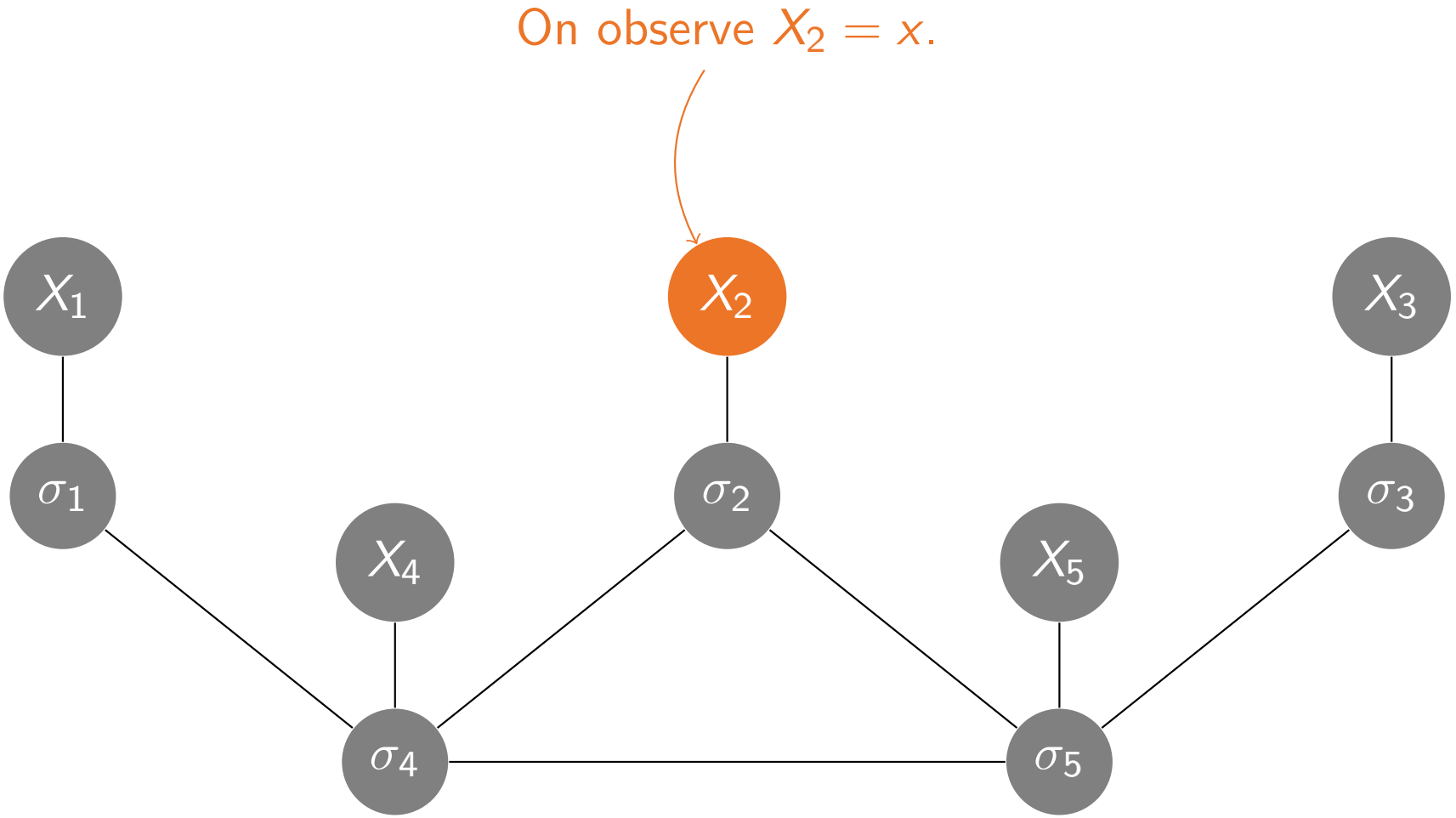


Notre approximation



MRF de variables binaires

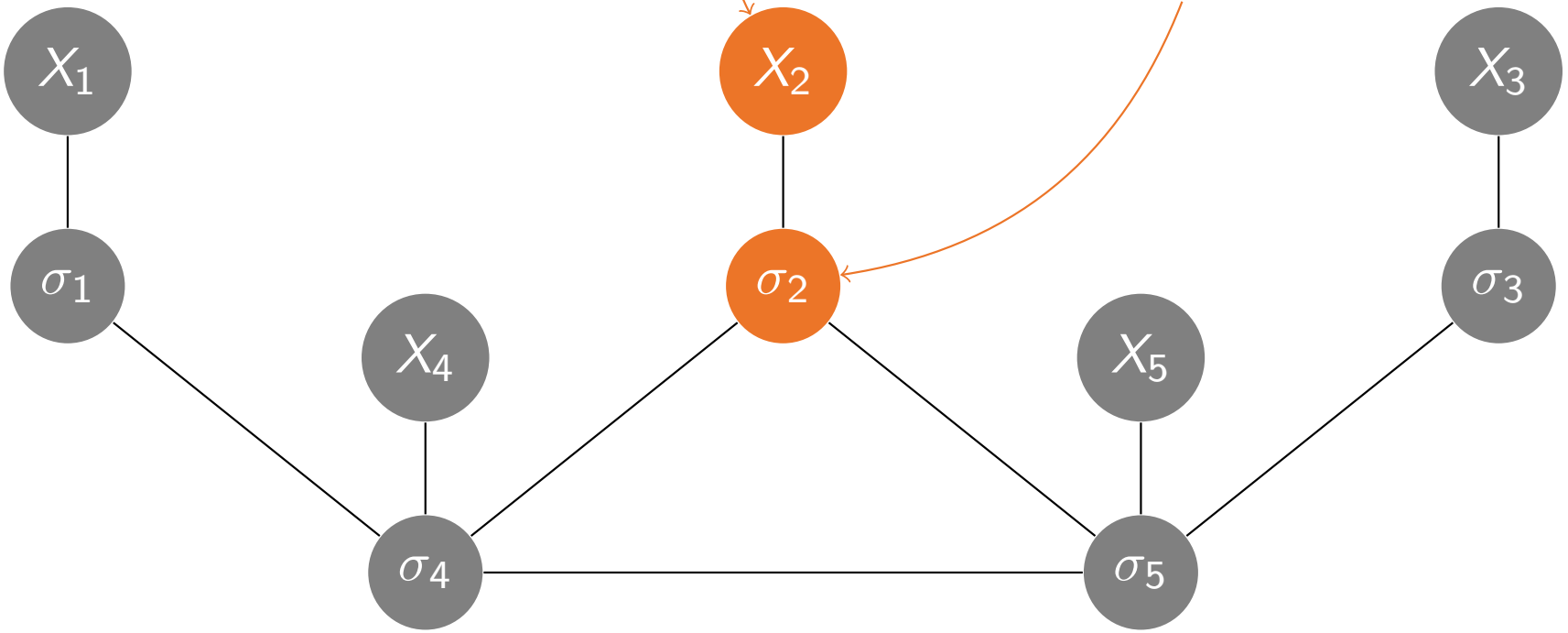
Notre approximation



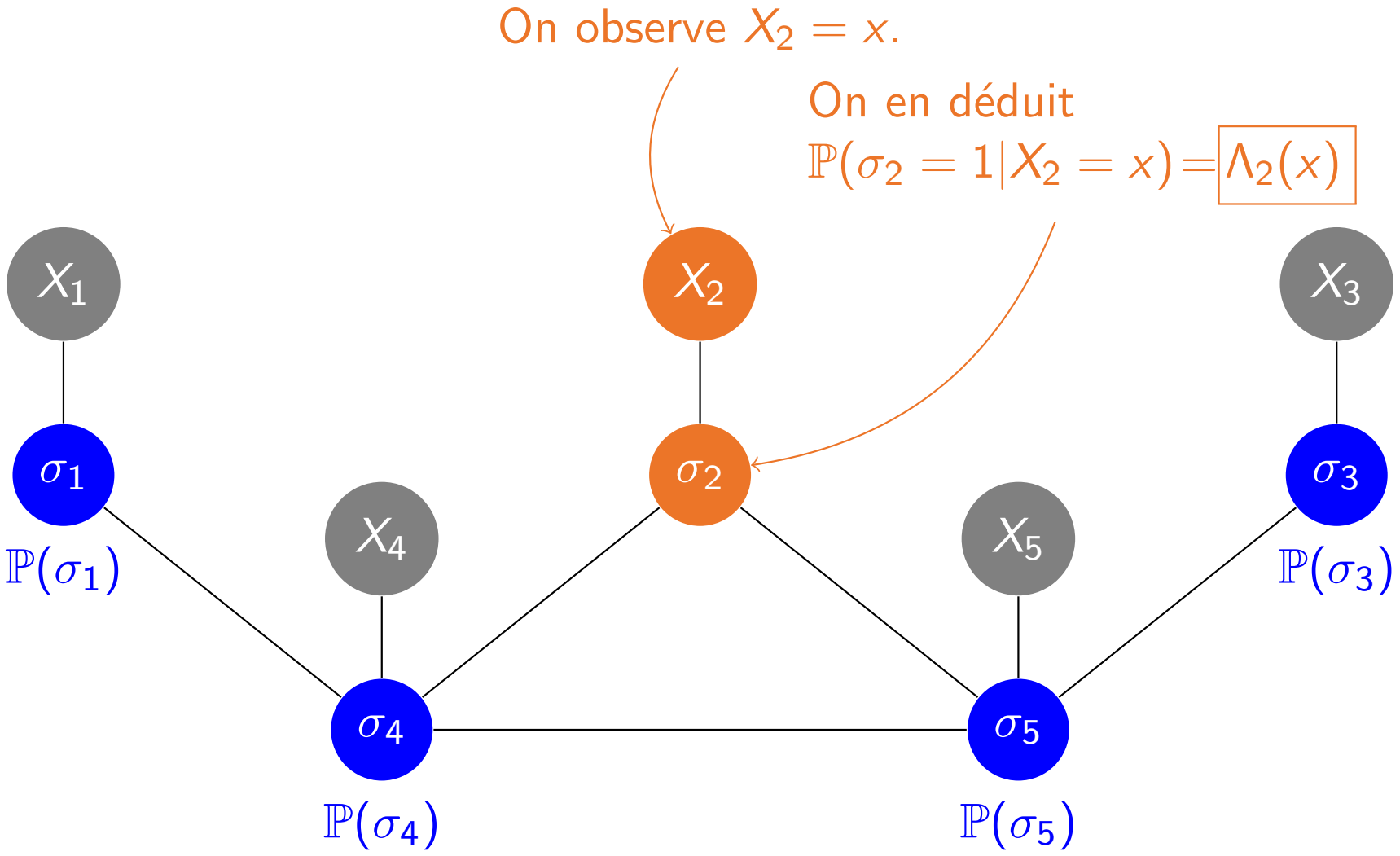
Notre approximation

On observe $X_2 = x$.

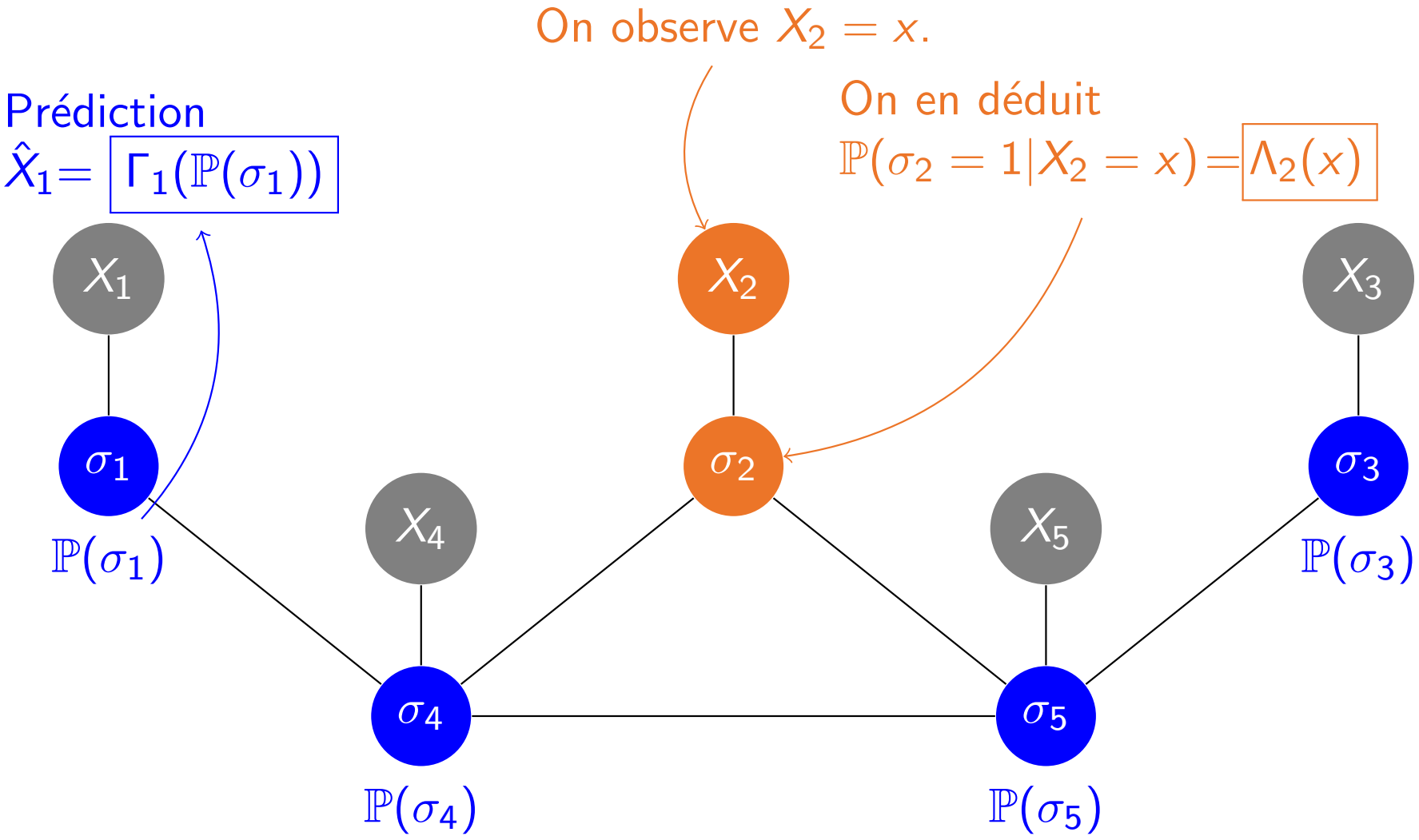
On en déduit
$\mathbb{P}(\sigma_2 = 1 | X_2 = x) = \boxed{\Lambda_2(x)}$



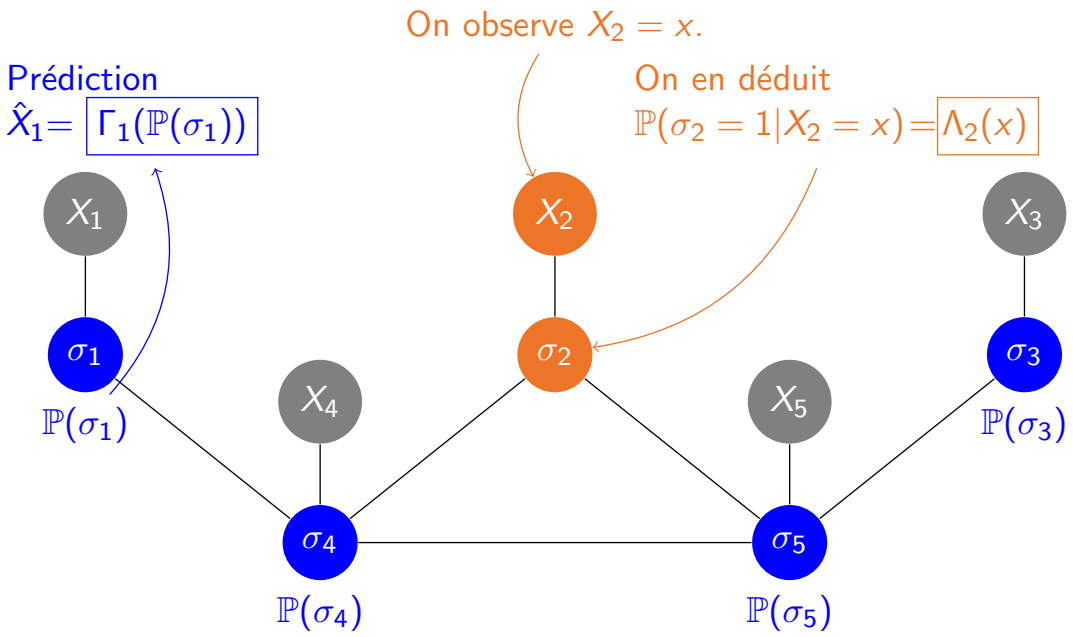
Notre approximation



Notre approximation



Plan de la présentation



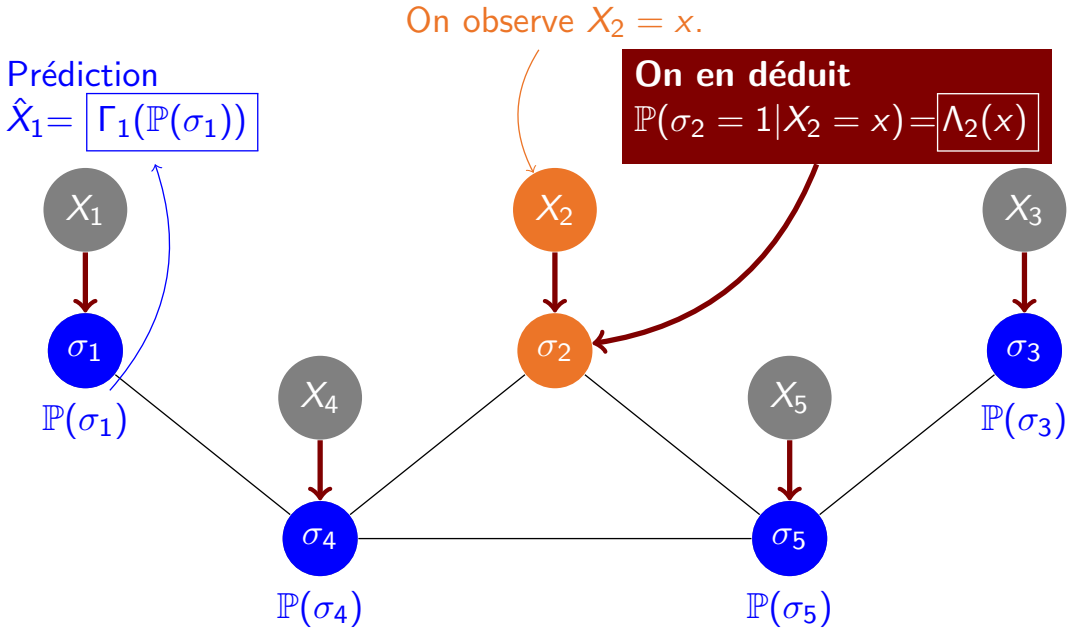
I. *Définition des variables latentes*
a. *Fonction d'encodage*
b. Fonction de décodage

II. Estimation du modèle d'Ising

III. Algorithme d'inférence

IV. Expérimentations numériques

Plan de la présentation



I. *Définition des variables latentes*
a. *Fonction d'encodage*
b. Fonction de décodage

II. Estimation du modèle d'Ising

III. Algorithme d'inférence

IV. Expérimentations numériques

Définition des variables binaires

Choisir Λ est équivalent à définir σ

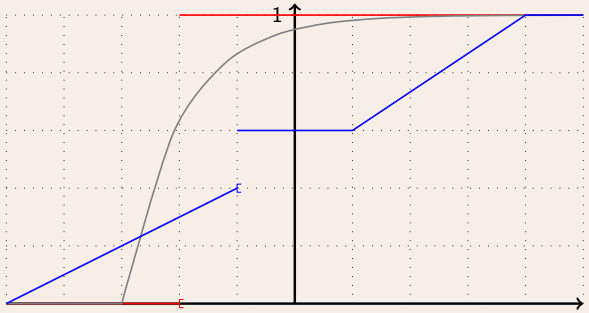
- $\Lambda(x) = \mathbb{P}(\sigma = 1 | X = x)$
- $\mathbb{P}(\sigma = 1) = \int_x \Lambda(x) dF_X(x)$, avec $F_X(x) = \mathbb{P}(X \leq x)$

Critères

- Information mutuelle,
- Entropie.

Contraintes

- Fonction croissante (de 0 à 1),
- Continue à droite, limitée à gauche (càdlàg).



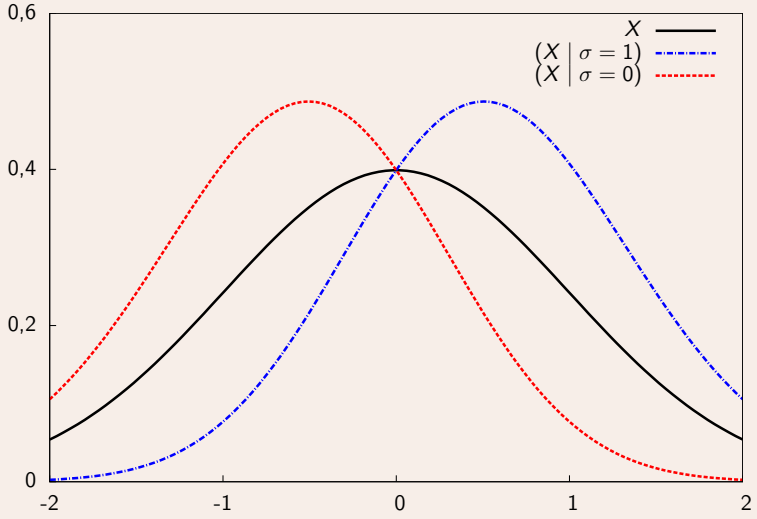
Interprétation stochastique

Mixture avec un ordre stochastique

Lois conditionnelles : $F_X^s(x) = \mathbb{P}(X \leq x \mid \sigma = s)$

$$dF_X = \mathbb{P}(\sigma = 1) dF_X^1 + \mathbb{P}(\sigma = 0) dF_X^0$$

avec $F^1 \leq F^0$.



Seuil aléatoire

- Λ càdlàg, croissante de 0 à 1.
- $\Rightarrow \exists Y$ v.a.r $\mid \Lambda(x) = \mathbb{P}(Y \leq x) = F_Y(x)$.

$$\sigma \stackrel{\text{def}}{=} \mathbb{1}_{\{\mathbf{Y} \leq \mathbf{X}\}}$$

Choix de Λ : l'information mutuelle comme critère

Problème variationnel

- $\Lambda_{\mathrm{MI}} \stackrel{\text{def}}{=} \mathrm{argmax}_{\Lambda}\, I(\sigma, X) = \mathbb{1}_{\{x \geq q_X^{0.5}\}}$.
- preuve :

$$I(X, \sigma) = H(\mathbb{P}(\sigma = 1)) - \int_x H(\Lambda(x)) dF_X(x),$$

avec $H(x) = -x \log x - (1-x)\log(1-x)$. Terme de droite nul pour $\Lambda(x) \in \{0, 1\}$, terme de gauche maximal pour $\mathbb{P}(\sigma = 1) = 0.5$.

σ est une fonction déterministe de X.

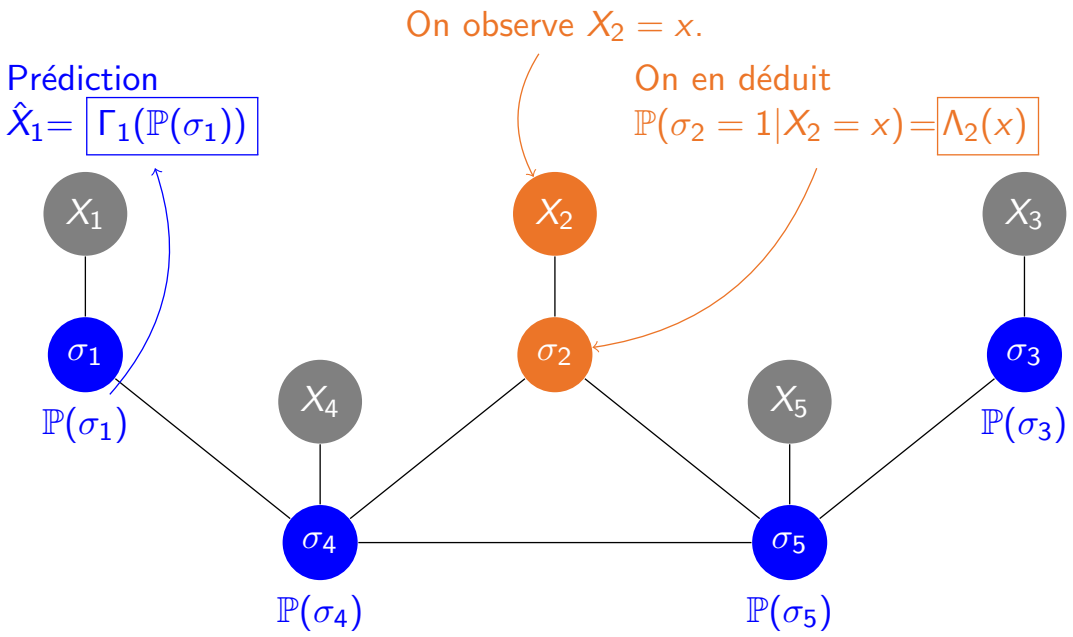
Choix de Λ : le critère entropique

Problème variationnel : maximisation de l'entropie de $\Lambda(X)$

- $\text{argmax}_\Lambda \; S(\Lambda) = F_X(x)$ (fonction de répartition de X).
- preuve : l'entropie est maximale pour une variable uniforme sur $[0, 1]$ et la fonction de répartition transforme X en une v.a. uniforme.

σ est une variable aléatoire à X fixée.

Fonction de décodage Γ



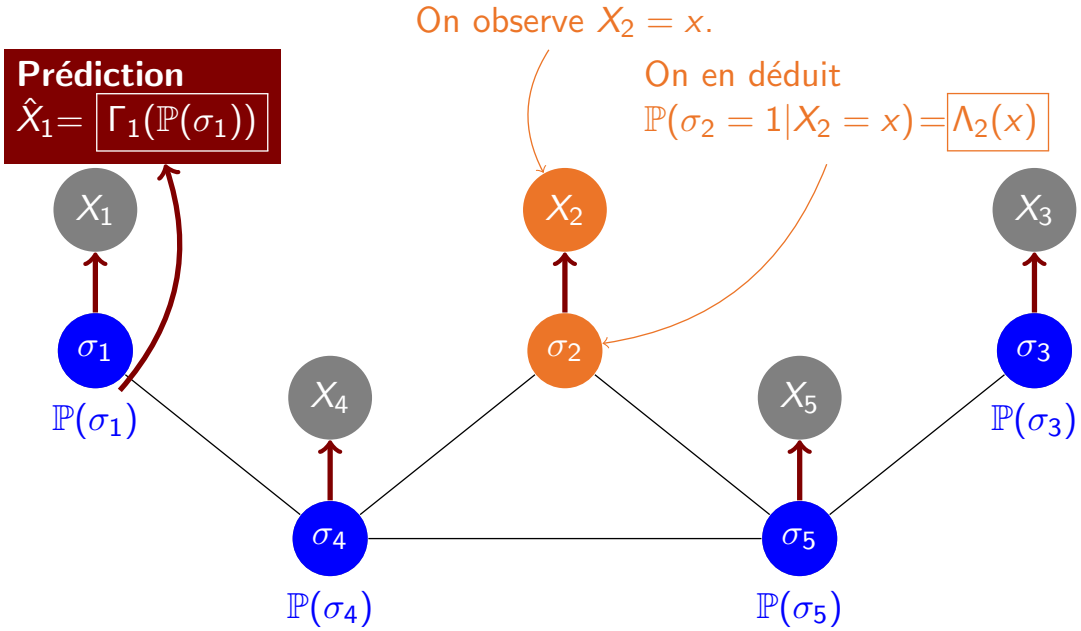
I. *Définition des variables latentes*
a. Fonction d'encodage
b. *Fonction de décodage*

II. Estimation du modèle d'Ising

III. Algorithme d'inférence

IV. Expérimentations numériques

Fonction de décodage Γ



I. *Définition des variables latentes*
a. Fonction d'encodage
b. *Fonction de décodage*

II. Estimation du modèle d'Ising

III. Algorithme d'inférence

IV. Expérimentations numériques

Fonction de décodage Γ

Si Λ est inversible : « maximum de vraisemblance ».

- On peut choisir $\Gamma = \Lambda^{-1}$.
- $\Lambda^{-1}(b)$ correspond à la seule valeur de X telle que $(\sigma | X = x)$ soit distribué selon b.

Sinon : approche bayésienne.

- En utilisant la règle de Jeffrey, on obtient la distribution $\widehat{F}$ de X mise à jour

$$\widehat{F}(x) = bF^1(x) + (1-b)F^0(x),$$

  avec $F^s(x) = \mathbb{P}(X \leq x | \sigma = s)$.
- On calcule le prédicteur souhaité de $\widehat{F}$ (moyenne, médiane,...).
- On peut faire cela que Λ soit inversible ou non.

Fonction de décodage Γ

Cas de la fonction $\Lambda_{\mathrm{MI}} = \mathbb{1}_{\{x \geq q_X^{0.5}\}}$

- Fonction non inversible ▶ approche bayésienne
- Médiane de $\widehat{F}$:

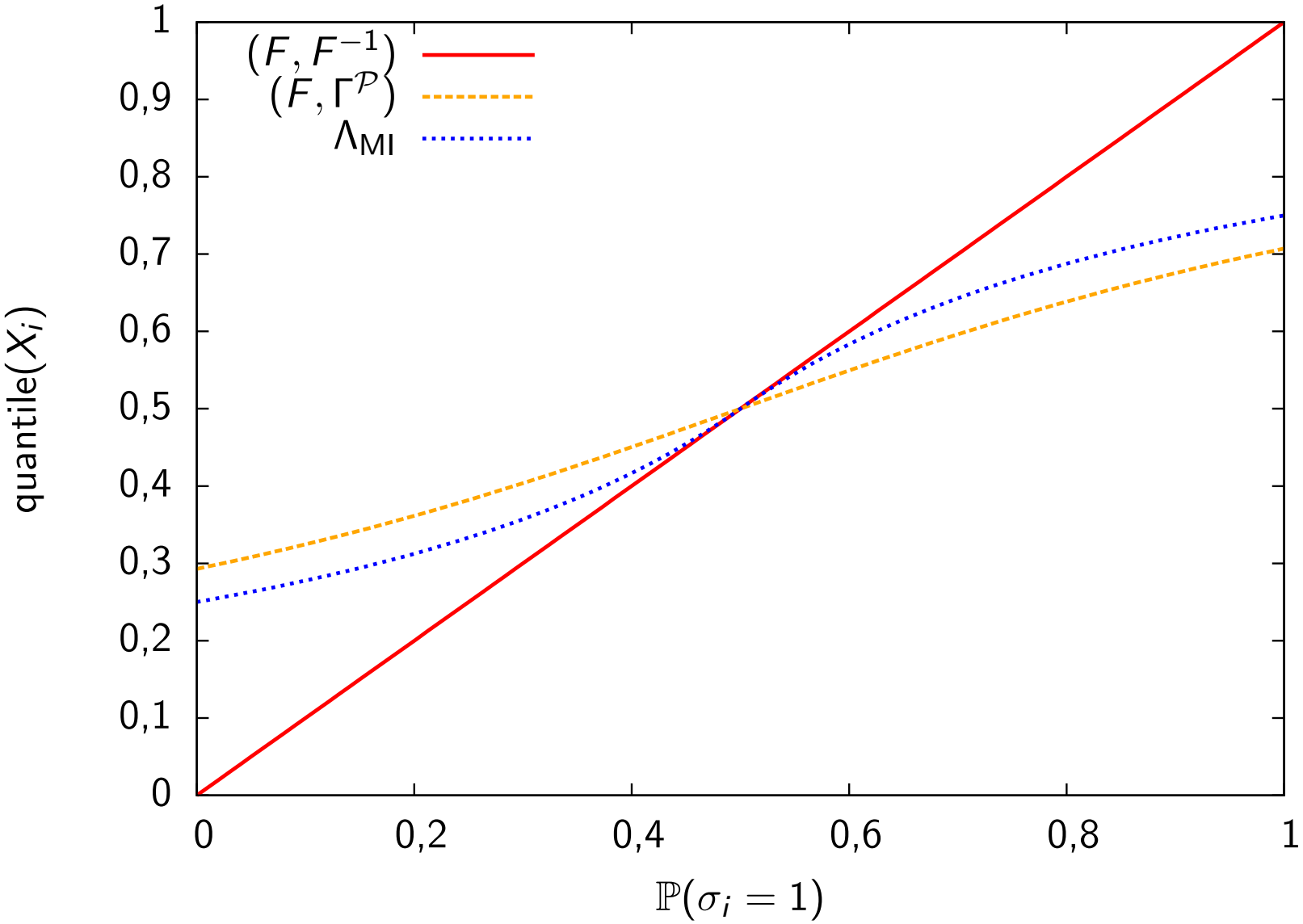
$$\Gamma^{\mathcal{P}}(b) = \begin{cases} F^{-1}\left(\frac{1}{4(1-b)}\right), & \text{si } b \leq \frac{1}{2}, \\ F^{-1}\left(\frac{4b-1}{4b}\right), & \text{si } b > \frac{1}{2}. \end{cases}$$

Cas de la fonction de répartition $\Lambda_{\mathrm{S}} = F$

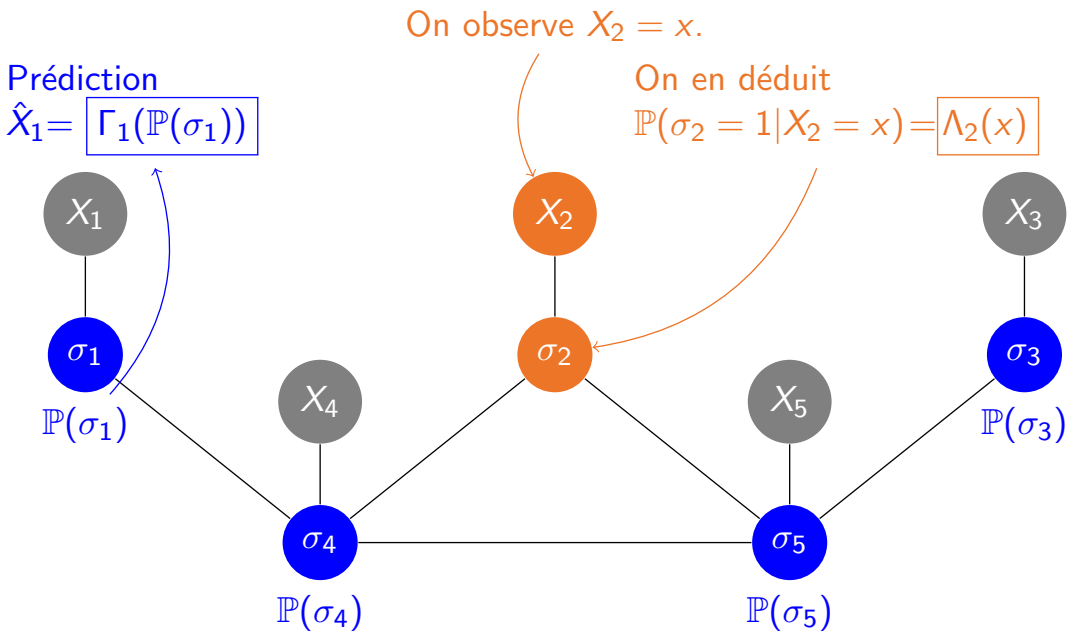
- Maximum de vraisemblance :
  - prédiction en absence d'information : $F_X^{-1}(\mathbb{P}(\sigma = 1)) = q_X^{0.5}$,
  - optimal seulement pour une erreur L^1.
- Approche bayésienne :

$$\Gamma^{\mathcal{P}}(b) = F^{-1}\left(\frac{2(b-1) + \sqrt{(2b-1)^2 + 1}}{4b-2}\right).$$

Fonction de décodage Γ



Estimation du modèle d'Ising



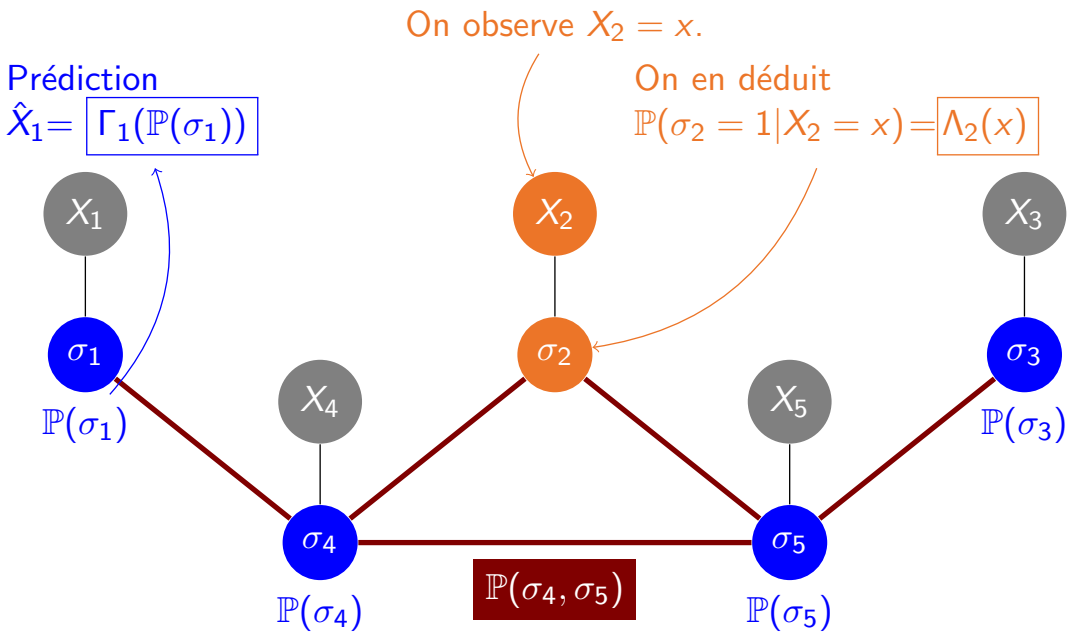
I. Définition des variables latentes
a. Fonction d'encodage
b. Fonction de décodage

II. *Estimation du modèle d'Ising*

III. Algorithme d'inférence

IV. Expérimentations numériques

Estimation du modèle d'Ising



I. Définition des variables latentes
a. Fonction d'encodage
b. Fonction de décodage

II. *Estimation du modèle d'Ising*

III. Algorithme d'inférence

IV. Expérimentations numériques

Estimation du modèle d'Ising

Observations historiques

- Pour un ensemble de paires de variables $(i,j) \in \mathbb{E}$, on a $N_{i,j}$ observations indépendantes

$$(X_i, X_j) = (x_i^k, x_j^k), k \in \{1, \ldots, N_{i,j}\}$$

- On suppose la structure du graphe connue.
- On utilise la fonction Λ empirique.

Loi des paires de variables $\mathbb{P}(\sigma_i, \sigma_j)$

- Le choix de Λ fixe les lois marginales :

$$\mathbb{P}(\sigma_i = 1) = \int_x \Lambda_i(x) dF_x(x).$$

- Reste un paramètre de corrélation : $\mathrm{cov}(\sigma_i, \sigma_j)$.

Estimation du paramètre de corrélation

Méthodes des moments

- en imposant $\mathbb{E}[\Lambda_1(X_1)\Lambda_2(X_2)] = \langle \Lambda_1(X_1)\Lambda_2(X_2) \rangle$:

$$\mathrm{cov}(\sigma_1, \sigma_2) = \widehat{\mathrm{cov}}(\Lambda_1(X_1), \Lambda_2(X_2)) \prod_{i \in \{1,2\}} \frac{\mathrm{var}(\sigma_i)}{\widehat{\mathrm{var}}(\Lambda_i(X_i))}.$$

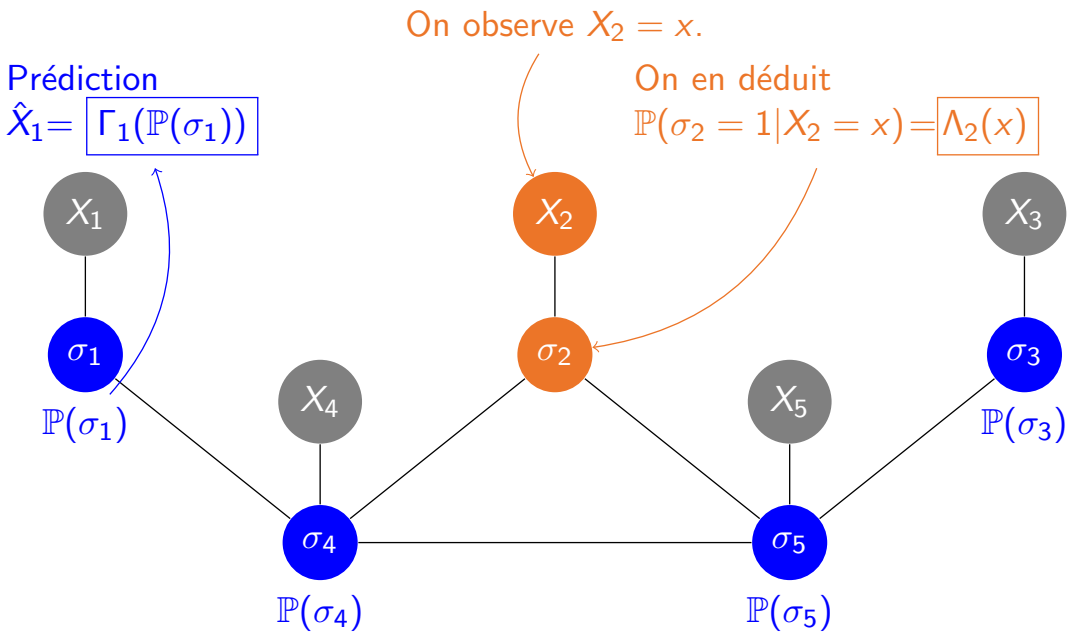
- formule explicite mais problème de saturation des bornes.

Maximisation de la vraisemblance

- pas de formule explicite, maximisation par l'algorithme « Expectation-Maximization » (EM).
- estimation plus robuste, moins de problèmes de saturation.

▶ Remarque : les deux approches sont strictement équivalentes dans le cas de la fonction Λ_{MI}.

Algorithme d'inférence



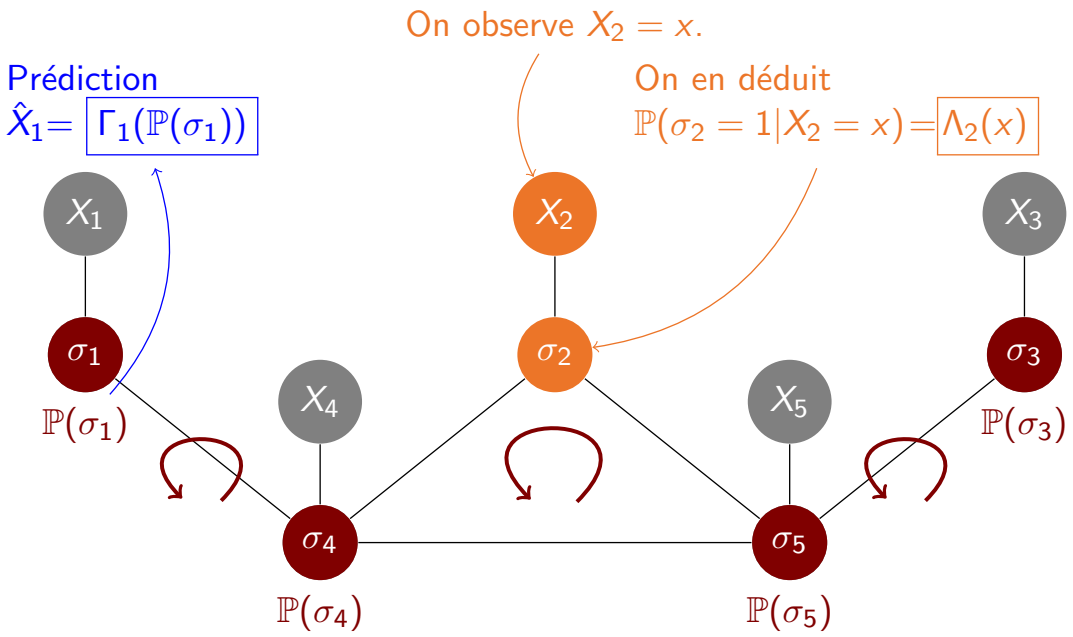
I. Définition des variables latentes
a. Fonction d'encodage
b. Fonction de décodage

II. Estimation du modèle d'Ising

III. *Algorithme d'inférence*

IV. Expérimentations numériques

Algorithme d'inférence



I. Définition des variables latentes
a. Fonction d'encodage
b. Fonction de décodage

II. Estimation du modèle d'Ising

III. *Algorithme d'inférence*

IV. Expérimentations numériques

Belief Propagation (Pearl, 1982)

Objectif : calcul approché des lois marginales $\mathbb{P}(\sigma_i)$

- À partir d'une forme produit :

$$\mathbb{P}(\boldsymbol{\sigma}) = \prod_{(ij)} \psi_{ij}(\sigma_i, \sigma_j) \prod_i \phi_i(\sigma_i)$$

Belief Propagation (Pearl, 1982)

Objectif : calcul approché des lois marginales $\mathbb{P}(\sigma_i)$

- À partir d'une forme produit :

$$\mathbb{P}(\boldsymbol{\sigma}) = \prod_{(ij)} \psi_{ij}(\sigma_i, \sigma_j) \prod_i \phi_i(\sigma_i)$$

Principe

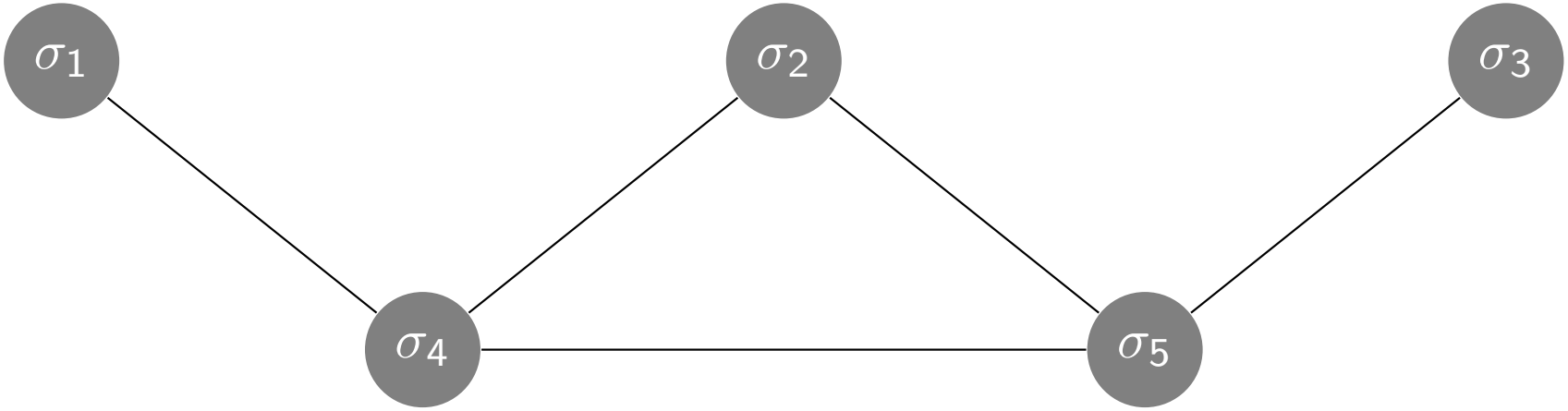
- Méthode itérative par passage de messages.
- Marginales exactes si le graphe ne comporte pas de cycle.
- Minimisation d'une pseudo-distance aux vraies marginales (Yedidia *et al.*, 2005).
- S'il y a convergence, celle-ci est rapide.
- Popularité dans de nombreux domaines.

▶ Résultat sur la stabilité des points fixes

Belief Propagation

Un algorithme de passage de messages

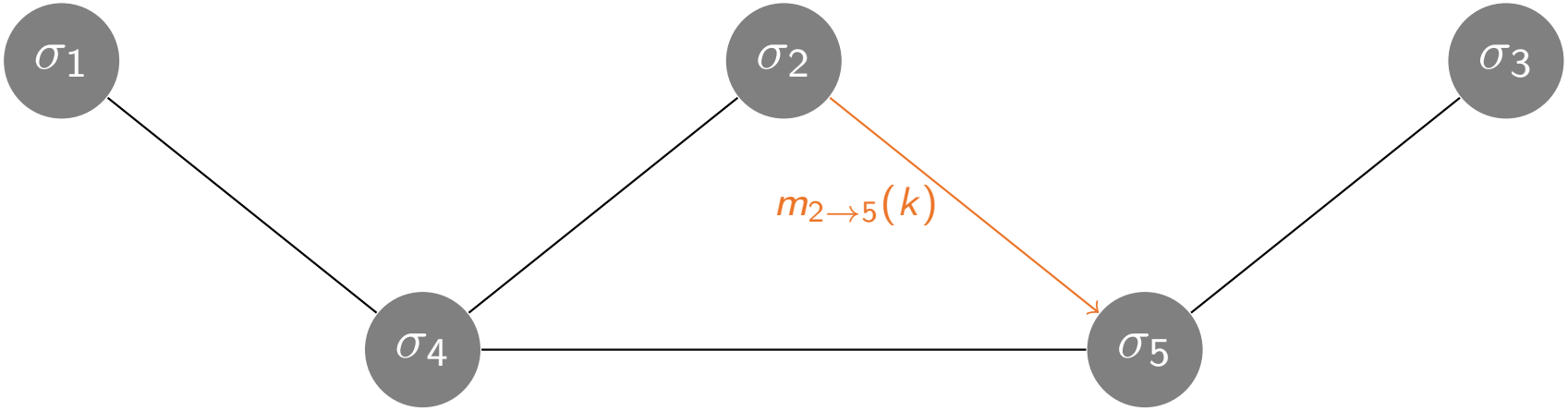
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

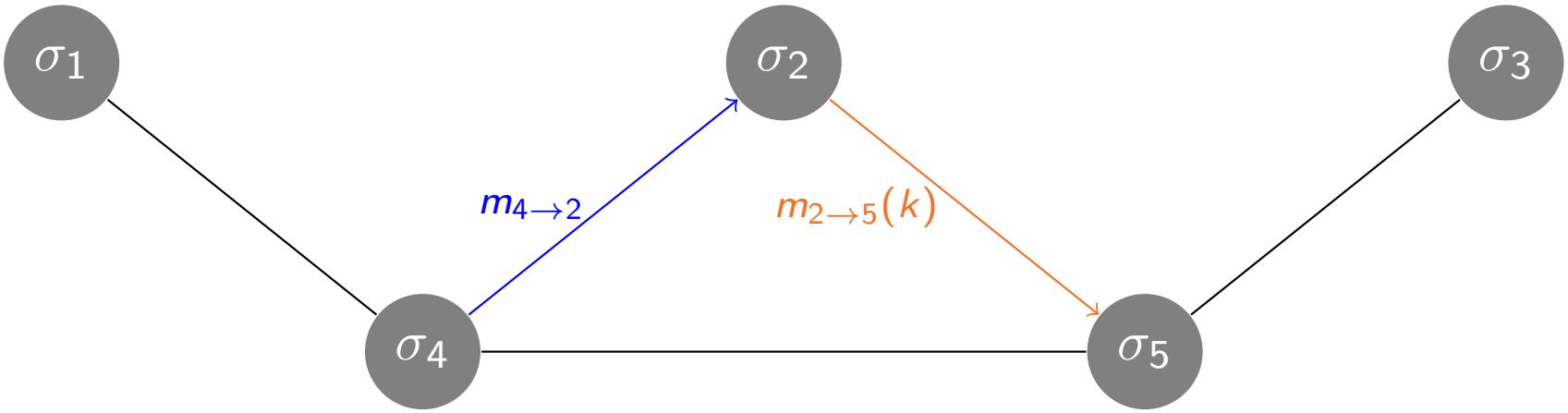
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

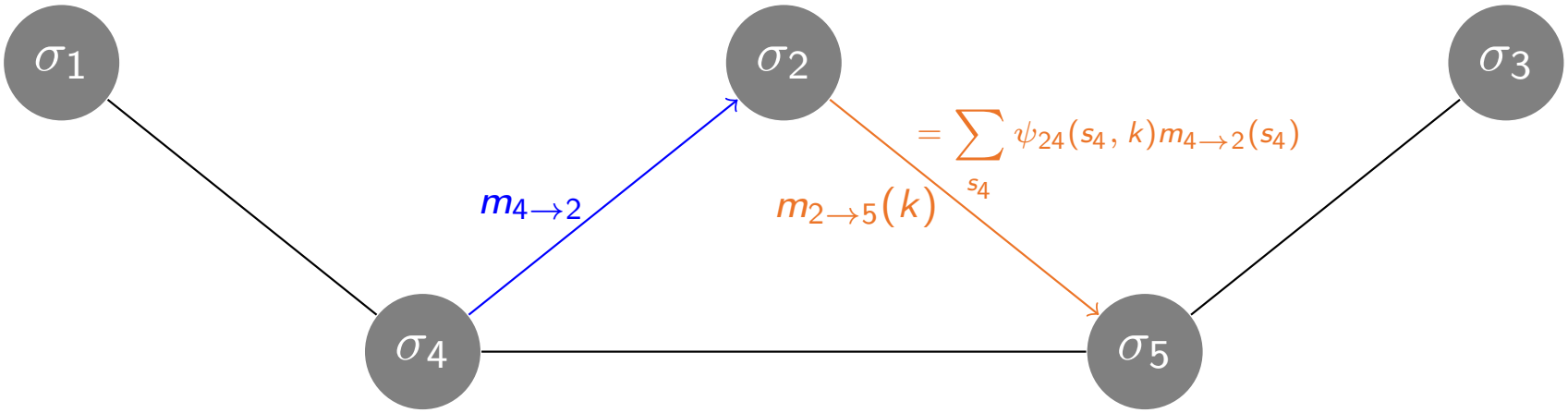
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

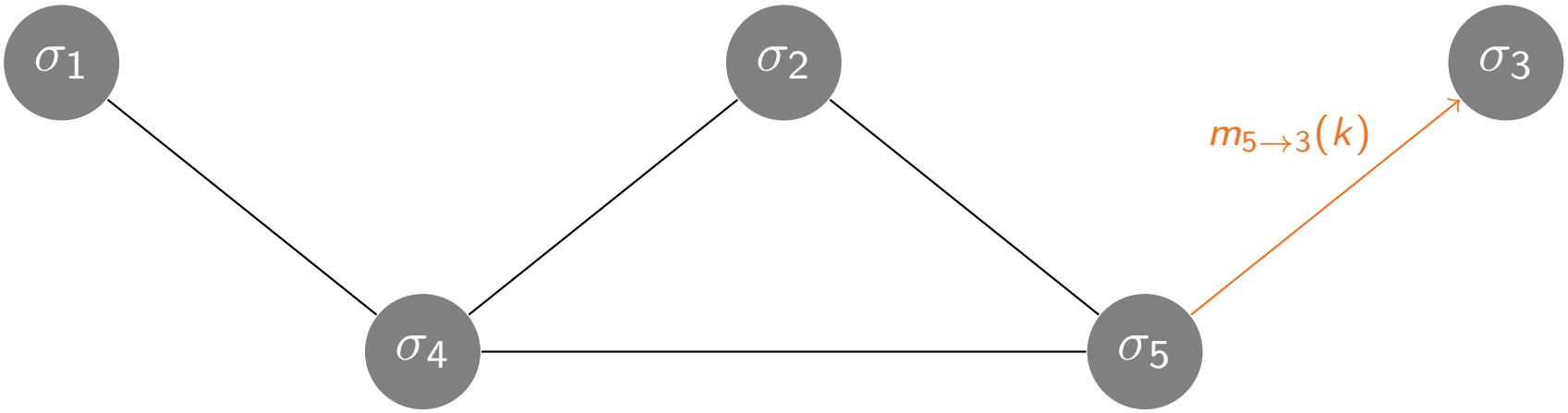
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

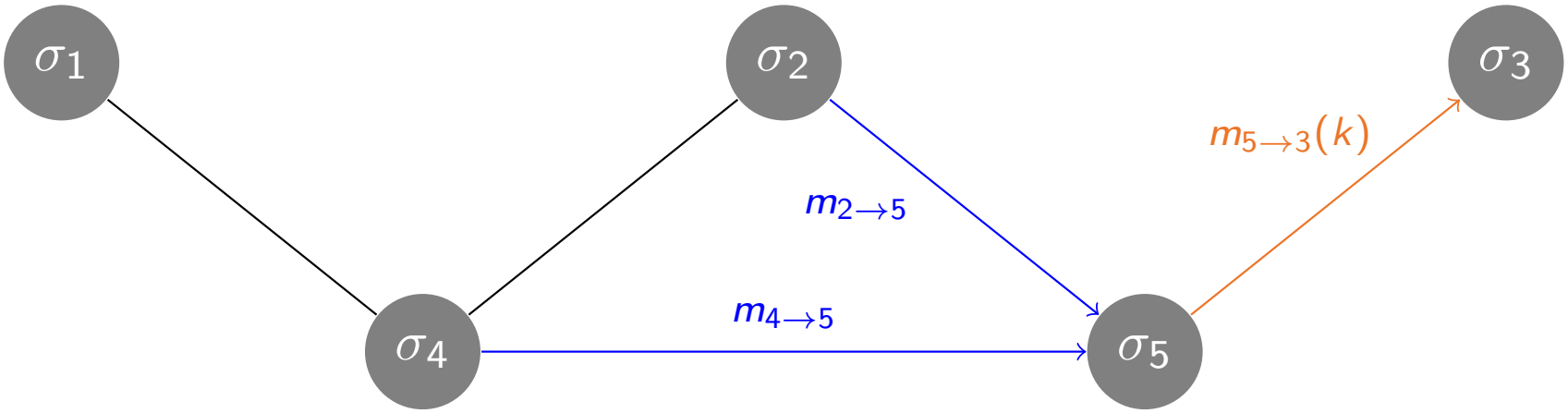
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

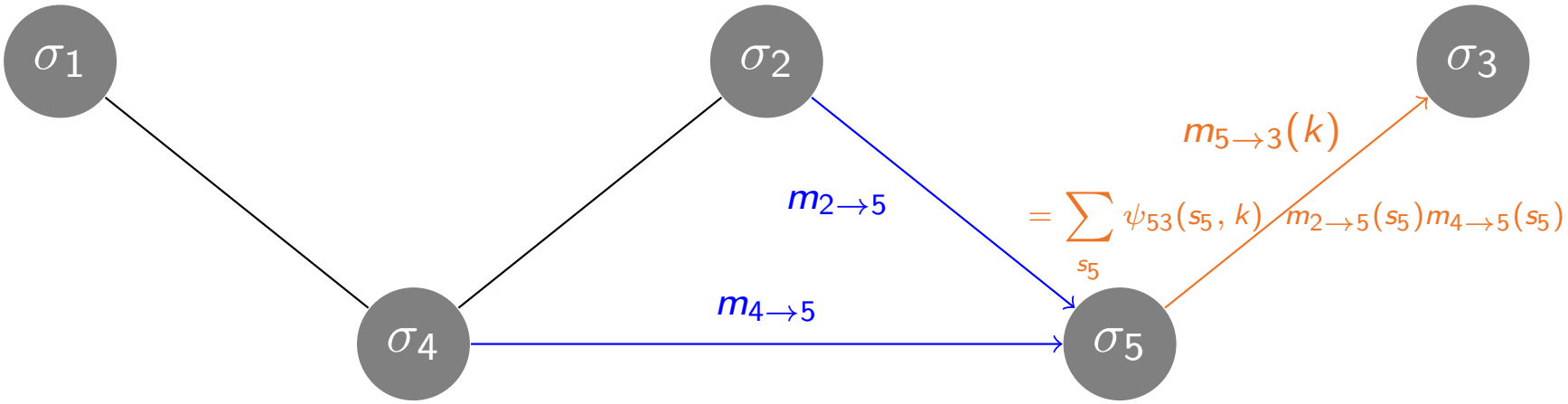
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

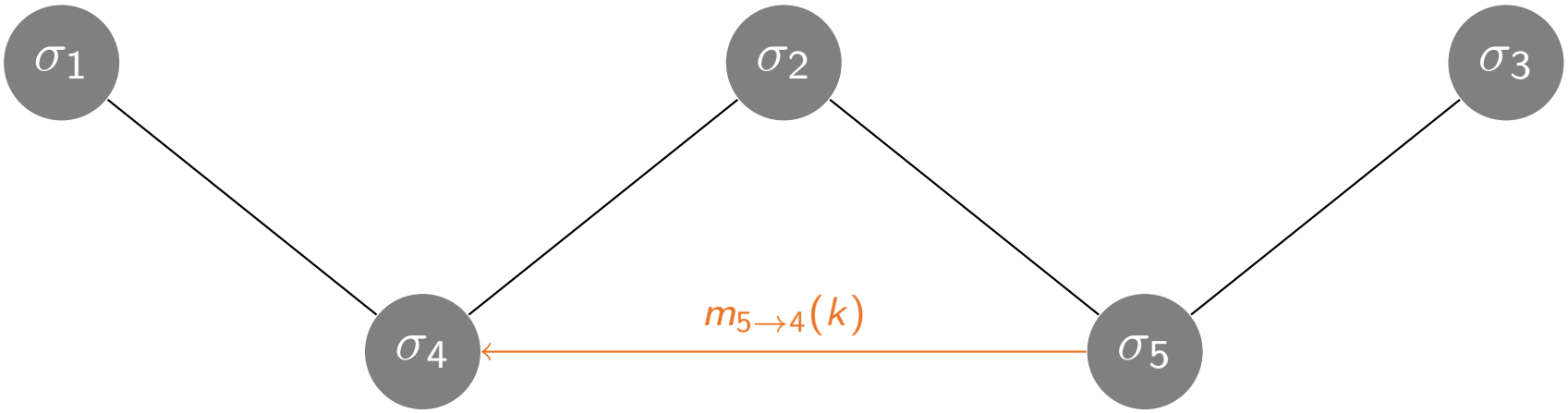
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

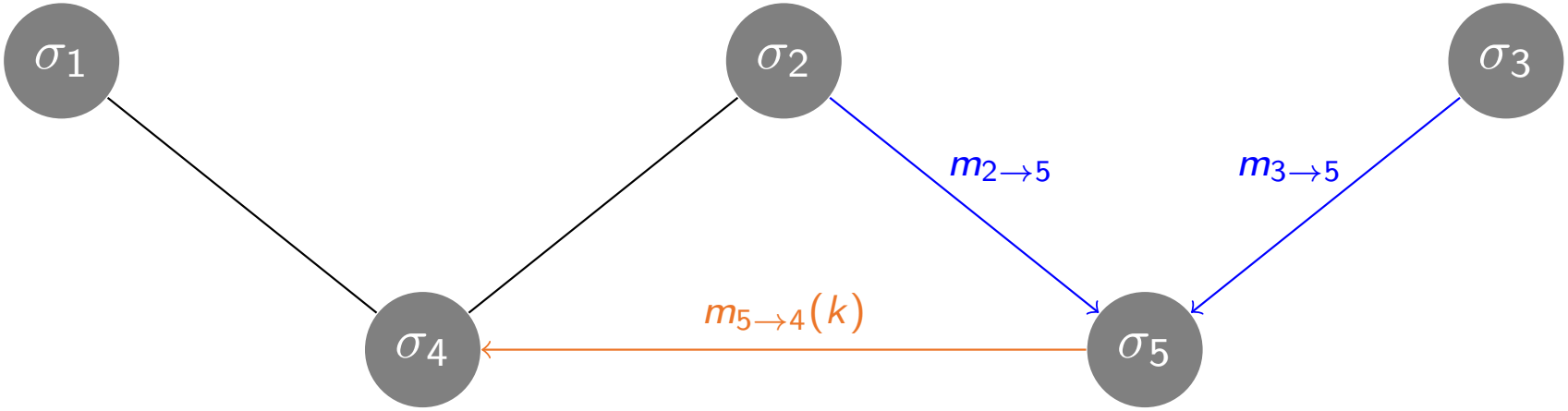
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

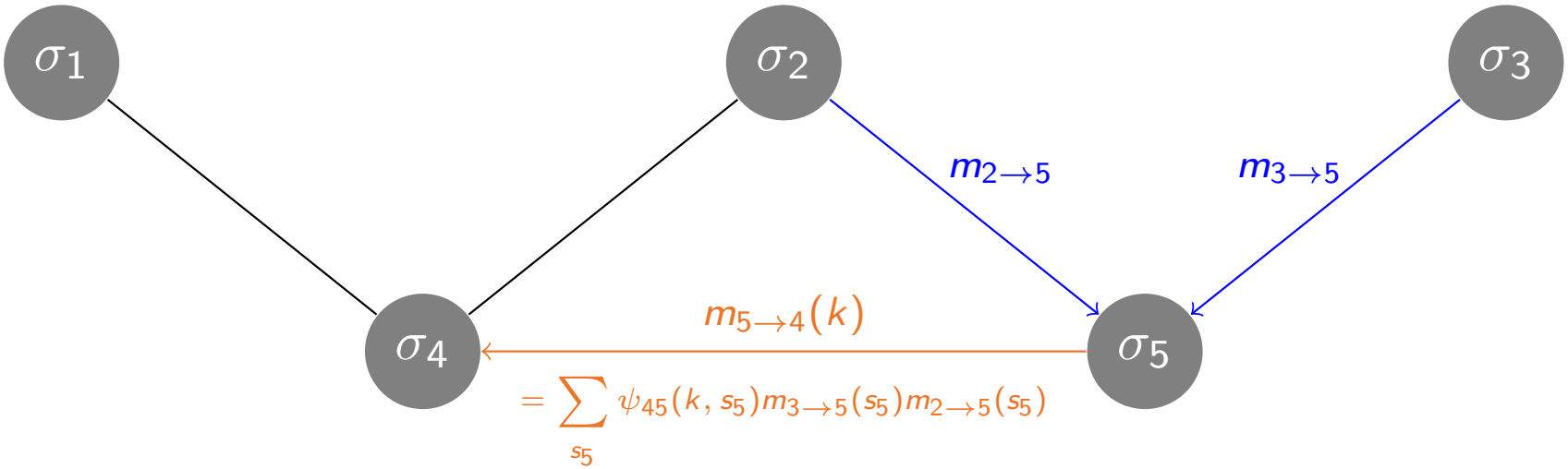
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

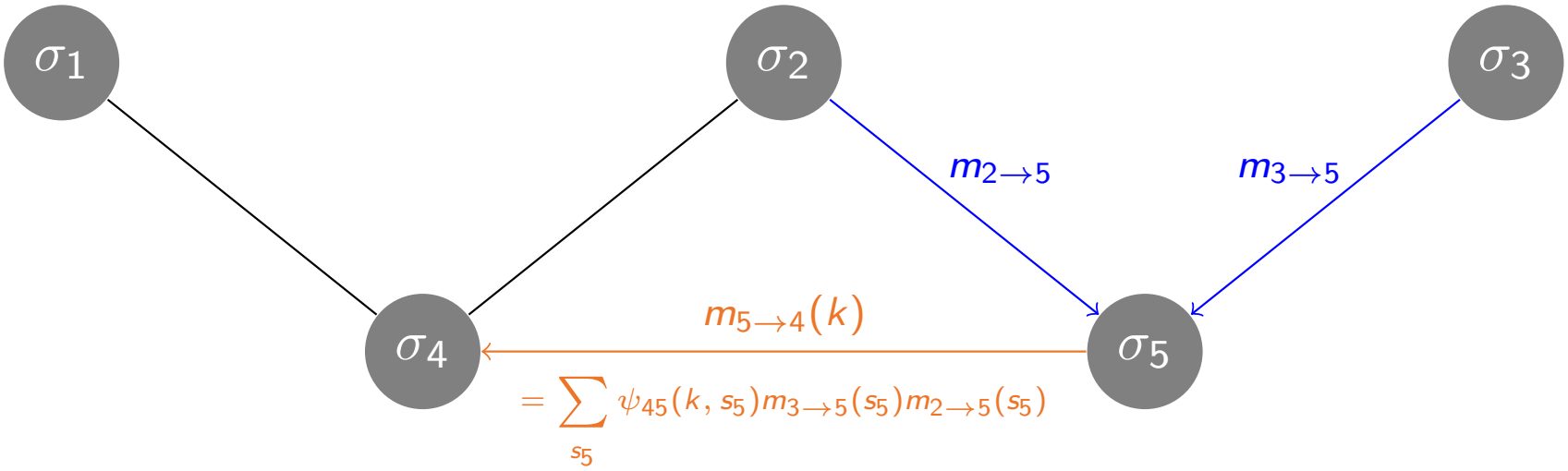
$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



Belief Propagation

Un algorithme de passage de messages

$$m_{i\to j}(s_j) \propto \sum_{s_i} \psi_{ij}(s_i, s_j)\phi_i(s_i) \prod_{k\in\partial i, k\neq j} m_{k\to i}(s_i)$$



$$b_i(x_i) \propto \phi_i(x_i) \prod_{j\in\partial i} m_{j\to i}(x_i)$$

Approximation des lois marginales

Belief Propagation

Comment introduire nos observations ?

- Une observation de X_i se traduit en probabilité sur σ_i.
- Fixer la valeur de σ_i est naturel avec BP mais pas sa probabilité.

Belief Propagation

Comment introduire nos observations ?

- Une observation de X_i se traduit en probabilité sur σ_i.
- Fixer la valeur de σ_i est naturel avec BP mais pas sa probabilité.

Repartons du problème de minimisation

$\{$Points fixes stables de BP$\} \subset \{$Minima locaux d'une approximation de $\mathrm{KL}(b||\mathbb{P})\}$

$$\min_b \sum_{\boldsymbol{\sigma}} b(\boldsymbol{\sigma}) \log \frac{b(\boldsymbol{\sigma})}{\mathbb{P}(\boldsymbol{\sigma})}$$

sous contraintes

$$b(\boldsymbol{\sigma}) = \prod_{(ij)} \frac{b_{ij}(\sigma_i, \sigma_j)}{b_i(\sigma_i) b_j(\sigma_j)} \prod_i b_i(\sigma_i), \quad \sum_{\sigma_j} b_{ij}(\sigma_i, \sigma_j) = b_i(\sigma_i), \quad \sum_{\sigma_j} b_j(\sigma_j) = 1.$$

Les mises à jour sont obtenues à partir des points stationnaires du Lagrangien de ce problème.

Mirror Belief Propagation

Nouvelles contraintes

Pour chaque observation $X_i = x_i$, on ajoute une contrainte du type

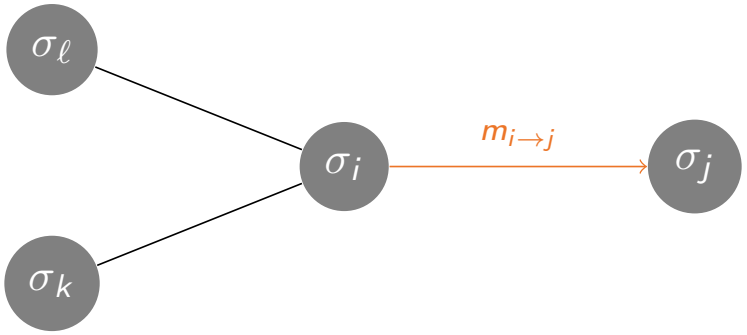
$$b_i(\sigma_i = 1) = \Lambda(x_i) \stackrel{\text{def}}{=} b_i^*(1), \quad b_i(\sigma_i = 0) = 1 - \Lambda(x_i) \stackrel{\text{def}}{=} b_i^*(0)$$

Version modifiée de Belief Propagation :

- $m_{i \to j}(\sigma_j)$ inchangé si X_i n'est pas observé
- sinon :

$$m_{i \to j}(\sigma_j) \propto \sum_{\sigma_i} \varphi_{ij}(\sigma_i, \sigma_j) \frac{b_i^*(\sigma_i)}{m_{j \to i}(\sigma_i)}$$

Le nœud i ne transfère plus l'information.



Mirror Belief Propagation

Nouvelles contraintes

Pour chaque observation $X_i = x_i$, on ajoute une contrainte du type

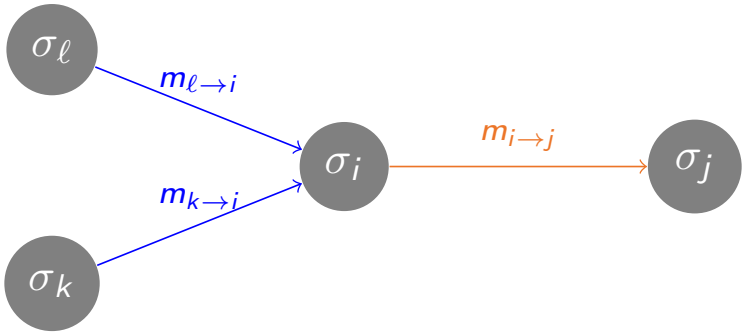
$$b_i(\sigma_i = 1) = \Lambda(x_i) \stackrel{\text{def}}{=} b_i^*(1), \quad b_i(\sigma_i = 0) = 1 - \Lambda(x_i) \stackrel{\text{def}}{=} b_i^*(0)$$

Version modifiée de Belief Propagation :

- $m_{i \to j}(\sigma_j)$ inchangé si X_i n'est pas observé
- sinon :

$$m_{i \to j}(\sigma_j) \propto \sum_{\sigma_i} \varphi_{ij}(\sigma_i, \sigma_j) \frac{b_i^*(\sigma_i)}{m_{j \to i}(\sigma_i)}$$

Le nœud i ne transfère plus l'information.



Mirror Belief Propagation

Nouvelles contraintes

Pour chaque observation $X_i = x_i$, on ajoute une contrainte du type

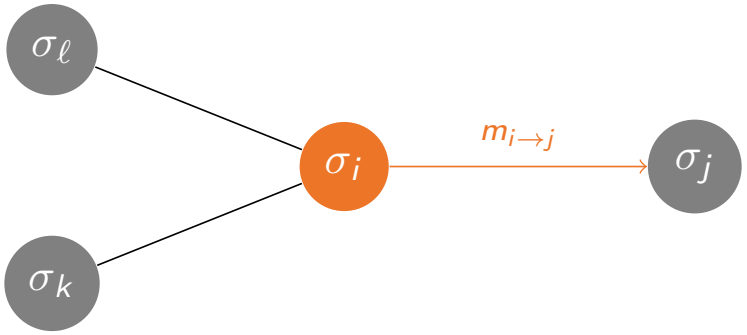
$$b_i(\sigma_i = 1) = \Lambda(x_i) \stackrel{\text{def}}{=} b_i^*(1), \quad b_i(\sigma_i = 0) = 1 - \Lambda(x_i) \stackrel{\text{def}}{=} b_i^*(0)$$

Version modifiée de Belief Propagation :

- $m_{i \to j}(\sigma_j)$ inchangé si X_i n'est pas observé
- sinon :

$$m_{i \to j}(\sigma_j) \propto \sum_{\sigma_i} \varphi_{ij}(\sigma_i, \sigma_j) \frac{b_i^*(\sigma_i)}{m_{j \to i}(\sigma_i)}$$

Le nœud i ne transfère plus l'information.



Mirror Belief Propagation

Nouvelles contraintes

Pour chaque observation $X_i = x_i$, on ajoute une contrainte du type

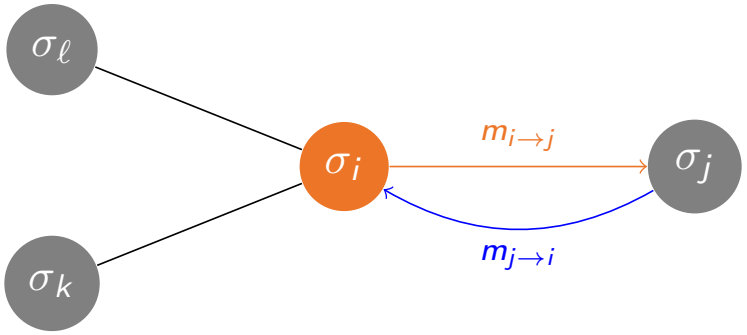
$$b_i(\sigma_i = 1) = \Lambda(x_i) \stackrel{\text{def}}{=} b_i^*(1), \quad b_i(\sigma_i = 0) = 1 - \Lambda(x_i) \stackrel{\text{def}}{=} b_i^*(0)$$

Version modifiée de Belief Propagation :

- $m_{i \to j}(\sigma_j)$ inchangé si X_i n'est pas observé
- sinon :

$$m_{i \to j}(\sigma_j) \propto \sum_{\sigma_i} \varphi_{ij}(\sigma_i, \sigma_j) \frac{b_i^*(\sigma_i)}{m_{j \to i}(\sigma_i)}$$

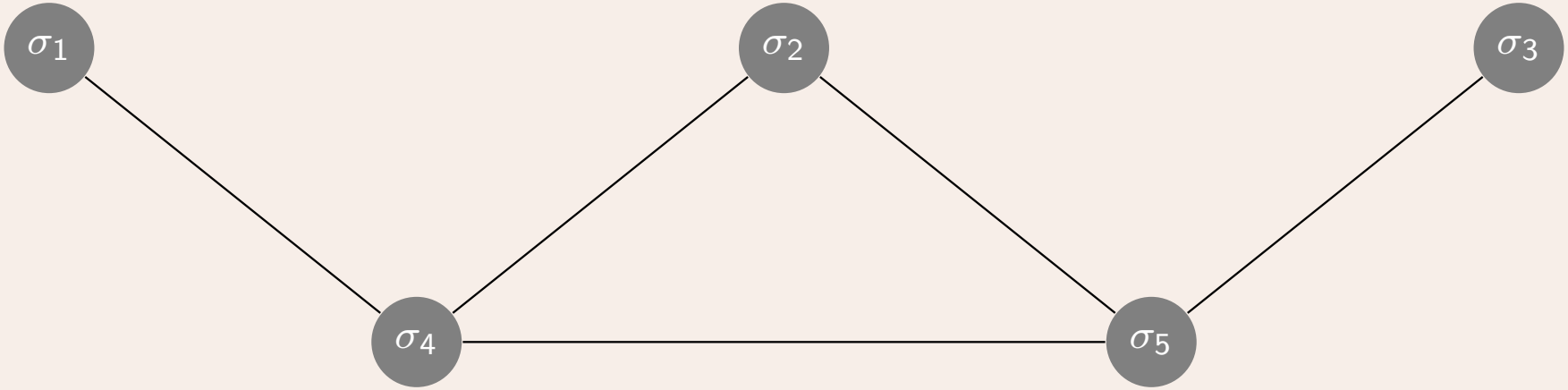
Le nœud i ne transfère plus l'information.



Stabilité de l'algorithme mirror BP

Segmentation du graphe

- Fixer une distribution marginale a l'effet de couper le graphe à ce sommet.



Une condition suffisante de convergence.

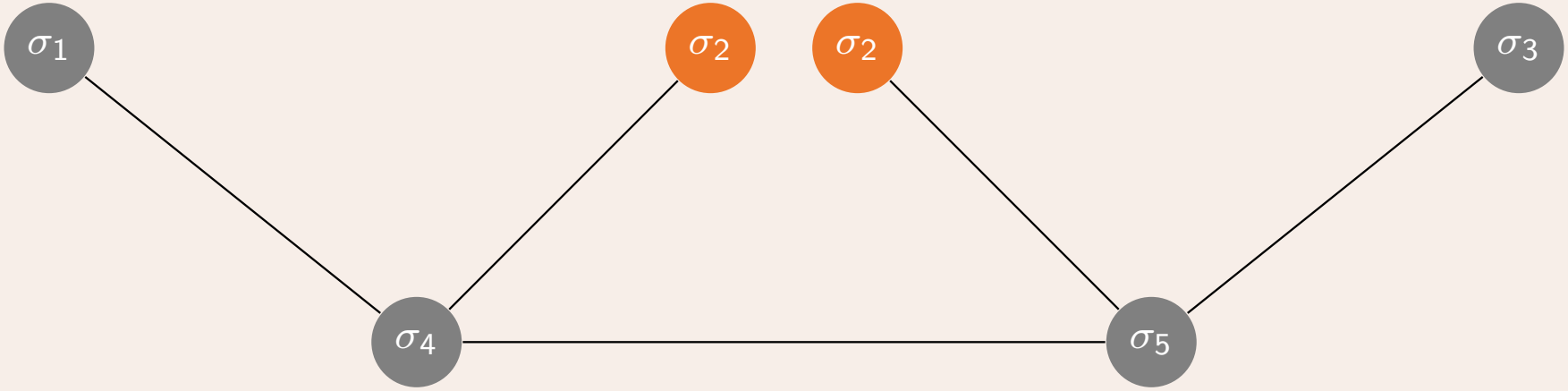
Lorsque le graphe résultant est formé d'arbres déconnectés contenant chacun au plus deux variables observées, l'algorithme converge.

Empiriquement : convergence rapide en général.

Stabilité de l'algorithme mirror BP

Segmentation du graphe

- Fixer une distribution marginale a l'effet de couper le graphe à ce sommet.



Une condition suffisante de convergence.

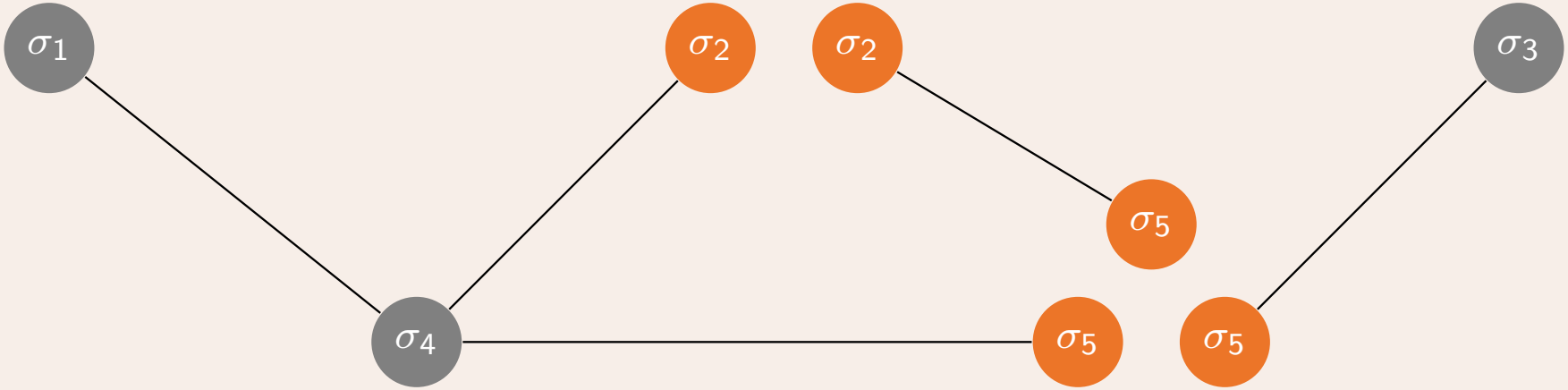
Lorsque le graphe résultant est formé d'arbres déconnectés contenant chacun au plus deux variables observées, l'algorithme converge.

Empiriquement : convergence rapide en général.

Stabilité de l'algorithme mirror BP

Segmentation du graphe

- Fixer une distribution marginale a l'effet de couper le graphe à ce sommet.



Une condition suffisante de convergence.

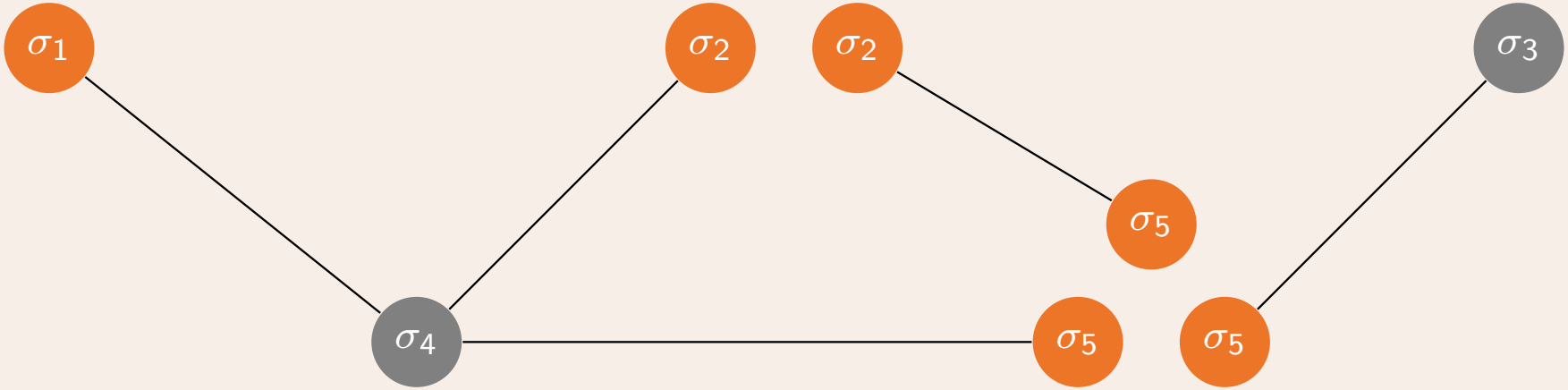
Lorsque le graphe résultant est formé d'arbres déconnectés contenant chacun au plus deux variables observées, l'algorithme converge.

Empiriquement : convergence rapide en général.

Stabilité de l'algorithme mirror BP

Segmentation du graphe

- Fixer une distribution marginale a l'effet de couper le graphe à ce sommet.

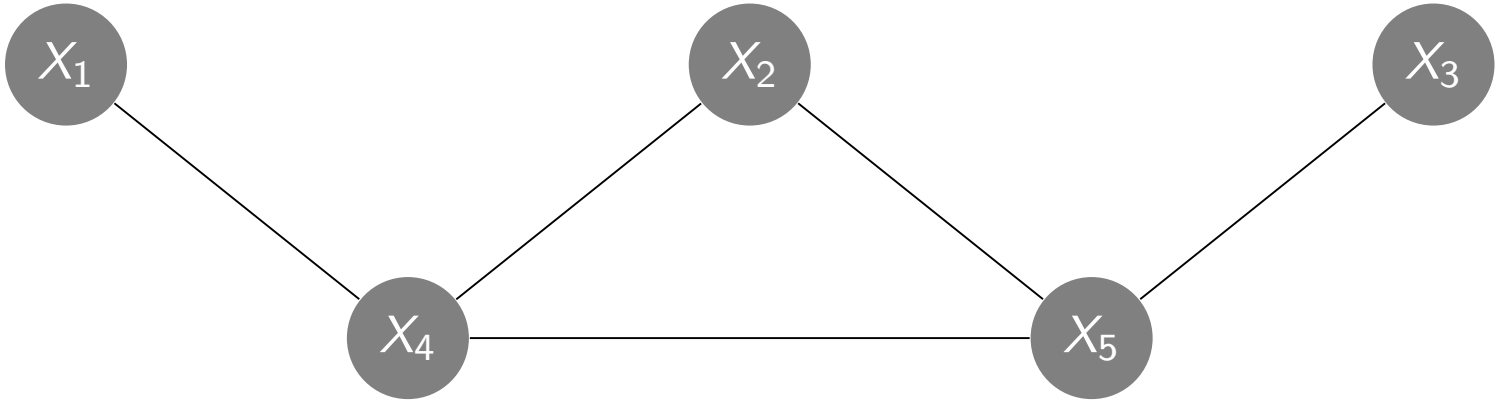


Une condition suffisante de convergence.

Lorsque le graphe résultant est formé d'arbres déconnectés contenant chacun au plus deux variables observées, l'algorithme converge.

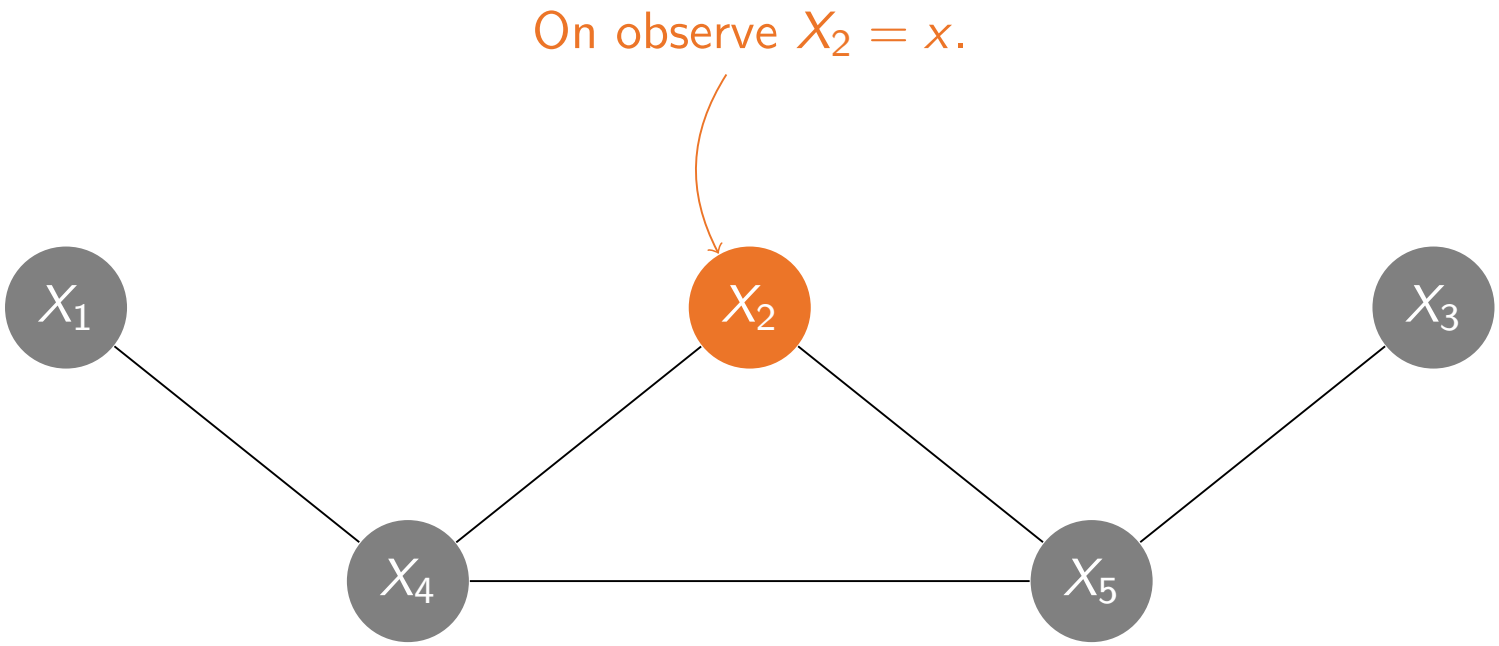
Empiriquement : convergence rapide en général.

Parenthèse : le cas gaussien.

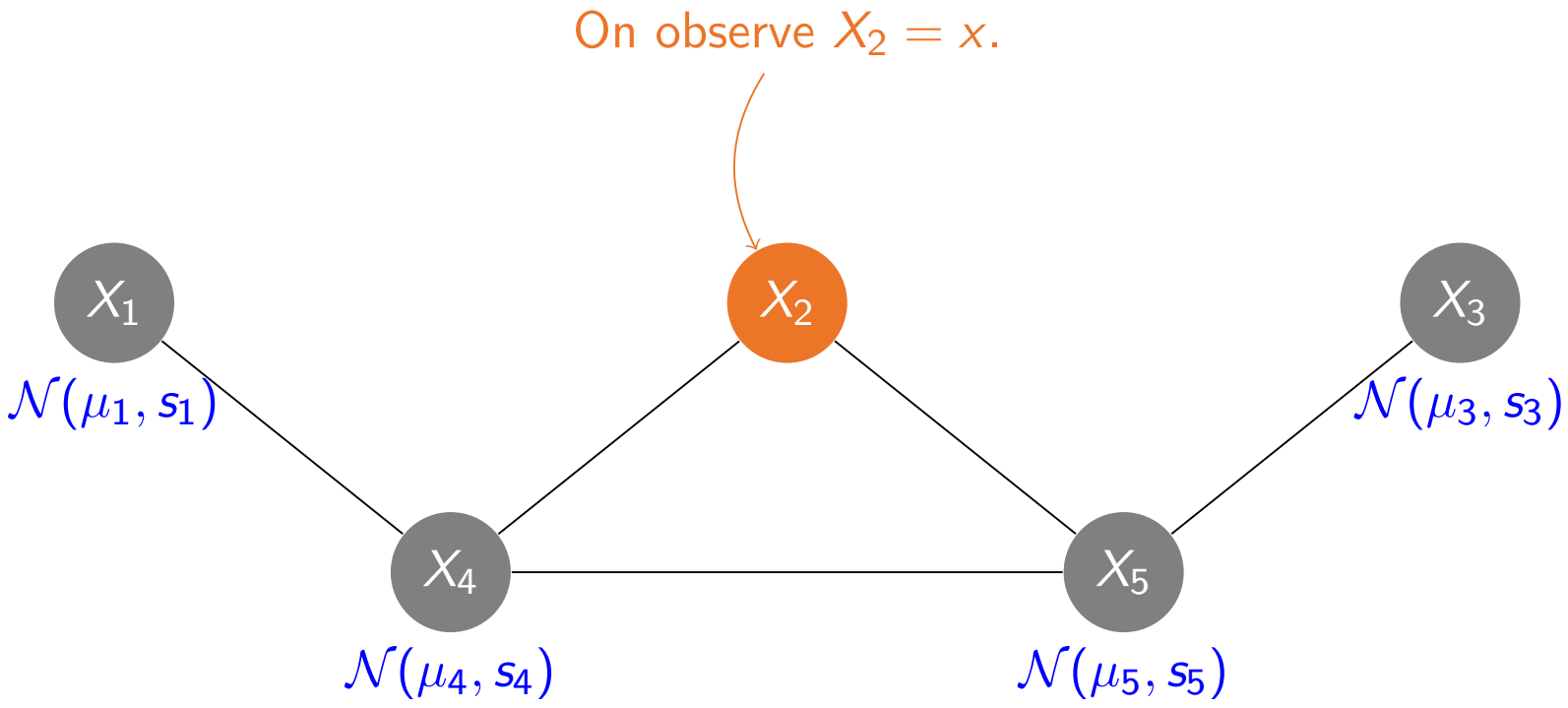


MRF de variables gaussiennes

Parenthèse : le cas gaussien.

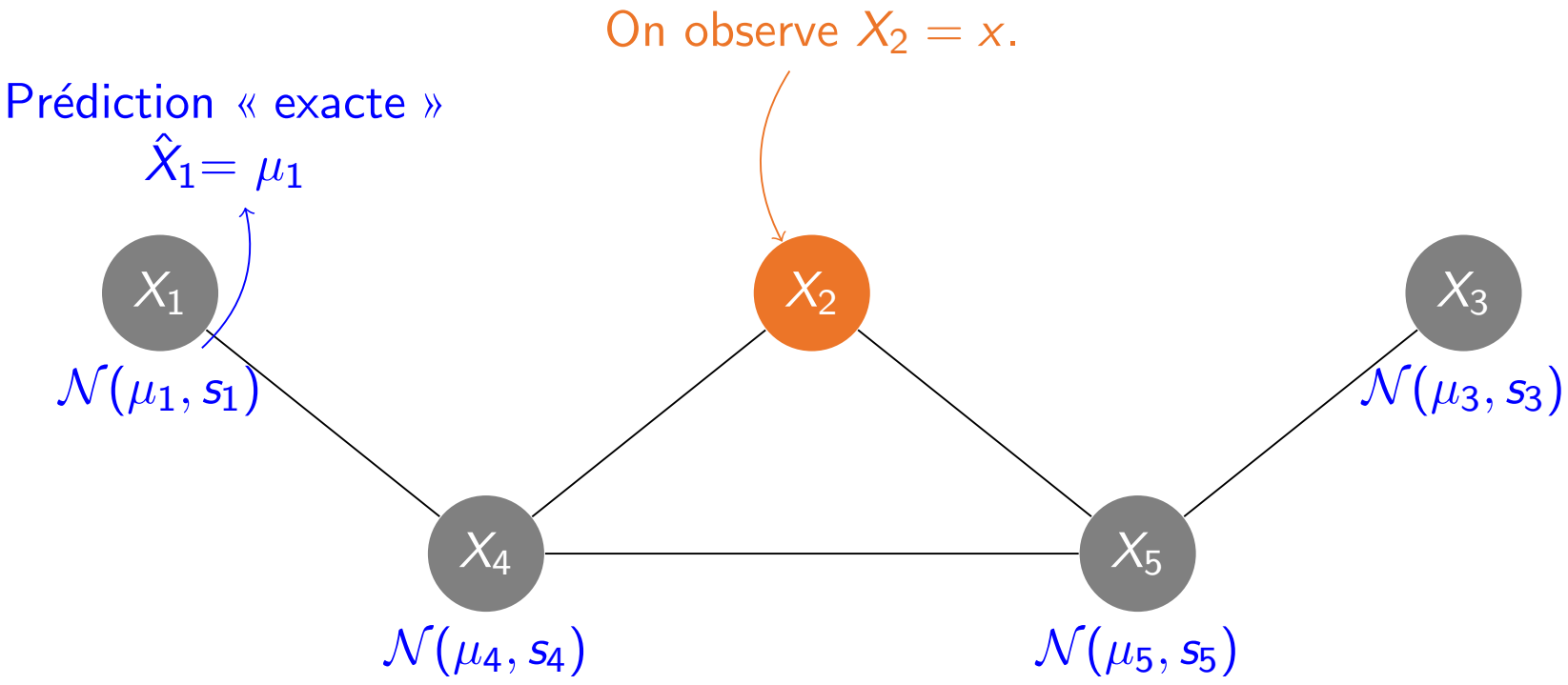


Parenthèse : le cas gaussien.



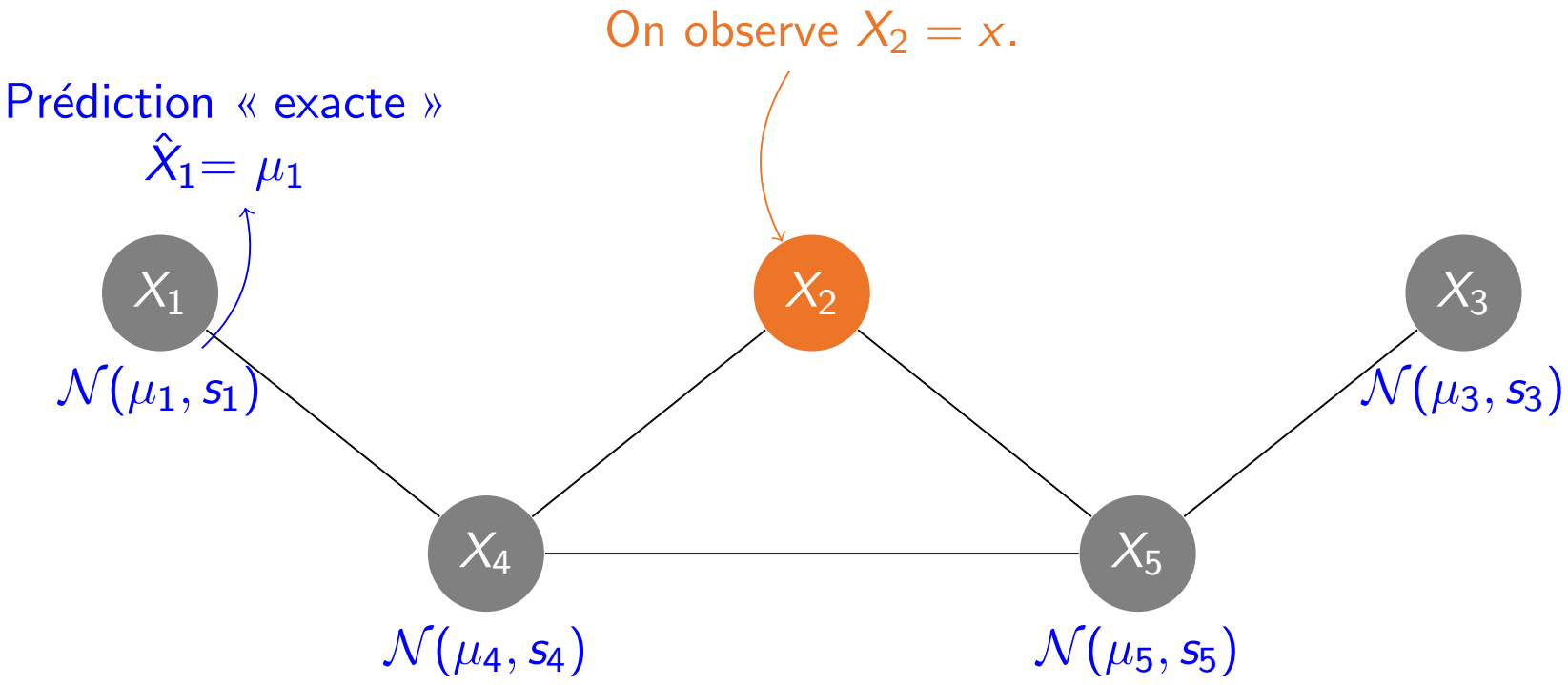
Gaussian Belief Propagation

Parenthèse : le cas gaussien.



Gaussian Belief Propagation

Parenthèse : le cas gaussien.



Gaussian Belief Propagation

Problème d'estimation : modèle de vraisemblance maximale compatible avec GaBP

Résumé

- 3 choix possibles pour l'encodage/décodage :
  - la fonction de seuil Λ_{MI} et le décodage bayésien,
  - la fonction de répartition F avec le décodage « direct » F^{-1},
  - la fonction de répartition F avec le décodage bayésien $\Gamma^{\mathcal{P}}$.
- Estimation des marginales de paires par l'approche EM.
- Inférence sur le modèle d'Ising par l'algorithme mirror BP.

Résumé

- 3 choix possibles pour l'encodage/décodage :
  - la fonction de seuil Λ_{MI} et le décodage bayésien,
  - la fonction de répartition F avec le décodage « direct » F^{-1},
  - la fonction de répartition F avec le décodage bayésien $\Gamma^{\mathcal{P}}$.
- Estimation des marginales de paires par l'approche EM.
- Inférence sur le modèle d'Ising par l'algorithme mirror BP.

Expérience de décimation

- On répète l'expérience ci-dessous pour obtenir le comportement des différents estimateurs.
- Pour une configuration donnée du réseau :
  - on révèle les variables dans un ordre aléatoire,
  - on prédit les variables non observées par les différentes méthodes décrites,
  - on calcule l'erreur (L^1) de prédiction sur ces variables.

Création du modèle

Données synthétiques

- permet de se comparer à la solution exacte du problème,
- ... et de compenser l'absence de données réelles...

Création du modèle

Données synthétiques

- permet de se comparer à la solution exacte du problème,
- ... et de compenser l'absence de données réelles...

Procédure de génération des modèles

1. On choisit un graphe de dépendance $\mathcal{G}$

Création du modèle

Données synthétiques

- permet de se comparer à la solution exacte du problème,
- ... et de compenser l'absence de données réelles...

Procédure de génération des modèles

1. On choisit un graphe de dépendance $\mathcal{G}$
2. On génère (aléatoirement) une matrice de précision $\mathbf{A}$ de support $\mathcal{G}$,

Création du modèle

Données synthétiques

- permet de se comparer à la solution exacte du problème,
- ... et de compenser l'absence de données réelles...

Procédure de génération des modèles

1. On choisit un graphe de dépendance $\mathcal{G}$
2. On génère (aléatoirement) une matrice de précision $\mathbf{A}$ de support $\mathcal{G}$,
3. On génère des réalisations du vecteur gaussien $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1})$,

Création du modèle

Données synthétiques

- permet de se comparer à la solution exacte du problème,
- ... et de compenser l'absence de données réelles...

Procédure de génération des modèles

1. On choisit un graphe de dépendance $\mathcal{G}$
2. On génère (aléatoirement) une matrice de précision $\mathbf{A}$ de support $\mathcal{G}$,
3. On génère des réalisations du vecteur gaussien $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1})$,
4. On transforme $\mathbf{Y}$ pour obtenir des distributions marginales choisies.

Création du modèle

Données synthétiques

- permet de se comparer à la solution exacte du problème,
- ... et de compenser l'absence de données réelles...

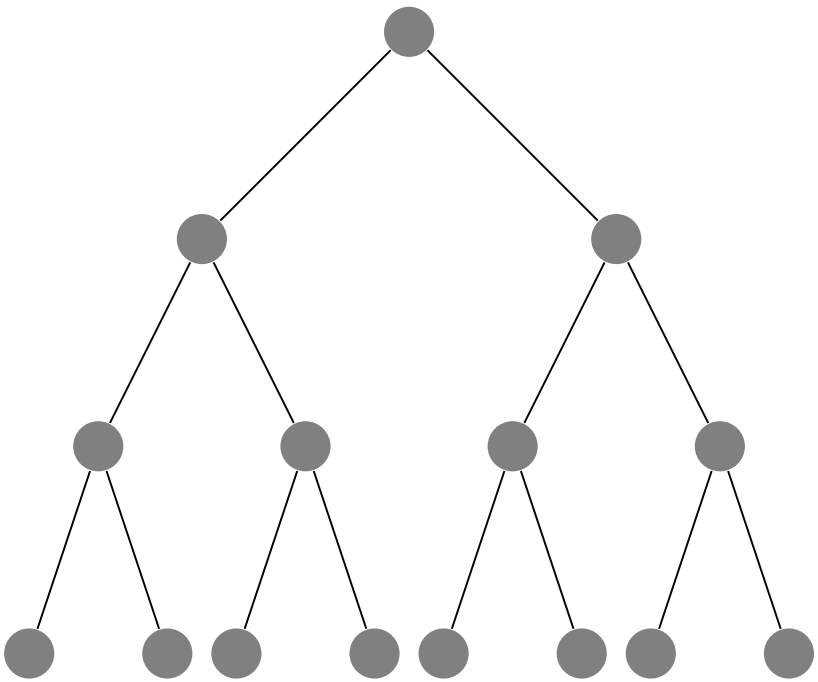
Procédure de génération des modèles

1. On choisit un graphe de dépendance $\mathcal{G}$
2. On génère (aléatoirement) une matrice de précision $\mathbf{A}$ de support $\mathcal{G}$,
3. On génère des réalisations du vecteur gaussien $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1})$,
4. On transforme $\mathbf{Y}$ pour obtenir des distributions marginales choisies.

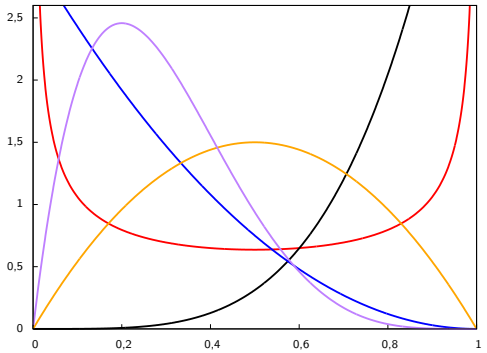
Types de graphes

- Arbres réguliers,
- Modèle jouet d'une agglomération.

Arbres réguliers



- Distributions marginales $\beta(a, b)$ des variables.



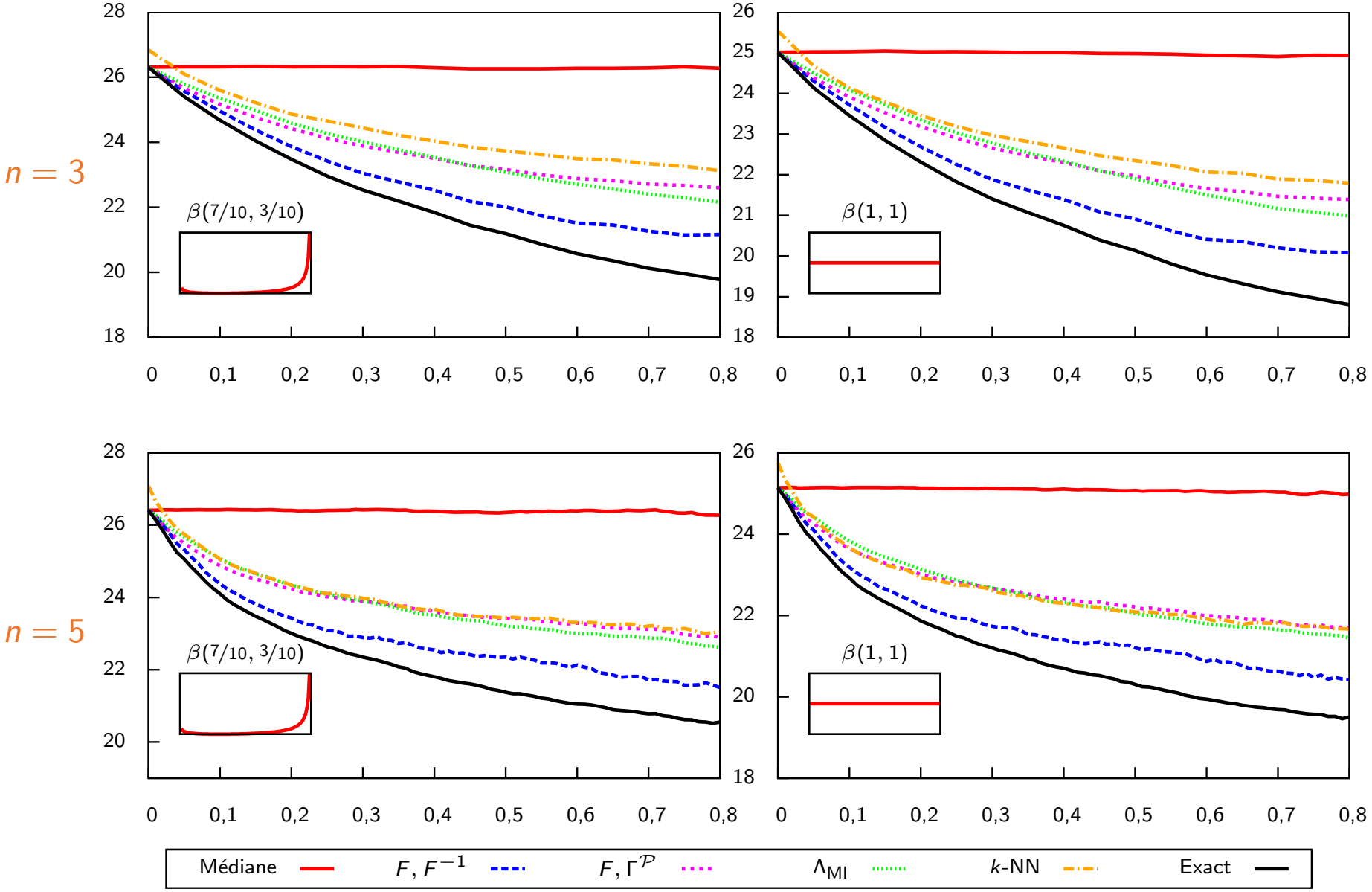
- Modèle exact, « paires indépendantes » :

$$\mathbb{P}(\boldsymbol{\sigma}) = \prod_{ij} \frac{\mathbb{P}(\sigma_i, \sigma_j)}{\mathbb{P}(\sigma_i)\mathbb{P}(\sigma_j)} \prod_i \mathbb{P}(\sigma_i)$$

- Belief Propagation réalise l'inférence exacte sur un arbre.

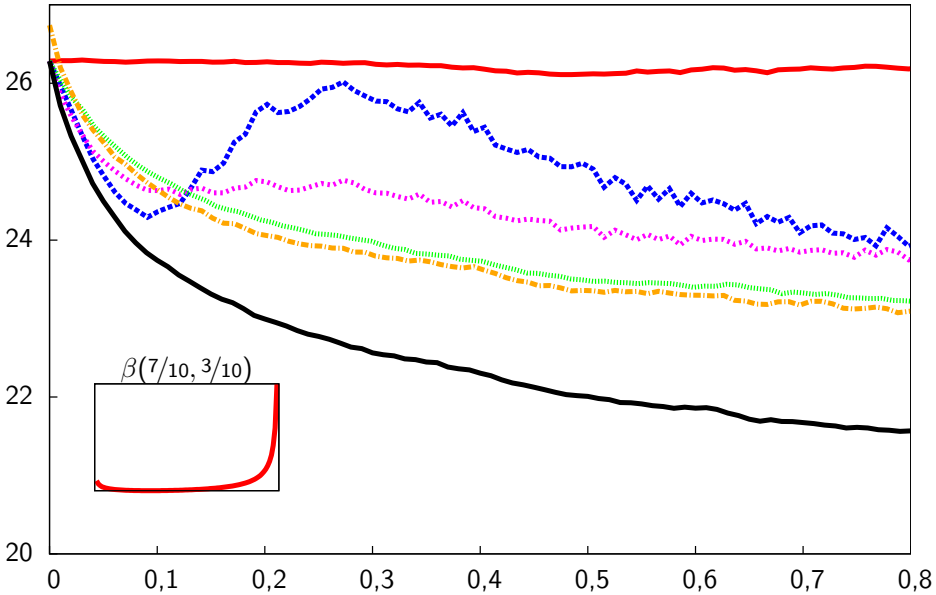
Effet encodage/décodage.

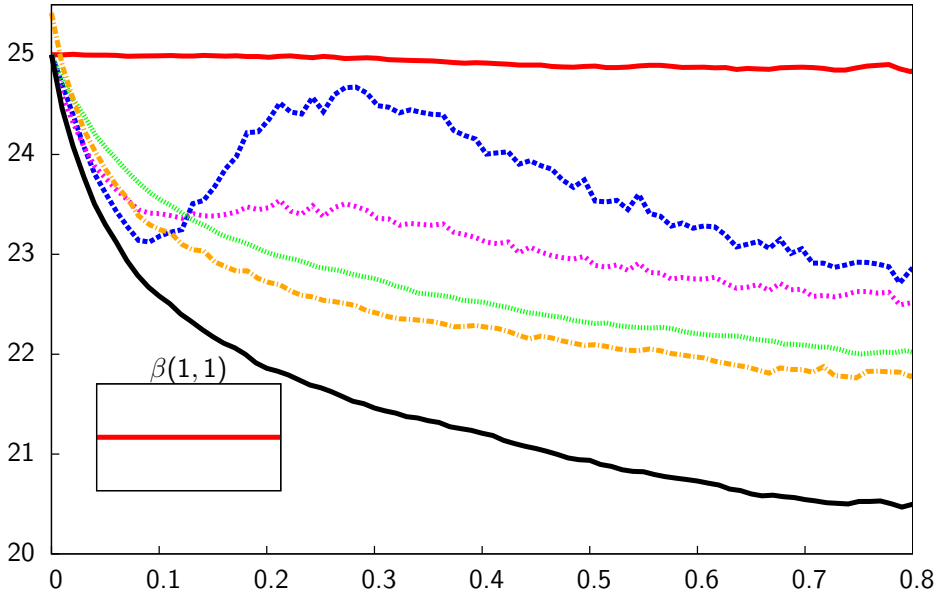
Arbres réguliers : résultats



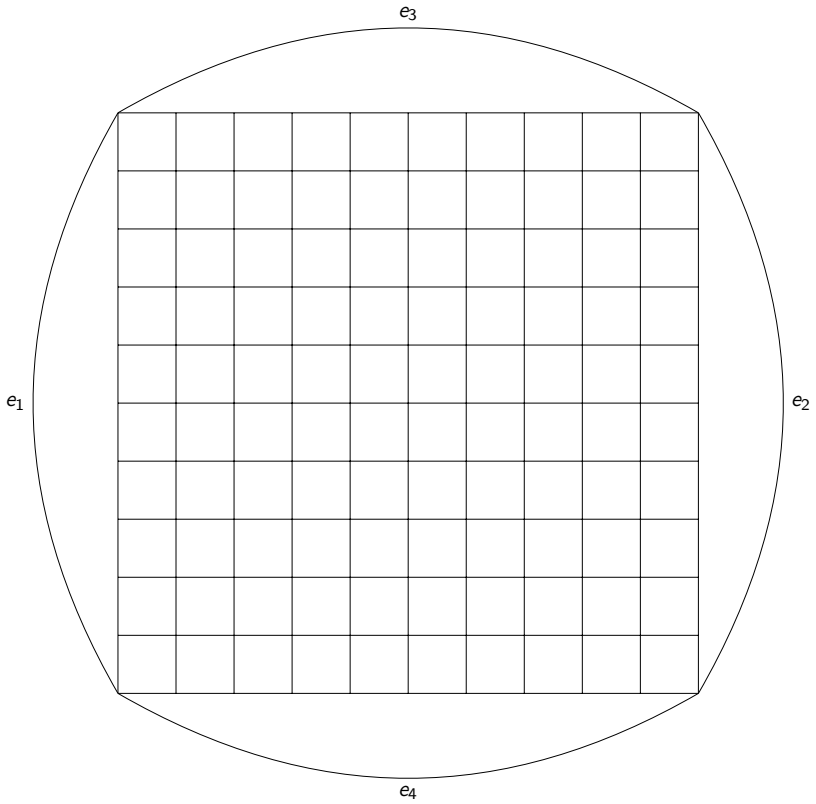
Arbres réguliers : résultats

$n = 10$





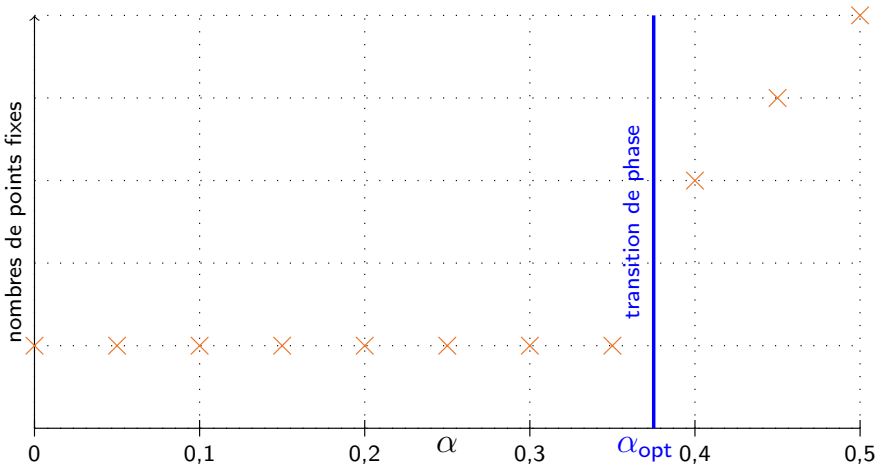
Réseau urbain



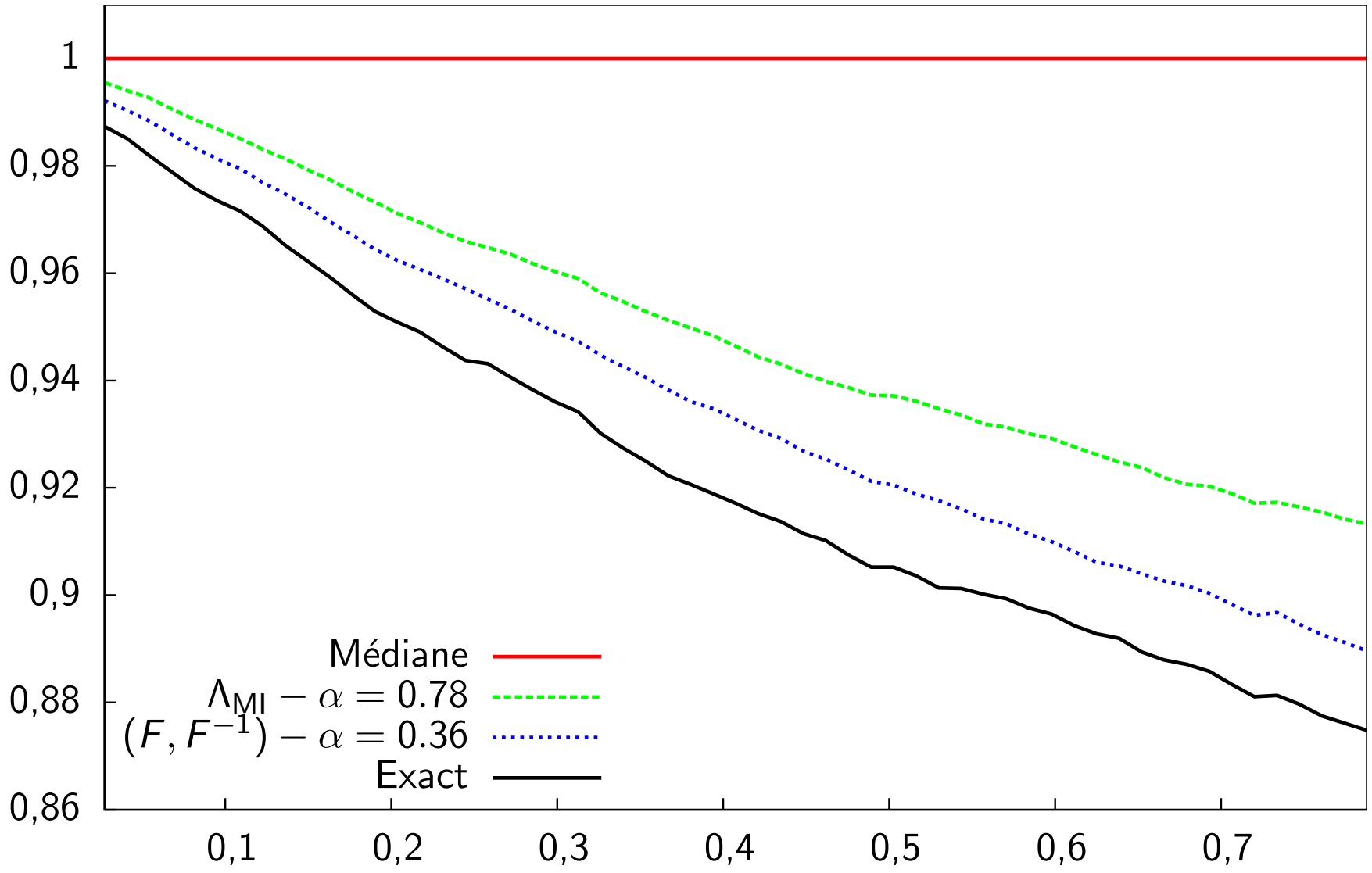
- Dépendances fixées par les routages possibles.
- Distributions marginales des temps de trajet issues de données d'autoroutes australiennes.
- Estimation du modèle :

$$\mathbb{P}(\boldsymbol{\sigma}) = \prod_{ij} \left(\frac{\mathbb{P}(\sigma_i, \sigma_j)}{\mathbb{P}(\sigma_i)\mathbb{P}(\sigma_j)} \right)^{\alpha} \prod_{i} \mathbb{P}(\sigma_i)$$

- Calibration du paramètre α :



Réseau urbain : résultats



Conclusion :

- Méthodes simples permettant une résolution approchée du problème à partir
  - d'une matrice de covariance incomplète,
  - d'une fonction de répartition par variable,
- Contraintes non classiques prises en compte (compatibilité avec BP).
- Résultats théoriques sur l'algorithme BP.

Conclusion :

- Méthodes simples permettant une résolution approchée du problème à partir
  - d'une matrice de covariance incomplète,
  - d'une fonction de répartition par variable,
- Contraintes non classiques prises en compte (compatibilité avec BP).
- Résultats théoriques sur l'algorithme BP.

Perspectives :

- Extension à des modèles à plus d'états.
- Estimation du graphe du modèle d'Ising.
- Ajout de variables macroscopiques (tests en cours).
- Test sur données réelles.
- Approfondir l'approche gaussienne.

Merci de votre attention !

Densité : $f_{\mathbf{A}}(\mathbf{x}) = \sqrt{\frac{\det(\mathbf{A})}{(2\pi)^n}} \exp\left(-\frac{1}{2}\langle \mathbf{x}, \mathbf{A}\mathbf{x}\rangle\right)$

Vraisemblance (cas d'observations complètes) :
$\mathcal{L}_{\mathbf{A}}(\{\mathbf{x}^k\}_{k\in\{1\ldots \mathrm{K}\}}) = \log\det \mathbf{A} + \mathrm{Tr}(\mathbf{A}\widehat{\mathbf{\Sigma}}) = \int \log(f_{\mathbf{A}}(\mathbf{x}))\hat{\mu}^c(\mathbf{x})d\mathbf{x}$, avec
$\hat{\mu}^c(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}} \prod_i \mathbb{1}_{\{x_i = x_i^k\}}$.

Vraisemblance (cas d'observations incomplètes) :
$\mathcal{L}_{\mathbf{A}}(\{\mathbf{x}^k\}_{k\in\{1\ldots \mathrm{K}\}}) = \frac{1}{2}\sum_{k=1}^{\mathrm{K}} \left(\log\det \mathbf{\Sigma}_{\mathbb{V}_k}^{-1} - \left\langle \mathbf{x}^k, \mathbf{\Sigma}_{\mathbb{V}_k}^{-1}\mathbf{x}^k\right\rangle\right)$.

Vraisemblance associée à une mesure empirique $\hat{\mu}$:
$\mathcal{L}(\hat{\mu}, \mathbf{A}) = \frac{1}{2}\log\det(\mathbf{A}) - \frac{1}{2}\sum_{(i,j)} A_{ij}\mathbb{E}_{\hat{\mu}^{ij}}\left[X_i X_j\right]$.

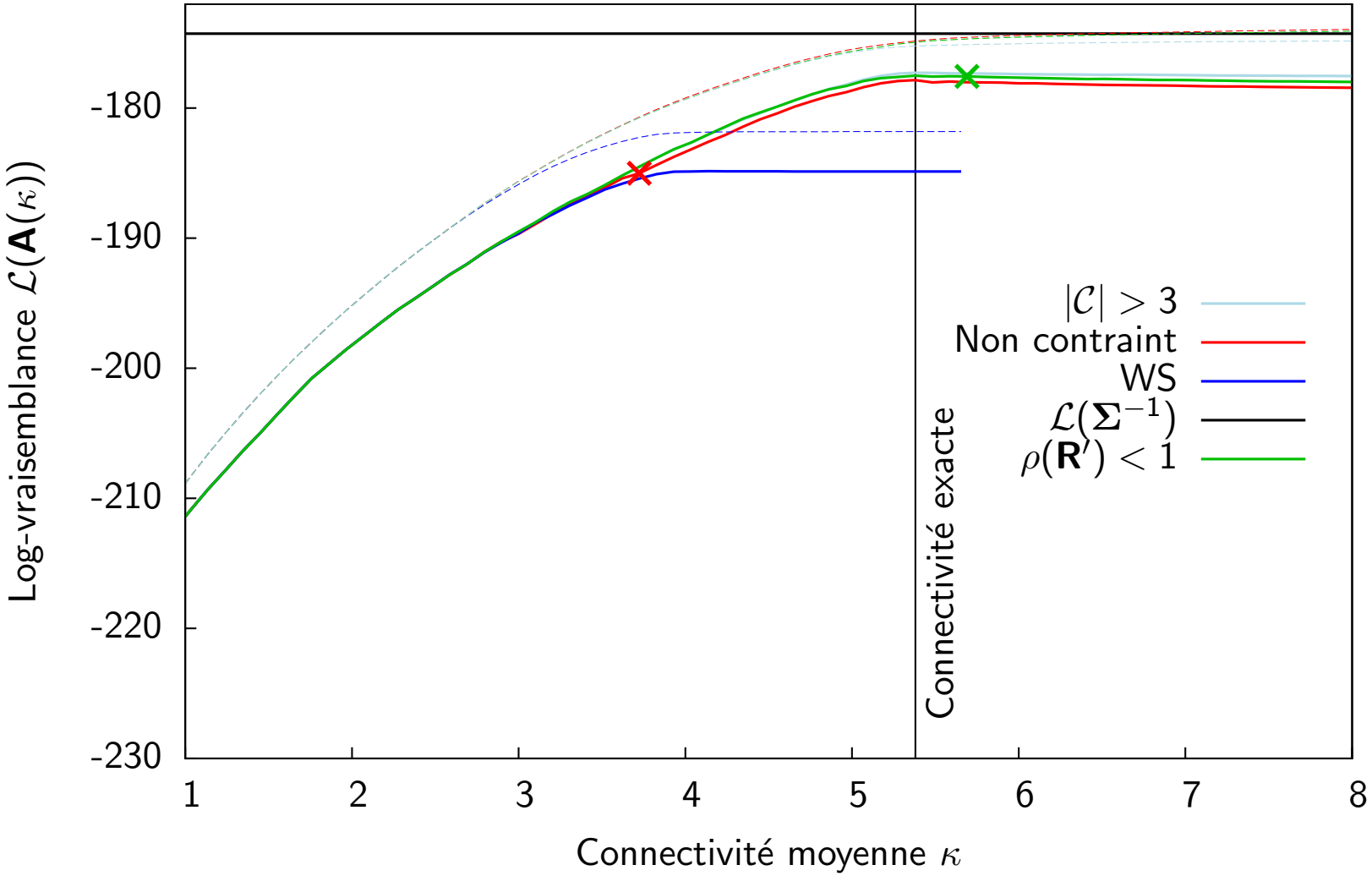
Définition de la mesure empirique $\hat{\mu}$:
$\hat{\mu}(\mathbf{x}) = \frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}} \prod_{i\in V_k} \mathbb{1}_{\{x_i = x_i^k\}}\ \nu(\mathbf{x}_{V\setminus V_k} \mid \mathbf{x}_{V_k})$.

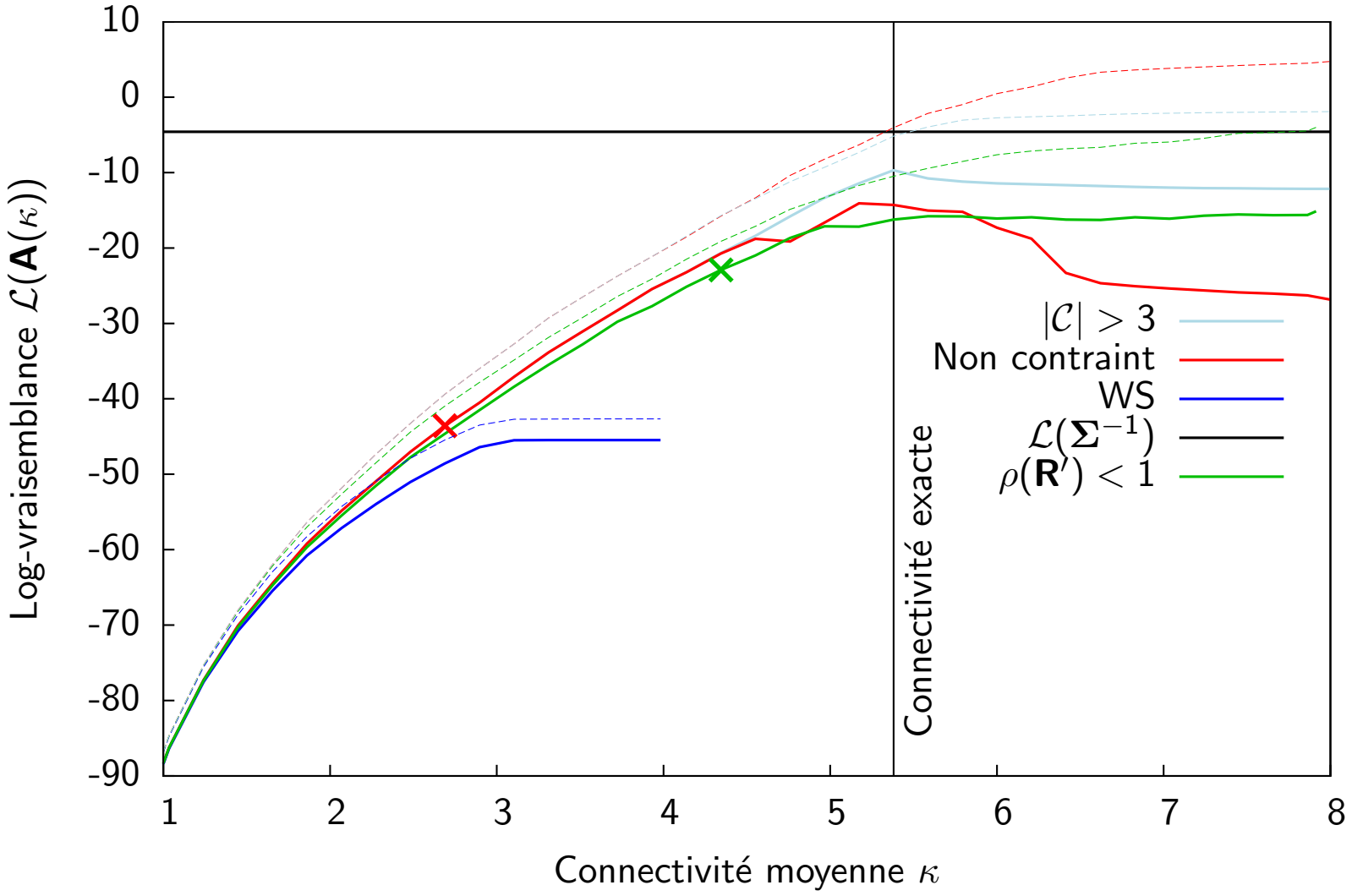
Perturbation optimale d'une paire (i, j) :

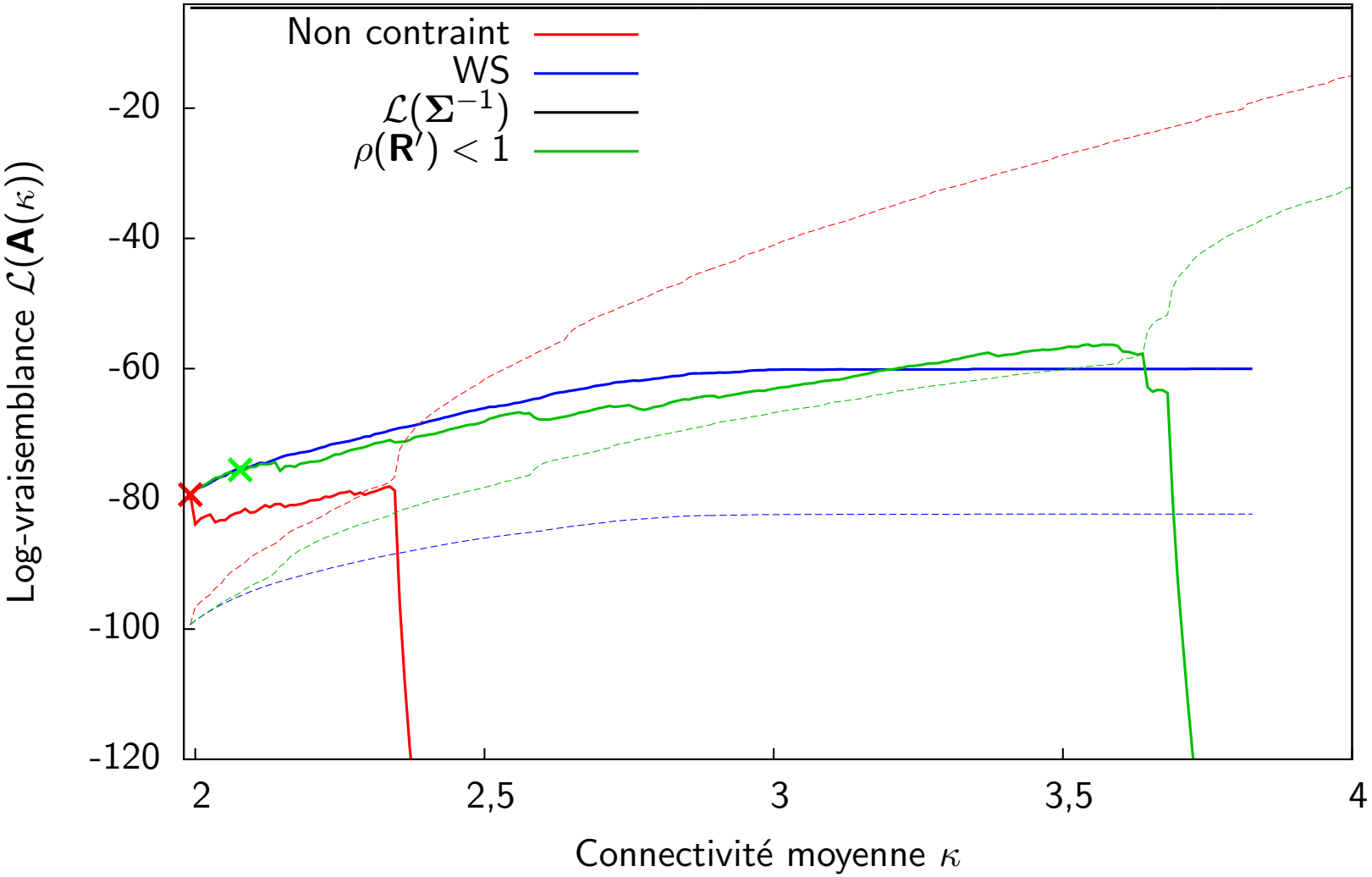
$$\mathbf{A}' = \mathbf{A} + \left[\widehat{\mathbf{\Sigma}}_{\{i,j\}}^{-1} - \mathbf{\Sigma}_{\{i,j\}}^{-1} \right]$$

Variation de vraisemblance associée :

$$\Delta \mathcal{L}^{ij} = \frac{1}{2} \left(\mathrm{Tr} \left(\widehat{\mathbf{\Sigma}}_{\{i,j\}} \mathbf{\Sigma}_{\{i,j\}}^{-1} \right) - 2 - \log \frac{\det \widehat{\mathbf{\Sigma}}_{\{i,j\}}}{\det \mathbf{\Sigma}_{\{i,j\}}} \right)$$


-180
-190
-200
-210
-220
-230
1
2
3
4
5
6
7
8
Log-vraisemblance $\mathcal{L}(\mathbf{A}(\kappa))$
Connectivité moyenne κ
Connectivité exacte
$|\mathcal{C}| > 3$
Non contraint
WS
$\mathcal{L}(\mathbf{\Sigma}^{-1})$
$\rho(\mathbf{R}') < 1$


10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
1
2
3
4
5
6
7
8
Log-vraisemblance $\mathcal{L}(\mathbf{A}(\kappa))$
Connectivité moyenne κ
Connectivité exacte
$|\mathcal{C}| > 3$
Non contraint
WS
$\mathcal{L}(\mathbf{\Sigma}^{-1})$
$\rho(\mathbf{R}') < 1$


Non contraint
WS
$\mathcal{L}(\mathbf{\Sigma}^{-1})$
$\rho(\mathbf{R}') < 1$
-20
-40
-60
-80
-100
-120
2
2,5
3
3,5
4
Log-vraisemblance $\mathcal{L}(\mathbf{A}(\kappa))$
Connectivité moyenne κ