

Réseaux Bayésiens à Densités Conditionnelles

9èmes Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes

Santiago Cortijo, Marvin Lasserre, Christophe Gonzales

31 mai 2018

Sorbonne Université, CNRS, Laboratoire d'informatique de Paris 6, LIP6



1. Contexte général
2. Réseaux bayésiens à densités conditionnelles tronquées (ctdBN)
3. Réseaux bayésiens à densités conditionnelles (cdBN)
4. Conclusion

Contexte général

- Les **réseaux bayésiens** permettent de **raisonner avec l'incertain**.

Réseaux bayésiens et variables continues

- Les réseaux bayésiens permettent de raisonner avec l'incertain.
- **Problème !** Ils ne peuvent traiter qu'avec des variables aléatoires discrètes.

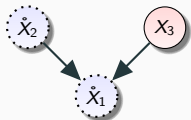
Réseaux bayésiens et variables continues

- Les réseaux bayésiens permettent de raisonner avec l'incertain.
- **Problème !** Ils ne peuvent traiter qu'avec des variables aléatoires discrètes.
- Plusieurs approches proposées pour des variables mixtes :

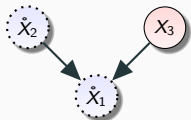
Réseaux bayésiens et variables continues

- Les **réseaux bayésiens** permettent de **raisonner avec l'incertain**.
- **Problème !** Ils ne peuvent traiter qu'avec des **variables aléatoires discrètes**.
- Plusieurs approches proposées pour des variables mixtes :
 - Modèles gaussiens conditionnels (**CG**) :

- Exemple :



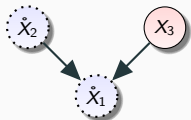
- Exemple :



- [Lauritzen, 1992] :

$$P(\dot{X}_1 | \dot{X}_2, X_3 = x_3) = \mathcal{N}(\alpha(x_3) + \beta(x_3)\dot{X}_2, \sigma(x_3))$$

- Exemple :

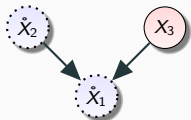


- [Lauritzen, 1992] :

$$P(\dot{X}_1 | \dot{X}_2, X_3 = x_3) = \mathcal{N}(\alpha(x_3) + \beta(x_3)\dot{X}_2, \sigma(x_3))$$

- Ne peut pas exprimer $P(X_2 | \dot{X}_1)$.

- Exemple :



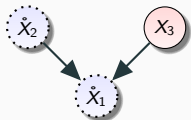
- [Lauritzen, 1992] :

$$P(\dot{X}_1 | \dot{X}_2, X_3 = x_3) = \mathcal{N}(\alpha(x_3) + \beta(x_3)\dot{X}_2, \sigma(x_3))$$

- Ne peut pas exprimer $P(X_2 | \dot{X}_1)$. [Lerner et al., 2001]

Modèles gaussiens conditionnels

- Exemple :



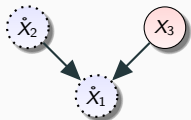
- [Lauritzen, 1992] :

$$P(\dot{X}_1 | \dot{X}_2, X_3 = x_3) = \mathcal{N}(\alpha(x_3) + \beta(x_3)\dot{X}_2, \sigma(x_3))$$

- Ne peut pas exprimer $P(X_2 | \dot{X}_1)$. [Lerner et al., 2001]
- Relation **linéaire** entre les variables.

Modèles gaussiens conditionnels

- Exemple :



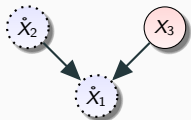
- [Lauritzen, 1992] :

$$P(\dot{X}_1 | \dot{X}_2, X_3 = x_3) = \mathcal{N}(\alpha(x_3) + \beta(x_3)\dot{X}_2, \sigma(x_3))$$

- Ne peut pas exprimer $P(X_2 | \dot{X}_1)$. [Lerner et al., 2001]
- Relation **linéaire** entre les variables.
- **Avantage** : Inférence et apprentissage **rapides**.

Modèles gaussiens conditionnels

- Exemple :



- [Lauritzen, 1992] :

$$P(\dot{X}_1 | \dot{X}_2, X_3 = x_3) = \mathcal{N}(\alpha(x_3) + \beta(x_3)\dot{X}_2, \sigma(x_3))$$

- Ne peut pas exprimer $P(X_2 | \dot{X}_1)$. [Lerner et al., 2001]
- Relation **linéaire** entre les variables.
- **Avantage** : Inférence et apprentissage **rapides**.
- **Inconvénient** : **Hypothèse forte** sur le modèle.

Réseaux bayésiens et variables continues

- Les réseaux bayésiens permettent de raisonner avec l'incertain.
- **Problème !** Ils ne peuvent traiter qu'avec des variables aléatoires discrètes.
- Plusieurs approches proposées pour des variables mixtes :
 - Modèles gaussiens conditionnels (CG) : Peu expressifs mais inférence et apprentissage rapides.

- Les réseaux bayésiens permettent de raisonner avec l'incertain.
- **Problème !** Ils ne peuvent traiter qu'avec des variables aléatoires discrètes.
- Plusieurs approches proposées pour des variables mixtes :
 - Modèles gaussiens conditionnels (CG) : Peu expressifs mais inférence et apprentissage rapides.
 - Modèles à base de mixtures (MTE, MoP, MTBF) :

- **Mixture d'exponentielles (MTE)**, [Moral et al., 2001] :

$$p(\dot{X}_1 | X_2 = x_2) = a_0(x_2) + \sum_{i=1}^k a_i(x_2) \exp(b_i(x_2)\dot{x}_1)$$

- **Mixture d'exponentielles (MTE)**, [Moral et al., 2001] :

$$p(\dot{X}_1 | X_2 = x_2) = a_0(x_2) + \sum_{i=1}^k a_i(x_2) \exp(b_i(x_2)\dot{X}_1)$$

- **Avantage** : modèle **expressif** !

- **Mixture d'exponentielles (MTE)**, [Moral et al., 2001] :

$$p(\dot{X}_1 | X_2 = x_2) = a_0(x_2) + \sum_{i=1}^k a_i(x_2) \exp(b_i(x_2)\dot{x}_1)$$

- **Avantage** : modèle **expressif** !

$$\phi_1(\dot{x}_1) = \begin{cases} 3 + e^{\dot{x}_1} & \text{si } 0 < \dot{x}_1 \leq 2 \\ 2 + e^{-2x_1} & \text{si } 2 < \dot{x}_1 < 4 \end{cases}$$

$$\phi_2(\dot{x}_2) = \begin{cases} 1 + e^{4\dot{x}_2} & \text{si } 0 < \dot{x}_2 \leq 3 \\ 3 + e^{-\dot{x}_2} & \text{si } 3 < \dot{x}_2 < 5 \end{cases}$$

- **Mixture d'exponentielles (MTE)**, [Moral et al., 2001] :

$$p(\dot{X}_1 | X_2 = x_2) = a_0(x_2) + \sum_{i=1}^k a_i(x_2) \exp(b_i(x_2)\dot{x}_1)$$

- **Avantage** : modèle **expressif** !

$$\phi(\dot{x}_1, \dot{x}_2) = \phi_1(\dot{x}_1)\phi_2(\dot{x}_2)$$

- **Mixture d'exponentielles (MTE)**, [Moral et al., 2001] :

$$p(\dot{X}_1 | X_2 = x_2) = a_0(x_2) + \sum_{i=1}^k a_i(x_2) \exp(b_i(x_2)\dot{x}_1)$$

- **Avantage** : modèle **expressif** !

$$\phi(\dot{x}_1, \dot{x}_2) = \begin{cases} 3 + 3e^{4\dot{x}_2} + e^{\dot{x}_1} + e^{\dot{x}_1+4\dot{x}_2} & \text{si } 0 < \dot{x}_1 \leq 2, 0 < \dot{x}_2 \leq 3 \\ 9 + 3e^{-\dot{x}_2} + 3e^{\dot{x}_1} + e^{\dot{x}_1-\dot{x}_2} & \text{si } 0 < \dot{x}_1 \leq 2, 3 < \dot{x}_2 < 5 \\ 2 + 2e^{4\dot{x}_2} + e^{-2\dot{x}_1} + e^{-2\dot{x}_1+4\dot{x}_2} & \text{si } 2 < \dot{x}_1 < 4, 0 < \dot{x}_2 \leq 3 \\ 6 + 2e^{-\dot{x}_2} + 3e^{2\dot{x}_1} + e^{-2\dot{x}_1-\dot{x}_2} & \text{si } 2 < \dot{x}_1 < 4, 3 < \dot{x}_2 < 5 \end{cases}$$

- **Mixture d'exponentielles (MTE)**, [Moral et al., 2001] :

$$p(\dot{X}_1 | X_2 = x_2) = a_0(x_2) + \sum_{i=1}^k a_i(x_2) \exp(b_i(x_2)\dot{x}_1)$$

- **Avantage** : modèle **expressif** !

$$\phi(\dot{x}_1, \dot{x}_2) = \begin{cases} 3 + 3e^{4\dot{x}_2} + e^{\dot{x}_1} + e^{\dot{x}_1+4\dot{x}_2} & \text{si } 0 < \dot{x}_1 \leq 2, 0 < \dot{x}_2 \leq 3 \\ 9 + 3e^{-\dot{x}_2} + 3e^{\dot{x}_1} + e^{\dot{x}_1-\dot{x}_2} & \text{si } 0 < \dot{x}_1 \leq 2, 3 < \dot{x}_2 < 5 \\ 2 + 2e^{4\dot{x}_2} + e^{-2\dot{x}_1} + e^{-2\dot{x}_1+4\dot{x}_2} & \text{si } 2 < \dot{x}_1 < 4, 0 < \dot{x}_2 \leq 3 \\ 6 + 2e^{-\dot{x}_2} + 3e^{2\dot{x}_1} + e^{-2\dot{x}_1-\dot{x}_2} & \text{si } 2 < \dot{x}_1 < 4, 3 < \dot{x}_2 < 5 \end{cases}$$

- **Inconvénient** : inférence trop **longue**...

- **Mixture d'exponentielles (MTE)**, [Moral et al., 2001] :

$$p(\dot{X}_1 | X_2 = x_2) = a_0(x_2) + \sum_{i=1}^k a_i(x_2) \exp(b_i(x_2)\dot{x}_1)$$

- **Avantage** : modèle **expressif** !

$$\phi(\dot{x}_1, \dot{x}_2) = \begin{cases} 3 + 3e^{4\dot{x}_2} + e^{\dot{x}_1} + e^{\dot{x}_1+4\dot{x}_2} & \text{si } 0 < \dot{x}_1 \leq 2, 0 < \dot{x}_2 \leq 3 \\ 9 + 3e^{-\dot{x}_2} + 3e^{\dot{x}_1} + e^{\dot{x}_1-\dot{x}_2} & \text{si } 0 < \dot{x}_1 \leq 2, 3 < \dot{x}_2 < 5 \\ 2 + 2e^{4\dot{x}_2} + e^{-2\dot{x}_1} + e^{-2\dot{x}_1+4\dot{x}_2} & \text{si } 2 < \dot{x}_1 < 4, 0 < \dot{x}_2 \leq 3 \\ 6 + 2e^{-\dot{x}_2} + 3e^{2\dot{x}_1} + e^{-2\dot{x}_1-\dot{x}_2} & \text{si } 2 < \dot{x}_1 < 4, 3 < \dot{x}_2 < 5 \end{cases}$$

- **Inconvénient** : inférence trop **longue**...
- Autres modèles : **Mixture de polynômes (MoP)** [Shenoy, 2011] et **Mixture de fonctions de bases tronquées (MTBF)** [Langseth et al., 2012].

Réseaux bayésiens et variables continues

- Les réseaux bayésiens permettent de raisonner avec l'incertain.
- **Problème !** Ils ne peuvent traiter qu'avec des variables aléatoires discrètes.
- Plusieurs approches proposées pour des variables mixtes :
 - Modèles gaussiens conditionnels (CG) : Peu expressifs mais inférence et apprentissage rapides.
 - Modèles à base de mixtures (MTE, MoP, MTBF) : Très précis mais inférence et apprentissage long.

Réseaux bayésiens et variables continues

- Les réseaux bayésiens permettent de raisonner avec l'incertain.
- **Problème !** Ils ne peuvent traiter qu'avec des variables aléatoires discrètes.
- Plusieurs approches proposées pour des variables mixtes :
 - Modèles gaussiens conditionnels (CG) : Peu expressifs mais inférence et apprentissage rapides.
 - Modèles à base de mixtures (MTE, MoP, MTBF) : Très précis mais inférence et apprentissage long.
 - **Discrétisation** des variables continues :

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

Discrétisation des variables continues

- Exemple : $\hat{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

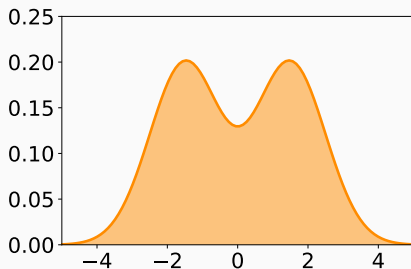


Figure 1 – Densité de probabilité de \hat{X}

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

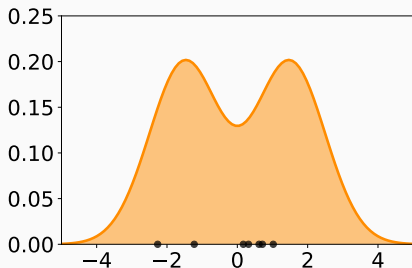


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}
1	0.17
2	1.02
3	0.62
4	-1.23
5	-2.27
6	0.32
7	0.71

Table 1 – Échantillon

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

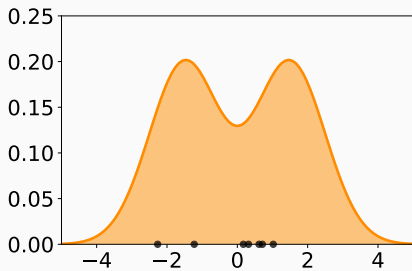


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}
1	0.17
2	1.02
3	0.62
4	-1.23
5	-2.27
6	0.32
7	0.71

Table 1 – Échantillon

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, \dots, g\} \quad \Delta_{\dot{X}}(\dot{x}) = \begin{cases} 0 & \text{si } \dot{x} < t_1, \\ k & \text{si } t_k \leq \dot{x} < t_{k+1}, \\ g & \text{si } \dot{x} \geq t_g \end{cases}$$

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

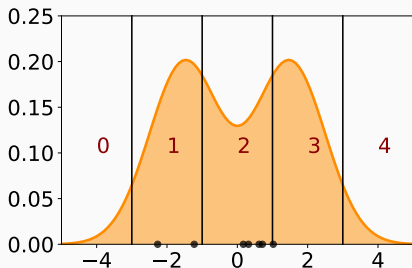


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	3
4	-1.23	2
5	-2.27	2
6	0.32	3
7	0.71	3

Table 1 – Échantillon

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, \dots, 4\} \quad \Delta_{\dot{X}}(\dot{x}) = \begin{cases} 0 & \text{si } \dot{x} < -3, \\ 1 & \text{si } -3 \leq \dot{x} < -1, \\ 2 & \text{si } -1 \leq \dot{x} < 1, \\ 3 & \text{si } 1 \leq \dot{x} < 3, \\ 4 & \text{si } \dot{x} \geq 3 \end{cases}$$

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

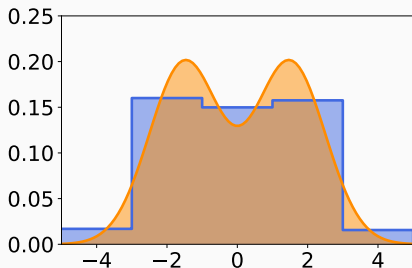


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	3
4	-1.23	2
5	-2.27	2
6	0.32	3
7	0.71	3

Table 1 – Échantillon

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, \dots, 4\} \quad \Delta_{\dot{X}}(\dot{x}) = \begin{cases} 0 & \text{si } \dot{x} < -3, \\ 1 & \text{si } -3 \leq \dot{x} < -1, \\ 2 & \text{si } -1 \leq \dot{x} < 1, \\ 3 & \text{si } 1 \leq \dot{x} < 3, \\ 4 & \text{si } \dot{x} \geq 3 \end{cases}$$

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

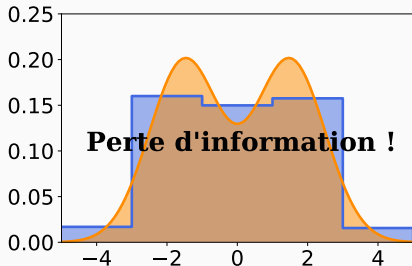


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	3
4	-1.23	2
5	-2.27	2
6	0.32	3
7	0.71	3

Table 1 – Échantillon

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, \dots, 4\} \quad \Delta_{\dot{X}}(\dot{x}) = \begin{cases} 0 & \text{si } \dot{x} < -3, \\ 1 & \text{si } -3 \leq \dot{x} < -1, \\ 2 & \text{si } -1 \leq \dot{x} < 1, \\ 3 & \text{si } 1 \leq \dot{x} < 3, \\ 4 & \text{si } \dot{x} \geq 3 \end{cases}$$

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

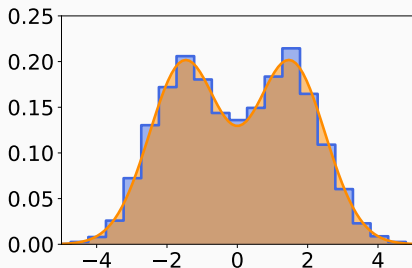


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	3
4	-1.23	2
5	-2.27	2
6	0.32	3
7	0.71	3

Table 1 – Échantillon

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, \dots, 19\}$$

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

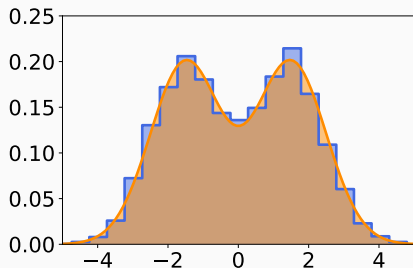


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	3
4	-1.23	2
5	-2.27	2
6	0.32	3
7	0.71	3

Table 1 – Échantillon

- Limités à quelques *cutpoints* (~ 5) pour inférence et apprentissage rapide.

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

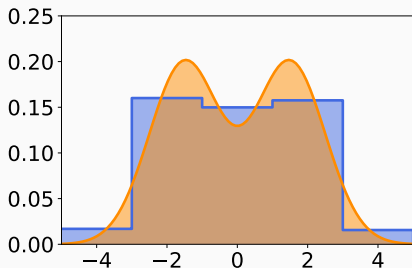


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	3
4	-1.23	2
5	-2.27	2
6	0.32	3
7	0.71	3

Table 1 – Échantillon

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, \dots, 4\} \quad \Delta_{\dot{X}}(\dot{x}) = \begin{cases} 0 & \text{si } \dot{x} < -3, \\ 1 & \text{si } -3 \leq \dot{x} < -1, \\ 2 & \text{si } -1 \leq \dot{x} < 1, \\ 3 & \text{si } 1 \leq \dot{x} < 3, \\ 4 & \text{si } \dot{x} \geq 3 \end{cases}$$

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

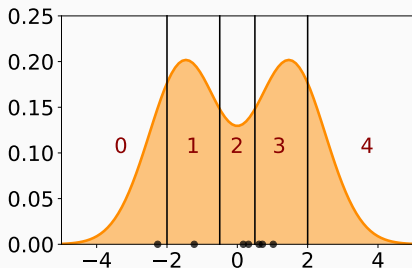


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	4
4	-1.23	2
5	-2.27	1
6	0.32	3
7	0.71	4

Table 1 – Échantillon

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, \dots, 4\} \quad \Delta_{\dot{X}}(\dot{x}) = \begin{cases} 0 & \text{si } \dot{x} < -2, \\ 1 & \text{si } -2 \leq \dot{x} < -1, \\ 2 & \text{si } -1 \leq \dot{x} < 1, \\ 3 & \text{si } 1 \leq \dot{x} < 2, \\ 4 & \text{si } \dot{x} \geq 2 \end{cases}$$

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

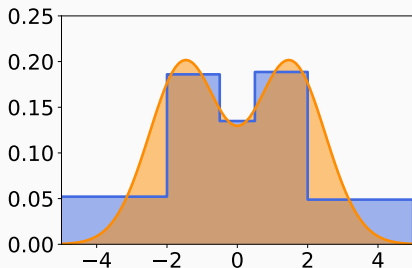


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	4
4	-1.23	2
5	-2.27	1
6	0.32	3
7	0.71	4

Table 1 – Échantillon

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, \dots, 4\} \quad \Delta_{\dot{X}}(\dot{x}) = \begin{cases} 0 & \text{si } \dot{x} < -2, \\ 1 & \text{si } -2 \leq \dot{x} < -1, \\ 2 & \text{si } -1 \leq \dot{x} < 1, \\ 3 & \text{si } 1 \leq \dot{x} < 2, \\ 4 & \text{si } \dot{x} \geq 2 \end{cases}$$

Discrétisation des variables continues

- Exemple : $\dot{X} \sim \frac{1}{2}\mathcal{N}(-3/2, 1) + \frac{1}{2}\mathcal{N}(3/2, 1)$

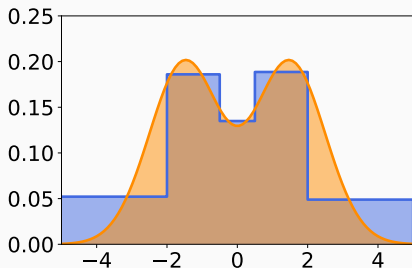


Figure 1 – Densité de probabilité de \dot{X}

n°	\dot{X}	X
1	0.17	3
2	1.02	4
3	0.62	4
4	-1.23	2
5	-2.27	1
6	0.32	3
7	0.71	4

Table 1 – Échantillon

- Limités à quelques *cutpoints* (~ 5) pour inférence et apprentissage rapide.
- Comment placer les *cutpoints* pour minimiser la perte d'information ?

Réseaux bayésiens et variables continues

- Les **réseaux bayésiens** permettent de **raisonner avec l'incertain**.
- **Problème !** Ils ne peuvent traiter qu'avec des **variables aléatoires discrètes**.
- Plusieurs approches proposées pour des variables mixtes :
 - Modèles gaussiens conditionnels (**CG**) : **Peu expressifs** mais **inférence et apprentissage rapides**.
 - Modèles à base de mixtures (**MTE, MoP, MTBF**) : **Très précis** mais **inférence et apprentissage long**.
 - **Discrétisation** des variables continues : **Perte d'information** ou **trop de cutpoints**.

Réseaux bayésiens et variables continues

- Les **réseaux bayésiens** permettent de **raisonner avec l'incertain**.
- **Problème !** Ils ne peuvent traiter qu'avec des **variables aléatoires discrètes**.
- Plusieurs approches proposées pour des variables mixtes :
 - Modèles gaussiens conditionnels (**CG**) : **Peu expressifs** mais **inférence et apprentissage rapides**.
 - Modèles à base de mixtures (**MTE, MoP, MTBF**) : **Très précis** mais **inférence et apprentissage long**.
 - **Discrétisation** des variables continues : **Perte d'information** ou **trop de cutpoints**.
- (Cortijo, Gonzales, 2018) Réseaux bayésiens à densités tronquées (**ctdBN**) : Bon compromis entre expressivité et complexité.

Réseaux bayésiens à densités conditionnelles tronquées (ctdBN)

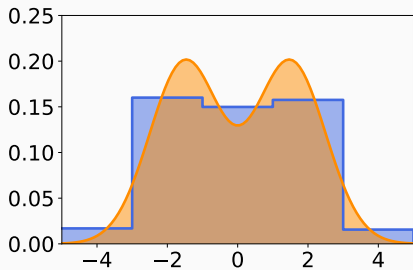


Figure 2 – Densité de probabilité de \hat{X}

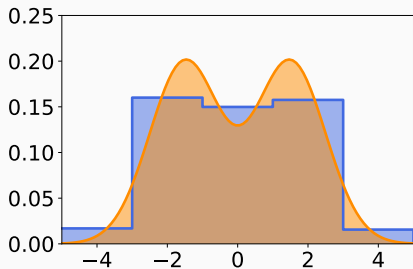


Figure 2 – Densité de probabilité de \hat{X}

- Diminuer la **perte d'information** et le nombre de **cutpoints**.

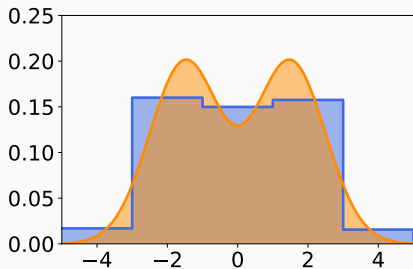


Figure 2 – Densité de probabilité de \hat{X}

- Diminuer la **perte d'information** et le nombre de **cutpoints**.
- **Solution** : **densités conditionnelles tronquées**.

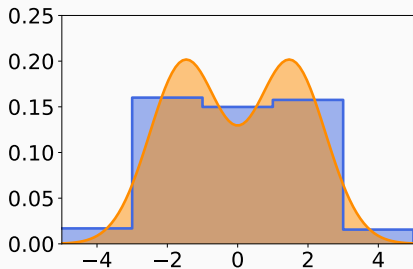


Figure 2 – Densité de probabilité de \dot{X}

- Fonction $f(\dot{X}_i|X_i) : \Omega_{\dot{X}_i} \times \Omega_{X_i} \mapsto \mathbb{R}_0^+$ telle que :

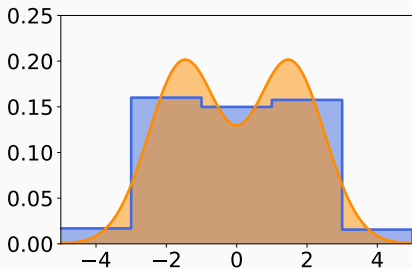


Figure 2 – Densité de probabilité de \hat{X}

- Fonction $f(\hat{X}_i|X_i) : \Omega_{\hat{X}_i} \times \Omega_{X_i} \mapsto \mathbb{R}_0^+$ telle que :
 1. $f(\hat{x}_i|x_i) = 0$ si $\hat{x}_i \notin [t_{x_i}, t_{x_i+1})$

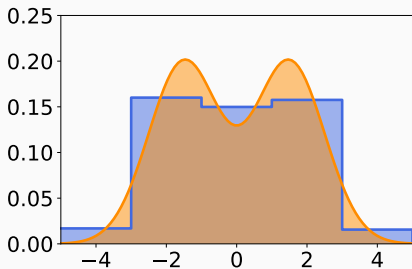


Figure 2 – Densité de probabilité de \dot{X}

- Fonction $f(\dot{X}_i|X_i) : \Omega_{\dot{X}_i} \times \Omega_{X_i} \mapsto \mathbb{R}_0^+$ telle que :
 1. $f(\dot{x}_i|x_i) = 0$ si $\dot{x}_i \notin [t_{x_i}, t_{x_i+1})$
 2. L'équation suivante est satisfaite :

$$\int_{t_{x_i}}^{t_{x_i+1}} f(\dot{x}_i|x_i) d\dot{x}_i = 1,$$

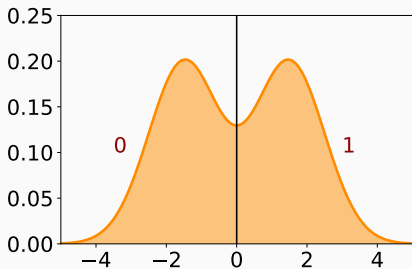


Figure 3 – Densité de probabilité de \check{X}

$$\Delta_{\check{X}} : \Omega_{\check{X}} \rightarrow \{0, 1\} \quad \Delta_{\check{X}}(\check{x}) = \begin{cases} 0 & \text{si } \check{x} < 0, \\ 1 & \text{si } \check{x} \geq 0 \end{cases}$$

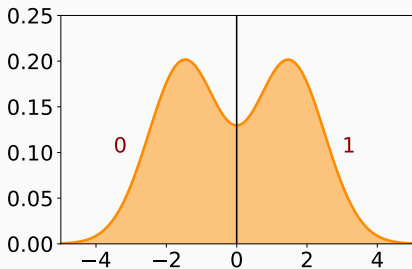


Figure 3 – Densité de probabilité de \check{X}

$$\Delta_{\check{X}} : \Omega_{\check{X}} \rightarrow \{0, 1\} \quad \Delta_{\check{X}}(\check{x}) = \begin{cases} 0 & \text{si } \check{x} < 0, \\ 1 & \text{si } \check{x} \geq 0 \end{cases}$$

$$f(\check{x} | X = i) = \mathcal{N}(\mu_i, \sigma_i)$$

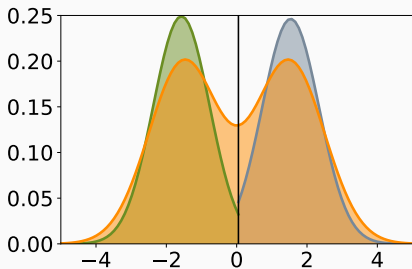


Figure 4 – Densité de probabilité de \dot{X}

$$\Delta_{\dot{X}} : \Omega_{\dot{X}} \rightarrow \{0, 1\} \quad \Delta_{\dot{X}}(\dot{x}) = \begin{cases} 0 & \text{si } \dot{x} < 0, \\ 1 & \text{si } \dot{x} \geq 0 \end{cases}$$

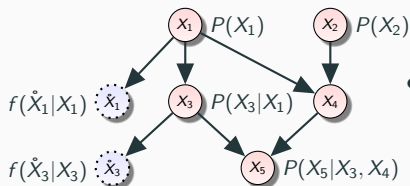
$$f(\dot{x}|X = i) = \mathcal{N}(\mu_i, \sigma_i)$$

BN à densités conditionnelles tronquées (ctdBN)

- Représentation d'une fonction de probabilité jointe h sur \mathbf{X}_D et $\dot{\mathbf{X}}_C$ au travers :
 1. D'un réseaux bayésiens sur \mathbf{X}_D
 2. D'une fonction de **densité conditionnelle tronquée** $f(\dot{X}_i|X_i)$ pour chaque $\dot{X}_i \in \dot{\mathbf{X}}_C$

tels que :

$$h(\mathbf{X}_D, \dot{\mathbf{X}}_C) = \underbrace{\prod_{X \in \mathbf{X}_D} P(X|\text{Pa}(X))}_{\text{Partie discrète}} \underbrace{\prod_{\dot{X} \in \dot{\mathbf{X}}_C} f(\dot{X}|X)}_{\text{Partie continue}}$$



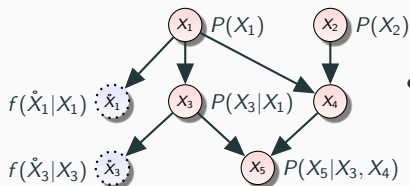
- Remarque : \dot{X}_i n'a pas d'enfant et $\text{Pa}(\dot{X}_i) = \{X_i\} \rightarrow$ inférence rapide.

BN à densités conditionnelles tronquées (ctdBN)

- Représentation d'une fonction de probabilité jointe h sur \mathbf{X}_D et $\dot{\mathbf{X}}_C$ au travers :
 1. D'un réseaux bayésiens sur \mathbf{X}_D
 2. D'une fonction de **densité conditionnelle tronquée** $f(\dot{X}_i|X_i)$ pour chaque $\dot{X}_i \in \dot{\mathbf{X}}_C$

tels que :

$$h(\mathbf{X}_D, \dot{\mathbf{X}}_C) = \underbrace{\prod_{X \in \mathbf{X}_D} P(X|\text{Pa}(X))}_{\text{Partie discrète}} \underbrace{\prod_{\dot{X} \in \dot{\mathbf{X}}_C} f(\dot{X}|X)}_{\text{Partie continue}}$$



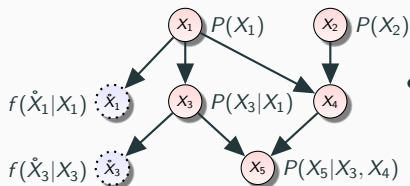
- Remarque : \dot{X}_i n'a pas d'enfant et $\text{Pa}(\dot{X}_i) = \{X_i\} \rightarrow$ inférence rapide.

BN à densités conditionnelles tronquées (ctdBN)

- Représentation d'une fonction de probabilité jointe h sur \mathbf{X}_D et $\dot{\mathbf{X}}_C$ au travers :
 1. D'un réseaux bayésiens sur \mathbf{X}_D
 2. D'une fonction de **densité conditionnelle tronquée** $f(\dot{X}_i|X_i)$ pour chaque $\dot{X}_i \in \dot{\mathbf{X}}_C$

tels que :

$$h(\mathbf{X}_D, \dot{\mathbf{X}}_C) = \underbrace{\prod_{X \in \mathbf{X}_D} P(X|\text{Pa}(X))}_{\text{Partie discrète}} \underbrace{\prod_{\dot{X} \in \dot{\mathbf{X}}_C} f(\dot{X}|X)}_{\text{Partie continue}}$$



- Remarque : \dot{X}_i n'a pas d'enfant et $\text{Pa}(\dot{X}_i) = \{X_i\} \rightarrow$ inférence rapide.

Apprentissage des ctdBNs

- Apprentissage **conjoint** de la **structure** et de la **discrétisation** :

$$(\mathcal{G}^*, \Delta^*) = \text{Argmax}_{\mathcal{G}, \Delta} p(\mathcal{G}, \Delta | \mathring{\mathcal{D}}) \stackrel{\text{Bayes}}{=} \text{Argmax}_{\mathcal{G}, \Delta} p(\mathring{\mathcal{D}} | \mathcal{G}, \Delta) p(\mathcal{G}, \Delta)$$

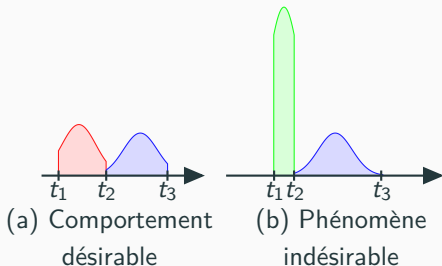
- $p(\mathring{\mathcal{D}} | \mathcal{G}, \Delta)$ fait intervenir $f(\mathring{x}_i^{(j)} | x_i^{(j)}, \theta, \Delta)$

Apprentissage des ctdBNs

- Apprentissage **conjoint** de la **structure** et de la **discrétisation** :

$$(\mathcal{G}^*, \Delta^*) = \text{Argmax}_{\mathcal{G}, \Delta} p(\mathcal{G}, \Delta | \mathcal{D}) \stackrel{\text{Bayes}}{=} \text{Argmax}_{\mathcal{G}, \Delta} p(\mathcal{D} | \mathcal{G}, \Delta) p(\mathcal{G}, \Delta)$$

- $p(\mathcal{D} | \mathcal{G}, \Delta)$ fait intervenir $f(\hat{x}_i^{(j)} | x_i^{(j)}, \theta, \Delta)$
- Si intervalle $[t_k, t_{k+1}[$ rendu **arbitrairement petit** autour de $\hat{x}_i^{(j)}$,
 $f(\hat{x}_i^{(j)} | x_i^{(j)}, \theta, \Delta)$ **arbitrairement grand** $\Rightarrow p(\mathcal{D} | \mathcal{G}, \Delta) \rightarrow \infty$

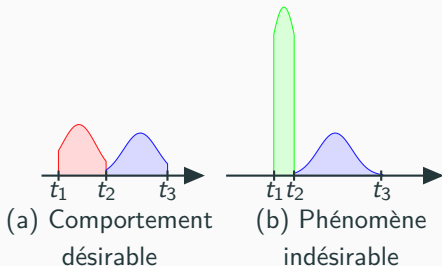


Apprentissage des ctdBNs

- Apprentissage **conjoint** de la **structure** et de la **discrétisation** :

$$(\mathcal{G}^*, \Delta^*) = \text{Argmax}_{\mathcal{G}, \Delta} p(\mathcal{G}, \Delta | \mathcal{D}) \stackrel{\text{Bayes}}{=} \text{Argmax}_{\mathcal{G}, \Delta} p(\mathcal{D} | \mathcal{G}, \Delta) p(\mathcal{G}, \Delta)$$

- $p(\mathcal{D} | \mathcal{G}, \Delta)$ fait intervenir $f(\hat{x}_i^{(j)} | x_i^{(j)}, \theta, \Delta)$
- Si intervalle $[t_k, t_{k+1}[$ rendu **arbitrairement petit** autour de $\hat{x}_i^{(j)}$, $f(\hat{x}_i^{(j)} | x_i^{(j)}, \theta, \Delta)$ **arbitrairement grand** $\Rightarrow p(\mathcal{D} | \mathcal{G}, \Delta) \rightarrow \infty$



- Discontinuités** dans la densité \rightarrow propriété **non désirable**.

Réseaux bayésiens à densités conditionnelles (cdBN)

- Représentation d'une fonction de probabilité jointe h sur \mathbf{X}_D et $\dot{\mathbf{X}}_C$:

$$h(\mathbf{X}_D, \dot{\mathbf{X}}_C) = \prod_{X \in \mathbf{X}_D} P(X | \mathbf{Pa}(X)) \prod_{\dot{X} \in \dot{\mathbf{X}}_C} f(\dot{X} | X)$$

- Représentation d'une fonction de probabilité jointe h sur \mathbf{X}_D et \mathbf{X}_C :

$$h(\mathbf{X}_D, \mathbf{X}_C) = \prod_{X \in \mathbf{X}_D} P(X | \text{Pa}(X)) \prod_{\dot{X} \in \mathbf{X}_C} f(\dot{X} | X)$$

- $f(\dot{X}_i | X_i)$ est maintenant une densité conditionnelle (**non-tronquée**) :

$$f(\dot{X}_i | X_i = x_i) : \Omega_{\dot{X}_i} \mapsto \mathbb{R}_0^+ \quad \text{tel que} \quad \int_{\dot{X}_i \in \Omega_{\dot{X}_i}} f(\dot{X}_i | x_i) d\dot{X}_i = 1$$

- Représentation d'une fonction de probabilité jointe h sur \mathbf{X}_D et $\dot{\mathbf{X}}_C$:

$$h(\mathbf{X}_D, \dot{\mathbf{X}}_C) = \prod_{X \in \mathbf{X}_D} P(X | \text{Pa}(X)) \prod_{\dot{X} \in \dot{\mathbf{X}}_C} f(\dot{X} | X)$$

- $f(\dot{X}_i | X_i)$ est maintenant une densité conditionnelle (**non-tronquée**) :

$$f(\dot{X}_i | X_i = x_i) : \Omega_{\dot{X}_i} \mapsto \mathbb{R}_0^+ \quad \text{tel que} \quad \int_{\dot{X}_i \in \Omega_{\dot{X}_i}} f(\dot{X}_i | x_i) d\dot{X}_i = 1$$

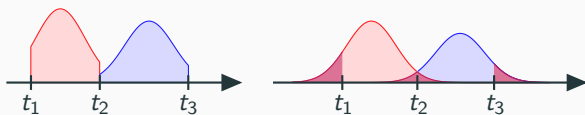


Figure 5 – Densités pour les ctdBNs v.s. densités pour les cdBNs.

- Représentation d'une fonction de probabilité jointe h sur \mathbf{X}_D et $\dot{\mathbf{X}}_C$:

$$h(\mathbf{X}_D, \dot{\mathbf{X}}_C) = \prod_{X \in \mathbf{X}_D} P(X | \text{Pa}(X)) \prod_{\dot{X} \in \dot{\mathbf{X}}_C} f(\dot{X} | X)$$

- $f(\dot{X}_i | X_i)$ est maintenant une densité conditionnelle (**non-tronquée**) :

$$f(\dot{X}_i | X_i = x_i) : \Omega_{\dot{X}_i} \mapsto \mathbb{R}_0^+ \quad \text{tel que} \quad \int_{\dot{X}_i \in \Omega_{\dot{X}_i}} f(\dot{X}_i | x_i) d\dot{X}_i = 1$$

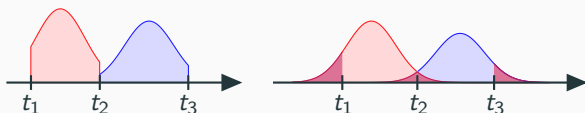


Figure 5 – Densités pour les ctdBNs v.s. densités pour les cdBNs.

- Il n'y a **plus de discrétisation** : les variables $\mathbf{X}_C = \{X_{d+1}, \dots, X_n\}$ sont maintenant des **variables cachées**.

- Représentation d'une fonction de probabilité jointe h sur \mathbf{X}_D et $\dot{\mathbf{X}}_C$:

$$h(\mathbf{X}_D, \dot{\mathbf{X}}_C) = \prod_{X \in \mathbf{X}_D} P(X | \text{Pa}(X)) \prod_{\dot{X} \in \dot{\mathbf{X}}_C} f(\dot{X} | X)$$

- $f(\dot{X}_i | X_i)$ est maintenant une densité conditionnelle (**non-tronquée**) :

$$f(\dot{X}_i | X_i = x_i) : \Omega_{\dot{X}_i} \mapsto \mathbb{R}_0^+ \quad \text{tel que} \quad \int_{\dot{X}_i \in \Omega_{\dot{X}_i}} f(\dot{X}_i | x_i) d\dot{X}_i = 1$$

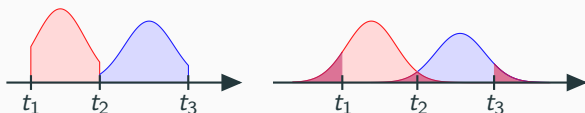


Figure 5 – Densités pour les ctdBNs v.s. densités pour les cdBNs.

- Il n'y a **plus de discrétisation** : les variables $\mathbf{X}_C = \{X_{d+1}, \dots, X_n\}$ sont maintenant des **variables cachées**.
- Apprentissage d'un modèle avec des variables cachées : **EM structurel (SEM)** [Friedman, 1998].

Apprentissage des cdBNs : l'algorithme

Input: Une base de données \mathring{D}

Output: Un cdBN $\mathcal{B} = (\mathcal{G}, \Theta)$

Apprentissage des cdBNs : l'algorithme

Input: Une base de données \mathring{D}

Output: Un cdBN $\mathcal{B} = (\mathcal{G}, \Theta)$

1 $t \leftarrow 0$

2 $\mathcal{G}_0 \leftarrow$ une structure initiale de cdBN

3 $\Theta_0 \leftarrow$ Paramètres MAP de \mathcal{G}_0 étant donné \mathring{D}

Apprentissage des cdBNs : l'algorithme

Input: Une base de données $\mathring{\mathcal{D}}$

Output: Un cdBN $\mathcal{B} = (\mathcal{G}, \Theta)$

```
1  $t \leftarrow 0$ 
2  $\mathcal{G}_0 \leftarrow$  une structure initiale de cdBN
3  $\Theta_0 \leftarrow$  Paramètres MAP de  $\mathcal{G}_0$  étant donné  $\mathring{\mathcal{D}}$ 
4 repeat
    // E step
5   foreach  $\mathcal{G}$  dans le voisinage  $\mathbf{N}(\mathcal{G}_t)$  de  $\mathcal{G}_t$  do
6      $\lfloor$  Affecter score  $Sc(\mathcal{G}) = \log p(\mathcal{G}|\mathring{\mathcal{D}})$  à  $\mathcal{G}$ 
    // M step
7    $\mathcal{G}_{t+1} \leftarrow \text{Argmax}_{\mathcal{G} \in \mathbf{N}(\mathcal{G}_t)} Sc(\mathcal{G})$ 
8    $\Theta_{t+1} \leftarrow$  Paramètres MAP de  $\mathcal{G}_{t+1}$  étant donné  $\mathring{\mathcal{D}}$ 
9    $t \leftarrow t + 1$ 
10 until  $\mathcal{G}_{t+1} = \mathcal{G}_t$ ;
```

Apprentissage des cdBNs : l'algorithme

Input: Une base de données $\mathring{\mathcal{D}}$

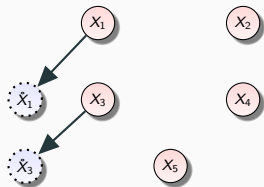
Output: Un cdBN $\mathcal{B} = (\mathcal{G}, \Theta)$

```
1  $t \leftarrow 0$ 
2  $\mathcal{G}_0 \leftarrow$  une structure initiale de cdBN
3  $\Theta_0 \leftarrow$  Paramètres MAP de  $\mathcal{G}_0$  étant donné  $\mathring{\mathcal{D}}$ 
4 repeat
    // E step
5   foreach  $\mathcal{G}$  dans le voisinage  $\mathbf{N}(\mathcal{G}_t)$  de  $\mathcal{G}_t$  do
6      $\lfloor$  Affecter score  $Sc(\mathcal{G}) = \log p(\mathcal{G}|\mathring{\mathcal{D}})$  à  $\mathcal{G}$ 
    // M step
7    $\mathcal{G}_{t+1} \leftarrow \text{Argmax}_{\mathcal{G} \in \mathbf{N}(\mathcal{G}_t)} Sc(\mathcal{G})$ 
8    $\Theta_{t+1} \leftarrow$  Paramètres MAP de  $\mathcal{G}_{t+1}$  étant donné  $\mathring{\mathcal{D}}$ 
9    $t \leftarrow t + 1$ 
10 until  $\mathcal{G}_{t+1} = \mathcal{G}_t$ ;
11 return cdBN  $\mathcal{B} = (\mathcal{G}_t, \Theta_t)$ 
```

Apprentissage des cdBNs : quelques détails

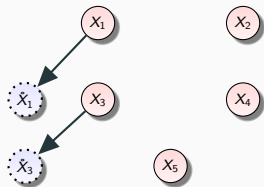
Apprentissage des cdBNs : quelques détails

- Structure initiale : $\mathcal{G}_0 = (\mathbf{X}, \mathcal{A}_0)$ où $\mathcal{A}_0 = \{(X_i, \dot{X}_i) : i = d + 1, \dots, n\}$



Apprentissage des cdBNs : quelques détails

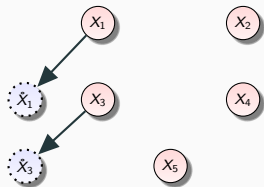
- Structure initiale : $\mathcal{G}_0 = (\mathbf{X}, \mathcal{A}_0)$ où
 $\mathcal{A}_0 = \{(X_i, \hat{X}_i) : i = d + 1, \dots, n\}$



- $Sc(\mathcal{G}) = \log p(\mathcal{G}|\hat{\mathcal{D}}) \stackrel{\text{Bayes}}{=} \log p(\hat{\mathcal{D}}|\mathcal{G}) + \log(p(\mathcal{G})/p(\mathcal{D}))$

Apprentissage des cdBNs : quelques détails

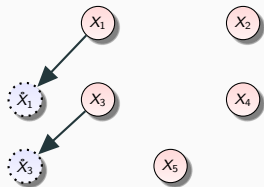
- Structure initiale : $\mathcal{G}_0 = (\mathbf{X}, \mathcal{A}_0)$ où $\mathcal{A}_0 = \{(X_i, \dot{X}_i) : i = d + 1, \dots, n\}$



- $Sc(\mathcal{G}) = \log p(\mathcal{G}|\hat{\mathcal{D}}) \stackrel{\text{Bayes}}{=} \log p(\hat{\mathcal{D}}|\mathcal{G}) + \log(p(\mathcal{G})/p(\mathcal{D}))$
- $Sc(\mathcal{G}) \approx \sum_{\mathbf{x}_C^{(\mathcal{D})}} P(\mathbf{x}_C^{(\mathcal{D})} | \mathbf{x}_D^{(\mathcal{D})}, \dot{\mathbf{x}}_C^{(\mathcal{D})}, \mathcal{G}_t, \Theta_t) \times \left[\sum_{i=1}^n Sc(X_i | \mathbf{Pa}(X_i)) + \sum_{i=d+1}^n Sc(\dot{X}_i | X_i) \right],$

Apprentissage des cdBNs : quelques détails

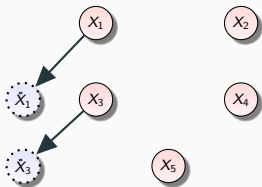
- Structure initiale : $\mathcal{G}_0 = (\mathbf{X}, \mathcal{A}_0)$ où $\mathcal{A}_0 = \{(X_i, \dot{X}_i) : i = d + 1, \dots, n\}$



- $Sc(\mathcal{G}) = \log p(\mathcal{G}|\hat{\mathcal{D}}) \stackrel{\text{Bayes}}{=} \log p(\hat{\mathcal{D}}|\mathcal{G}) + \log(p(\mathcal{G})/p(\mathcal{D}))$
- $Sc(\mathcal{G}) \approx \sum_{\mathbf{x}_c^{(\mathcal{D})}} P(\mathbf{x}_c^{(\mathcal{D})} | \mathbf{x}_D^{(\mathcal{D})}, \dot{\mathbf{x}}_c^{(\mathcal{D})}, \mathcal{G}_t, \Theta_t) \times \left[\sum_{i=1}^n Sc(X_i | \mathbf{Pa}(X_i)) + \sum_{i=d+1}^n Sc(\dot{X}_i | X_i) \right],$
- $Sc(X_i | \mathbf{Pa}(X_i)) \rightarrow$ calcul classique.

Apprentissage des cdBNs : quelques détails

- Structure initiale : $\mathcal{G}_0 = (\mathbf{X}, \mathcal{A}_0)$ où $\mathcal{A}_0 = \{(X_i, \dot{X}_i) : i = d + 1, \dots, n\}$



- $Sc(\mathcal{G}) = \log p(\mathcal{G}|\hat{\mathcal{D}}) \stackrel{\text{Bayes}}{=} \log p(\hat{\mathcal{D}}|\mathcal{G}) + \log(p(\mathcal{G})/p(\mathcal{D}))$
- $Sc(\mathcal{G}) \approx \sum_{\mathbf{x}_C^{(\mathcal{D})}} P(\mathbf{x}_C^{(\mathcal{D})} | \mathbf{x}_D^{(\mathcal{D})}, \dot{\mathbf{x}}_C^{(\mathcal{D})}, \mathcal{G}_t, \Theta_t) \times \left[\sum_{i=1}^n Sc(X_i | \mathbf{Pa}(X_i)) + \sum_{i=d+1}^n Sc(\dot{X}_i | X_i) \right],$
- $Sc(X_i | \mathbf{Pa}(X_i)) \rightarrow$ calcul classique.
- $Sc(\dot{X}_i | X_i)$ identique pour n'importe quelle structure $\mathcal{G} \rightarrow$ peut être ignoré.

Conclusion

- **Difficultés** pour étendre les réseaux bayésiens à des variables à la fois **discrètes et continues** : modèle soit pas assez expressif mais rapide en inférence et apprentissage soit modèle très expressif mais inférence et apprentissage long.

- **Difficultés** pour étendre les réseaux bayésiens à des variables à la fois **discrètes et continues** : modèle soit pas assez expressif mais rapide en inférence et apprentissage soit modèle très expressif mais inférence et apprentissage long.
- Compromis entre les deux : réseaux bayésiens à densités tronquées (**ctdBN**). Mais problème au niveau de l'**apprentissage conjoint de la structure et de la discrétisation**.

- **Difficultés** pour étendre les réseaux bayésiens à des variables à la fois **discrètes et continues** : modèle soit pas assez expressif mais rapide en inférence et apprentissage soit modèle très expressif mais inférence et apprentissage long.
- Compromis entre les deux : réseaux bayésiens à densités tronquées (**ctdBN**). Mais problème au niveau de l'**apprentissage conjoint de la structure et de la discrétisation**.
- Solution : éviter la discrétisation avec les **réseaux bayésiens à densité conditionnelles (cdBN)**.

- **Difficultés** pour étendre les réseaux bayésiens à des variables à la fois **discrètes et continues** : modèle soit pas assez expressif mais rapide en inférence et apprentissage soit modèle très expressif mais inférence et apprentissage long.
- Compromis entre les deux : réseaux bayésiens à densités tronquées (**ctdBN**). Mais problème au niveau de l'**apprentissage conjoint de la structure et de la discrétisation**.
- Solution : éviter la discrétisation avec les **réseaux bayésiens à densité conditionnelles (cdBN)**.
- Proposition d'un **algorithme d'apprentissage** pour les cdBNs.

- Tester l'algorithme sur des cas concrets (**aGrUM**).

- Tester l'algorithme sur des cas concrets (**aGrUM**).
- Beaucoup d'inférence avec SEM, comment faire ?

- Tester l'algorithme sur des cas concrets (**aGrUM**).
- Beaucoup d'inférence avec SEM, comment faire ?
- Quel est l'influence de la taille du domaine des variables cachées ?

- Tester l'algorithme sur des cas concrets (**aGrUM**).
- Beaucoup d'inférence avec SEM, comment faire ?
- Quel est l'influence de la taille du domaine des variables cachées ?
- Score pour modèle gaussien, peut-on en dériver d'autres ?

- Tester l'algorithme sur des cas concrets (**aGrUM**).
- Beaucoup d'inférence avec SEM, comment faire ?
- Quel est l'influence de la taille du domaine des variables cachées ?
- Score pour modèle gaussien, peut-on en dériver d'autres ?
- Comparaison avec d'autres modèles.

Merci de votre attention !



Friedman, N. (1998).

The Bayesian structural EM algorithm.

In *Proc. of Uncertainty in Artificial Intelligence*, pages 129–138.



Langseth, H., Nielsen, T., Rumí, R., and Salmerón, A. (2012).

Mixtures of truncated basis functions.

International Journal of Approximate Reasoning, 53(2) :212–227.



Lauritzen, S. (1992).

Propagation of probabilities, means and variances in mixed graphical association models.

Journal of the American Statistical Association, 87 :1098–1108.



Lerner, U., Segal, E., and Koller, D. (2001).

Exact inference in networks with discrete children of continuous parents.

In *Proceedings of UAI'01*, pages 319–328.



Moral, S., Rumí, R., and Salmerón, A. (2001).

Mixtures of truncated exponentials in hybrid Bayesian networks.

In *Proceedings of ECSQARU'01*, volume 2143 of *LNAI*, pages 156–167.



Shenoy, P. (2011).

A re-definition of mixtures of polynomials for inference in hybrid Bayesian networks.

In *Proceedings of ECSQARU'11*, volume 6717 of *LNCS*, pages 98–109.