

Utilisation des réseaux bayésiens et de l'approche de Fenton pour l'estimation de probabilité d'occurrence d'événements

Quynh Anh Do Hoang, Jérémie Guiochet, Mohamed Kaâniche, David Powell

▶ **To cite this version:**

Quynh Anh Do Hoang, Jérémie Guiochet, Mohamed Kaâniche, David Powell. Utilisation des réseaux bayésiens et de l'approche de Fenton pour l'estimation de probabilité d'occurrence d'événements. 2013. hal-00823620

HAL Id: hal-00823620
https://hal.science/hal-00823620

Submitted on 17 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Utilisation des réseaux bayésiens et de l'approche de Fenton pour l'estimation de probabilité d'occurrence d'événements

Rapport LAAS-CNRS
Numéro N° 13077

Quynh Anh DO HOANG, Jérémie GUIOCHET,
Mohamed KAÂNICHE et David POWELL

CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France

Univ de Toulouse, INP, UPS, LAAS, F-31400 Toulouse, France

Mai 2013

Résumé

Les réseaux bayésiens sont utilisés dans plusieurs domaines pour estimer la probabilité d'un événement non observable à partir des événements observables. Pour chacun des nœuds du réseau bayésien, une table de probabilité est souvent définie manuellement. Pour une table plus complexe, N. E. Fenton propose d'utiliser des expressions logiques ou des lois de probabilités. À travers quatre exemples de réseaux bayésiens, nous effectuons une étude approfondie des calculs effectués en les vérifiants grâce à l'outil AgenaRisk.

Mots-clés : réseaux bayésiens, estimation de probabilité, loi Normale Tronquée, AgenaRisk

Table des matières

1 Introduction **5**

2 Étude d'un modèle simple à trois états **5**
2.1 Cas 1 : nœuds booléens, TdP du nœud fils renseignée manuellement 6
2.2 Cas 2 : nœuds valués (3 valeurs) et TdP du nœud fils renseignée manuellement 7
2.3 Cas 3 : nœuds valués (3 valeurs) et TdP du nœud fils renseignée par expression 9
2.4 Cas 4 : nœuds valués (5 valeurs) et TdP du nœud fils renseignée par loi Normale Tronquée . 11

3 Exploration des différentes techniques proposées par AgenaRisk pour renseigner les Tables de Probabilité en utilisant la loi Normale Tronquée **15**
3.1 Espérance μ_{Mean} 16
3.2 Espérance μ_{Min} avec le nombre de nœuds n 16
3.3 Espérance μ_{Max} avec le nombre de nœuds n 16
3.4 Espérance $\mu_{MixMinMax}$ 16

4 Technique d'échantillonnage pour calculer la distribution connaissant l'état des nœuds parents **17**

5 Utilisation d'une Table de Probabilité partitionnée **18**

6 Conclusion **20**

Table des figures

1 Modèle étudié : 2 nœuds parents (A,B) et 1 nœud fils (C) 5
2 Cas 1 : Probabilité totale de C 6
3 Cas 1 : Probabilité de C sachant A 7
4 Cas 1 : Probabilité de A sachant C 7
5 Cas 2 : Probabilité totale de C 8
6 Cas 2 : Probabilité de C sachant A 9
7 Cas 2 : Probabilité de A sachant C 9
8 Cas 3 : Probabilité totale de C 11
9 Cas 3 : Probabilité de C sachant A 11
10 Cas 3 : Probabilité de A sachant C 11
11 Cas 4 : Probabilité totale de C 14
12 Cas 4 : Probabilité de C sachant A 15
13 Cas 4 : Probabilité de A sachant C 15
14 Les 4 formules de calcul de moyenne 17
15 9 distributions de probabilité du nœud fils C pour A_{high} et B_{low} 18
16 La moyenne des distributions est celle du nœud fils C pour A_{high} et B_{low} 18
17 Résultats du cas 4 retrouvés par calculs théoriques 18
18 Le cas quasi-déterministe avec différentes approches 20
19 La fonction de densité de la loi normale 21
20 Table des valeurs de la fonction de répartition de la loi Normale centrée réduite. 23

Liste des tableaux

1 Cas 1 : Tables de Probabilité des nœuds renseignées manuellement 6
2 Cas 2 : Tables de Probabilité des nœuds renseignées manuellement 8
3 Cas 3 : Tables de Probabilité des nœuds A et B renseignées manuellement 10
4 Cas 3 : Table de Probabilité du nœud C obtenue à partir de l'expression 10
5 Cas 4 : Tables de Probabilité des nœuds A et B renseignées manuellement 12
6 Valeurs correspondantes pour les 5 intervalles 12
7 Résultats fournis par l'outil AgenaRisk à partir des paramètres d'entrée 13
8 Différentes valeurs entre calcul théorique et l'outil AgenaRisk 14
9 Cas déterministe : Table de Probabilité du nœud C 19
10 Différentes manières pour remplir la TdP du nœud C 19

1 Introduction

Les réseaux bayésiens décrivent les relations cause à effets de façon probabiliste, permettant ainsi les calculs de probabilité conditionnelle à partir des tables de probabilité (TdP) de chacun des nœuds du réseau. Les valeurs d'une TdP sont souvent fixées manuellement. Dans ce rapport nous nous intéressons à la détermination d'une TdP d'une manière systématique. Plus particulièrement, nous étudions la proposition de Fenton [1] qui répartit les probabilités suivant plusieurs lois (dont la loi Normale Tronquée), et qui propose des outils d'estimation de la probabilité globale. Fenton propose également un outil logiciel AgenaRisk [2] pour mettre en œuvre cette approche. Cet outil sera utilisé ici (version 5.6.0) pour valider les calculs théoriques, ou comparer plusieurs méthodes de calcul. Dans un premier temps, nous observons les résultats données par l'outil à travers 4 cas d'étude classés du plus simple au plus complexe. Puis nous présentons les techniques que l'outil met en place pour remplir une TdP suivant une loi Normale Tronquée : le calcul de l'espérance et la répartition des valeurs probabilistes. Nous abordons ensuite l'application de plusieurs lois de probabilité à une même table avant de conclure ce rapport.

2 Étude d'un modèle simple à trois états

Dans les exemples qui suivent, nous considérons le modèle monté sur la Figure 1, où l'événement C dépend des événements A et B. Les TdP des nœuds parents A et B seront toujours rentrées manuellement. Celle du nœud fils C est remplie avec différentes méthodes. Les deux nœuds parents A et B influencent le nœud fils indépendamment, et sont indépendants.

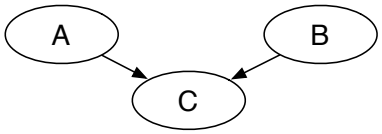


FIGURE 1 – Modèle étudié : 2 nœuds parents (A,B) et 1 nœud fils (C)

La méthode d'analyse suit généralement les étapes suivantes :

1. Dessin du graphe et choix du type des états des nœuds parmi :
   - Boolean - par ex. "Vrai", "Faux"
   - Labelled - par ex. "Rouge", "Jaune, "Vert"
   - Ranked (valué) - par ex. "Low", "Medium", "High"
   - Integer Interval - par ex. 0, 1, [2, 3], [4], [5 – infinity]
   - Continuous Interval - par ex. [0, 10], [10 – 20], [20 – infinity]
   - Discrete Numeric - par ex. un ensemble de valeur non ordonné -2, 0, 2.5, 3.6, 10
2. Saisie de la table de probabilité des états pour chacun des nœuds :
   - manuellement
   - par calcul arithmétique ou avec une expression
   - par calcul avec une loi de distribution
3. Calcul de la probabilité totale pour un nœud d'être dans son état, sans connaissance de l'état des autres nœuds (ex :$\mathbb{P}(C = vrai)$)
   - Application du calcul de la probabilité totale

   Énoncé : Si $((B_i)_{i\in I}$ est un système exhaustif (fini ou dénombrable) d'événements, et si quel que soit $i \in I, \mathbb{P}(B_i \neq 0)$ alors, pour tout événement A :

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

4. Calcul avec observation sur un nœud parent, par exemple $\mathbb{P}(C = vrai|A = vrai)$
   – Mise à jour de TdP suite à l'observation, application du calcul de la probabilité totale
5. Calcul avec observation sur un nœud enfant, par exemple $\mathbb{P}(A = vrai|C = vrai)$
   – Étant donné 2 événements A et B, le théorème de Bayes permet de déterminer la probabilité de *A sachant B*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\overline{A})\mathbb{P}(\overline{A})}$$

où $\overline{A}$ est le complémentaire de A.

2.1 Cas 1 : nœuds booléens, TdP du nœud fils renseignée manuellement

Connaissance des TdP de TOUS les nœuds (ex. : avec 2 valeurs : vrai, faux) :

1. Dessiner le graphe, choix du nœud de type booléen
2. Renseigner les tables manuellement.
   Dans cet exemple, C ne peut être VRAI si et seulement si au moins l'un des parents est VRAI.

nœud A	
Vrai	Faux
0,8	0,2

nœud B	
Vrai	Faux
0,7	0,3

nœud C				
	A Vrai		A Faux	
	B Vrai	B Faux	B Vrai	B Faux
C Vrai	1	1	1	0
C Faux	0	0	0	1

TABLE 1 – Cas 1 : Tables de Probabilité des nœuds renseignées manuellement

3. Calcul de $\mathbb{P}(C = Vrai)$ sans aucune observation
   Par souci de clarté, on notera $\mathbb{P}(A)$ pour $\mathbb{P}(A = Vrai)$ et $\mathbb{P}(\overline{A})$ pour $\mathbb{P}(A = Faux)$. On rappelle que A et B sont indépendants, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
   En appliquant la formule de probabilité totale sur C :

$$\mathbb{P}(C) = \mathbb{P}(C|A\cap B)\mathbb{P}(A)\mathbb{P}(B)+\mathbb{P}(C|A\cap\overline{B})\mathbb{P}(A)\mathbb{P}(\overline{B})+\mathbb{P}(C|\overline{A}\cap B)\mathbb{P}(\overline{A})\mathbb{P}(B)+\mathbb{P}(C|\overline{A}\cap\overline{B})\mathbb{P}(\overline{A})\mathbb{P}(\overline{B})$$

$$= 1 \times 0,8 \times 0,7 + 1 \times 0,8 \times 0,3 + 1 \times 0,2 \times 0,7 + 0 \times 0,2 \times 0,3 = 0,94$$

   Donc $\mathbb{P}(C) = 0,94$ et de même $\mathbb{P}(\overline{C}) = 1 \times 0,2 \times 0,3 = 0,06$.
   Avec l'outil AgenaRisk, on saisit les 3 tables précédentes et l'affichage donne pour C le résultat présenté dans la Figure 2, cohérent avec le calcul théorique.

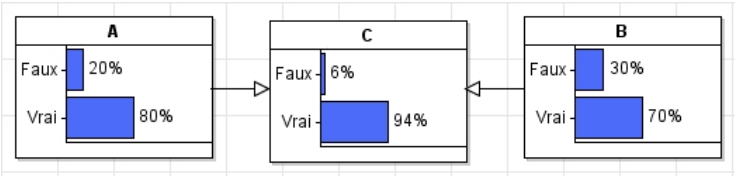


FIGURE 2 – Cas 1 : Probabilité totale de C

4. Calcul avec observation sur un nœud parent : $\mathbb{P}(C|A)$

   Ex. : Observation $A = VRAI$, seule la TdP de A est mise à jour :

A Vrai	A Faux
1	0

   La même formule de calcul nous donne en théorie :

$$\mathbb{P}(C|A) = 1 \times 1 \times 0,7 + 1 \times 1 \times 0,3 + 1 \times 0 \times 0,7 + 0 \times 0 \times 0,3 = 1$$

   De même, si l'observation est placée sur $B = VRAI$:

$$\mathbb{P}(C|B) = 1 \times 0,8 \times 1 + 1 \times 0,8 \times 0 + 1 \times 0,2 \times 1 + 0 \times 0,2 \times 0 = 1$$

   Avec l'outil, on place une observation $A = VRAI$, l'outil met à jour le graphe dans la Figure 3.

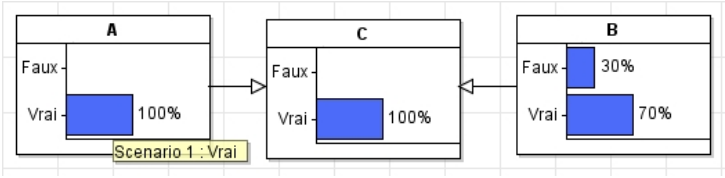


FIGURE 3 – Cas 1 : Probabilité de C sachant A

5. Calcul avec observation sur un nœud enfant : $\mathbb{P}(A|C)$
   Ne connaissant pas l'état de A, l'observation de C nous permet d'en estimer avec le théorème de Bayes :

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(C|A)\mathbb{P}(A)}{\mathbb{P}(C)}$$

   avec $\mathbb{P}(C|A) = 1$ (calcul plus haut) ; $\mathbb{P}(A) = 0,8$ (TdP) ; $\mathbb{P}(C) = 0,94$ (calcul plus haut).
   Alors :

$$\mathbb{P}(A|C) = \frac{1 \times 0,8}{0,94} = 0,85106 \qquad \mathbb{P}(B|C) = \frac{1 \times 0,7}{0,94} = 0,74468$$

   En plaçant une observation $C = VRAI$, l'outil met à jour le graphe dans la Figure 4.

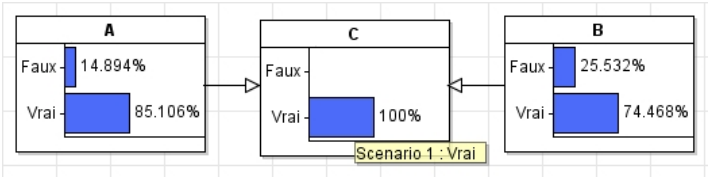


FIGURE 4 – Cas 1 : Probabilité de A sachant C

2.2 Cas 2 : nœuds valués (3 valeurs) et TdP du nœud fils renseignée manuellement

Connaissance des TdP de TOUS les nœuds à 3 valeurs (ex. : avec 3 valeurs : low, medium, high) :

1. Dessiner le graphe, choix du nœud de type "ranked" avec 3 états : low, medium, high.
   On modélise ici la « Qualité du travail » , conditionnée par la « Performance de l'équipe » (nœud A) et l'« Adéquation des outils » (nœud B). On suppose que la performance prime sur les outils.

Performance de l'équipe (A)		
low	medium	high
0,8	0,1	0,1

et

Adéquation des outils (B)		
low	medium	high
0,6	0,3	0,1

Qualité du travail (C)									
Perf. Equip. (A)	low			medium			high		
Adeq. Outils (B)	low	med.	high	low	med.	high	low	med.	high
C low	1	1	0	1	0	0	0	0	0
C med.	0	0	1	0	1	1	1	0	0
C high	0	0	0	0	0	0	0	1	1

TABLE 2 – Cas 2 : Tables de Probabilité des nœuds renseignées manuellement

2. Renseigner les tables manuellement.
   Dans cet exemple, l'influence de la « Performance de l'équipe » est plus importante que l'« Adéquation des outils » : même avec de très bons outils, une performance médiocre ne peut donner qu'un travail médiocre. Chaque colonne représente une combinaison des états de (A,B) suivant laquelle l'état de C est déterminée. On choisit pour cet exemple un seul état possible pour C, la probabilité est de 0 (si C n'est pas dans cet état) ou 1 (si C est dans cet état).
3. Calcul de $\mathbb{P}(C = high)$ sans aucune observation
   Pour tout $X_i, X_j \in [low, medium, high]$, on peut calculer la probabilité du travail de se trouver dans différents états de qualité sans connaissance sur la « Performance de l'équipe » ni l'« Adéquation des outils ».
   $$\mathbb{P}(C = X) = \sum \mathbb{P}(C = high | A = X_i \cap B = X_j)\mathbb{P}(A = X_i)\mathbb{P}(B = X_j)$$
   $\mathbb{P}(C = high) = 1 \times 0,3 \times 0,1 + 1 \times 0,1 \times 0,1 = 0,04$
   $\mathbb{P}(C = medium) = 1 \times 0,8 \times 0,1 + 1 \times 0,1 \times 0,3 + 1 \times 0,1 \times 0,1 + 1 \times 0,1 \times 0,6 = 0,18$
   $\mathbb{P}(C = low) = 1 \times 0,8 \times 0,6 + 1 \times 0,8 \times 0,3 + 1 \times 0,1 \times 0,6 = 0,78$
   On saisit les TdP précédentes, et l'outil affiche le résultat présenté dans la Figure 5, cohérent avec les calculs théoriques.

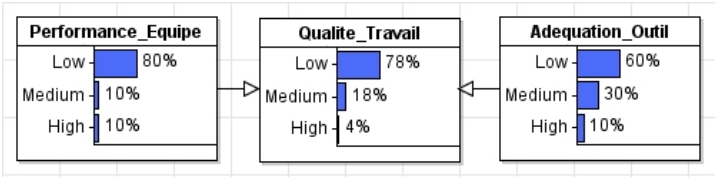


FIGURE 5 – Cas 2 : Probabilité totale de C

4. Calcul avec observation sur un nœud parent : $\mathbb{P}(C = high | A = high)$
   Observation $A = high$, seule la TdP de A est mise à jour :

A low	A med.	A high
0	0	1

   Le même calcul théorique nous donne : $\mathbb{P}(C = high | A = high) = 1 \times 1 \times 0,3 + 1 \times 1 \times 0,1 = 0,4$
   De même : $\mathbb{P}(C = medium | A = high) = 1 \times 1 \times 0,6 = 0,6$ et $\mathbb{P}(C = low | A = high) = 0$.
   En plaçant l'observation $A = high$, le logiciel met à jour les tables et affiche le résultat dans la Figure 6. On retrouve des résultats théoriques.
5. Calcul avec observation sur un nœud enfant : $\mathbb{P}(A = medium | C = medium)$
   Ne connaissant pas l'état de A, l'observation de C nous permet d'en estimer avec le théorème de

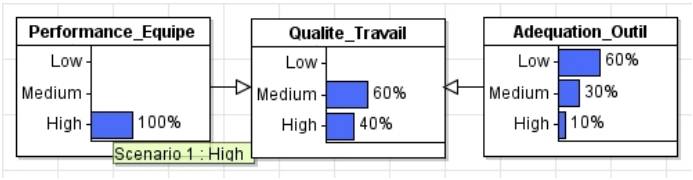


FIGURE 6 – Cas 2 : Probabilité de C sachant A

Bayes :

$$\mathbb{P}(A = medium|C = medium) = \frac{\mathbb{P}(C = medium|A = medium)\mathbb{P}(A = medium)}{\mathbb{P}(C = medium)}$$

avec $\mathbb{P}(C = medium|A = medium) = 1 \times 1 \times 0,3 + 1 \times 1 \times 0,1 = 0,4$; $\mathbb{P}(A = medium) = 0,1$ (TdP) ; $\mathbb{P}(C = medium) = 0,18$ (calcul plus haut). Alors :

$$\mathbb{P}(A = medium|C = medium) = \frac{0,4 \times 0,1}{0,18} = 0,222222$$

On obtient également $\mathbb{P}(C = medium|B = high) = 1 \times 0,8 \times 1 + 1 \times 0,1 \times 1 = 0,9$ puis

$$\mathbb{P}(B = high|C = medium) = \frac{\mathbb{P}(C = medium|B = high)\mathbb{P}(B = high)}{\mathbb{P}(C = medium)} = \frac{0,9 \times 0,1}{0,18} = 0,5$$

Avec l'outil, l'observation $C = medium$ affiche les mêmes résultats théoriques dans la Figure 7.

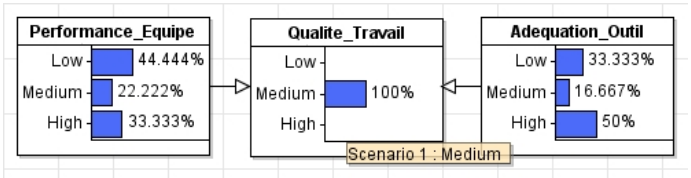


FIGURE 7 – Cas 2 : Probabilité de A sachant C

2.3 Cas 3 : nœuds valués (3 valeurs) et TdP du nœud fils renseignée par expression

Connaissance des TdP de TOUS les nœuds à 5 valeurs (ex. : avec 5 états : very low, low, medium, high, very high) :

1. Dessiner le graphe, choix du nœud de type "ranked" avec 5 états : very low, low, medium, high, very high
   Modélisons le cas d'une implémentation de code, l'« Efficacité » du code demande à la fois une bonne « Implémentation » et un bon « Algorithme ». Si l'un des deux est mauvais, l'« Efficacité » l'est aussi.
2. Renseigner la table du nœud fils avec une expression
   Les TdP des nœuds parents sont toujours renseignées manuellement, le changement ne concerne que la TdP du nœud fils. Les probabilités sont choisies de manière arbitraire pour illustrer les calculs.
   Le nœud C comportant 125 valeurs (25 colonnes de 5 états chacune). Pour remplir cette TdP, on choisit une expression arithmétique qui décrit le mieux l'influence des nœuds parents sur le nœud

Implémentation (A)				
very low	low	medium	high	very high
0,1	0,1	0,25	0,3	0,25

Algorithme (B)				
very low	low	medium	high	very high
0,1	0,25	0,3	0,25	0,1

TABLE 3 – Cas 3 : Tables de Probabilité des nœuds A et B renseignées manuellement

C : on se base sur la plus mauvaise valeur des nœuds parents pour déterminer celle du nœud fils : *Efficacité = min(Implémentation, Algorithme).*
Par exemple : $min(Implementation_{low}; Algorithme_{high}) = Efficacite_{low}$. Ainsi, dans la colonne où on trouve A_{low} et B_{high}, la possibilité que C se trouve dans l'état *low* est de 1 (probabilité de 0 à 1) et 0 ailleurs. En procédant de la même manière, l'outil remplit automatiquement toute la TdP dont voici un extrait (table 4) :

Efficacité (C)										
A	low					medium				
B	v.low	low	med.	high	v.high	v.low	low	med.	high	v.high
C v.low	1	0	0	0	0	1	0	0	0	0
C low	0	1	1	1	1	0	1	0	0	0
C med.	0	0	0	0	0	0	0	1	1	1
C high	0	0	0	0	0	0	0	0	0	0
C v.high	0	0	0	0	0	0	0	0	0	0

TABLE 4 – Cas 3 : Table de Probabilité du nœud C obtenue à partir de l'expression

3. Calcul de $\mathbb{P}(C = v.high)$ sans aucune observation
Le seul cas où $(C = v.high) \neq 0$ est $(A_{v.high}, B_{v.high})$, sa probabilité totale est :
$\mathbb{P}(C = v.high) = \mathbb{P}(C = v.high|A = v.high, B = v.high)\mathbb{P}(A = v.high)\mathbb{P}(B = v.high) = 1 \times 0,25 \times 0,1 = 0,025$

Pour $\mathbb{P}(C = medium) \neq 0$: les cas sont $(A_{medium}, B_{v.high})$, (A_{medium}, B_{high}), (A_{high}, B_{medium}), $(A_{v.high}, B_{medium})$ et (A_{medium}, B_{medium}).
En application numérique, cela donne :

$$\mathbb{P}(C = medium) = 1 \times 0,25 \times 0,1 + 1 \times 0,25 \times 0,25 + 1 \times 0,3 \times 0,3 + 1 \times 0,25 \times 0,3 + 1 \times 0,25 \times 0,3$$

$$\mathbb{P}(C = medium) = 0,025 + 0,0625 + 0,09 + 0,075 + 0,075 = 0,3275$$

Avec l'aide de l'outil, le même résultat est présenté dans la Figure 8.

4. Calcul avec observation sur un nœud parent : $\mathbb{P}(C = high|A = High)$
Le fait de placer une observation sur un nœud parent entraîne la modification de la valeur numérique utilisé pour calculer la probabilité totale. Sachant $(A = high)$, alors $\mathbb{P}(A = low) = \mathbb{P}(A = medium) = 0$, nous gardons seuls les cas $(A_{high}, B_{v.high})$ et (A_{high}, B_{high}).

$$\mathbb{P}(C = high) = 1 \times 1 \times 0,1 + 1 \times 1 \times 0,25 = 0,35$$

L'outil présente les probabilités de C sachant $A = high$ dans la Figure 9.

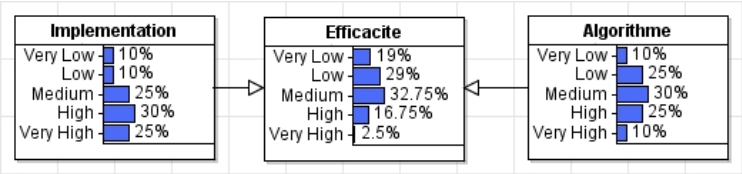


FIGURE 8 – Cas 3 : Probabilité totale de C

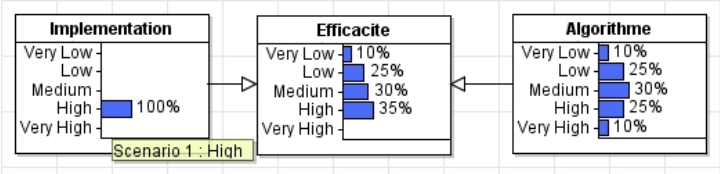


FIGURE 9 – Cas 3 : Probabilité de C sachant A

5. Calcul avec observation sur un nœud enfant : $\mathbb{P}(A = high|C = low)$
   Appliquons le théorème de Bayes :

$$\mathbb{P}(A = high|C = low) = \frac{\mathbb{P}(C = low|A = high)\mathbb{P}(A = high)}{\mathbb{P}(C = low)}$$

   Figure 9 nous donne $\mathbb{P}(C = low|A = high) = 0,25$, TdP de A donne $\mathbb{P}(A = high) = 0,3$; Figure 8 donne $\mathbb{P}(C = low) = 0,29$. Alors :

$$\mathbb{P}(A = high|C = low) = \frac{0,25 \times 0,3}{0,29} = 0,25862$$

   Figure 10 résume les autres probabilités de A sachant $C = low$.

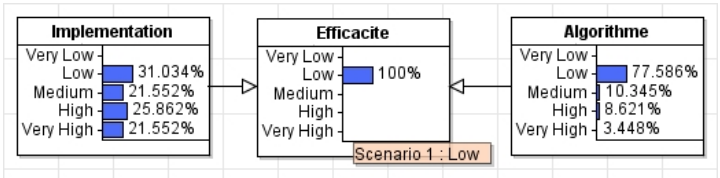


FIGURE 10 – Cas 3 : Probabilité de A sachant C

2.4 Cas 4 : nœuds valués (5 valeurs) et TdP du nœud fils renseignée par loi Normale Tronquée

Connaissance des TdP de TOUS les nœuds à 5 valeurs (ex. : avec 5 états : very low, low, medium, high, very high) :

1. Dessiner le graphe, choix du nœud de type "ranked" avec 5 états : very low, low, medium, high, very high
   Nous modélisons le même exemple que dans la section précédente, à savoir : C représente l'Efficacité d'un code avec un eImplémentation A et un Algorithme B.

Implémentation (A)				
very low	low	medium	high	very high
0,1	0,1	0,25	0,3	0,25

Algorithme (B)				
very low	low	medium	high	very high
0,1	0,25	0,3	0,25	0,1

TABLE 5 – Cas 4 : Tables de Probabilité des nœuds A et B renseignées manuellement

2. Renseigner la table du nœud fils avec loi Normale Tronquée
   Les TdP des nœuds parents sont toujours renseignées manuellement.
   Cas du nœud C :
   Le logiciel permet de remplir la TdP avec une loi Normale Tronquée pour ce type de nœud valué. Cette loi a été choisie car elle se rapproche de la modélisation qu'on souhaite obtenir : une valeur probable ajustée par une incertitude en fonction de la corrélation entre les nœuds parents et le nœud fils. Elle est tronquée sur l'intervalle [0 ;1] qui correspond aux valeurs d'une probabilité. Pour appliquer le calcul, l'outil a besoin de 4 paramètres :
   (a) Le poids du premier nœud parent w_A (le poids représente l'influence sur le nœud fils et intervient dans le calcul de la moyenne)
   (b) Le poids du second nœud parent w_B(le poids représente l'influence sur le nœud fils et intervient dans le calcul de la moyenne)
   (c) La variance σ^2 dont l'inverse agit comme indice de crédibilité (meilleure indice signifie meilleure corrélation)
   (d) L'espérance μ. L'outil propose 4 formules pour calculer cette espérance (μ_{Mean}, μ_{Min}, μ_{Max} et $\mu_{MixMinMax}$). (voir Section 4)

   Il y a 25 combinaisons possibles des états des nœuds A et B, conduisant aux 25 colonnes dans la TdP du nœud C. Connaissant l'état du nœud A et l'état du nœud B spécifique à une colonne, l'outil AgenaRisk calcule l'espérance μ correspondante et applique la loi Normale Tronquée pour déterminer la probabilité du nœud C d'être dans chacun des états : very low, low, medium, high, very high. De la même manière, l'outil remplit la table colonne par colonne.
   Afin de retrouver ces résultats à partir des calculs théoriques, on commence par chercher comment l'outil calcule l'espérance μ, c'est-à-dire quelles sont les valeurs attribuées à chacun des nœuds parents. On prend comme hypothèse que pour chaque état, c'est la valeur médiane de l'intervalle qui est utilisée dans le calcul de l'espérance (table 6).

Nom	very low	low	medium	high	very high
Intervalle	0 - 0,2	0,2 - 0,4	0,4 - 0,6	0,6 - 0,8	0,8 - 1
Valeur médiane	0,1	0,3	0,5	0,7	0,9

TABLE 6 – Valeurs correspondantes pour les 5 intervalles

Prenons $\sigma^2 = 0,01$ et $w_A = w_B = 1$ et la moyenne pondérée :

$$\mu_{Mean} = \frac{A_{etat} \times w_A + B_{etat} \times w_B}{w_A + w_B}$$

– cas de (A_{high}, B_{low}) :

$$\mu_{Mean} = \frac{A_{high} \times w_A + B_{low} \times w_B}{w_A + w_B} = \frac{0,7 \times 1 + 0,3 \times 1}{1 + 1} = 0,5$$

– cas de (A_{medium}, B_{high}) :

$$\mu_{Mean} = \frac{A_{medium} \times w_A + B_{high} \times w_B}{w_A + w_B} = \frac{0,5 \times 1 + 0,7 \times 1}{1+1} = 0,6$$

– cas de $(A_{v.high}, B_{high})$:

$$\mu_{Mean} = \frac{A_{v.high} \times w_A + B_{high} \times w_B}{w_A + w_B} = \frac{0,9 \times 1 + 0,7 \times 1}{1+1} = 0,8$$

A partir de cette valeur de μ_{Mean}, on demande à l'outil la répartition des probabilité dans 5 états de C pour chacun des cas (table 7) en lui indiquant l'état des nœuds parents. Par la suite, on utilise le calcul théorique pour retrouver la répartition de C dans le cas $\mu_{Mean} = 0,5$. Si on retrouve les mêmes résultats fournis par l'outil pour (A_{high}, B_{low}), l'hypothèse que l'outil base son calcul sur la valeur médiane est pertinente.

Paramètres : $\sigma^2 = 0,01$ et $w_A = w_B = 1$			
nœudA/nœudB	high/low	medium/high	v.high/high
μ_{Mean} supposé	0,5	0,6	0,8
C v.low	0.00350	1.5074E-4	2.0841E-8
C low	0.18264	0.03661	1.5156E-4
C med.	0.62769	0.46326	0.03705
C high	0.18264	0.46332	0.47572
C v.high	0.00350	0.03663	0.48706

TABLE 7 – Résultats fournis par l'outil AgenaRisk à partir des paramètres d'entrée

En considérant la valeur $\mu_{Mean} = 0,5$, on applique le calcul théorique de la loi Normale Tronquée (voir détail en Annexe 6). Par changement de variable, $Z = \frac{X-\mu}{\sigma} = \frac{X-0,5}{0,1}$ suit la loi $\mathcal{N}(0,1)$.

$$\mathbb{P}(X \leq 0) = \mathbb{P}(\frac{X-0,5}{0,1} \leq \frac{0-0,5}{0,1}) = \Phi(-5) = 1 - \Phi(5) = 0,0001$$

Ainsi, on obtient :

$$\mathbb{P}(X \leq 0,2) = \Phi(\frac{X-0,5}{0,1} \leq \frac{0,2-0,5}{0,1}) = \Phi(\frac{0,2-0,5}{0,1}) = \Phi(-3) = 1 - \Phi(3) = 0,0013$$

$$\mathbb{P}(X \leq 0,4) = \Phi(\frac{X-0,5}{0,1} \leq \frac{0,4-0,5}{0,1}) = \Phi(\frac{0,4-0,5}{0,1}) = \Phi(-1) = 1 - \Phi(1) = 0,1587$$

$$\mathbb{P}(X \leq 0,6) = \Phi(\frac{X-0,5}{0,1} \leq \frac{0,6-0,5}{0,1}) = \Phi(\frac{0,6-0,5}{0,1}) = \Phi(1) = 0,8413$$

$$\mathbb{P}(X \leq 0,8) = \Phi(\frac{X-0,5}{0,1} \leq \frac{0,8-0,5}{0,1}) = \Phi(\frac{0,8-0,5}{0,1}) = \Phi(3) = 0,9987$$

$$\mathbb{P}(X \leq 1) = \Phi(\frac{X-0,5}{0,1} \leq \frac{1-0,5}{0,1}) = \Phi(\frac{1-0,5}{0,1}) = \Phi(5) = 0,9999$$

La répartition des états du nœud C pour (A_{high}, B_{low}) est alors théoriquement :

Very low : $\mathbb{P}(0 \leq X \leq 0,2 | 0 \leq X \leq 1) = \frac{\Phi(-3)-\Phi(-5)}{\Phi(5)-\Phi(-5)} = \frac{0,0013-0,0001}{0,9999+0,9999-1} = \frac{0,0012}{0,9998} = 0,0012$
Low : $\mathbb{P}(0,2 \leq X \leq 0,4 | 0 \leq X \leq 1) = \frac{\Phi(-1)-\Phi(-3)}{\Phi(5)-\Phi(-5)} = \frac{0,1574}{0,9998} = 0,1574$

Medium : $\mathbb{P}(0,4 \leq X \leq 0,6 | 0 \leq X \leq 1) = \frac{\Phi(1)-\Phi(-1)}{\Phi(5)-\Phi(-5)} = \frac{0,6826}{0,9998} = 0,6828$
High : $\mathbb{P}(0,6 \leq X \leq 0,8 | 0 \leq X \leq 1) = \frac{\Phi(3)-\Phi(1)}{\Phi(5)-\Phi(-5)} = \frac{0,1574}{0,9998} = 0,1574$
Very High : $\mathbb{P}(0,8 \leq X \leq 1 | 0 \leq X \leq 1) = \frac{\Phi(5)-\Phi(3)}{\Phi(5)-\Phi(-5)} = \frac{0,0013}{0,9998} = 0,0012$

Répartition des états du nœud C ($w_A = w_B = 1$ et $\sigma = 0,1$)		
	Théorie ($\mu = 0,5$)	Avec l'outil (A_{high}, B_{low})
Very Low	0,0012	0.00350
Low	0,1574	0.18264
Medium	0,6828	0.62769
High	0,1574	0.18264
Very High	0,0012	0.00350

TABLE 8 – Différentes valeurs entre calcul théorique et l'outil AgenaRisk

On peut observer sur la table 8 que les résultats théoriques (en utilisant μ_{Mean} supposé) sont différents de ceux calculés par l'outil AgenaRisk. Il s'avère que c'est notre hypothèse sur l'utilisation de la valeur médiane (présenté table 6) qui induit les différences entre notre calcul théorique et les résultats de l'outil. En effet, l'outil AgenaRisk au lieu de prendre une seule valeur, par exemple 0.7 pour le niveau *high*, utilise 3 valeurs [1] 0.6, 0.7 et 0.8. Ainsi, à partir d'un état de A (par exemple *high*) et un état de B (par exemple *low*), l'outil calcule 9 combinaisons possibles, et pour chacune de ces combinaisons il calcule l'espérance μ_{Mean}, puis la distribution de probabilité suivant la loi Normale Tronquée. Puis pour un état donné de C sachant l'état de A et B, la probabilité est la moyenne de ces 9 répartitions. Le détail de cet exemple de calcul est donné en Section 4. Dans la suite de ce rapport, on utilise la TdP fourni par l'outil.

3. Calcul de $\mathbb{P}(C = medium)$ sans aucune observation
Pour tout $X_i, X_j \in [v.low, low, medium, high, v.high]$:

$$\mathbb{P}(C = medium) = \sum_{X_i, X_j} \mathbb{P}(C = high | A = X_i \cap B = X_j)\mathbb{P}(A = X_i)\mathbb{P}(B = X_j)$$

A partir des tables générées, on calcule la probabilité totale comme dans les cas précédents. L'outil affiche les valeurs calculées dans la Figure 11 : $\mathbb{P}(C = v.low) = 0,4581$, $\mathbb{P}(C = low) = 0,18468$, $\mathbb{P}(C = medium) = 0,35588$, $\mathbb{P}(C = high) = 0,30772$, $\mathbb{P}(C = v.high) = 0,10592$.

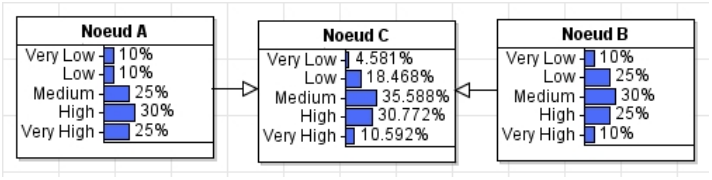


FIGURE 11 – Cas 4 : Probabilité totale de C

4. Calcul avec observation sur un nœud parent : $\mathbb{P}(C = high | A = low)$
En observant $\mathbb{P}(A = low)$, on recalcule la probabilité de C en tenant compte uniquement des combinaisons (A_{low}, B_X) avec $X \in [v.low, low, medium, high, v.high]$. Le résultat est reporté dans la Figure 12 :

1. En réalité l'outil utilise 5 valeurs par défaut, mais c'est configurable

$\mathbb{P}(C = high|A = low) = 0,10387$, $\mathbb{P}(C = v.high|A = low) = 0,00$, $\mathbb{P}(C = medium|A = low) = 0,39166$, $\mathbb{P}(C = low|A = low) = 0,39333$, $\mathbb{P}(C = v.low|A = low) = 0,10655$.

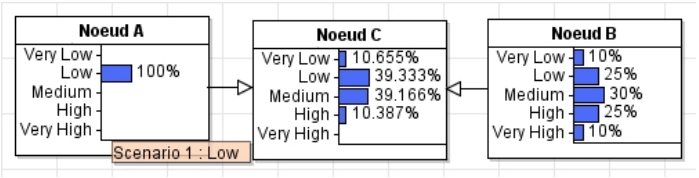


FIGURE 12 – Cas 4 : Probabilité de C sachant A

5. Calcul avec observation sur un nœud enfant : $\mathbb{P}(A = high|C = low)$
   La méthode employée est identique au cas 3. Ainsi en observant $\mathbb{P}(A = high)$ et pour tout $X_i, X_j \in [v.low, low, medium, high, v.high]$:

$$\mathbb{P}(C = low|A = high) = \sum \mathbb{P}(C = low)\mathbb{P}(A = high)\mathbb{P}(B = X_i) = 0,10387$$

   Ensuite, on utilise le théorème de Bayes pour calculer (voir Figure 13) :

$$\mathbb{P}(A = high|C = low) = \frac{\mathbb{P}(C = low|A = high)\mathbb{P}(A = high)}{\mathbb{P}(C = low)} = \frac{0,10387 \times 0,3}{0,18468} = 0,16873$$

   $\mathbb{P}(A = v.high|C = low) = 0,3859$, $\mathbb{P}(A = medium|C = low) = 0,32909$
   $\mathbb{P}(A = low|C = low) = 0,21298$, $\mathbb{P}(A = v.low|C = low) = 0,25061$

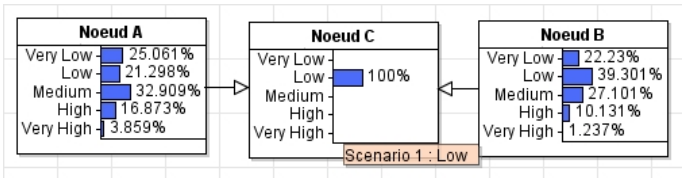


FIGURE 13 – Cas 4 : Probabilité de A sachant C

3 Exploration des différentes techniques proposées par AgenaRisk pour renseigner les Tables de Probabilité en utilisant la loi Normale Tronquée

Comme présenté dans le cas 4 du rapport, on peut indiquer à l'outil de remplir la Table de Probabilité du nœud fils C suivant une loi de probabilité. Cette partie présente en détail la démarche de l'outil pour remplir la table à partir des connaissances sur ses nœuds parents. La loi Normale Tronquée est présentée en détail ici mais il existe d'autres manières pour indiquer à l'outil comment remplir la TdP (voir Section 5).

Une loi Normale tronquée est caractérisée par son espérance μ, sa variance σ^2 et l'intervalle de l'étude après la double troncature. Dans le cadre de l'étude, l'intervalle est [0,1]. Il reste donc à préciser les valeurs de μ et σ^2 de manière arbitraire ou suivant des règles de calculs.

Cas de 2 nœuds parents et un nœud fils. L'outil propose 4 façons pour calculer l'espérance μ : μ_{Mean}, μ_{Min}, μ_{Max} et $\mu_{MixMinMax}$. Ces formules prennent en compte l'influence des nœuds parents sur le nœud fils via 2 paramètres w_A et w_B - poids des nœuds parents A et B.

Ces 4 méthodes suffisent pour couvrir la plupart des cas rencontrés (d'après les auteurs). Toutefois, il est possible d'utiliser d'autres formules arithmétiques ou logiques pour les cas spécifiques.

3.1 Espérance μ_{Mean}

$$\mu_{Mean} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i} = \frac{w_A X_A + w_B X_B}{w_A + w_B}$$

Cette formule représente l'espérance comme la moyenne pondérée des nœuds parents, plus le poids w est grand, plus l'influence du nœud parent est importante.

3.2 Espérance μ_{Min} avec le nombre de nœuds n

$$\mu_{Min} = MIN\left[\frac{w_i X_i + \sum_{i \neq j}^{n} X_j}{w_i + (n-1)}\right] = MIN\left[\frac{w_A X_A + X_B}{w_A + 2 - 1}, \frac{w_B X_B + X_A}{w_B + 2 - 1}\right]$$

n représente le nombre de nœuds parents (dans notre exemple $n = 2$). Cette formule génère une espérance μ_{Min} plus grande que $\mu = min(X_A, X_B)$ mais reste inférieure à μ_{Mean}. Dans le cas de poids unitaire, on retrouve μ_{Mean}.

3.3 Espérance μ_{Max} avec le nombre de nœuds n

$$\mu_{Max} = MAX\left[\frac{w_i X_i + \sum_{i \neq j}^{n} X_j}{w_i + (n-1)}\right] = MAX\left[\frac{w_A X_A + X_B}{w_A + 2 - 1}, \frac{w_B X_B + X_A}{w_B + 2 - 1}\right]$$

De même, μ_{Max} donne une valeur intermédiaire entre μ_{Mean} et $\mu = max(X_A, X_B)$. Dans le cas de poids unitaire, on retrouve μ_{Mean}.

3.4 Espérance $\mu_{MixMinMax}$

Pour cette formule, les nœuds parents n'ont pas toujours le même poids, ceci change quand le nœud parent A passe dans un autre état où sa valeur est plus grande (ou plus petite) que celle du parent B. Par exemple, si $(X_A \leq X_B)$ alors w_{min} sera attribué à X_A et w_{max} à X_B.

$$\mu_{MixMinMax} = \frac{w_{min} MIN(X_i) + w_{max} MAX(X_i)}{w_{min} + w_{max}} = \frac{w_{min} X_A + w_{max} X_B}{w_{min} + w_{max}}$$

La Figure 14 compare les différents résultats obtenus avec ces 4 méthodes de calcul de moyenne.

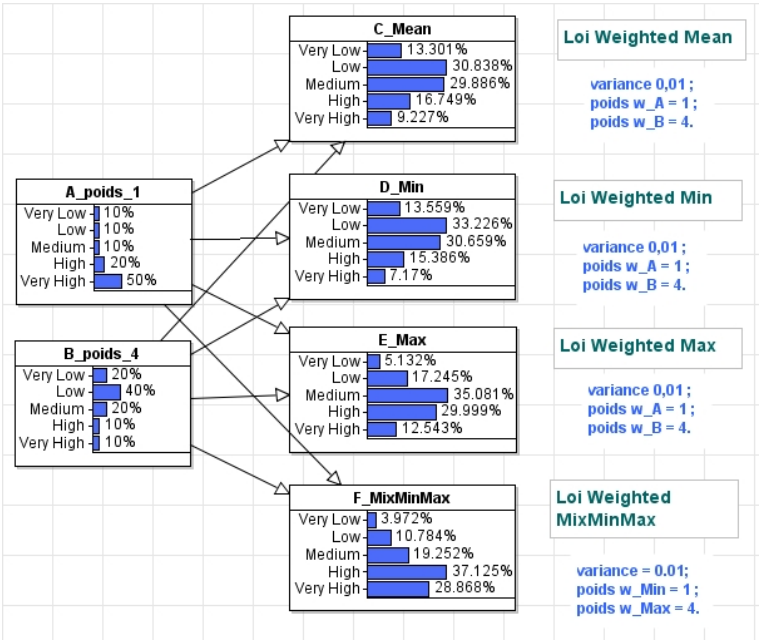


FIGURE 14 – Les 4 formules de calcul de moyenne

4 Technique d'échantillonnage pour calculer la distribution connaissant l'état des nœuds parents

Le calcul de l'espérance μ fait appel à un paramètre X_{parent}. Or, dans le cas d'un nœud valué, un état est représenté par un intervalle continu, et donc il faut déterminer une valeur X à utiliser pour représenter cet intervalle. Dans le cas 4 de ce rapport, nous avons présenté une méthode qui utilise la valeur médiane (voir table 8) pour représenter l'intervalle, mais nous détaillons par la suite un calcul différent qui est celui effectué effectivement par l'outil. En effet, l'outil AgenaRisk, au lieu de prendre une valeur unique pour un niveau (*high*, par exemple), utilise 3 valeurs 0.6, 0.7 et 0.8. Le nombre de valeurs est paramétré par la variable « taille de l'échantillon ». L'outil utilise par défaut 5 valeurs, mais dans la suite nous prendrons 3 valeurs pour illustrer la méthode. Ainsi, à partir d'un état de A (par exemple *high*) et un état de B (par exemple *low*), l'outil calcule 9 combinaisons possibles.

Dans l'exemple, on souhaite calculer la probabilité liée à chaque état de C sachant A_{high} et B_{low}. A se trouve alors dans l'intervalle [0,6 ; 0,8] et B dans [0,2 ; 0,4]. Sachant que la taille de l'échantillon est de 3, pour chaque intervalle, on choisit 3 valeurs à intervalles réguliers (la borne inférieure, la moyenne et la borne supérieure). Cela fait 9 combinaisons de valeur à partir de A_{high} et B_{low}. On procède alors au calcul de l'espérance μ_{Mean}. Par exemple pour ces valeurs $X_A = 0,8$ $(high)$ et $X_B = 0,2$ (low), sachant que les poids des parents est $w_A = w_B = 1$ on a :

$$\mu_{Mean} = \frac{w_A X_A + w_B X_B}{w_A + w_B} = \frac{1 \times 0,8 + 1 \times 0,2}{1 + 1} = 0,5$$

On peut donc appliquer la loi Normale Tronquée sur [0 ;1] pour $\mu = 0,5$ (et on fixe par exemple $\sigma 2 = 0,01$). En faisant de même pour les 8 autres combinaisons, on obtient 9 distributions de probabilités (Figure 15). La distribution résultant de la moyenne de ces 9 distributions est celle que l'outil donne au nœud C sachant A_{high} et B_{low} (Figure 16) pour une taille de l'échantillon à 3.

A_High	0.6	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8
B_Low	0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4
espérance	**0.4**	**0.45**	**0.5**	**0.45**	**0.5**	**0.55**	**0.5**	**0.55**	**0.6**
variance	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Les 9 distributions obtenues par calcul TNormal(espérance, variance) tronquée sur [0,1]									
C v.low	0.02271918	0.006206289	0.0013496121	0.006206289	0.0013496121	0.00023261088	0.0013496121	0.00023261088	0.000031671258
C low	0.47726497	0.3023289	0.15730545	0.3023289	0.15730545	0.0665748	0.15730545	0.0665748	0.02271918
C medium	0.47726497	0.6246574	0.6826899	0.6246574	0.6826899	0.6246574	0.6826899	0.6246574	0.47726497
C high	0.02271918	0.0665748	0.15730545	0.0665748	0.15730545	0.3023289	0.15730545	0.3023289	0.47726497
C v.high	0.000031671258	0.00023261088	0.0013496121	0.00023261088	0.0013496121	0.006206289	0.0013496121	0.006206289	0.02271918

FIGURE 15 – 9 distributions de probabilité du nœud fils C pour A_{high} et B_{low}

	Somme des 9	Moyenne des 9	Avec l'outil
C v.low	0.039677487318	0.004408609702	0.0044086096
C low	1.7097079	0.189967544444444	0.18996754
C medium	5.50122924	0.611247693333333	0.61124766
C high	1.7097079	0.189967544444444	0.18996754
C v.high	0.039677487318	0.004408609702	0.0044086096

FIGURE 16 – La moyenne des distributions est celle du nœud fils C pour A_{high} et B_{low}

En augmentant la taille de l'échantillon à 5, on obtient 25 distributions dont la moyenne est la distribution donnée par l'outil dans le cas 4 (Figure 17).

	Somme des 25	Moyenne des 25	Avec l'outil
C v.low	0.087643064538	0.00350572258152	0.0035057226
C low	4.56619433	0.1826477732	0.18264778
C medium	15.69232504	0.6276930016	0.627693
C high	4.56619433	0.1826477732	0.18264778
C v.high	0.087643064538	0.00350572258152	0.0035057226

FIGURE 17 – Résultats du cas 4 retrouvés par calculs théoriques

Il est à noter que la moyenne des distributions est différente de la distribution ayant pour l'espérance la moyenne des espérances $\mu_{moy} = 0,5$ (voire table 8).

5 Utilisation d'une Table de Probabilité partitionnée

Cette partie met en avant une possibilité qu'offre l'outil : celle de pouvoir préciser une méthode de répartition de probabilité différente pour chacune des combinaisons possibles des nœuds parents (chaque combinaison correspond à une colonne dans la TdP du nœud fils). La méthode de répartition choisie peut être une loi probabiliste (loi Normale Tronquée, loi de Poisson, loi binomiale...), une expression arithmétique (addition, soustraction, logarithme, sinus, cosinus...) ou une expression logique (AND, OR, XOR...). Le choix de la méthode dépend du type de nœud et le résultat souhaité.

Par la suite, on revient sur le cas déterministe (cas 1) pour illustrer cette possibilité. Au lieu de rentrer manuellement la TdP du nœud fils, on laisse l'outil compléter la table à partir de la loi Normale Tronquée (pour le type de nœud valué) ou une expression logique. Ainsi, on choisit pour chacune des 4 combinaisons des paramètres qui conviennent. Table 9 rappelle la TdP du nœud fils renseignée manuellement dans le cas 1, Figure 18 le résultat correspondant (cas d'une table manuelle) qui sert comme référence.

Reprenons les 4 combinaisons des nœuds parents : A_{Vr}/B_{Vr}, A_{Vr}/B_{Fx}, A_{Fx}/B_{Vr}, A_{Fx}/B_{Fx} (4 colonnes de la TdP du nœud C). A chaque fois, il faut favoriser le parent (en attribuant un poids fort) qui se

nœud C				
	A Vrai		A Faux	
	B Vrai	B Faux	B Vrai	B Faux
C Vrai	1	1	1	0
C Faux	0	0	0	1

TABLE 9 – Cas déterministe : Table de Probabilité du nœud C

trouve dans l'état *VRAI* avec très peu d'incertitude (donc faible variance). Dans le cas d'un nœud valué, l'état *FAUX* correspond à l'intervalle [0 - 0,5] et *VRAI* à [0,5 - 1] donc on utilise par exemple μ_{Max} en cas de loi Normale Tronquée ou $max(X_A, X_B)$ en cas d'une expression. Concrètement, on choisit pour chaque combinaison (avec une très petite variance σ^2) :

- A_{Vrai}/B_{Vrai} : μ_{Mean} ou μ_{Min} ou μ_{Max} avec poids égal ou $max(X_A, X_B)$.
- A_{Vrai}/B_{Faux} : μ_{Max} avec grand w_A ou $max(X_A, X_B)$.
- A_{Faux}/B_{Vrai} : μ_{Max} avec grand w_B ou $max(X_A, X_B)$.
- A_{Faux}/B_{Faux} : μ_{Mean} ou μ_{Min} ou μ_{Max} avec poids égal ou $max(X_A, X_B)$.

La Figure 18 compare les 4 approches pour représenter ce cas quasi-déterministe. Les nœuds parents étant identiques, le résultat change en fonction de la méthode indiquée à l'outil (Table 10) :

- Table manuelle : TdP renseignée manuellement, on retrouve le cas 1 détaillé dans ce rapport.
- Expression $max(X_A, X_B)$: TdP renseignée à l'aide de l'expression $max(X_A, X_B)$, le résultat de l'expression est binaire (0 ou 1), on obtient la même TdP que le cas précédent.
- Table partitionnée : TdP renseignée à l'aide de la loi Normale Tronquée, l'espérance est calculée avec la formule μ_{Max} ou μ_{Mean} suivant le cas. Les poids des nœuds parents diffèrent d'une colonne à l'autre.
- Loi Weighted MixMinMax : TdP renseignée à l'aide de la loi Normale Tronquée, l'espérance est calculée avec la formule $\mu_{MixMinMax}$. Les nœuds parents n'ont pas toujours le même poids.

Loi de distribution	A_{Vrai}/B_{Vrai}	A_{Vrai}/B_{Faux}	A_{Faux}/B_{Vrai}	A_{Faux}/B_{Faux}
Table manuelle	voir cas 1			
$X_C = max(X_A, X_B)$	résultat identique au cas 1			
Loi Normale Tronquée	$w_A = w_B = 1$ espérance μ_{Mean}	$w_A = 5$, $w_B = 0$ espérance μ_{Max}	$w_A = 0$, $w_B = 5$ espérance μ_{Max}	$w_A = w_B = 1$ espérance μ_{Mean}
	variance $\sigma^2 = 0,0005$			
Loi Normale Tronquée	$w_{Max} = 5$, $w_{Min} = 0$ variance $\sigma^2 = 0,0005$, espérance $\mu_{MixMinMax}$			

TABLE 10 – Différentes manières pour remplir la TdP du nœud C

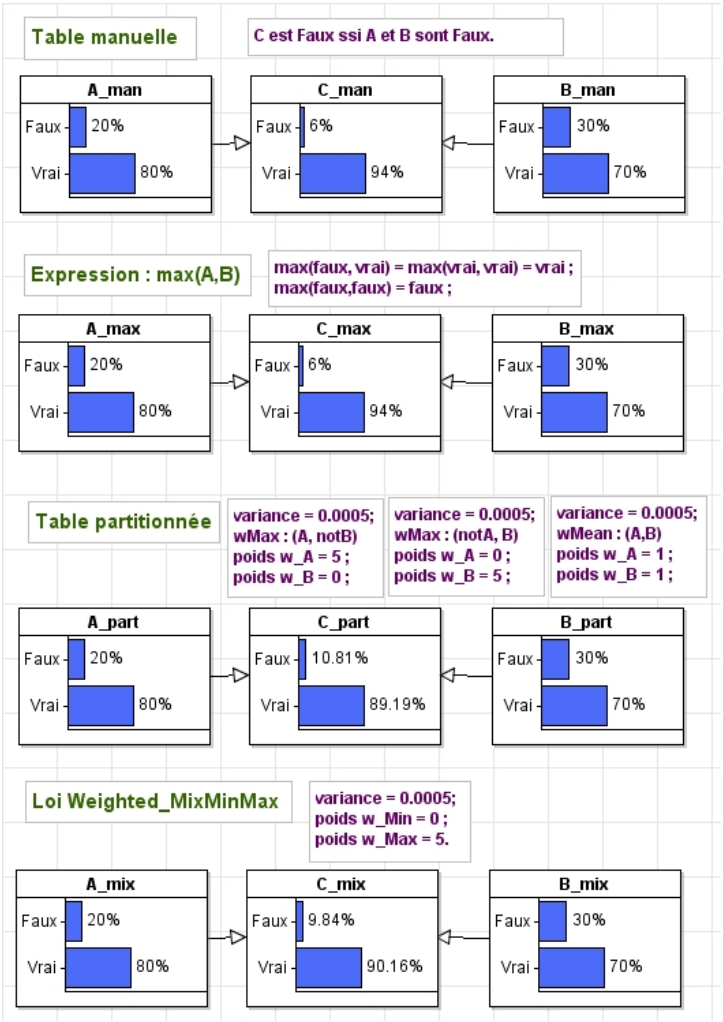


FIGURE 18 – Le cas quasi-déterministe avec différentes approches

6 Conclusion

Les exemples de réseaux bayésiens étudiés avec l'outil AgenaRisk et présentés dans ce rapport nous ont permis de comprendre la technique pour renseigner une Table de Probabilité en utilisant une loi de probabilité. Cette technique permet de gagner du temps quand la table est de grande taille tout en gardant une cohérence entre l'avis d'un expert et la répartition des valeurs probabilistes.

Références

[1] N. E. Fenton, M. Neil and J. Gallan, "Using Ranked nodes to model qualitative judgements in Bayesian Networks" in *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 10, pp. 1420-1432, Oct. 2007

[2] http ://www.agenarisk.com, La version 5.6.0 est utilisée dans ce rapport.

ANNEXE

Calcul théorique de la loi Normale Tronquée

La loi Normale est une loi de probabilité absolument continue qui dépend de deux paramètres : son espérance réelle μ et son écart-type réel positif σ. Cette loi est également appelée loi gaussienne, loi de Gauss ou loi de Laplace.

Quand $\mu = 0$ et $\sigma = 1$, la loi Normale est dite **centrée réduite**.

La variance et l'écart-type σ de la variable aléatoire X qui suit une loi Normale est définie par :

$$Var(X) = E[(X - \mu)^2]$$

$$\sigma = \sqrt{Var(X)}$$

La densité de probabilité de la loi Normale est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Pour une loi Normale centrée réduite ($\mu = 0$ et $\sigma = 1$), la fonction de densité est donnée par :

$$f(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}t^2}$$

La courbe de la densité d'une loi Normale est appelée courbe de Gauss (Figure 19).

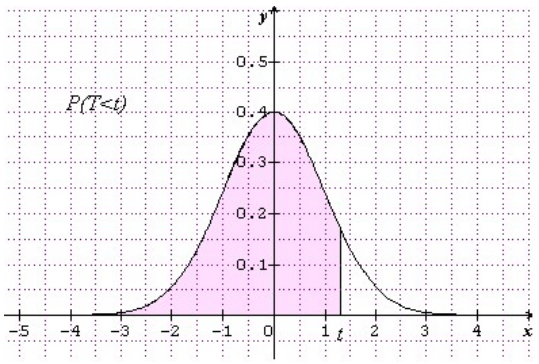


FIGURE 19 – La fonction de densité de la loi normale

La fonction de répartition de Φ d'une variable aléatoire t de la loi Normale centrée réduite est définie pour tout nombre réel a, $-\infty < a < \infty$, par :

$$\Phi(a) = \mathbb{P}(t \leq a) = \int_{-\infty}^{a} f(t)\, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \mathrm{e}^{\left(-\frac{1}{2}t^2\right)}\, dt$$

Application pour calcul de probabilité :

$$\mathbb{P}(t \leq a) = \Phi(a) = \int_{-\infty}^{a} f(t)\, dt$$

$$\mathbb{P}(a \leq t \leq b) = \Phi(b) - \Phi(a) = \int_{a}^{b} f(t)\, dt$$

Si on sait en plus que cette variable aléatoire t appartient à l'itervalle [0,1] (c'est le cas pour une probabilité), alors :

$$\mathbb{P}(a \leq t \leq b | 0 \leq t \leq 1) = \frac{\Phi(b) - \Phi(a)}{\Phi(1) - \Phi(0)} = \frac{\int_a^b f(t)\,dt}{\int_0^1 f(t)\,dt}$$

On dit alors que la loi Normale est tronquée sur l'intervalle [0,1].

Le calcul direct des probabilités en utilisant sa fonction de répartition n'étant pas possible, les staticiens se sont basés sur la table de valeurs de la fonction de répartition $\Phi(x) = \mathbb{P}(X \leq x)$, lorsque X suit la loi Normale centrée réduite $\mathcal{N}(0,1)$. Un exemple de cette table [2] se trouve à la Figure 20. La ligne donne la valeur de X à une décimale et la colonne donne la valeur de la deuxième décimale. Par exemple, la probabilité $\mathbb{P}(X = 2,63)$ se trouve à l'intersection de la ligne 2,60 et la colonne 0,03 : $\mathbb{P}(X = 2,63) = 0,9957$. En cas de valeur négative, comme la courbe est symétrique, on peut calculer : $\mathbb{P}(X) = 1 - \mathbb{P}(-X)$

Exemple d'application :

$$\mathbb{P}(0,2 \leq X \leq 0,4 | 0 \leq X \leq 1) = \frac{\Phi(0,4) - \Phi(0,2)}{\Phi(1) - \Phi(0)}$$

$$\mathbb{P}(0,2 \leq X \leq 0,4 | 0 \leq X \leq 1) = \frac{0,6554 - 0,5793}{0,8413 - 0,5} = \frac{0,0761}{0,3413} = 0,22297$$

Si l'on considère maintenant une variable normale X de paramètres μ et σ^2 quelconques, la variable $Z = \frac{X-\mu}{\sigma}$ sera normale centrée réduite. Par conséquent, on peut exprimer la fonction de répartition de X de la manière suivant :

$$F_X(a) = \mathbb{P}(X \leq a) = \mathbb{P}(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}) = \Phi(\frac{a-\mu}{\sigma})$$

Exemple d'application : On veut calculer la probabilité $\mathbb{P}(0,2 \leq X \leq 0,4 | 0 \leq X \leq 1)$ qui suit la loi normale de paramètres $\mu = 0,5$ et $\sigma = 0,1$, c'est à dire $\mathcal{N}(0.5, 0.1)$. Ce sera donc la variable $Z = \frac{X-0,5}{0,1}$ qui suit la loi normale centrée réduite. Alors :

$$\mathbb{P}(0,2 \leq X \leq 0,4 | 0 \leq X \leq 1) = \frac{\Phi(\frac{0,4-0,5}{0,1}) - \Phi(\frac{0,2-0,5}{0,1})}{\Phi(\frac{1-0,5}{0,1}) - \Phi(\frac{0-0,5}{0,1})}$$

$$\mathbb{P}(0,2 \leq X \leq 0,4 | 0 \leq X \leq 1) = \frac{\Phi(-1) - \Phi(-3)}{\Phi(5) - \Phi(-5)} = \frac{1 - \Phi(1) - 1 + \Phi(3)}{\Phi(5) - 1 + \Phi(5)}$$

$$\mathbb{P}(0,2 \leq X \leq 0,4 | 0 \leq X \leq 1) = \frac{\Phi(3) - \Phi(1)}{\Phi(5) + \Phi(5) - 1} = \frac{0,9987 - 0,8413}{0,9999 + 0,9999 - 1} = \frac{0,1574}{0,9998} = 0,1574$$

2. Extrait du livre « Initiation au probabilités ». Sheldon ROSS. Presses polytechniques et universitaires romandes.

Tableau 5.4 Aire $\Phi(x)$ située sous la densité normale standard à gauche de x

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	8413	.8438	.8461	.8485	.8508	.8531	.8554	.8557	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

FIGURE 20 – Table des valeurs de la fonction de répartition de la loi Normale centrée réduite.