

Construction de Fonctions de Croyance Prédicatives par une Approche Fréquentiste



Thierry Denœux

Université de Technologie de Compiègne

HEUDIASYC (UMR CNRS 6599)

<http://www.hds.utc.fr/~tdenoeux>



Problème posé

- Nombreux outils pour le raisonnement dans un environnement où la connaissance de certaines variables est modélisée par des fonctions de croyance :
 - Règles de combinaison, marginalisation, extension vide, conditionnement, déconditionnement,...
 - Algorithmes de propagation dans les réseaux évidentiels (*valuation-based networks*)
 - Théorème de Bayes Généralisé
 - ...
- Problème : Origine des fonctions de croyance manipulées ?
 - Opinions d'experts (élicitation)
 - **Données (observations)** → lien avec des modèles statistiques ?



Exemple : application du TBG

- **Théorème de Bayes Généralisé (Smets, 1978)**
 - $\Omega = \{\omega_1, \dots, \omega_J\}$ ensemble d'états (classes, maladies, ...)
 - Variable $X \in \mathcal{X}$ (attribut, symptôme)
 - On connaît $m^{\mathcal{X}}[\omega_j]$, $j=1, \dots, J$.
 - On observe $X \in B \subseteq \mathcal{X}$
 - On cherche $m^{\Omega}[X \in B]$. Solution :

$$pl^{\Omega}[X \in B](A) = 1 - \prod_{\omega_j \in A} (1 - pl^{\mathcal{X}}[\omega_j](B)) \quad \forall A \subseteq \Omega$$

- **Problème : origine des $m^{\mathcal{X}}[\omega_j]$, $j=1, \dots, J$?**
 - Le plus souvent, on dispose d'un ensemble d'observations (échantillon) x_1, \dots, x_n de la variable X , pour chaque état ω_j
 - Comment en déduire $m^{\mathcal{X}}[\omega_j]$?



Formalisation du problème

- Soit une grandeur X définie sur $\mathcal{X} = \{\xi_1, \dots, \xi_k\}$, générée par un processus répétable \rightarrow **variable aléatoire de loi inconnue** \mathbb{P}_X
- Signification de $\mathbb{P}_X(A)$, $A \subseteq \Omega$: **fréquence limite** de réalisation de l'événement A au cours d'une suite infinie de répétitions

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

- On observe une réalisation $\mathbf{x} = (x_1, \dots, x_n)$ d'un échantillon iid $\mathbf{X} = (X_1, \dots, X_n)$, de variable aléatoire parente X
- Problème : construire une fonction de croyance $\text{bel}^x[\mathbf{x}]$ relative à une prochaine réalisation de la variable X

fonction de croyance prédictive



Exemple

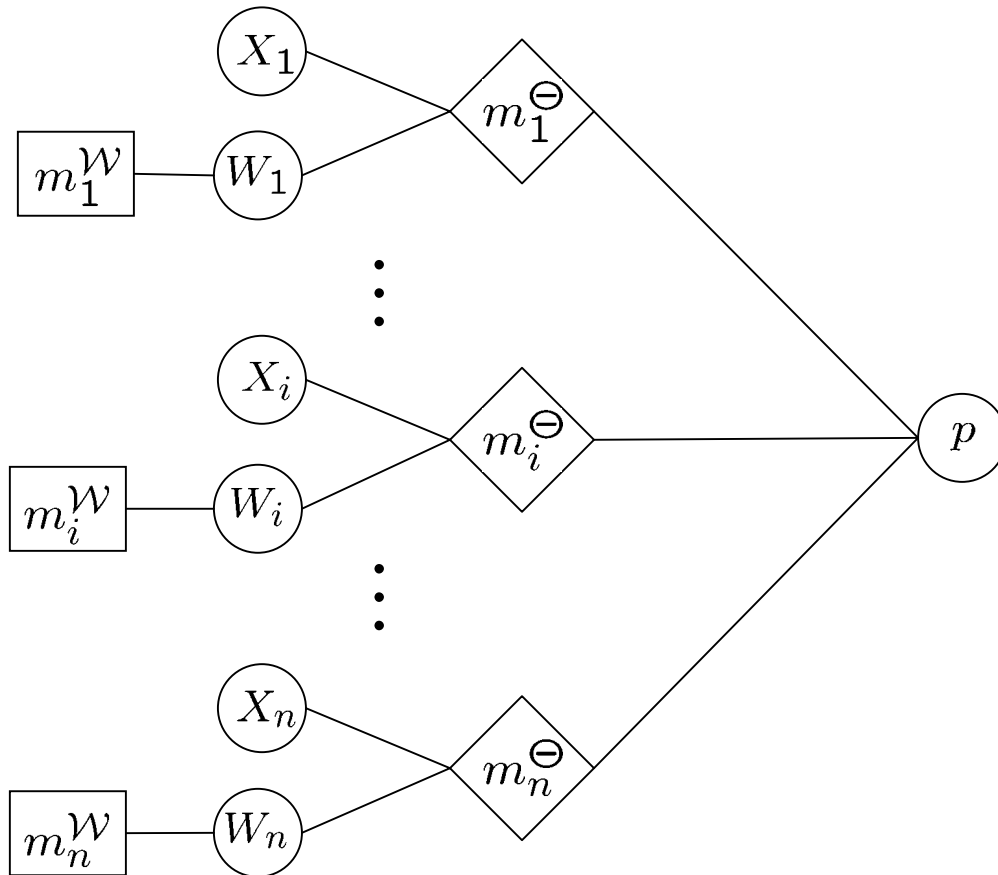
- Soit une urne contenant des boules blanches (ξ_1), rouges (ξ_2) et noires (ξ_3) en proportions p_1 , p_2 et p_3 .
- Soit X =couleur d'une boule tirée au hasard.
- **Fonction de croyance relative à X ?**
- Deux cas :
 - \mathbb{P}_X (composition de l'urne) est connue : $\text{bel}^X = \mathbb{P}_X$
 - \mathbb{P}_X est inconnue, mais on a effectué n tirages avec remise : on dispose d'un échantillon iid X_1, \dots, X_n de X
- Exemple : sur 10 tirages, 5 boules blanches, 3 boules rouges, et 2 boules noires $\rightarrow \text{bel}^X(\{\xi_2\})$?
- Solution triviale : $\text{bel}^X(\{\xi_2\}) = 3/10$, **non satisfaisante car ne tient pas compte de la taille de l'échantillon.**



Solution de Dempster (1966), $K=2$

- Cas $\mathcal{X}=\{0,1\}$, $X \sim \mathcal{B}(p)$, $p \in \mathcal{P}=[0,1]$.
- X_1, \dots, X_n échantillon iid
- Hypothèse : à chaque X_i est une variable $W_i \sim \mathcal{U}_{[0,1]}$, t.q.
$$X_i = 1 \Leftrightarrow W_i \leq p$$
- L'équation précédente définit une contrainte sur les variables X_i , W_i et p , qui peut être représentée par une fonction de croyance m_i^Θ sur $\Theta = \mathcal{X} \times \mathcal{W} \times \mathcal{P}$.
- On observe $X_i = x_i$, $i=1, \dots, n$
- Croyance sur \mathcal{P} :
 - Combinaison des $m_i^{\mathcal{W}}$ (bayésiennes) et des m_i^Θ ,
 - Conditionnement par $X_i = x_i$, $i=1, \dots, n$
 - marginalisation sur \mathcal{P}

Représentation graphique





Solution de Dempster (1966), $K=2$

- Solution (inférence sur p) : soit $N = \sum_{i=1}^n x_i$,

$$\begin{aligned} m^{\mathcal{P}}(a, b) &= \frac{n!}{(N-1)!(n-N-1)!} a^{N-1} (1-b)^{n-N-1} & 0 < N < n \\ m^{\mathcal{P}}(0, b) &= n(1-b)^{n-1} & N = 0 \\ m^{\mathcal{P}}(a, 1) &= na^{n-1} & N = n, \end{aligned}$$

- Prédiction : nouvelle variable X et variable pivotale associée W
- Combinaison de $m^{\mathcal{P}}$ avec m^{Θ} , puis marginalisation sur \mathcal{X}

$$\begin{aligned} m^{\mathcal{X}}(\{1\}) &= \frac{N}{n+1} \\ m^{\mathcal{X}}(\{0\}) &= \frac{n-N}{n+1} \\ m^{\mathcal{X}}(\mathcal{X}) &= \frac{1}{n+1}, \end{aligned}$$



Discussion

- Solution rigoureuse
- Retrouvée par
 - Smets (1994), sans faire appel aux variables pivotales W_i , dans le cadre du TBM
 - Kohlas et Monney (1995) dans la théorie des Hints
- Limitation
 - Difficulté du passage à $K > 2$
 - Cas trinomial ($K=3$) traité partiellement par Dempster (1968) et Smets (1994)
 - $K > 3$??
- Proposition d'une nouvelle approche reposant sur un principe différent



Nouvelle approche (1)

- Problème :
 - Soit $bel^{\mathcal{X}}[\mathbf{x}_n]$ la fonction de croyance sur X après observation d'une réalisation $\mathbf{x}_n = (x_1, \dots, x_n)$ de $\mathbf{X}_n = (X_1, \dots, X_n)$.
 - Quelles propriétés $bel^{\mathcal{X}}[\mathbf{x}_n]$ doit-elle satisfaire relativement à la loi de probabilité \mathbb{P}_X de X ?
- **Principe de Hacking** (1965) : Si les probabilités (fréquences) $\mathbb{P}_X(A)$ sont connues, alors $bel^{\mathcal{X}} = \mathbb{P}_X$.
- Version « faible » du principe de Hacking :

$$\forall A \subseteq \mathcal{X}, \quad bel^{\mathcal{X}}[\mathbf{X}_n](A) \xrightarrow{P} \mathbb{P}_X(A), \quad \text{qd } n \rightarrow \infty,$$

(Propriété P_1)



Nouvelle approche (2)

- Principe d'incertitude maximum (*Least Commitment Principle*) : Pour n fini, $bel^{\mathcal{X}}$ doit être moins informative que \mathbb{P}_X
- La condition $bel^{\mathcal{X}} \leq \mathbb{P}_X$ est trop restrictive :
 - Ex : si on a tiré n_1 boules blanches et n_2 boules noires, les probabilités $\mathbb{P}_X(\{\xi_1\})$ et $\mathbb{P}_X(\{\xi_2\})$ peuvent être arbitrairement petites $\rightarrow bel^{\mathcal{X}}(\mathcal{X})=1$.
- Condition plus faible :

$$\mathbb{P} \left(bel^{\mathcal{X}}[\mathbf{X}_n](A) \leq \mathbb{P}_X(A), \forall A \subset \mathcal{X} \right) \geq 1 - \alpha.$$

(Propriété P_2)



Illustration de la propriété P_2

$$\begin{aligned}x_n = x_1, \dots, x_n &\rightarrow \text{bel}^x[x_n] \\x'_n = x_1, \dots, x_n &\rightarrow \text{bel}^x[x'_n] \\x''_n = x_1, \dots, x_n &\rightarrow \text{bel}^x[x''_n] \\&\vdots\end{aligned}$$

Si le nombre de d'échantillons tend vers l'infini, la proportion de fonctions de croyance prédictives bel^x vérifiant $\text{bel}^x \leq \mathbb{P}_x$ tend vers une valeur au moins égale à $1-\alpha$.

Moyen d'obtenir cette propriété : **régions de confiance sur les paramètres de la loi multinomiale**



Région de confiance multinomiale

- Rappel : X_1, \dots, X_n échantillon iid de v.a. parente $X \in \mathcal{X} = \{\xi_1, \dots, \xi_K\}$.
- Soit N_k = nombre d'occurrences de ξ_i
- $\mathbf{N} = (N_1, \dots, N_K) \sim \mathcal{M}(p_1, \dots, p_K)$ avec $p_k = \mathbb{P}(\{\xi_k\})$.
- $S(\mathbf{N})$ est une **région de confiance de niveau $1-\alpha$** sur $\mathbf{p} = (p_1, \dots, p_K)$ ssi

$$\mathbb{P}(S(\mathbf{N}) \ni \mathbf{p}) \geq 1-\alpha$$

- Cas particulier : intervalles de confiance simultanés :

$$S(\mathbf{N}) = [P_1^-, P_1^+] \times \dots \times [P_K^-, P_K^+],$$

Intervalle de confiance simultanés de Goodman (1965)

$$P_k^- = \frac{b + 2N_k - \sqrt{\Delta_k}}{2(n + b)}$$

$$P_k^+ = \frac{b + 2N_k + \sqrt{\Delta_k}}{2(n + b)},$$

avec

$$\Delta_k = b \left(b + \frac{4N_k(n - N_k)}{n} \right).$$

$$\text{et } b = \chi_{1;1-\alpha/K}^2.$$



Exemple

- 220 patients psychiatriques classés en 4 catégories.
- $N=(91,49,37,43)$
- $1-\alpha=0.95$

Diagnosis	n_k/n	p_k^-	p_k^+
Neurotic	0.41	0.33	0.50
Depressed	0.22	0.16	0.30
Schizophrenic	0.17	0.11	0.24
Personality disorder	0.20	0.14	0.27

Des régions de confiance aux probabilités imprécises

- Une région de confiance $S(N)$ définit
 - Un ensemble de valeurs possibles pour le paramètre $p = (p_1, \dots, p_K)$
 - Une famille de mesures de probabilités
- Cette famille est cohérente au sens de Walley : elle est définie par son enveloppe inférieure

$$P^-(A) = \max \left(\sum_{\xi_k \in A} P_k^-, 1 - \sum_{\xi_k \notin A} P_k^+ \right), \quad \forall A \subseteq \mathcal{X}$$



Propriétés de P^-

- Propriété 1

$$P^-(A) \xrightarrow{P} \mathbb{P}_X(A) \text{ qd } n \rightarrow \infty, \forall A \subseteq \mathcal{X}$$

- Propriété 2

$$\mathbb{P}(P^- \leq \mathbb{P}_X) \geq 1 - \alpha,$$

- P^- est une capacité d'ordre 2 :

$$P^-(A \cup B) \geq P^-(A) + P^-(B) - P^-(A \cap B), \quad \forall A, B \subseteq \mathcal{X}.$$

- P^- n'est pas, en général, une fonction de croyance pour $K > 2$

Des probabilités imprécises aux fonctions de croyance (1)

- $K=2$: P^- est une fonction de croyance

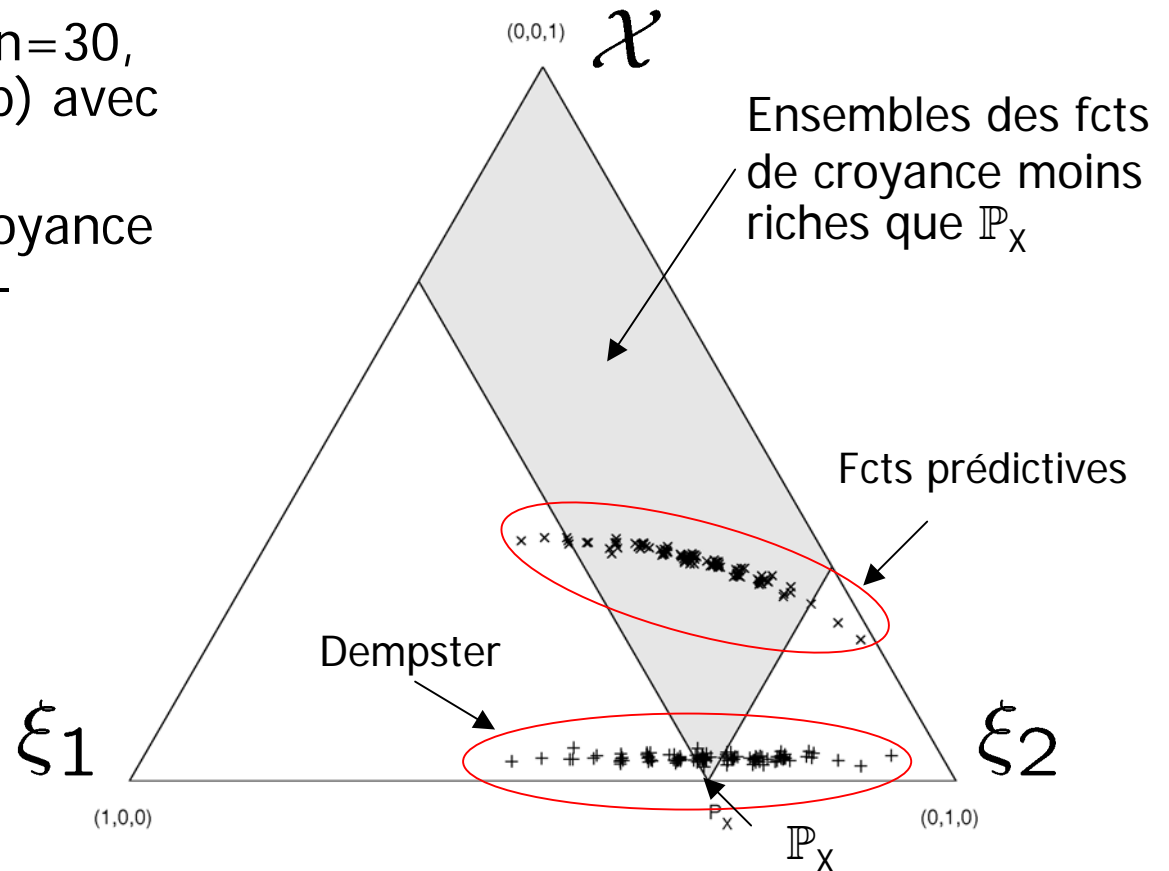
$$\begin{aligned}m^{\mathcal{X}}(\{\xi_1\}) &= P_1^- && \sim \hat{p} - u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\m^{\mathcal{X}}(\{\xi_2\}) &= P_2^- && \sim 1 - \hat{p} - u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\m^{\mathcal{X}}(\mathcal{X}) &= 1 - P_1^- - P_2^- && \sim 2u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\end{aligned}$$

avec $\hat{p} = N_1/n$

- Remarque : $m^{\mathcal{X}}(\mathcal{X})$ tend vers 0 en $n^{-1/2}$, contre n^{-1} dans la solution de Dempster: prix à payer pour la propriété 2 !

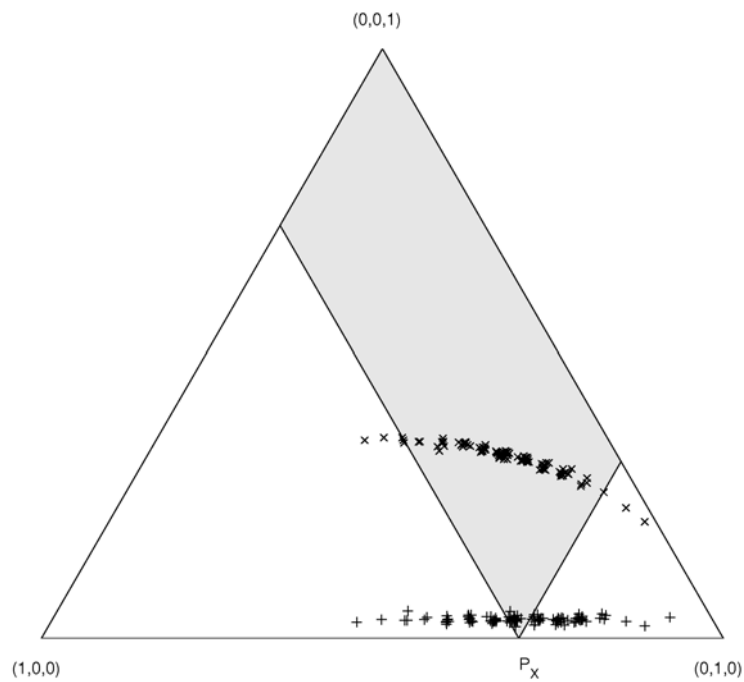
Exemple (1)

- 100 réalisations d'un échantillon iid de taille $n=30$, de v.a. parente $X \sim \mathcal{B}(p)$ avec $p=0.3$
- \rightarrow 100 fonctions de croyance prédictives au niveau $1-\alpha=0.95$

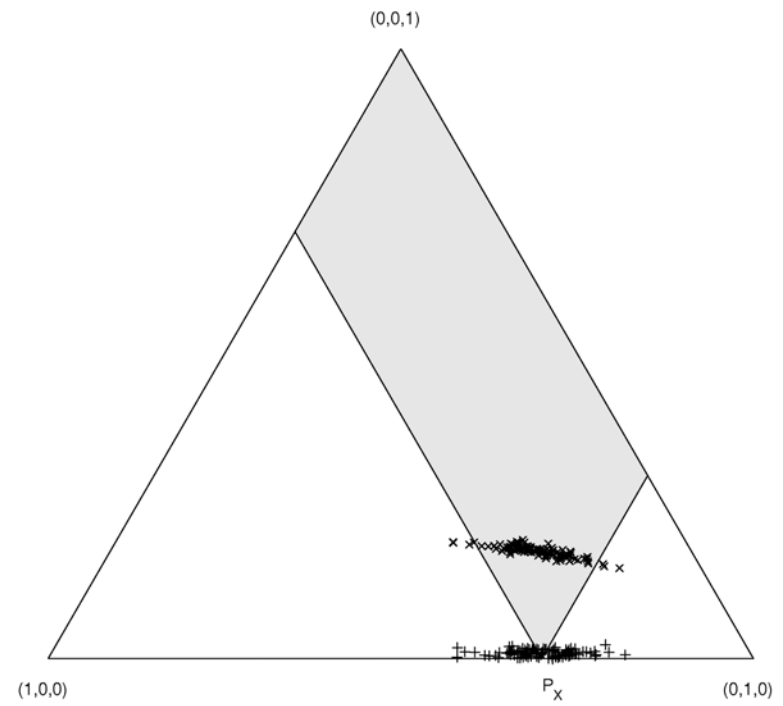


Exemple (2)

n=30



n=100



Des probabilités imprécises aux fonctions de croyance (2)

- $K > 2$: P^- n'est pas, en général, une fonction de croyance.
- Recherche de la fonction de croyance la plus riche vérifiant $bel^x \leq P^-$. On a bien alors

$$\mathbb{P}(bel^x \leq \mathbb{P}_X) \geq \mathbb{P}(P^- \leq \mathbb{P}_X) \geq 1 - \alpha.$$

- Méthode :
 - Définition d'un critère d'informativité
 - Maximisation du critère sous la contrainte $bel^x \leq P^-$



Problème d'optimisation

$$\max_{m^{\mathcal{X}}} \sum_{A \subseteq \mathcal{X}} \text{bel}^{\mathcal{X}}(A) = 2^K \sum_{B \subseteq \mathcal{X}} 2^{-|B|} m^{\mathcal{X}}(B)$$

sous les contraintes :

$$\sum_{B \subseteq A} m^{\mathcal{X}}(B) \leq P^-(A), \quad \forall A \subseteq \mathcal{X},$$

$$\sum_{A \subseteq \mathcal{X}} m^{\mathcal{X}}(A) = 1,$$

$$m^{\mathcal{X}}(A) \geq 0, \quad \forall A \subseteq \mathcal{X}.$$

Exemple : données psychiatriques

A	$P^-(A)$	$bel^{\mathcal{X}^*}(A)$	$m^{\mathcal{X}^*}(A)$
$\{\xi_1\}$	0.33	0.33	0.33
$\{\xi_2\}$	0.16	0.14	0.14
$\{\xi_1, \xi_2\}$	0.50	0.50	0.021
$\{\xi_3\}$	0.11	0.097	0.097
$\{\xi_1, \xi_3\}$	0.45	0.45	0.020
$\{\xi_2, \xi_3\}$	0.28	0.28	0.036
$\{\xi_1, \xi_2, \xi_3\}$	0.73	0.69	0.040
$\{\xi_4\}$	0.14	0.12	0.12
$\{\xi_1, \xi_4\}$	0.47	0.47	0.02
$\{\xi_2, \xi_4\}$	0.30	0.30	0.035
$\{\xi_1, \xi_2, \xi_4\}$	0.76	0.72	0.045
$\{\xi_3, \xi_4\}$	0.25	0.25	0.035
$\{\xi_1, \xi_3, \xi_4\}$	0.70	0.66	0.038
$\{\xi_2, \xi_3, \xi_4\}$	0.50	0.48	0.019
\mathcal{X}	1	1	0



Remarques

- Soit $bel^{\mathcal{X}^*}$ une solution du programme linéaire précédent. On a bien

$$bel_n^{\mathcal{X}^*} \xrightarrow{P} \mathbb{P}_{\mathcal{X}} \text{ qd } n \rightarrow \infty.$$

- Cas où K est très grand : **approximations**
 - Algorithm IRM (Hall et Lawry, 2004)
 - Solution triviale :

$$m_n^{\mathcal{X}^\circ}(\{\xi_k\}) = P_k^-, \quad k = 1, \dots, K$$

$$m_n^{\mathcal{X}^\circ}(\mathcal{X}) = 1 - \sum_{k=1}^K P_k^- ,$$



Cas où \mathcal{X} est ordonné

- On suppose que \mathcal{X} est ordonné : $\xi_1 < \dots < \xi_K$
- Dans ce cas, on peut restreindre les éléments focaux de $m^{\mathcal{X}}$ aux intervalles $A_{k,r} = \{\xi_k, \dots, \xi_r\}$.
- Sous ces contraintes supplémentaires, la solution du problème d'optimisation précédent est :

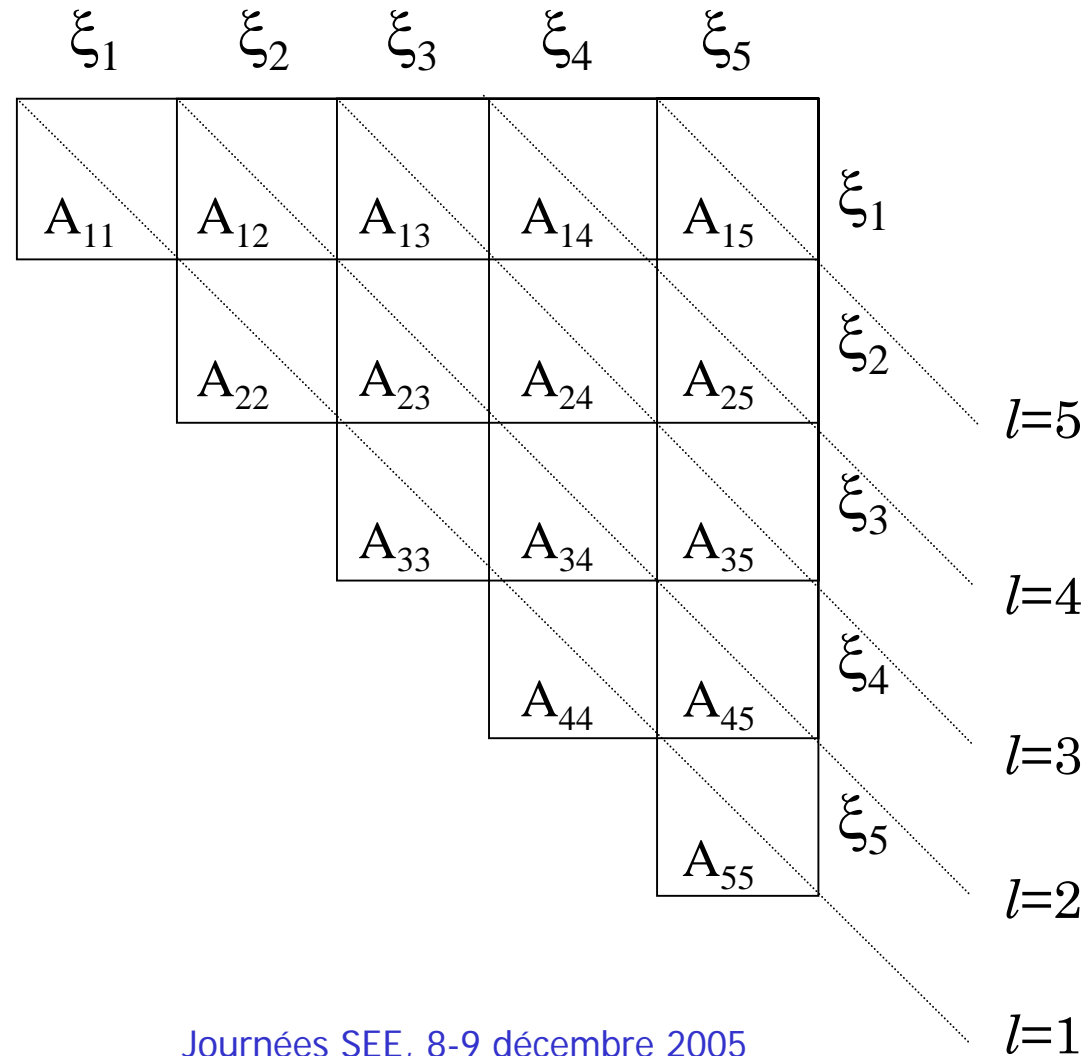
$$m^{\mathcal{X}^*}(A_{k,k}) = P_k^-,$$

$$m^{\mathcal{X}^*}(A_{k,k+1}) = P^-(A_{k,k+1}) \\ - P^-(A_{k+1,k+1}) - P^-(A_{k,k}),$$

$$m^{\mathcal{X}^*}(A_{k,r}) = P^-(A_{k,r}) - P^-(A_{k+1,r}) \\ - P^-(A_{k,r-1}) + P^-(A_{k+1,r-1}) \text{ pour } r > k + 1$$



Représentation





Exemple (1)

- Précipitations mensuelles en janvier en Arizona, période 1895-2004 (en pouces)

classe ξ_k	n_k	n_k/n	p_k^-	p_k^+
< 0.75	48	0.44	0.32	0.56
$[0.75, 1.25)$	17	0.15	0.085	0.27
$[1.25, 1.75)$	19	0.17	0.098	0.29
$[1.75, 2.25)$	11	0.10	0.047	0.20
$[2.25, 2.75)$	6	0.055	0.020	0.14
≥ 2.75	9	0.082	0.035	0.18

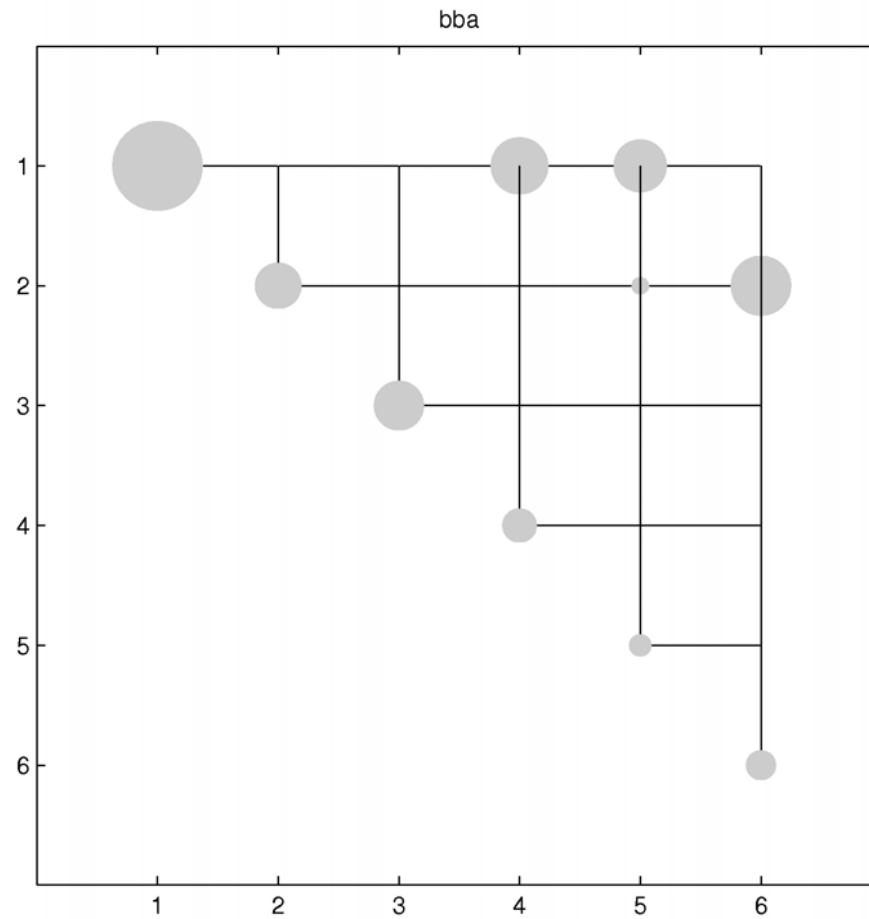
- Degré de croyance dans l'hypothèse selon laquelle le volume de précipitations en Arizona en janvier prochain excédera 2.25 pouces ?



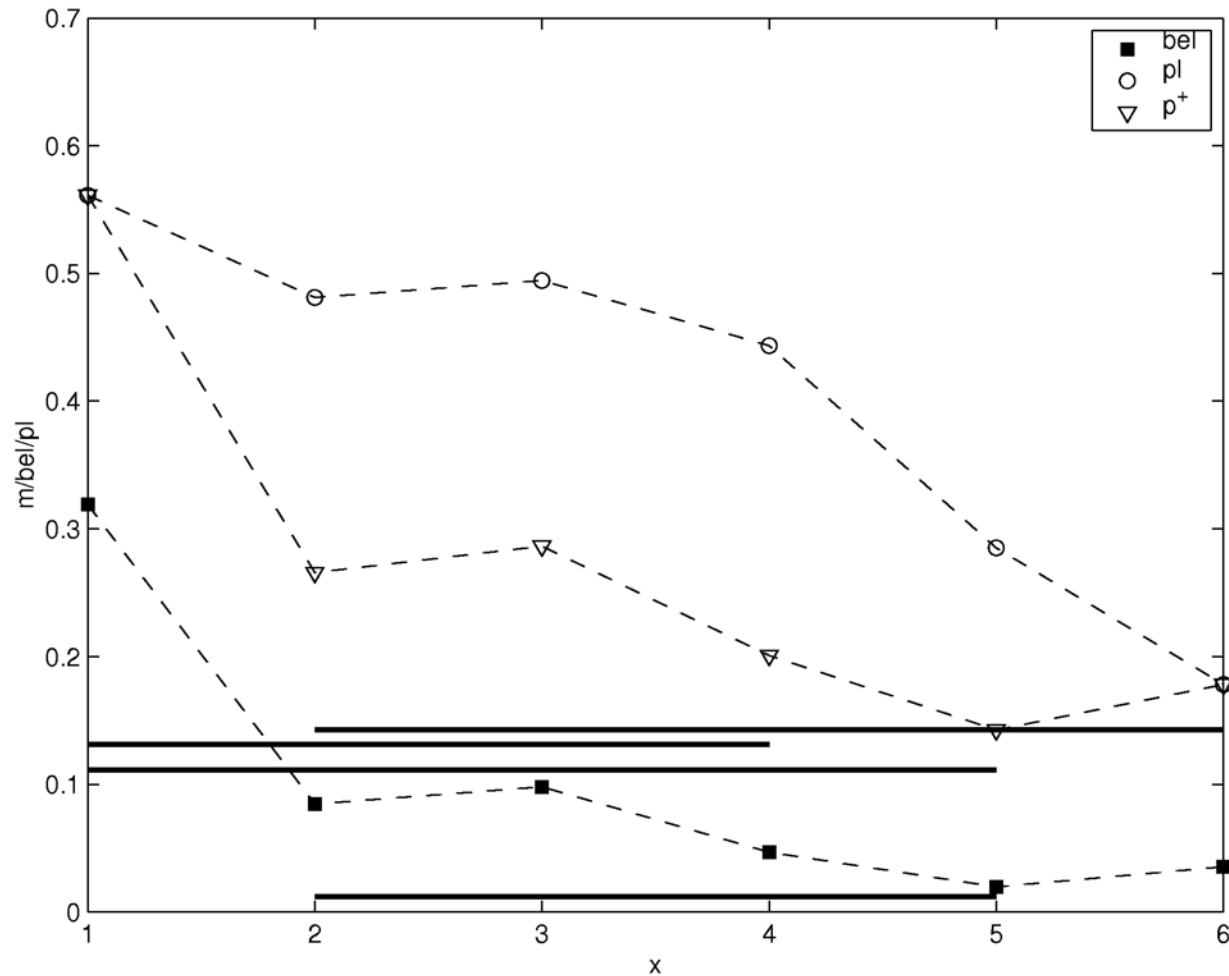
Exemple (2)

$m(A_{k,r})$	1	2	3	4	5	6
1	0.32	0	0	0.13	0.11	0
2	-	0.085	0	0	0.012	0.14
3	-	-	0.098	0	0	0
4	-	-	-	0.047	0	0
5	-	-	-	-	0.020	0
6	-	-	-	-	-	0.035

Exemple (3)



Exemple (4)





Exemple (5)

- Réponse à la question initiale ($1-\alpha=0.95$) :

$$\begin{aligned} \text{bel}(X \geq 2.25) &= \text{bel}^{\mathcal{X}^*}(\{\xi_5, \xi_6\}) \\ &= m^{\mathcal{X}^*}(\{\xi_5\}) + m^{\mathcal{X}^*}(\{\xi_6\}) + m^{\mathcal{X}^*}(\{\xi_5, \xi_6\}) \\ &= 0.020 + 0.035 + 0 = 0.055, \end{aligned}$$

$$\begin{aligned} \text{pl}(X \geq 2.25) &= \text{pl}^{\mathcal{X}^*}(\{\xi_5, \xi_6\}) \\ &= 0.020 + 0.035 + 0.012 + 0.14 + 0.11 \\ &= 0.317 \end{aligned}$$

- Propriété : dans 95 % des cas (pour 95 % des échantillons), l'intervalle $[\text{bel}^{\mathcal{X}}(A), \text{pl}^{\mathcal{X}}(A)]$ encadre la vraie fréquence de réalisation $\mathbb{P}_{\mathcal{X}}(A)$ de A.



Conclusion

- Une méthode facilement implémentable, permettant de **construire une fonction de croyance relative à une variable aléatoire discrète X , sur la base d'une réalisation d'un échantillon iid de X .**
- La solution possède des **propriétés bien définies relativement à la loi de probabilité inconnue de X** :
 - Convergence quand $n \rightarrow \infty$ (principe de Hacking)
 - Dominance avec une probabilité $\geq 1-\alpha$
- Expression analytique dans des cas simples
 - $K=2, K=3$
 - Restriction à des intervalles (cas où \mathcal{X} est ordonné)
- Extension en cours au cas où X est une variable continue.