Notions et mesures de cohérence bayésienne entre informations *a priori* et données de retour d'expérience

Nicolas Bousquet

Séminaire FIMA,
16 Mars 2006.

Introduction

Un exemple introductif

Contexte : données de durée de vie d'un composant $\sum$ de centrale nucléaire supposées suivre un modèle $\mathcal{M}(\theta)$.

temps réels : 134.9, 152.1, 133.7, 114.8, 110, 129, 78.7, 72.8, 132.2, 91.8
temps censurés (à droite) : 70, 159.5, 98.5, 167.2, 66.8, 95.3, 80.9, 83.2

On cherche à estimer θ avec une précision ordonnée par un organisme de contrôle (Agence de sûreté nucléaire).

Méthodes fréquentistes classiques (Newton-Raphson, EM, etc.) insuffisantes à estimer θ avec assez de précision.

Ex : modèle de Weibull $\mathcal{W}(\eta, \beta)$

$$\hat{\eta}_n = 140.8 \ +/-7.3,$$
$$\hat{\beta}_n = 4.51 \ +/-1.9$$

Appel au **bayésien subjectif** : experts EDF et Westinghouse.

Suite aux premières défaillances, les matériaux de $\sum$ ont été améliorés.

L'expertise prend en compte une certaine évolution technique, de façon difficilement ou non quantifiable $\Rightarrow$ valeurs typiques > 200.

Problème : l'expertise est en fort décalage avec les données.

Une façon objective de détecter ce décalage ?

Problème général & Motivations

On note $X \sim \mathcal{M}(\theta)$, $\theta \in \Theta \subset \mathbb{R}^d$) la variable aléatoire d'intérêt (durée de vie, etc.). Soit $p(x|\theta)$ sa densité.

θ est supposé aléatoire.

On connaît une source d'information subjective sur $\sum$, X, θ ....
... informations que l'on transforme en une loi *a priori* de densité $\pi(\theta)$.

La détection d'un conflit entre les domaines de confiance pour θ apportés

- par la vraisemblance $\mathcal{L}(\mathbf{X_n}; \theta)$ des données $\mathbf{X_n} = (X_1, \ldots, X_n)$,
- par la densité *a priori* $\pi(\theta)$

est un problème important pour l'emploi **concret** du bayésien subjectif.

D'un point de vue **théorique**, aucune raison de remettre en question le modèle {vraisemblance + *a priori*} (cohérence interne).

Dans l'industrie, bayésien employé comme palliatif au fréquentiste : on ne remet pas en cause la vraisemblance.

D'un point de vue pratique, la connaissance d'un tel conflit $\Rightarrow$

1. décalage temporel, technique ... *symétrique* entre expertise et données ;
2. remise en cause de π si les données sont bonnes ;
3. détection de données "polluant" la connaissance (si consensus d'expert fort) ;
4. mise en garde sur le résultat *a posteriori*.

Travaux précédents

1. De Finetti (1961), Dawid (1973), Hill (1974) : méthodes automatiques de rejet de données atypiques ($\rightarrow \infty$).
2. O'Hagan (1979, 1988, 1990, 2003) : comparaisons des queues des fonctions $\theta \rightarrow \pi(\theta)$ et $\theta \rightarrow \mathcal{L}(\mathbf{X_n}; \theta)$ (*credence orders*).
3. West (1984) et Lucas (1990) : même idée de comparaison de *scores* liés à la "lourdeur" des fonctions de θ.

   En général, les outils développés servent à évaluer la robustesse de l'a posteriori vis-à-vis de variations a priori.

4. Usureau & Dumon (2001) : une méthode "ingénieur" (produit de convolution entre $\mathcal{L}(\mathbf{X_n}; \theta)$ et $\pi(\theta)$) sans signification statistique.

Objectif

Élaborer une statistique simple $\mathcal{C}oh(\pi|\mathbf{X_n})$ mesurant la cohérence entre *a priori* π et données $\mathbf{X_n}$, avec la règle du type

$$\mathcal{C}oh(\pi|\mathbf{X_n}) \leq 1 \quad \Leftrightarrow \quad \text{cohérence.}$$

Construction d'une règle de cohérence

On considère qu'une connaissance minimale , produite à partir des données, est modélisée par π^{MIA}.

Idéalement, π^{MIA} est non informative.

$$\text{Alors} \quad \pi^{MIA}(\theta|\mathbf{X_n}) \quad = \quad \frac{\pi^{MIA}(\theta)\mathcal{L}(\mathbf{X_n};\theta)}{\int_\Theta \pi^{MIA}(\theta)\mathcal{L}(\mathbf{X_n};\theta)}$$

"mime" la vraisemblance et transmet approximativement la même information que les données.

Hypothèse : $\pi^{MIA}(\theta|\mathbf{X_n})$ modélise l'avis d'un expert "parfaitement" en accord avec les données.

Choix d'*a priori* non informatifs $\pi^{MIA} = \pi^J$

π^J doit être tel que $\pi^J(.|\mathbf{X_n})$ a une certaine validité fréquentiste.

Soit $\tilde{\theta}_n(\alpha)$ le quantile d'ordre α *a posteriori* sur n observations

$$P_\theta\left(\theta \leq \tilde{\theta}_n(\alpha)\right) = \underbrace{P\left(\theta \leq \tilde{\theta}_n(\alpha) \,|\, \mathbf{X_n}\right)}_{\alpha} + O\left(n^{-i/2}\right)$$

$\Rightarrow$ *coverage matching priors* [Peers, Mukerjee, Datta, Ghosh (1993, 1995, 1996)], solutions d'une équation différentielle.

$$\begin{array}{lcl} Dim\Theta = 1 & \Leftrightarrow & \text{Jeffreys} \\ Dim\Theta = 2 & \Rightarrow & \text{Reference prior (Berger-Bernardo, 1992)} \end{array}$$

L'éventuelle incohérence de π avec $\mathbf{X_n}$ s'exprime alors par la mesure d'une distance $D\{\pi^{MIA}(.|\mathbf{X_n}) \mid\mid \pi\}$.

Plus cette distance est grande, moins il y a cohérence.

La distance maximale de cohérence est atteinte quand $\pi = \pi^{MIA}$.

Définition

On définit la statistique de cohérence

$$\mathcal{C}oh(\pi|\mathbf{X_n}) = \frac{D\{\pi^{MIA}(.|\mathbf{X_n}) \mid\mid \pi\}}{D\{\pi^{MIA}(.|\mathbf{X_n}) \mid\mid \pi^{MIA}\}}$$

et la règle

$$\mathcal{C}oh(\pi|\mathbf{X_n}) \leq 1 \quad \Leftrightarrow \quad \text{cohérence.}$$

Choix d'une distance $D(\pi_1||\pi_2)$

Une distance en terme de quantité d'information portée par π_1 et π_2 (classe de Ali-Silvey).

Kullback-Leibler : $D(\pi_1||\pi_2) = KL(\pi_1||\pi_2) = \int_\Theta \pi_1(\theta) \log \frac{\pi_1(\theta)}{\pi_2(\theta)} \, d\theta.$

1. la plus naturelle (information de Shannon) ;
2. reflète un *regret* informatif dû au choix de π_2 lorsque le meilleur est π_1 ;
3. propriétés analytiques supérieures aux autres (Chernoff) ;
4. rend $\mathcal{C}oh$ indépendante du choix de paramétrisation θ.

Élicitation de π^{MIA}

Difficultés

π^{MIA} représente la connaissance minimale que l'on peut placer dans un *a priori*, cohérent avec les données.

Idéalement, $\pi^{MIA} = \pi^J$ où $\pi^J(\theta) = \sqrt{\det I(\theta)} =$ **Jeffreys** (indépendance de paramétrisation), où $I(\theta)$ est la matrice de Fisher de $\mathcal{M}(\theta)$.

Excepté pour des modèles $\mathcal{M}(\theta)$ **discrets**, π^J est toujours impropre

$$\int_\Theta \pi^J(\theta)\, d\theta \quad = \quad \infty.$$

et le dénominateur de $\mathcal{C}oh$ est alors défini à une constante près(linéairement).

Exemple : Modèle multinomial avec *a priori* de Dirichlet.

Soit $X \sim \mathcal{M}(m, p_1, \ldots, p_d)$ avec $m \in \mathbb{N}$, $p_i > 0\ \forall i \in \{1, \ldots, d\}$ et $\sum_{i=1}^d p_i = 1$. Soient $(m_1, \ldots, m_d) \in [0, m]^d$ les données observées avec $\sum_{i=1}^d m_i = m$. Alors $P\{X = (m_1, \ldots, m_d)\} = m! \prod_{i=1}^d p_i^{m_i} / \prod_{i=1}^d m_i!$.

Alors Jeffreys $= \pi^J(p_1, \ldots, p_d) \propto 1/\prod_{i=1}^d p_i$. Celui-ci est proprement défini (Dirichlet) :

$$\pi^J(p_1, \ldots, p_d) = \mathcal{D}ir(1/2, \ldots, 1/2).$$

En conséquence
$\pi^J(m, p_1, \ldots, p_d | m_1, \ldots, m_d) = \mathcal{D}ir(m_1 + 1/2, \ldots, m_d + 1/2)$.

Supposons avoir choisi l'*a priori* $\pi(p_1, \ldots, p_d)$ comme une densité de la loi de Dirichlet $\mathcal{D}ir(a_1, \ldots, a_d)$.

Alors

$$Coh(\pi|m_1,\ldots,m_d) =$$

$$\frac{\log \frac{\Gamma(\alpha)}{\Gamma\left(\sum_{j=1}^{d} a_j\right)} + \sum_{j=1}^{d} \log \frac{\Gamma(a_j)}{\Gamma(\alpha_j)} + \sum_{j=1}^{d} [\alpha_j - a_j]\left[\Psi(\alpha_j) - \Psi(a_j)\right]}{\log \frac{\Gamma(\alpha)}{\Gamma(d/2)} + d\sqrt{\pi}/2 - \sum_{i=1}^{d} \log\Gamma(\alpha_i) + \sum_{i=1}^{d} m_i\Psi(\alpha_i) + m(\gamma + 2\log 2)}$$

avec Ψ la fonction digamma, γ la constante d'Euler ($\simeq 0.577215$), $\alpha_i = 1/2 + m_i$ et $\alpha = \sum_{i=1}^{d} \alpha_i$.

Démarche à suivre

Il faut donc des règles pour trouver π^{MIA} propre, élicité d'après les données (ajustement de π^J).

1. On construit de façon **objective** un certain nombre de candidats MIA formant l'ensemble $\Pi^{MIA} = \{\pi_i^{MIA},\ i \in I\}$.
    - maximum d'entropie sous contraintes (Jaynes) ;
    - *a priori* localement uniformes sur un domaine de confiance de la vraisemblance (Box & Tiao) ;
    - *Maximum Likelihood Priors* (Hartigan)
    - *Posterior priors* (Berger & Pérez) ...
2. On sélectionne le moins informatif de tous.

Élicitation de *Posterior priors*

Les *a priori* élicités sont *objectifs* et ne dépendent que des données.

Les *a priori* élicités sont toujours propres.

Pas de choix de contrainte(s) à faire (maximum d'entropie).

Notion de MTS (1)

Posterior prior : construction liée à la notion d'*échantillon d'apprentissage minimal* (MTS) [Berger & Perrichi, 1996, 2002].

Soit π^J un *a priori* non informatif.

Un MTS $X(l)$ est une quantité minimale de données dans $\mathbf{X_n} = (X_1, \ldots, X_n)$ telle que $\pi^J(\theta|X(l))$ est propre.

Un MTS est donc constitué du nombre de données minimal pour lequel θ est identifiable .

Notion de MTS (2)

Originellement, les MTS sont introduits dans le contexte de la **sélection de modèles**.

Soient 2 modèles $p_1(x|\theta_1)$ et $p_2(x|\theta_2)$. On cherche à calculer le facteur de Bayes

$$B_{ji}^J = \frac{m_j(x)}{m_i(x)} = \frac{\int p_j(x|\theta_j)\pi_j^J(\theta_j)\ d\theta_j}{\int p_i(x|\theta_i)\pi_i^J(\theta_i)\ d\theta_i}.$$

où π_i^J et π_j^J sont les plus neutres possibles (*default priors*).

But : améliorer les tests fréquentistes (rapport de vraisemblance, utilisation des p-values, pondération de la dimension, etc.)

Problème : $(\pi_j^J,\ \pi_i^J)$ et donc B_{ji}^J définis à une constante près.

Notion de MTS (3)

Soit $X(l)$ un MTS et $X(-l)$ le reste de l'échantillon.

Berger et Perrichi (1996) : $B_{ji}^{AI} = B_{ji}^{J} \sum_{k=1}^{L} B_{ij}^{J}(X(l_k))$ où

$$B_{ij}^{J}(X(l)) = \frac{\int p_i(X(-l)|\theta_i)\pi_i^J(\theta_i|X(l))\ d\theta_i}{\int p_j(X(-l)|\theta_j)\pi_j^J(\theta_j|X(l))\ d\theta_j}$$

est asymptotiquement équivalent à un facteur de Bayes "propre"

A priori intrinsèques propres π_i et π_j modélisant raisonnablement l'ignorance.

Généralisation des *(expected) Posterior priors*

Pérez et Berger (1998, 2002)

Candidats MIA : on fait le choix

1. d'un *a priori* non informatif π^J (*coverage matching prior*),
2. d'une mesure g_i pour les MTS $X(l)$ (par exemple prédictive, empirique, etc.).

On élicite le candidat π_i^{MIA} comme

$$\pi_i^{MIA}(\theta) \quad \propto \quad \int \pi^J(\theta|X(l))\; g_i(X(l))\; dX(l).$$

Cohérence arithmétique

Définition

$$\mathcal{C}oh_A(\pi|\mathbf{X_n}) = \frac{1}{L}\sum_{i=1}^{L}\frac{KL\left\{\pi^J\left(.|X(I_i),\mathbf{X_n}\right)\ ||\ \pi\right\}}{KL\left\{\pi^J\left(.|X(I_i),\mathbf{X_n}\right)\ ||\ \pi^J\left(.|X(I_i)\right)\right\}}.$$

Une version "Leave-one-MTS-out" :

$$\mathcal{C}oh_A(\pi|\mathbf{X_n}) = \frac{1}{L}\sum_{i=1}^{L}\frac{KL\left\{\pi^J\left(.|\mathbf{X_n}\right)\ ||\ \pi\right\}}{KL\left\{\pi^J\left(.|\mathbf{X_n}\right)\ ||\ \pi^J\left(.|X(I_i)\right)\right\}}.$$

Si π^J existe, permet de pallier l'absence de candidats MIA simples à éliciter, explicites, etc.

Règles de sélection

Candidats MIA obtenus : π_i^{MIA}, $i \in I$.
On veut sélectionner le “moins informatif”.

Règle 1 : **Regret prédictif** *a posteriori* .

$$\mathcal{R}_n(i) \quad = \quad \int KL\left\{\pi^J(.|\mathbf{Y_n}) \,||\, \pi_i^{MIA}(.|\mathbf{Y_n})\right\} \; m_i(\mathbf{Y_n}) \; d\mathbf{Y_n}$$

avec $m_1(x) = \int_\Theta p(x|\theta)\pi_i^{MIA}(\theta)$.

Normalité asymptotique (Clarke, 1999) : $\mathcal{R}_n(i) \xrightarrow{n\to\infty} 0$.

On sélectionne le candidat avec la meilleure convergence vers 0 :

$$\pi^{MIA} \quad = \quad \arg\min_{i\in I} \lim_{n\to\infty} \sum_{j\in I, j\neq i} \frac{\mathcal{R}_n(i)}{\mathcal{R}_n(j)}. \qquad (C_1)$$

Règle 2 : **Gain prédictif en information** .

$$\mathcal{K}_n(i) = \int_{S^n} KL\left\{\pi_i^{MIA}(.|\mathbf{Y_n}) \,||\, \pi_i^{MIA}\right\} \; m_i(\mathbf{Y_n}) \, d\mathbf{Y_n}.$$

La distance de Kullback entre *a priori* et *a posteriori* = information transmise par les données (Règle de Bernardo,1979).

On cherche la divergence la plus rapide quand $n \longrightarrow \infty$.

$$\pi^{MIA} = \arg\min_{i \in I} \lim_{n \to \infty} \sum_{j \in I, j \neq i} \frac{\mathcal{K}_n(j)}{\mathcal{K}_n(i)}. \qquad (C_2)$$

Une idée de test de sélection (critère moins général)

1. Soit π une densité *a priori* définie sur Θ.
2. Soit des simulations marginales $\mathbf{Y_n} = (Y_1, \ldots, Y_n)$ ($\Leftrightarrow \pi$ cohérent avec $\mathbf{Y_n}$).
3. Pour tout candidat MIA π_i^{MIA}, soit β_i la puissance du test

$$\mathcal{C}oh_i(\pi|\mathbf{Y_n}) - 1 \leq 0.$$

4. On sélectionne $\pi^{MIA} = \arg\max_{i \in I} \beta_i$.

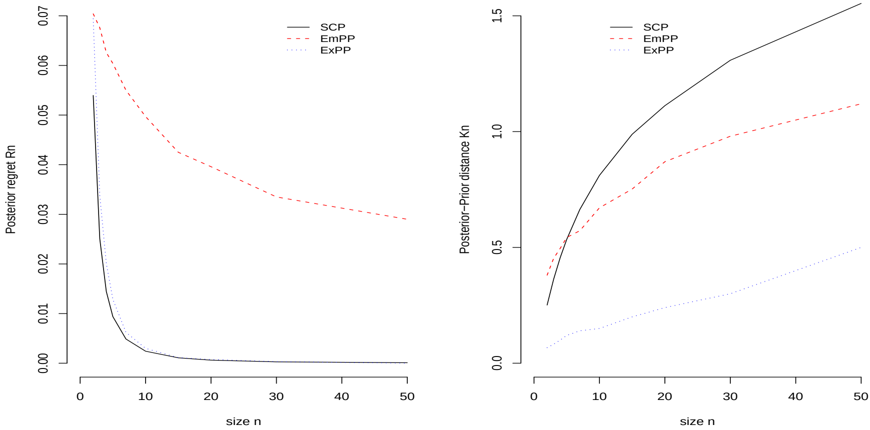
Exemples des modèles exponentiel et de Weibull

Modèle exponentiel

$\mathbf{X_n} = (X_1, \ldots, X_n)$ de densité $p(x|\theta) = \theta \exp(-\theta x)$.
$\pi^J(\theta) \propto \theta^{-1}$ (Jeffreys).
1 MTS est composé d'une seule donnée (non censurée)

Candidats MIA :

1. *Conjugué exhaustif (SCP)* : $\pi_1^{MIA}(\theta) = \pi^J(\theta | \sum_{i=1}^n X_i/n)$.
   Propriété : $\mathcal{R}_n(1) \sim (3n)^{-1}$.
2. *Posterior Prior empirique (EmPP)* : $\pi_2^{MIA}(\theta) = \frac{1}{L} \sum_{i=1}^{L} \pi^J(\theta | X_i)$.
3. *Expected Posterior Prior (ExPP)* : $\pi_3^{MIA}(\theta) = \frac{\hat{\theta}_n}{(\hat{\theta}_n + \theta)^2}$.



$\mathcal{R}_n(i)$ (gauche) et $\mathcal{K}_n(i)$ $\Rightarrow$ sélectionnent SCP (*a priori* conjugué)

Cohérence vis-à-vis d'un échantillon simulé

$$\mathbf{X_n} = (142, 143, 470, 419, 185, 84, 8, 27, 573, 17) \sim \mathcal{E}(\theta_0 = 1/150).$$

MLE $\hat{\theta}_n = 1/207 < \theta_0$.

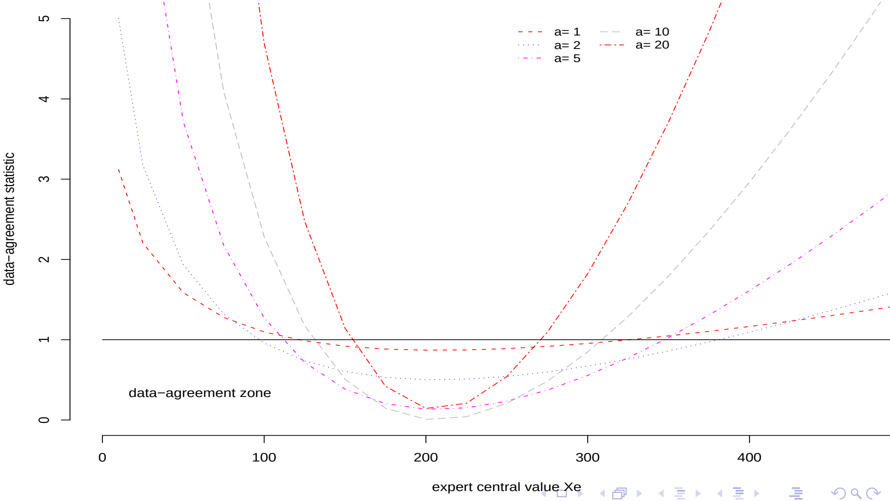
On désire évaluer la cohérence de $\pi(\theta) = \mathcal{G}(a, aX_e)$ (conjugué)

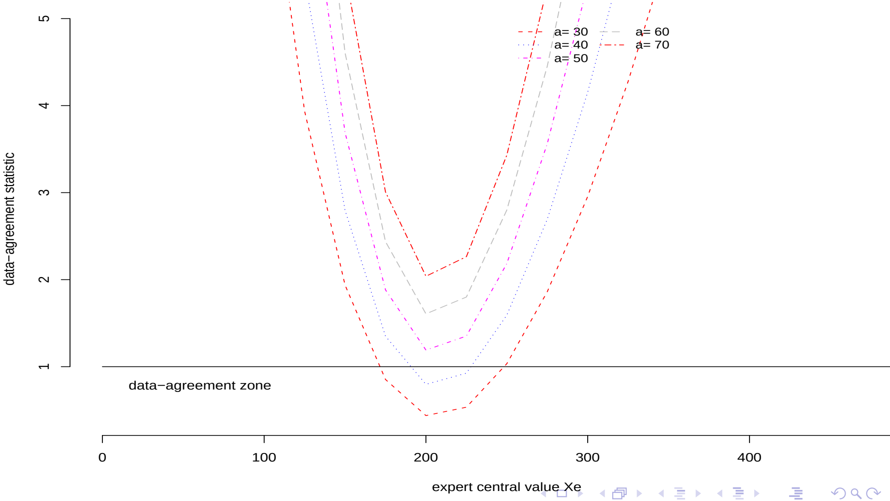
1. $X_e =$ avis central d'expert $(= 1/E[\theta])$
2. $a \in \mathbb{N}^*$ correspond à une taille d'échantillon *fictif* $\tilde{\mathbf{X}}_\mathbf{a}$ de moyenne X_e.

$$\text{En effet} \quad \pi(\theta) \quad \propto \quad \pi^J(\theta)\mathcal{L}(\tilde{\mathbf{X}}_\mathbf{a}, \theta)$$

Le "meilleur" *a priori* possible est en $X_e = \hat{\theta}_n^{-1} = 207$, pour $a = 10$.

A priori π		SCP	EmPP	ExPP	$Coh_A(\pi; \mathbf{X_n})$
$a = 5$	$X_e = 10$	12.97	8.750	9.823	8.805
	$X_e = 150$	0.444	0.364	0.235	0.316
	$X_e = 207$	0.156	0.097	0.015	0.118
	$X_e = 300$	0.638	0.282	0.402	0.437
	$X_e = 500$	3.444	1.827	2.543	2.323
$a = 10$	$X_e = 10$	25.63	17.24	17.95	17.40
	$X_e = 150$	0.580	0.560	0.402	0.424
	$X_e = 207$	0	0.021	0.004	0.003
	$X_e = 300$	0.966	0.386	0.679	0.665
	$X_e = 500$	6.580	3.481	4.645	4.436


a= 1
a= 2
a= 5
a= 10
a= 20
data−agreement statistic
data−agreement zone
expert central value Xe
0
1
2
3
4
5
0
100
200
300
400


data−agreement statistic
5
4
3
2
1
a= 30
a= 40
a= 50
a= 60
a= 70
data−agreement zone
0
100
200
300
400
expert central value Xe

Modèle de Weibull

$\mathbf{X_n} = (X_1, \ldots, X_n)$ de densité $p(x|\theta) = \beta\eta^{-\beta}x^{\beta-1}\exp(-\eta^{-\beta}x^{\beta})$.

Reference prior (Sun, 1997) : $\pi^J(\eta, \beta) \propto (\eta\beta)^{-1}$.

Un MTS $= (X_i, X_j) > 1$, $(X_i \neq X_i)$.

Le candidat EmPP est explicite.

On peut se faciliter la tâche en reparamétrisant :

$$\left\{ \begin{array}{lll} \eta & \rightarrow & \mu = \eta^{-\beta}, \\ \beta & \rightarrow & \beta \end{array} \right.$$

On obtient alors

$$\pi_1^{MIA}(\mu, \beta) = \frac{1}{L_1 L_2} \sum_{i=1}^{L_1} \sum_{i=1}^{L_2} \pi^{ij}(\mu|\beta)\ \pi^{ij}(\beta),$$

$$\begin{aligned} \text{avec} \quad \pi^{ij}(\mu|\beta) &= \mathcal{G}\left(2, X_i^\beta + X_j^\beta\right), \\ \pi^{ij}(\beta) &= \frac{(X_i X_j)^{\beta-2}}{2|\log X_i/X_j| \left(X_i^\beta + X_j^\beta\right)^2}. \end{aligned}$$

Facilite les calculs *a posteriori*.

On reprend les données EDF (âge de remplacement de matériels équipant le circuit secondaire)

temps réels : 134.9, 152.1, 133.7, 114.8, 110.0, 129.0, 78.7, 72.8, 132.2, 91.8
temps censurés (à droite) : 70.0, 159.5, 98.5, 167.2, 66.8, 95.3, 80.9, 83.2

MLE : $(\hat{\eta}_n, \hat{\beta}_n) = (140.8, 4.51)$ avec $\hat{\sigma}_n = (7.3, 1.9)$.

Expertises	intervalle à (5%,95%)	valeur médiane
EDF	(200,300)	250
Westinghouse	(100,500)	250

Idée : 4 *a priori* proches et indépendants pour η et β

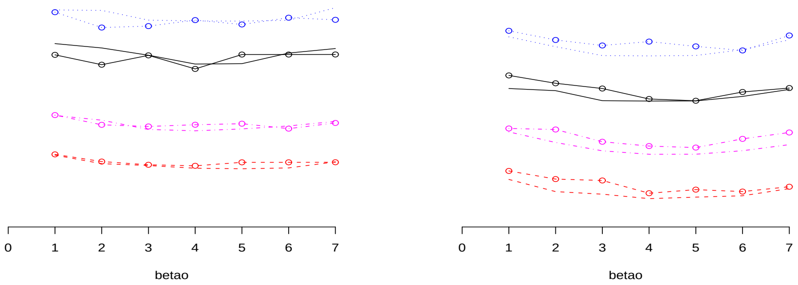
On fait varier $\pi(\beta)$ principalement dans $[1,5]$ avec σ_β variable.

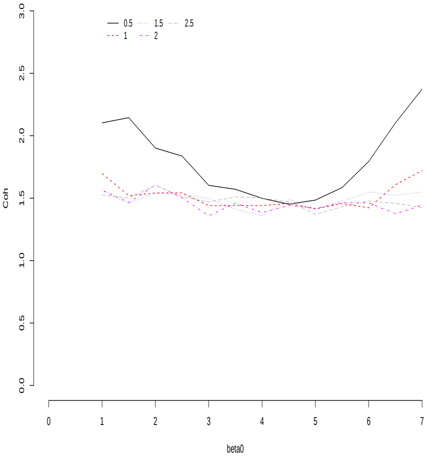
$$\begin{aligned}
\pi_1(\beta|\beta_0,\alpha_\beta) &= \mathcal{G}(\alpha_\beta, \alpha_\beta\beta_0^{-1}), \\
\pi_2(\beta|\beta_0,\alpha_\beta) &= \mathcal{N}^+\left(\beta_0, \sigma_\beta^2 = \frac{\beta_0^2}{\alpha_\beta}\right)
\end{aligned}$$

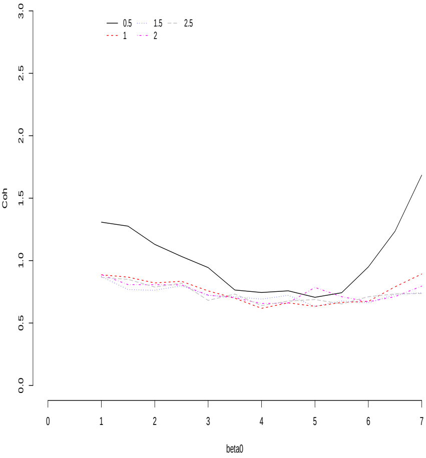
On choisit $\tilde{\beta} = 3$ et on décale "à gauche" l'interv. de confiance pour η.

$$\begin{aligned}
\pi_1(\eta|t,\alpha_\eta) &= \mathcal{G}\left(\alpha_\eta, \alpha_\eta \frac{\Gamma(1+1/\tilde{\beta})}{t}\right), \\
\pi_2(\eta|t,\alpha_\eta) &= \mathcal{N}^+\left(\frac{t}{\Gamma(1+1/\tilde{\beta})}, \sigma_\eta^2 = \frac{t^2}{\alpha_\eta\ \Gamma^2(1+1/\tilde{\beta})}\right).
\end{aligned}$$

Expertise	intervalle (5%,95%)	valeur médiane	t	α_η	σ_η
EDF	(224,336)	280	252	60	36.4
Westinghouse	(112,560)	280	269	4.5	142.0




0.5
1.5
2.5
1
2
Coh
beta0
0 1 2 3 4 5 6 7
0.0 0.5 1.0 1.5 2.0 2.5 3.0


0.5
1.5
2.5
1
2
Coh
beta0
0 1 2 3 4 5 6 7
0.0 0.5 1.0 1.5 2.0 2.5 3.0

Conclusions

- Il existe des modélisations *a priori* permettant à l'expertise Westinghouse d'être cohérente avec les données (large variances pour β).
- On détecte toujours un conflit entre l'expertise EDF et les données.
- Modélisation *a priori* plus fine (conditionnelle)
- Présence très influente d'anciennes données $\Rightarrow$ nécessité d'*a priori* "forts".

$\Rightarrow$ Pondérer l'expertise EDF, préciser l'origine des données.

Conclusions et problèmes de calibration

1. Un préalable à l'inférence subjective.
2. Un jugement de l'expert ou des données facilité.
3. Permet de d'utiliser et sélectionner des *a priori* raisonnables.
4. Ouvre une voie vers de nombreux problèmes de calibration.

- Lorsqu'on cherche à **calibrer** π sans information particulière sur l'expert, on propose de choisir π telle que

$$\mathcal{C}oh(\pi|\mathbf{X_n}) = 1.$$

avec des contraintes centrales (ordre 1) sur π (reflétant l'avis quantitatif de l'expert).

$\Rightarrow$ Compromis raisonnable biais-variance.

- Candidats MIA (*posterior priors*) parfois encore trop informatifs $\Rightarrow$ la règle peut rejeter des *a priori* moins informatifs.

$\Rightarrow$ pas intéressants pour le problème (n'apportent pas d'info *a posteriori*).
$\Rightarrow$ doivent être automatiquement acceptés.

Il faut définir une **quantité minimale d'information** (nb. ou fraction de données fictives, etc.).

- Exemple : on suppose qu'un expert ne peut pas apporter moins d'une fraction $1/q$ d'information apportée par une donnée où $q \in \mathbb{N}^*$.
- $\Leftrightarrow$ q informations minimales subjectives "valent" 1 donnée réelle.
- Le *posterior prior* correspondant serait alors

$$\pi(\theta) \quad \propto \quad \int \pi^J(\theta) \left[\mathcal{L}(X(I), \theta)\right]^{1/mq} \; g(X(I)) \; dX(I)$$

- $\left[\mathcal{L}(X(I), \theta)\right]^{1/mq}$ = vraisemblance fictive, où m est la taille des MTS du problème.